



Grado en Análisis de Negocios / Business Analytics

ESTUDIO COMPARATIVO SOBRE CORRUPCIÓN EN VENEZUELA, EL SALVADOR Y URUGUAY

Clave: 201801166

Autora: Tatiana López Buendía

Directora: Susana Josefa Gago Rodríguez

Resumen

Este Trabajo de Fin de Grado realiza un análisis comparativo de los factores que explican la corrupción en tres países de América Latina: Venezuela, El Salvador y Uruguay. Para ello, se construyó una base de datos a partir de variables económicas, sociales, institucionales y específicas sobre corrupción y sobornos, extraídas de fuentes y encuestas de acceso público del Banco Mundial como Enterprise Surveys, Worldwide Governance Indicators, World Development Indicators y ESG Data.

La metodología empleada combina técnicas de análisis estadístico y Machine Learning – como Random Forest (Classifier y Regressor), Regresión Logística y Regresión Lineal Ridge – aplicadas tanto a nivel global como de forma específica por país.

Entre los principales hallazgos, se observa una asociación positiva entre la corrupción y factores como la desigualdad de ingresos, la pobreza extrema y la debilidad institucional, mientras que variables como el acceso a servicios básicos, la alfabetización y la calidad regulatoria actúan como factores protectores.

El estudio pone de relieve la influencia del contexto nacional en el impacto de cada variable y la necesidad de diseñar políticas públicas diferenciadas y adaptadas a las particularidades de cada país.

Palabras clave: Corrupción, Soborno, Venezuela, El Salvador, Uruguay, Machine Learning, Random Forest, Regresión Logística, Regresión Lineal, Factores Económicos, Factores Sociales, Factores Institucionales, Factores de Gobernanza.

Abstract

This thesis conducts a comparative analysis of the factors that explain corruption in three Latin American countries: Venezuela, El Salvador, and Uruguay. To this end, a database was constructed from economic, social, institutional, and specific variables on corruption and bribery, extracted from publicly available sources and surveys from the World Bank, such as Enterprise Surveys, Worldwide Governance Indicators, World Development Indicators, and ESG Data.

The methodology used combines statistical analysis and machine learning techniques – such as Random Forest (Classifier and Regressor), Logistic Regression, and Linear Ridge Regression – applied both globally and in a country-specific manner.

Among the main findings, a positive association is observed between corruption and factors such as income inequality, extreme poverty, and institutional weakness, while variables such as access to basic services, literacy, and regulatory quality act as protective factors.

The study highlights the influence of the national context on the impact of each variable and the need to design differentiated public policies adapted to the specificities of each country.

Keywords: Corruption, Bribery, Venezuela, El Salvador, Uruguay, Machine Learning, Random Forest, Logistic Regression, Linear Regression, Economic Factors, Social Factors, Institutional Factors, Governance Factors.

Índice

1.	Introducción.....	11
1.1.	Contextualización del trabajo	11
1.2.	Motivación del trabajo.....	12
1.3.	Objetivos.....	13
2.	Marco teórico.....	14
2.1.	Corrupción y crecimiento económico	14
2.2.	Corrupción y consecuencias sociales	16
2.3.	Conclusión del marco teórico	18
3.	Metodología.....	19
3.1.	Enfoque metodológico.....	19
3.2.	Fuentes de datos	19
3.3.	Técnicas empleadas	20
4.	Obtención de datos	21
4.1.	Fuente de datos	21
4.2.	Selección de variables clave	25
4.3.	Justificación de las variables seleccionadas	30
5.	Limpieza y análisis exploratorio de datos	32
5.1.	Descripción general	32
5.2.	Limpieza inicial de valores nulos	33
5.3.	Imputación de valores nulos mediante KNNImputer	35
5.4.	Estadísticas descriptivas	38
5.5.	Correlaciones entre variables.....	41
5.6.	Distribuciones.....	45
6.	Desarrollo de modelos	53
6.1.	Modelo de Random Forest	53

6.2.	Modelos de Regresión Logística y Regresión Lineal.....	73
7.	Discusión de los resultados	91
7.1.	Random Forest Classifier	91
7.2.	Regresión Logística	92
7.3.	Random Forest Regressor – Análisis por país.....	95
7.4.	Regresión Lineal Ridge – Análisis por país	98
7.5.	Comparación de resultados entre modelos	101
7.6.	Conclusión del análisis	103
8.	Conclusiones y trabajos futuros.....	106
9.	Referencias bibliográficas	111
10.	ANEXO: Declaración Uso IAG	115

Índice de Figuras

Figura 1: Interfaz de selección de variables para la base de datos Enterprise Surveys..	23
Figura 2: Interfaz de selección de variables para la base de datos World Development Indicators.	23
Figura 3: Interfaz de selección de variables para la base de datos Worldwide Governance Indicators.	24
Figura 4: Interfaz de selección de variables para la base de datos Environment Social and Governance (ESG) Data.	24
Figura 5: Nombres de las columnas de la base de datos.	32
Figura 6: Tipos de datos y valores nulos para cada país.	33
Figura 7: Porcentaje de valores nulos por año.	34
Figura 8: Valores nulos y porcentaje de valores nulos para cada país tras la eliminación de la primera década (1960-1969).	35
Figura 9: Evolución del error cuadrático medio (RMSE) según el número de vecinos (k) para Venezuela en la imputación con KNNImputer.	36
Figura 10: Evolución del error cuadrático medio (RMSE) según el número de vecinos (k) para El Salvador en la imputación con KNNImputer.	37
Figura 11: Evolución del error cuadrático medio (RMSE) según el número de vecinos (k) para Uruguay en la imputación con KNNImputer.	37
Figura 12: Mapa de calor de la matriz de correlaciones entre las variables clave.	43
Figura 13: Mapa de calor de la matriz de correlaciones entre las variables clave para Venezuela.	44
Figura 14: Mapa de calor de la matriz de correlaciones entre las variables clave para El Salvador.	44
Figura 15: Mapa de calor de la matriz de correlaciones entre las variables clave para Uruguay.	45
Figura 16: Histogramas de las variables de corrupción.	46
Figura 17: Histogramas de las variables de corrupción por país.	46
Figura 18: Histogramas de las variables económicas.	48
Figura 19: Histogramas de las variables económicas por país.	49
Figura 20: Histogramas de las variables sociales.	50
Figura 21: Histogramas de las variables sociales por país.	51

Figura 22: Histograma de las variables institucionales.	51
Figura 23: Histograma de las variables institucionales por país.	52
Figura 24: Dimensiones del conjunto de datos total del que se parte en el desarrollo de los modelos.	55
Figura 25: Mediana de la variable Bribery Incidence.	56
Figura 26: Nuevas variables dummies para los países.	56
Figura 27: Dimensiones de los subconjuntos X e y.	57
Figura 28: Resultado de aplicar la técnica SMOTE sobre el conjunto de entrenamiento.	58
Figura 29: Mejores hiperparámetros para el modelo de Random Forest Classifier.	60
Figura 30: Creación del modelo de Random Forest Classifier con los hiperparámetros óptimos.	60
Figura 31: Matriz de confusión del modelo de Random Forest Classifier.	62
Figura 32: Curva ROC del modelo de Random Forest Classifier.	62
Figura 33: Resultados de la validación cruzada para Random Forest Classifier.	63
Figura 34: Informe de clasificación con validación cruzada para Random Forest Classifier.	63
Figura 35: Matriz de confusión para validación cruzada en Random Forest Classifier.	64
Figura 36: Curva ROC del modelo de Random Forest Classifier con validación cruzada.	64
Figura 37: Orden de las variables por importancia en el modelo de Random Forest Classifier.	65
Figura 38: Top 10 variables más importantes según el modelo de Random Forest Classifier.	66
Figura 39: Distribución real de Bribery Class por país.	68
Figura 40: Proporción de predicciones de Bribery Class por país con Random Forest Classifier.	68
Figura 41: Importancia relativa de las variables de país en el modelo de Random Forest Classifier.	69
Figura 42: Métricas del modelo Random Forest con regresión por país.	70
Figura 43: Top 15 Variables más importantes para Venezuela según el modelo de Regresión de Random Forest.	70

Figura 44: Top 15 Variables más importantes para El Salvador según el modelo de Regresión de Random Forest.....	71
Figura 45: Top 15 Variables más importantes para Uruguay según el modelo de Regresión de Random Forest.....	72
Figura 46: Comparación de las 5 variables más importantes por país según el modelo de regresión de Random Forest.	72
Figura 47: Mejores hiperparámetros para el modelo de Regresión Logística.	76
Figura 48: Creación del modelo de Regresión Logística con los hiperparámetros óptimos.	76
Figura 49: Matriz de confusión del modelo de Regresión Logística.....	77
Figura 50: Curva ROC del modelo de Regresión Logística.....	78
Figura 51: Resultados de la validación cruzada con Regresión Logística.	79
Figura 52:Informe de clasificación con validación cruzada para Regresión Logística..	79
Figura 53: Matriz de confusión para validación cruzada en Regresión Logística.....	80
Figura 54: Curva ROC con la validación cruzada para Regresión Logística.....	80
Figura 55: Orden de las variables por importancia absoluta en el modelo de Regresión Logística.	81
Figura 56: Orden por importancia de las variables que aumentan la probabilidad de corrupción en el modelo de Regresión Logística.	83
Figura 57: Orden por importancia de las variables que disminuyen la probabilidad de corrupción en el modelo de Regresión Logística.	84
Figura 58: Métricas del modelo Regresión Lineal por país.....	85
Figura 59: Top 15 Variables más importantes para Venezuela según el modelo de Regresión Lineal.....	87
Figura 60: Top 15 Variables más importantes para El Salvador según el modelo de Regresión Lineal.....	88
Figura 61: Top 15 Variables más importantes para Uruguay según el modelo de Regresión Lineal.....	89
Figura 62: Comparación de las 5 variables más importantes por país según el modelo de Regresión Lineal. El símbolo (+) indica que la variable contribuye a aumentar la corrupción; (-) indica que está asociada a una menor corrupción.	89

Índice de Tablas

Tabla 1: Estadísticas descriptivas de variables de soborno y corrupción.....	38
Tabla 2: Estadísticas descriptivas de variables de gobernanza.....	39
Tabla 3: Estadísticas descriptivas de factores económicos.	40
Tabla 4: Informe de clasificación sobre el conjunto de prueba del modelo de Random Forest Classifier con todos los países.	61
Tabla 5: Informe de clasificación sobre el conjunto de prueba del modelo de Regresión Logística con todos los países.	77

1. Introducción

1.1. Contextualización del trabajo

La corrupción constituye un fenómeno estructural que debilita el funcionamiento institucional, genera ineficiencias económicas y reproduce desigualdades sociales. Su impacto en América Latina ha sido ampliamente documentado, afectando tanto a la competitividad empresarial como a la calidad de vida de la población.

En este contexto, el presente Trabajo de Fin de Grado se centra en el análisis comparativo de las dinámicas de soborno en tres países latinoamericanos con perfiles institucionales, sociales y económicos claramente diferenciados: Venezuela, El Salvador y Uruguay.

La selección de estos tres países se basa en los resultados del Índice de Percepción de la Corrupción (CPI) publicado por (Transparency International, 2024), una de las métricas más reconocidas a nivel internacional para evaluar la corrupción en el sector público.

El CPI asigna a cada país una puntuación que va de 0 a 100, donde 0 indica un nivel extremadamente alto de corrupción percibida y 100 representa un entorno percibido como completamente transparente. Esta puntuación se construye a partir de múltiples fuentes de datos, incluyendo evaluaciones de expertos y encuestas de opinión a empresarios, lo que permite captar la percepción generalizada sobre el abuso de poder, el nepotismo, el soborno y la integridad institucional.

En el informe del año 2023 se asigna a Venezuela una puntuación de 13/100, ubicándola entre los países más corruptos del mundo; a El Salvador, 31/100, reflejando una situación intermedia; y a Uruguay, 73/100, consolidándose como el país más transparente de América Latina.

Esta variabilidad en los niveles de corrupción permite construir una base de datos para benchmarking ideal para los propósitos de este estudio. Su carácter heterogéneo constituye una herramienta clave para identificar posibles patrones comparativos de baja, media y alta corrupción, analizar diferencias estructurales e interpretar cómo los factores institucionales, económicos y sociales se relacionan con la incidencia del soborno.

Adicionalmente, este enfoque comparativo también persigue identificar buenas prácticas, limitaciones institucionales y oportunidades para el diseño de políticas públicas más efectivas en la lucha contra la corrupción.

1.2. Motivación del trabajo

La motivación principal de este estudio radica en comprender cómo y por qué surgen las prácticas corruptas – en particular, los sobornos – en contextos institucionales diversos. El contraste entre Venezuela, con altos niveles de corrupción, El Salvador, en una situación intermedia, y Uruguay, como referente regional, permite explorar la interacción entre factores como la pobreza, la desigualdad, la informalidad económica, la efectividad gubernamental y el acceso a servicios públicos.

Además, la aplicación de técnicas de *Machine Learning* sobre bases de datos oficiales del Banco Mundial, como *Enterprise Surveys*, *ESG Data*, *Worldwide Governance Indicators* y *World Development Indicators*, ilustra cómo las técnicas de Business Analytics estudiadas en el grado permiten incorporar un enfoque cuantitativo y facilitan un análisis innovador de fenómeno con tintes éticos y económicos: la corrupción, completando los estudios previos que, tradicionalmente, han abordado esta lacra tanto desde perspectivas cualitativas y también cuantitativas.

Este trabajo busca demostrar que el estudio de **la corrupción puede ser facilitado contribuyendo a que sea analizada de forma objetiva y sistemática mediante el análisis y modelado de datos estructurados**, con el fin de extraer conclusiones empíricas robustas que contribuyan a una mejor comprensión del fenómeno y orienten futuras políticas públicas.

1.3. Objetivos

El objetivo general del trabajo es identificar y analizar los factores que explican la incidencia de los sobornos en Venezuela, El Salvador y Uruguay, mediante un enfoque cuantitativo y comparativo basado en datos estructurados y técnicas de aprendizaje automático.

Para alcanzar este propósito, se propone, en primer lugar, analizar la influencia de las variables económicas, sociales e institucionales sobre la incidencia del soborno. A continuación, se pretende comparar los patrones de soborno y corrupción detectados en los tres países seleccionados (Venezuela, El Salvador y Uruguay), con el fin de poner de relieve similitudes y diferencias estructurales en sus dinámicas.

Asimismo, se desarrollarán modelos de clasificación y regresión mediante algoritmos de *Machine Learning*, con el objetivo de identificar las variables más relevantes en la predicción de contextos de alta o baja corrupción y valorar cómo varía su peso relativo entre los tres países analizados.

Finalmente, se aspira a que los hallazgos obtenidos permitan contribuir con evidencia empírica a la formulación de recomendaciones de política pública orientadas a fortalecer la transparencia institucional y reducir las condiciones que favorecen las prácticas corruptas.

2. Marco teórico

Numerosos estudios han abordado el fenómeno de la corrupción desde una perspectiva multidimensional, analizando su relación con variables económicas, sociales e institucionales como las utilizadas en este estudio. En este contexto, la revisión del estado de la cuestión resulta esencial para situar el estudio comparativo entre Venezuela, El Salvador y Uruguay, y para establecer las bases sobre las cuales interpretar los resultados empíricos.

2.1. Corrupción y crecimiento económico

El efecto de la corrupción sobre el crecimiento económico ha sido objeto de un amplio debate académico. Tradicionalmente, se ha comparado su impacto al de un impuesto, con la diferencia esencial de que el pago de sobornos no genera ingresos públicos y, por tanto, no permite financiar bienes y servicios que podrían favorecer el desarrollo. Desde esta perspectiva, la corrupción no solo reduce la eficiencia económica, sino que también debilita la capacidad del Estado para cumplir funciones redistributivas y productivas.

Shleifer y Vishny (1993) argumentan que la corrupción puede ser incluso más perjudicial que los impuestos, ya que genera mayores costes de transacción debido a la incertidumbre, la opacidad y la imposibilidad de hacer cumplir legalmente los “contratos corruptos”. Además, los pagos de sobornos suelen estar sujetos a arbitrariedad, lo que introduce riesgos adicionales para las empresas y desalienta la inversión a largo plazo. Por otro lado, la llamada “hipótesis de la corrupción eficiente” sostiene que los sobornos podrían agilizar trámites administrativos en contextos de elevada burocracia, facilitando que las empresas más eficientes accedan más rápido a licencias o contratos (Lui, 1985).

No obstante, la evidencia empírica tiende a rechazar esta visión optimista. Mauro (1995), a través de estudios transnacionales, encuentra una correlación negativa entre los niveles de corrupción y el crecimiento del PIB. Este hallazgo refuerza la hipótesis de que la corrupción distorsiona el entorno competitivo, incrementa la incertidumbre y limita las oportunidades de crecimiento, especialmente para las pequeñas y medianas empresas. Adicionalmente, diversos análisis concluyen que en entornos donde los sobornos son frecuentes, las empresas tienden a crecer a un ritmo menor, lo que repercute

negativamente en la inversión, el empleo y la productividad general (Fisman & Svensson, 2007).

Más allá del debate sobre los costes de la corrupción, la discusión contemporánea se ha centrado también en el papel del Estado. En este contexto, Hopkin y Rodríguez-Pose (2007) distinguen dos corrientes contrapuestas: la visión del “*grabbing hand*” (mano que arrebatata), que considera que un Estado intervencionista y con amplias atribuciones regulatorias genera incentivos para prácticas corruptas, y la visión del “*helping hand*” (mano que ayuda), que sostiene que el Estado puede desempeñar un papel positivo en el desarrollo económico si se centra en funciones redistributivas y de provisión de servicios. Su análisis demuestra que es el exceso de regulación sobre el sector privado el que mejor predice la corrupción, mientras que altos niveles de gasto público en servicios sociales se asocian con menores niveles de prácticas corruptas, lo que sugiere que no es el tamaño del Estado en sí mismo lo que importa, sino la calidad y el enfoque de su intervención.

Esto sugiere que los sistemas de bienestar avanzados logran combinar un Estado fuerte en términos de redistribución y provisión de servicios con un entorno empresarial gobernado por reglas claras y transparentes. Por tanto, las políticas anticorrupción deberían centrarse más en simplificar y racionalizar los marcos regulatorios, especialmente aquellos que otorgan discrecionalidad a los funcionarios, que en reducir el tamaño del Estado en sí mismo.

Este enfoque es especialmente útil para el análisis comparativo de América Latina. Un estudio documental sobre Chile, Colombia, Costa Rica, Guatemala y México concluye que los países con mejores indicadores institucionales, como Estado de Derecho, Control del Ejecutivo, Índice de Democracia Liberal o PIB per cápita, presentan niveles significativamente más bajos de corrupción (Torres Fragoso, 2023). Chile y Costa Rica, en particular, destacan por sus sistemas institucionales robustos y sus bajos niveles de soborno, mientras que Colombia, Guatemala y México muestran patrones opuestos, con fragilidad institucional y corrupción más extendida.

De igual modo, en el estudio centrado en América Latina y el Caribe, se demuestra que los países con mayores niveles de corrupción tienden a registrar menores tasas de crecimiento económico, estableciendo una correlación negativa entre el índice de control de la corrupción y el PIB per cápita (Desfrancois & Pastas Gutiérrez, 2022).

Adicionalmente, en países en desarrollo, la corrupción también afecta el acceso al crédito formal. Según Amin y Motta (2021), en contextos con alta corrupción, las pequeñas y medianas empresas enfrentan mayores restricciones financieras, lo que las obliga a recurrir al financiamiento informal y limita su capacidad de expansión e inversión. Este fenómeno representa una barrera adicional para el crecimiento económico inclusivo.

En el ámbito de la inversión, la corrupción actúa como un obstáculo tanto para la inversión pública como privada. Vito Tanzi y Hamid Davood (1997) argumentan que la corrupción distorsiona la asignación del gasto público, perjudicando sectores clave como educación y salud.

Desde el punto de vista de la inversión extranjera directa, Wei (2000) demuestra que los niveles elevados de corrupción funcionan como un impuesto implícito que disuade la llegada de capital extranjero.

Por otro lado, también se ha demostrado que la corrupción puede incidir en la estabilidad macroeconómica, con estudios que encuentran evidencia de una relación positiva entre corrupción e inflación en países en desarrollo, argumentando que la corrupción debilita los mecanismos de control institucional sobre la política monetaria (Özşahin & Üçler, 2017).

Finalmente, en el plano institucional, los seis indicadores de gobernanza propuestos por el Banco Mundial (control de la corrupción, estado de derecho, estabilidad política, calidad regulatoria, voz y rendición de cuentas, y efectividad gubernamental) están altamente interrelacionados. Kaufmann, Kraay y Mastruzzi (2009) señalan que mejoras en estas dimensiones generan sinergias que fortalecen la calidad institucional y reducen las oportunidades para prácticas corruptas.

2.2. Corrupción y consecuencias sociales

Desde una perspectiva social, la corrupción genera efectos regresivos. Gupta, Davoodi y Alonso-Terme (2002) demuestran que un aumento en la corrupción incrementa la desigualdad del ingreso y reduce el acceso a servicios básicos como la salud, la educación y la justicia, afectando especialmente a los grupos más vulnerables. Además, según Vega (2025) existe una correlación general entre altos niveles de corrupción (medidos como

baja puntuación del Índice de Percepción de la Corrupción IPC) y valores elevados del Coeficiente de Gini (que mide la desigualdad de ingresos en una sociedad).

En materia de salud, Lewis (2006) y el Jameel (2023) advierten que la corrupción reduce la eficiencia del gasto sanitario y deteriora los indicadores de salud pública, especialmente en países con menor capacidad institucional. En el ámbito educativo, Cárdenas Denham (2012) documenta cómo la presencia de prácticas corruptas en la gestión de sistemas educativos reduce la probabilidad de lograr una adecuada distribución de oportunidades educativas.

Respecto a la provisión de los servicios públicos básicos como la electricidad, la corrupción impacta desproporcionadamente a los más desfavorecidos, aumentando los costes y reduciendo el acceso a servicios esenciales como la electricidad (World Bank, 2025). Esta situación se agrava en América Latina, donde, según el canal Los Tiempos (2023), millones de personas aún carecen de acceso a la electricidad, un problema que la corrupción puede agravar al desviar fondos de proyectos de infraestructura o encarecer su implementación.

Asimismo el informe del Banco Mundial elaborado por Loayza (2018) vincula la corrupción con el desempleo y la informalidad laboral. Según este estudio, los altos niveles de corrupción aumentan los costes y la incertidumbre para operar en la economía formal, lo que empuja a empresas y trabajadores hacia el sector informal de la corrupción. En este marco, el autor sostiene que cualquier política orientada a reducir la informalidad debe ir acompañada de esfuerzos por fortalecer la institucionalidad y aumentar la productividad laboral y empresarial.

Otro eje relevante de análisis es la relación entre corrupción y género. Diversas investigaciones señalan que las mujeres, enfrentan formas diferenciadas y desproporcionadas de corrupción.

Un informe universitario subraya que, en muchos contextos, la corrupción representa una barrera adicional para el acceso equitativo al empleo, la justicia o los servicios públicos, reforzando desigualdades de género existentes (U4 Anti-Corruption Resource Centre, 2024). También destaca que una de las formas más preocupantes y menos denunciadas es la corrupción sexual, una práctica en la que las autoridades o funcionarios exigen favores sexuales a cambio de servicios, licencias, acceso a educación, atención médica o

protección legal. Esta forma de soborno tiende a quedar fuera de las estadísticas oficiales y de los marcos normativos tradicionales, lo que dificulta su sanción.

Cabe destacar la hipótesis establecida según Haberson (2021), en la que establece que una mayor participación de mujeres en espacios políticos y administrativos suele estar asociada con menores niveles de corrupción, debido a estilos de liderazgo más inclusivos y orientados a la rendición de cuentas. No obstante, como advierten organismos como el Banco Mundial, la inclusión femenina debe ir acompañada de instituciones sólidas y entornos laborales libres de discriminación y violencia, para que su impacto anticorrupción sea efectivo.

2.3. Conclusión del marco teórico

Los estudios revisados evidencian que la corrupción es un fenómeno complejo y multidimensional, con impactos significativos en el crecimiento económico, el acceso al crédito, la inversión, la igualdad social, la calidad de los servicios públicos y la gobernanza institucional. También se pone de manifiesto que sus efectos no son homogéneos, afectando de forma diferenciada a determinados grupos, como las mujeres o las pequeñas empresas.

Este marco teórico ofrece una base sólida para interpretar los resultados del presente estudio. De esta forma, se persigue evaluar si los casos de Venezuela, El Salvador y Uruguay reflejan dinámicas de soborno que se alinean con la teoría existente, o si, por el contrario, aportan nuevas perspectivas sobre los factores determinantes de la corrupción.

3. Metodología

3.1. Enfoque metodológico

El presente estudio adopta un **método de investigación cuantitativo**, orientado al análisis de datos numéricos con el fin de identificar patrones, relaciones y comportamientos asociados a la corrupción, y en particular, a la práctica del soborno en tres países latinoamericanos: Venezuela, El Salvador y Uruguay. Este enfoque permite trabajar con grandes volúmenes de datos provenientes de fuentes oficiales – como las bases de datos del Banco Mundial – y aplicar herramientas estadísticas y computacionales para extraer conclusiones objetivas.

El **tipo de razonamiento empleado es inductivo**, ya que parte de la observación y el análisis de datos específicos – recopilados a partir de encuestas empresariales y otros indicadores económicos y sociales – para generar conclusiones generales sobre cómo diversos factores económicos, sociales e institucionales influyen en la prevalencia del soborno en Venezuela, El Salvador y Uruguay.

3.2. Fuentes de datos

Los datos utilizados en este trabajo proceden de fuentes oficiales y de acceso público del **Banco Mundial**. En concreto, se han utilizado las siguientes bases de datos:

- **Enterprise Surveys:** encuestas empresariales que recogen la experiencia de las empresas con prácticas corruptas, acceso a financiación, servicios públicos, entre otros.
- **Worldwide Governance Indicators (WGI):** indicadores agregados que miden aspectos como el Estado de derecho, la calidad regulatoria o el control de la corrupción.
- **World Development Indicators (WDI):** conjunto amplio de indicadores económicos y sociales, como PIB per cápita, índice de Gini o inversión extranjera directa.
- **ESG Data (Environmental, Social and Governance):** datos sobre factores sociales e institucionales relevantes para el análisis de gobernanza y desarrollo.

Además, se han tenido en cuenta los resultados del **Índice de Percepción de la Corrupción (CPI)** de *Transparency International*, que han servido como referencia comparativa.

Todas las bases de datos han sido integradas, limpiadas y transformadas en un único conjunto de datos homogéneo que permite el análisis comparativo por país y la aplicación de modelos de predicción.

3.3. Técnicas empleadas

El análisis se ha estructurado en dos grandes fases: una etapa inicial de Análisis Exploratorio de Datos (*EDA*) y una posterior fase de modelado predictivo mediante técnicas de *Machine Learning* supervisado.

Durante el EDA se realizaron descripciones estadísticas básicas, análisis de distribuciones, detección de valores atípicos y visualización de correlaciones entre variables. Esta etapa permitió identificar relaciones preliminares entre indicadores institucionales, económicos y sociales, así como diferencias significativas entre los países estudiados.

Posteriormente, se aplicaron técnicas de *Machine Learning* supervisado, en concreto, modelos de regresión y clasificación. Estos modelos permiten:

- **Detectar correlaciones** entre variables con los indicadores de soborno.
- **Identificar patrones** comunes en contextos de alta y baja corrupción.
- **Construir modelos predictivos**, con el objetivo de estimar la incidencia del soborno en función de determinadas condiciones estructurales.

Entre los algoritmos utilizados se incluyen el modelo de **Random Forest**, la **Regresión Logística** y la **Regresión Lineal**, que permiten evaluar la importancia relativa de cada variable en la predicción de los niveles de corrupción. Estos modelos se aplicaron tanto sobre el conjunto total de datos como de forma desagregada por país, con el fin de comparar las dinámicas específicas de Venezuela, El Salvador y Uruguay.

4. Obtención de datos

En este capítulo se describe el proceso de recopilación de los datos utilizados en el presente estudio comparativo sobre corrupción y sobornos en Venezuela, El Salvador y Uruguay. Dado el enfoque cuantitativo e inductivo del trabajo, la calidad, trazabilidad y fiabilidad de las fuentes de datos constituyen un aspecto esencial para garantizar la validez de los análisis posteriores.

La selección de las bases de datos se ha realizado en función de su relevancia temática y su reconocimiento internacional. Todas las fuentes provienen del portal oficial del Banco Mundial, que ofrece acceso libre y estructurado a indicadores económicos, sociales, institucionales y empresariales de alcance global. Estas bases permiten abordar la problemática de la corrupción desde una perspectiva integral, incorporando variables tanto de incidencia directa del soborno como de contexto político, institucional, económico y social.

A continuación, se detallan las fuentes utilizadas, el proceso de descarga desde la plataforma DataBank, así como la selección de variables clave. Todo ello con el objetivo de fundamentar un análisis riguroso y transparente.

4.1. Fuente de datos

Los datos utilizados en este estudio han sido obtenidos del portal oficial del **Banco Mundial**, a través del sistema *DataBank* (*DataBank | The World Bank, 2025*) .

El portal del Banco Mundial permite el libre acceso a bases de datos estructuradas a nivel mundial. Muchas de estas bases de datos provienen de **encuestas empresariales realizadas a nivel nacional**, mientras que otras consisten en **indicadores agregados elaborados por organismos internacionales** a partir de fuentes oficiales. Además, estas bases de datos están organizadas por temática, país, selección de indicadores y serie temporal.

Para este proyecto, se han obtenido los archivos CSV correspondientes a los países objeto de estudio: **Venezuela, El Salvador y Uruguay**. Así mismo, se han seleccionado las siguientes cuatro bases de datos:

- **Enterprise Surveys** (*Enterprise Surveys* | *DataBank*, 2025): Database que proviene directamente de encuestas empresariales realizadas por el Banco Mundial a empresas formales. Proporciona datos detallados sobre prácticas de gestión, corrupción, acceso a servicios y financiamiento, entre otros.
- **World Development Indicators (WDI)** (*World Development Indicators* | *DataBank*, 2025): Base de datos de indicadores económicos, sociales y demográficos, construida a partir de estadísticas oficiales nacionales e internacionales.
- **Worldwide Governance Indicators (WGI)** (*Worldwide Governance Indicators* | *DataBank*, 2025): Indicadores de gobernanza elaborados por el Banco Mundial a partir de múltiples fuentes, que incluyen encuestas de opinión y evaluaciones de expertos sobre aspectos como el control de la corrupción y la efectividad del gobierno.
- **Environment Social and Governance (ESG) Data** (*Environment Social and Governance (ESG) Data* | *DataBank*, 2025): Base de datos que contiene indicadores relacionados con sostenibilidad ambiental, factores sociales y gobernanza, publicados por el Banco Mundial a partir de datos institucionales y registros oficiales.

A continuación, las figuras: Figura 1, Figura 2, Figura 3 y Figura 4 muestran ejemplos del proceso de obtención y filtrado de datos desde la plataforma *DataBank* del Banco Mundial. En ellas se ilustra cómo se ha podido seleccionar la base de datos deseada, los países de interés, los indicadores económicos, sociales e institucionales, y el rango temporal correspondiente. De este modo, se evidencia la trazabilidad y la transparencia en el proceso de recopilación de datos utilizado en este estudio.

WORLD BANK GROUP

DataBank Enterprise Surveys ¹

Table

Variables Layout Styles Save Share Embed

Database Available 86 | Selected 1

Database preview ☒ ON ☐ OFF [?] Enter Keywords for

- ☐ Health Nutrition and Population Statistics by Wealth Quintile
- ☒ Enterprise Surveys
- ☐ Exporter Dynamics Database - Indicators at Country-Year Level
- ☐ Exporter Dynamics Database - Indicators at
- ☐ Landmine Contamination Casualties and Clearance (LC3D)
- ☐ Global Economic Monitor (GEM)
- ☐ Exporter Dynamics Database - Indicators at Country-Destination-Year Level
- ☐ Exporter Dynamics Database - Indicators at

Country Available 2 | Selected 3

Series Available 115 | Selected 115

Time Available 15 | Selected 15

Preview

Clear Selection | Add Country (3) | Add Series (115) | Add Time (15)

El Salvador

- Age of the establishment (years)
- Annual employment growth (%)
- Bribery depth (% of public transactions where a gift or informal payment was requested)
- Bribery incidence (percent of firms experiencing at least one bribe payment request)
- Capacity utilization (%)
- Days to clear direct exports through customs
- Days to clear imports from customs
- Days to obtain a construction-related permit
- Days to obtain an electrical connection (upon application)
- Days to obtain an import license
- Days to obtain an operating license
- If a generator is used, average proportion of electricity from a generator (%)

Figura 1: Interfaz de selección de variables para la base de datos Enterprise Surveys.

WORLD BANK GROUP

This page is in [English](#) [Español](#) [Français](#) [عربي](#) [中文](#)

DataBank World Development Indicators ¹

Table

Variables Layout Styles Save Share Embed

Database Available 86 | Selected 1

Database preview ☒ ON ☐ OFF [?] Enter Keywords for

- ☒ World Development Indicators
- ☐ Education Statistics - All Indicators
- ☐ Doing Business
- ☐ Millennium Development Goals
- ☐ Joint External Debt Hub
- ☐ Statistical Capacity Indicators
- ☐ Gender Statistics
- ☐ Health Nutrition and Population Statistics
- ☐ IDA Results Measurement System, Tier I Database - WDI
- ☐ Quarterly Public Sector Debt
- ☐ Jobs

Country Available 1 | Selected 3

Series Available 1496 | Selected 1496

Time Available 64 | Selected 64

Preview

Clear Selection | Add Country (3) | Add Series (1496) | Add Time (64)

Venezuela, RB ¹

- Adolescent fertility rate (births per 1,000 women ages 15-19)
- Agriculture, forestry, and fishing, value added (% of GDP)
- Annual freshwater withdrawals, total (% of internal resources)
- Births attended by skilled health staff (% of total)
- Contraceptive prevalence, any method (% of married women ages 15-49)
- Domestic credit provided by financial sector (% of GDP)
- Electric power consumption (kWh per capita)
- Energy use (kg of oil equivalent per capita)
- Exports of goods and services (% of GDP)
- External debt stocks, total (DOD, current US\$)
- Fertility rate, total (births per woman)
- Foreign direct investment, net inflows (BoP, current US\$)

Figura 2: Interfaz de selección de variables para la base de datos World Development Indicators.

WORLD BANK GROUP

DataBank | Worldwide Governance Indicators

Table

Variables | Layout | Styles | Save | Share | Embed

Database: Available 86 | Selected 1

Database preview: ON OFF Enter Keywords for

- Millennium Development Goals
- Joint External Debt Hub
- Global Economic Prospects
- Worldwide Governance Indicators**
- Health Nutrition and Population Statistics by Wealth Quintile
- Tier I Database – WDI
- Quarterly Public Sector Debt
- Jobs
- Africa Development Indicators
- Population estimates and projections
- Landmine Contamination Casualties and

Country: Available 214 | Selected 3

Series: Available 36 | Selected 36

Time: Available 25 | Selected 25

Preview

Clear Selection | Add Country (3) | Add Series (36) | Add Time (25)

El Salvador

- Control of Corruption: Estimate
- Control of Corruption: Number of Sources
- Control of Corruption: Percentile Rank
- Control of Corruption: Percentile Rank, Lower Bound of 90% Confidence Interval
- Control of Corruption: Percentile Rank, Upper Bound of 90% Confidence Interval
- Control of Corruption: Standard Error
- Government Effectiveness: Estimate
- Government Effectiveness: Number of Sources
- Government Effectiveness: Percentile Rank
- Government Effectiveness: Percentile Rank, Lower Bound of 90% Confidence Interval
- Government Effectiveness: Percentile Rank, Upper Bound of 90% Confidence Interval
- Government Effectiveness: Standard Error

Figura 3: Interfaz de selección de variables para la base de datos Worldwide Governance Indicators.

WORLD BANK GROUP

DataBank | Environment Social and Governance (ESG) Data

Table

Variables | Layout | Styles | Save | Share | Embed

Database: Available 86 | Selected 1

Database preview: ON OFF Enter Keywords for

- Health Equity and Financial Protection Indicators (HEFPI)
- Economic Fitness 2
- MIX Market
- Gender Disaggregated Labor Database (GDLD)
- Global Public Procurement
- Education Policy
- Global Financial Inclusion and Consumer Protection Survey
- Environment Social and Governance (ESG) Data**
- ICP 2017
- International Debt Statistics: DSSI
- Statistical Performance Indicators (SPI)
- Food Prices for Nutrition

Country: Available 1 | Selected 3

Series: Available 71 | Selected 71

Time: Available 64 | Selected 64

Preview

Clear Selection | Add Country (3) | Add Series (71) | Add Time (64)

El Salvador

- Access to clean fuels and technologies for cooking (% of population)
- Access to electricity (% of population)
- Adjusted savings: natural resources depletion (% of GNI)
- Adjusted savings: net forest depletion (% of GNI)
- Agricultural land (% of land area)
- Agriculture, forestry, and fishing, value added (% of GDP)
- Annual freshwater withdrawals, total (% of internal resources)
- Annualized average growth rate in per capita real survey mean consumption or income, total population (%)
- Cause of death, by communicable diseases and maternal, prenatal and nutrition conditions (% of total)
- Children in employment, total (% of children ages 7-14)
- CO2 emissions (metric tons per capita)
- Coastal protection
- Control of Corruption: Estimate

Figura 4: Interfaz de selección de variables para la base de datos Environment Social and Governance (ESG) Data.

4.2. Selección de variables clave

Con el objetivo de comprender cómo los factores políticos, institucionales y culturales influyen en la corrupción y el soborno, se ha realizado una selección rigurosa de las variables que permiten abordar esta problemática desde un enfoque cuantitativo, comparativo e inductivo.

Las variables provienen de las cuatro bases de datos descritas en el apartado 4.1: Enterprise Surveys, Worldwide Governance Indicators, World Development Indicators y ESG Data. En adición, podemos agrupar las variables seleccionadas en cuatro bloques temáticos que reflejan los principales ejes de análisis del estudio:

1. **Corrupción y soborno:** 2 indicadores provenientes de *Enterprise Surveys* que capturan de forma directa la existencia y percepción de prácticas corruptas entre las empresas y entidades públicas de los países analizados:
 - a. **Bribery Incidence:** La incidencia del soborno es el porcentaje de empresas que experimentan al menos una solicitud de pago de soborno en seis transacciones públicas relacionadas con acceso a servicios públicos, permisos, licencias e impuestos (*Glossary | DataBank, 2025*).
 - b. **Bribery Depth:** Porcentaje promedio de transacciones oficiales en las que se solicitó un soborno (*Glossary | DataBank, 2025*).
2. **Factores económicos:** 9 indicadores de desarrollo económico y entorno empresarial obtenidos de las bases de datos *World Development Indicators* y *ESG Data*. Estas variables permiten analizar el vínculo entre corrupción y desarrollo económico.
 - a. **GDP (US\$):** El PIB a precios de comprador es la suma del valor bruto añadido por todos los productores residentes de la economía, más los impuestos sobre los productos y menos los subsidios no incluidos en el valor de los productos. Se calcula sin deducir la depreciación de los activos fabricados ni el agotamiento y la degradación de los recursos naturales. Los datos se expresan en dólares estadounidenses. Las cifras del PIB en dólares se convierten a partir de las monedas nacionales utilizando los tipos de cambio oficiales de un año. (*Glossary | DataBank, 2025*).

- b. **Foreign Direct Investment (FDI) (US\$):** La inversión extranjera directa se refiere a los flujos de capital de inversión directa en la economía declarante. Es la suma del capital social, la reinversión de utilidades y otros capitales. La inversión directa es una categoría de inversión transfronteriza asociada con un residente de una economía que controla o ejerce un grado significativo de influencia en la gestión de una empresa residente en otra economía. La propiedad del 10% o más de las acciones ordinarias con derecho a voto es el criterio para determinar la existencia de una relación de inversión directa. Los datos están expresados en dólares estadounidenses (*Glossary | DataBank, 2025*).
- c. **Domestic credit to private sector (% of GDP):** Cuantifica el acceso al crédito del sector privado. El crédito interno al sector privado se refiere a los recursos financieros proporcionados por las sociedades financieras al sector privado. Estos recursos incluyen préstamos, compras de valores no participativos, créditos comerciales y otras cuentas por cobrar que generan un derecho de reembolso (*Glossary | DataBank, 2025*).
- d. **Gross domestic savings (% of GDP):** El ahorro interno bruto se calcula como el PIB menos el gasto de consumo final (consumo total) (*Glossary | DataBank, 2025*).
- e. **Gross capital formation (% of GDP):** Valor de las inversiones brutas en infraestructura, maquinaria y equipo. Puede verse afectado por la seguridad jurídica y la transparencia institucional (*Glossary | DataBank, 2025*).
- f. **Inflation, consumer prices (annual %):** La inflación, medida por el índice de precios al consumidor (IPC), refleja la variación porcentual anual del coste que el consumidor promedio debe asumir para adquirir una canasta de bienes y servicios. Esta variación puede ser fija o modificarse a intervalos específicos, como anualmente. Generalmente se utiliza la fórmula de Laspeyres (*Glossary | DataBank, 2025*).
- g. **Exports of goods and services (% of GDP):** Las exportaciones de bienes y servicios representan el valor de todos los bienes y demás servicios de mercado prestados al resto del mundo. Incluyen el valor de mercancías, fletes, seguros, transporte, viajes, regalías, derechos de licencia y otros

servicios, como servicios de comunicación, construcción, financieros, de información, empresariales, personales y gubernamentales. Excluyen la remuneración de los empleados, las rentas de inversión (anteriormente denominadas servicios de factores) y los pagos de transferencias (*Glossary | DataBank, 2025*).

- h. **Firms using banks to finance working capital (% of firms):** Las empresas que utilizan bancos para financiar capital de trabajo son el porcentaje de empresas que utilizan préstamos bancarios para financiar capital de trabajo (*Glossary | DataBank, 2025*).
- i. **Firms using banks to finance investment (% of firms):** Porcentaje de empresas que utilizan bancos para financiar inversiones (*Glossary | DataBank, 2025*).

3. **Factores sociales y de sostenibilidad:** 9 indicadores extraídos de las bases de datos *World Development Indicators* y *ESG Data*, que permiten analizar el efecto de la corrupción sobre el bienestar, los servicios básicos, la inclusión y la equidad social.

- a. **Urban population (% of total population):** La población urbana se refiere a las personas que viven en zonas urbanas, según la definición de las oficinas nacionales de estadística. Los datos son recopilados y procesados por la División de Población de las Naciones Unidas (*Glossary | DataBank, 2025*).
- b. **Unemployment (% of total of labor force):** El desempleo se refiere a la proporción de la fuerza laboral que no tiene trabajo pero que está disponible y busca empleo (*Glossary | DataBank, 2025*).
- c. **Health expenditure (% of GDP):** Gasto público y privado en salud como proporción del PIB, indicador de inversión social (*Glossary | DataBank, 2025*).
- d. **Poverty headcount ratio (\$1.90/day):** Porcentaje de la población que vive con menos de \$1.90 al día, umbral de pobreza extrema según el Banco Mundial (*Glossary | DataBank, 2025*).
- e. **Gini Index:** El índice de Gini mide el grado en que la distribución del ingreso (o, en algunos casos, el gasto de consumo) entre individuos u hogares dentro de una economía se desvía de una distribución

perfectamente equitativa. Una curva de Lorenz representa los porcentajes acumulados del ingreso total recibido en relación con el número acumulado de beneficiarios, comenzando por el individuo u hogar más pobre. El índice de Gini mide el área entre la curva de Lorenz y una línea hipotética de igualdad absoluta, expresada como porcentaje del área máxima bajo dicha línea. Por lo tanto, un índice de Gini de 0 representa igualdad perfecta, mientras que un índice de 100 implica desigualdad perfecta (*Glossary | DataBank, 2025*).

- f. **Female labor force participation (% of female population ages 15+):** Porcentaje de mujeres activas en el mercado laboral muestra el grado de participación de las mujeres en la fuerza laboral. La fuerza laboral está compuesta por personas de 15 años o más que aportan mano de obra para la producción de bienes y servicios durante un período específico. (*Glossary | DataBank, 2025*).
- g. **Literacy Rate (% of adults):** La tasa de alfabetización de adultos es el porcentaje de personas de 15 años o más que pueden leer y escribir y comprender una declaración breve y sencilla sobre su vida cotidiana (*Glossary | DataBank, 2025*).
- h. **Primary School Enrollment (% gross):** La tasa bruta de matriculación es la relación entre la matrícula total, independientemente de la edad, y la población del grupo de edad que oficialmente corresponde al nivel de educación indicado. La educación primaria proporciona a los niños habilidades básicas de lectura, escritura y matemáticas, además de una comprensión elemental de materias como historia, geografía, ciencias naturales, ciencias sociales, arte y música (*Glossary | DataBank, 2025*).
- i. **Access to electricity (% of population):** El acceso a la electricidad se refiere al porcentaje de la población con acceso a ella. Los datos de electrificación se recopilan de la industria, encuestas nacionales y fuentes internacionales (*Glossary | DataBank, 2025*).

- 4. **Factores de gobernanza:** 6 variables extraídas de *Worldwide Governance Indicators*, que permiten evaluar la calidad institucional de cada país, identificando entornos más o menos propensos a la corrupción.

- a. **Control of Corruption:** Indicador compuesto que refleja la capacidad de un país para controlar la corrupción. Evalúa la frecuencia de prácticas como el soborno, el desvío de fondos públicos, el uso indebido del poder para beneficios privados, y la captura del Estado por élites e intereses particulares (*Glossary | DataBank, 2025*).
- b. **Rule of Law:** Mide la confianza en las reglas de la sociedad y el cumplimiento de las mismas. Incluye indicadores sobre la calidad del sistema judicial, la efectividad del sistema policial, el respeto a los derechos de propiedad y el grado de cumplimiento de contratos. Un débil estado de derecho facilita la impunidad de prácticas corruptas (*Glossary | DataBank, 2025*).
- c. **Political Stability:** Indica la percepción de la probabilidad de que el gobierno sea desestabilizado o derrocado por medios inconstitucionales o violentos, incluidos el terrorismo y los conflictos sociales. Un entorno políticamente estable tiende a favorecer instituciones más fuertes y menos propensas a la corrupción (*Glossary | DataBank, 2025*).
- d. **Regulatory Quality:** Evalúa la capacidad del gobierno para formular e implementar políticas y regulaciones sólidas que permitan y promuevan el desarrollo del sector privado. La baja calidad regulatoria puede aumentar las oportunidades de corrupción, ya que permite discrecionalidad y favorece el soborno en procesos administrativos (*Glossary | DataBank, 2025*).
- e. **Voice and Accountability:** Refleja la percepción sobre el grado en que los ciudadanos pueden participar en la selección de su gobierno, así como la libertad de expresión, de asociación y de prensa (*Glossary | DataBank, 2025*).
- f. **Government Effectiveness:** Evalúa la calidad de los servicios públicos y la formulación e implementación de políticas públicas, permitiendo asociar la corrupción con el desempeño gubernamental (*Glossary | DataBank, 2025*).

4.3. Justificación de las variables seleccionadas

La selección de las **26 variables clave** ha seguido un **enfoque inductivo orientado al objetivo central del trabajo**: identificar las dinámicas de corrupción y soborno en relación con factores económicos, sociales, institucionales y ambientales en tres países latinoamericanos con niveles de corrupción contrastantes. Para ello, se han integrado indicadores que permiten un análisis holístico y comparativo, basado en evidencia cuantitativa.

El **Índice de Percepción de la Corrupción (CPI)** (Transparency International, 2024) asigna una puntuación de 0 (altamente corrupto) a 100 (muy transparente). La clasificación de los tres países analizados es la siguiente:

- **Venezuela** con una puntuación de **13/100**, ubicándola entre los países más corruptos del mundo. Esta calificación no solo refleja una alta incidencia de sobornos, sino también una institucionalidad debilitada y un entorno empresarial hostil.
- **El Salvador** con una puntuación de **31/100**, se sitúa en una posición intermedia. Es decir, presenta niveles significativos de corrupción, aunque menores que los de Venezuela. No obstante, bajo el liderazgo del presidente Nayib Bukele, se han implementado diversas reformas institucionales y tecnológicas que apuntan a una mejora en la eficiencia gubernamental y la lucha contra la corrupción.
- **Uruguay** con una puntuación de **73/100**, se posiciona como el país más transparente de América Latina, siendo un modelo regional en términos de gobernanza y control de la corrupción.

A partir de esta disparidad estructural entre países, se han incluido en el análisis **cuatro grupos de variables** que permiten abordar la corrupción de forma integral:

- **Variables de corrupción y soborno**, como la incidencia y profundidad del soborno, que capturan directamente la experiencia de las empresas frente a prácticas corruptas en sus interacciones con el sector público.
- **Factores económicos**, que permiten examinar cómo el entorno económico puede estar influido por la corrupción o, a su vez, generar condiciones que favorezcan su aparición.

- **Factores sociales y de sostenibilidad** que permiten explorar las consecuencias sociales de la corrupción y su impacto sobre el bienestar general y la equidad.
- **Factores de gobernanza** que permiten evaluar la solidez del estado frente a prácticas corruptas.

Este enfoque, basado en la vinculación entre conceptos clave y sus correspondientes variables cuantitativas, proporciona una base sólida para aplicar técnicas estadísticas y modelos de Machine Learning, todo ello orientado a **identificar patrones, correlaciones y posibles relaciones causales** entre las prácticas corruptas y el desarrollo económico, social e institucional de los países seleccionados.

Además, con el fin de consolidar la información y facilitar el posterior análisis estadístico, en el capítulo 5. Limpieza y análisis exploratorio de datos se explica cómo todas las variables seleccionadas han sido integradas, y estructuradas en un nuevo archivo Excel unificado. Este archivo final permite disponer de una base de datos coherente y lista para aplicar métodos analíticos.

5. Limpieza y análisis exploratorio de datos

Este capítulo tiene como objetivo garantizar la utilidad del conjunto de datos sobre el que se desarrollará el análisis comparativo entre Venezuela, El Salvador y Uruguay. Para ello, se lleva a cabo un proceso estructurado que incluye la limpieza de los datos recopilados en el capítulo 4. Obtención de datos, el tratamiento de valores nulos, y un análisis exploratorio detallado. Este último incluye tanto estadísticas descriptivas como análisis gráficos, con el fin de conocer el comportamiento de las variables seleccionadas, detectar patrones, asimetrías y posibles relaciones entre los diferentes bloques conceptuales (corrupción, gobernanza, desarrollo económico y bienestar social).

5.1. Descripción general

La base de datos cuenta con 192 filas y 28 columnas. En la Figura 5 podemos ver los nombres de las columnas de la base de datos, que se corresponden con las 26 variables seleccionadas para el estudio, junto con dos columnas adicionales: el nombre del país (Venezuela, El Salvador y Uruguay) y el año (1960-2023).

```
Index(['Country Name', 'Year', 'Bribery Incidence', 'Bribery Depth',
      'GDP (US$)', 'FDI (US$)', 'Domestic credit to private sector (% GDP)',
      'Gross domestic savings (% of GDP)', 'Gross capital formation (% GDP)',
      'Inflation, consumer prices (annual %)',
      'Exports of goods and services (% of GDP)',
      'Firms using banks to finance working capital (% of firms)',
      'Firms using banks to finance investment (% of firms)',
      'Urban population (% of total population)',
      'Unemployment (% of total of labor force)',
      'Health expenditure (% GDP)', 'Poverty headcount ratio ($1.90/day)',
      'Gini Index',
      'Female labor force participation (% of female population ages 15+)',
      'Literacy rate (% of adults)', 'Primary School Enrollment (% gross)',
      'Access to electricity (% population)', 'Control of Corruption',
      'Rule of Law', 'Political Stability', 'Regulatory Quality',
      'Voice and Accountability', 'Government Effectiveness'],
      dtype='object')
```

Figura 5: Nombres de las columnas de la base de datos.

Con el fin de facilitar el análisis y permitir una comparación clara entre países, se han generado tres subconjuntos de datos, uno por cada país: El Salvador, Uruguay y Venezuela. De esta forma, se trabaja con tres dataframes cada uno con 64 filas u observaciones (una por años) y 28 columnas o variables.

A continuación, se procede a examinar los tipos de datos y la presencia de valores nulos en la base de datos. En la Figura 6 se observa un resumen de los tipos de variables, junto con el número de valores nulos y su porcentaje para cada país.

Por ejemplo, se observa que las variables relacionadas con sobornos – *Bribery Incidence* y *Bribery Depth*— presentan un elevado porcentaje de valores nulos en los tres países, con tasas superiores al 90%. También destacan otras variables con altos niveles de ausencias, como *Children in employment*, *Health expenditure* o *Literacy rate*. En contraste, algunas variables como *GDP (US\$)*, *Gross capital formation* o *Urban population* tienen una cobertura mucho más completa. Con este análisis se refuerza la necesidad de aplicar técnicas avanzadas de imputación de valores nulos.

	Tipo de Dato	Nulos El Salvador	% El Salvador	Nulos Uruguay	% Uruguay	Nulos Venezuela	% Venezuela
Country Name	object	0	0.00	0	0.00	0	0.00
Year	int64	0	0.00	0	0.00	0	0.00
Bribery Incidence	float64	60	93.75	61	95.31	62	96.88
Bribery Depth	float64	61	95.31	61	95.31	62	96.88
GDP (US\$)	float64	5	7.81	0	0.00	9	14.06
FDI (US\$)	float64	12	18.75	10	15.62	10	15.62
Domestic credit to private sector (% GDP)	float64	41	64.06	0	0.00	10	15.62
Gross domestic savings (% of GDP)	float64	5	7.81	0	0.00	9	14.06
Gross capital formation (% GDP)	float64	5	7.81	0	0.00	9	14.06
Inflation, consumer prices (annual %)	float64	0	0.00	0	0.00	56	87.50
Exports of goods and services (% of GDP)	float64	5	7.81	0	0.00	9	14.06
Firms using banks to finance working capital (% of firms)	float64	60	93.75	61	95.31	62	96.88
Firms using banks to finance investment (% of firms)	float64	60	93.75	61	95.31	62	96.88
Urban population (% of total population)	float64	0	0.00	0	0.00	0	0.00
Unemployment (% of total of labor force)	float64	31	48.44	31	48.44	31	48.44
Health expenditure (% GDP)	float64	42	65.62	42	65.62	60	93.75
Poverty headcount ratio (\$1.90/day)	float64	36	56.25	34	53.12	51	79.69
Gini Index	float64	36	56.25	34	53.12	51	79.69
Female labor force participation (% of female population ages 15+)	float64	30	46.88	30	46.88	30	46.88
Literacy rate (% of adults)	float64	49	76.56	46	71.88	55	85.94
Primary School Enrollment (% gross)	float64	14	21.88	11	17.19	18	28.12
Access to electricity (% population)	float64	32	50.00	33	51.56	33	51.56
Control of Corruption	float64	39	60.94	39	60.94	39	60.94
Rule of Law	float64	39	60.94	39	60.94	39	60.94
Political Stability	float64	39	60.94	39	60.94	39	60.94
Regulatory Quality	float64	39	60.94	39	60.94	39	60.94
Voice and Accountability	float64	39	60.94	39	60.94	39	60.94
Government Effectiveness	float64	39	60.94	39	60.94	39	60.94

Figura 6: Tipos de datos y valores nulos para cada país.

5.2. Limpieza inicial de valores nulos

Para abordar la problemática de los valores nulos, el primer paso consistió en analizar la distribución de los mismos a lo largo del tiempo. En la Figura 7 se muestra el porcentaje de valores nulos por año. Se observa que en la primera década (1960-1969) el porcentaje de valores nulos es superior al 70%. Dado este elevado nivel de incompletitud, se optó por eliminar los años anteriores a 1970, con el objetivo de mejorar la calidad del análisis.

Year	Valores Nulos	Total Celdas	Porcentaje Nulos								
1960	63	78	80.77	1981	48	78	61.54	2003	17	78	21.79
1961	63	78	80.77	1982	53	78	67.95	2004	17	78	21.79
1962	63	78	80.77	1983	53	78	67.95	2005	17	78	21.79
1963	63	78	80.77	1984	53	78	67.95	2006	3	78	3.85
1964	63	78	80.77	1985	53	78	67.95	2007	16	78	20.51
1965	59	78	75.64	1986	54	78	69.23	2008	17	78	21.79
1966	59	78	75.64	1987	51	78	65.38	2009	15	78	19.23
1967	59	78	75.64	1988	53	78	67.95	2010	4	78	5.13
1968	59	78	75.64	1989	47	78	60.26	2011	15	78	19.23
1969	59	78	75.64	1990	50	78	64.10	2012	17	78	21.79
1970	54	78	69.23	1991	44	78	56.41	2013	16	78	20.51
1971	53	78	67.95	1992	39	78	50.00	2014	18	78	23.08
1972	53	78	67.95	1993	44	78	56.41	2015	20	78	25.64
1973	53	78	67.95	1994	47	78	60.26	2016	16	78	20.51
1974	53	78	67.95	1995	39	78	50.00	2017	18	78	23.08
1975	52	78	66.67	1996	22	78	28.21	2018	21	78	26.92
1976	53	78	67.95	1997	42	78	53.85	2019	22	78	28.21
1977	53	78	67.95	1998	21	78	26.92	2020	25	78	32.05
1978	53	78	67.95	1999	40	78	51.28	2021	24	78	30.77
1979	53	78	67.95	2000	20	78	25.64	2022	25	78	32.05
1980	53	78	67.95	2001	34	78	43.59	2023	30	78	38.46
				2002	17	78	21.79				

Figura 7: Porcentaje de valores nulos por año.

Una vez eliminados los años con mayor proporción de datos faltantes, se actualizó el análisis de valores nulos para el nuevo rango temporal comprendido entre 1970 y 2023. En la Figura 8 se presenta de nuevo el número y porcentaje de valores nulos por variable y país tras esta limpieza. Como se puede observar, esta limpieza ha reducido los niveles de incompletitud en múltiples variables clave.

Por ejemplo, en El Salvador, la variable *FDI (US\$)* pasó de un 18.75% de valores nulos a tan solo un 3.70%, y en Uruguay la variable *Primary School Enrollment (% gross)* descendió de un 17.19% a 1.85%. Asimismo, las 6 variables de gobernanza disminuyeron su proporción de valores ausentes del 60.94% al 53.70% tanto en El Salvador como en Uruguay y Venezuela.

Tras este proceso de depuración, se pasa a trabajar ahora con **54 observaciones por cada país**.

	Nulos El Salvador	% El Salvador	Nulos Uruguay	% Uruguay	Nulos Venezuela	% Venezuela
Country Name	0	0.00	0	0.00	0	0.00
Year	0	0.00	0	0.00	0	0.00
Bribery Incidence	50	92.59	51	94.44	52	96.30
Bribery Depth	51	94.44	51	94.44	52	96.30
GDP (US\$)	0	0.00	0	0.00	9	16.67
FDI (US\$)	2	3.70	0	0.00	0	0.00
Domestic credit to private sector (% GDP)	31	57.41	0	0.00	10	18.52
Gross domestic savings (% of GDP)	0	0.00	0	0.00	9	16.67
Gross capital formation (% GDP)	0	0.00	0	0.00	9	16.67
Inflation, consumer prices (annual %)	0	0.00	0	0.00	46	85.19
Exports of goods and services (% of GDP)	0	0.00	0	0.00	9	16.67
Firms using banks to finance working capital (% of firms)	50	92.59	51	94.44	52	96.30
Firms using banks to finance investment (% of firms)	50	92.59	51	94.44	52	96.30
Urban population (% of total population)	0	0.00	0	0.00	0	0.00
Unemployment (% of total of labor force)	21	38.89	21	38.89	21	38.89
Health expenditure (% GDP)	32	59.26	32	59.26	50	92.59
Poverty headcount ratio (\$1.90/day)	26	48.15	24	44.44	41	75.93
Gini Index	26	48.15	24	44.44	41	75.93
Female labor force participation (% of female population ages 15+)	20	37.04	20	37.04	20	37.04
Literacy rate (% of adults)	39	72.22	36	66.67	45	83.33
Primary School Enrollment (% gross)	4	7.41	1	1.85	8	14.81
Access to electricity (% population)	22	40.74	23	42.59	23	42.59
Control of Corruption	29	53.70	29	53.70	29	53.70
Rule of Law	29	53.70	29	53.70	29	53.70
Political Stability	29	53.70	29	53.70	29	53.70
Regulatory Quality	29	53.70	29	53.70	29	53.70
Voice and Accountability	29	53.70	29	53.70	29	53.70
Government Effectiveness	29	53.70	29	53.70	29	53.70

Figura 8: Valores nulos y porcentaje de valores nulos para cada país tras la eliminación de la primera década (1960-1969).

5.3. Imputación de valores nulos mediante KNNImputer

Una vez depurado el dataset inicial y eliminados los años con altos niveles de incompletitud (1960–1969), se procedió a la imputación de los valores nulos restantes para completar el conjunto de datos. Para ello, se utilizó el **algoritmo KNNImputer**, incluido en la biblioteca scikit-learn, el cual se basa en el método de los k vecinos más cercanos (K-Nearest Neighbors) (*KNNImputer*, 2025).

Esta técnica permite estimar los valores faltantes de una variable numérica a partir de los registros más similares en el conjunto de datos. La imputación se realiza mediante el valor medio de los k vecinos más cercanos, lo que permite preservar la coherencia multivariable del conjunto de datos (es decir, las relaciones conjuntas entre diferentes indicadores socioeconómicos).

Con el fin de identificar el valor óptimo de k para cada país, se ha llevado a cabo un procedimiento de **búsqueda de hiperparámetros**. Este consistió en eliminar aleatoriamente un 5% de los valores no nulos (de forma controlada), aplicar KNNImputer

con diferentes valores de k (de 1 a 10), y calcular el error cuadrático medio (RMSE) entre los valores reales y los imputados. Este procedimiento se realizó de forma independiente para cada país, evitando que las imputaciones se vean influenciadas por valores de países con contextos diferentes, lo que podría sesgar los resultados.

A partir de las gráficas obtenidas, se seleccionaron los siguientes valores de k para cada país:

- Venezuela: $k = 6$. En la Figura 9 podemos ver que $k = 6$ corresponde al punto de menor error antes de que el RMSE aumente de forma constante.
- El Salvador: $k = 8$. En la Figura 10 se observa una disminución marginal del RMSE en $k = 8$, indicando una tendencia a la estabilización. La elección de este valor evita recurrir al número máximo de vecinos, contribuyendo a una mayor eficiencia computacional y reduciendo el sobreajuste.
- Uruguay: $k = 5$. En la Figura 11 vemos que se alcanza el primer mínimo en $k = 5$, valor elegido para preservar un equilibrio entre precisión y generalización, ya que partir de ese valor el RMSE oscila de manera inestable, indicando que no se obtienen mejores significativas con más vecinos.

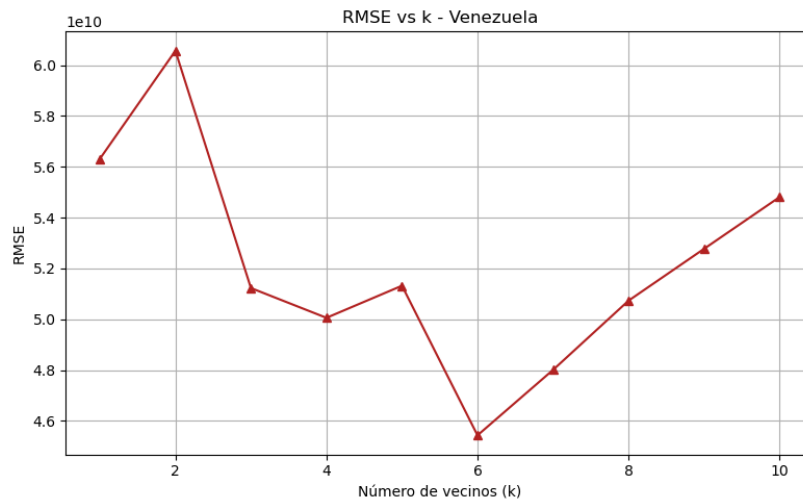


Figura 9: Evolución del error cuadrático medio (RMSE) según el número de vecinos (k) para Venezuela en la imputación con KNNImputer.

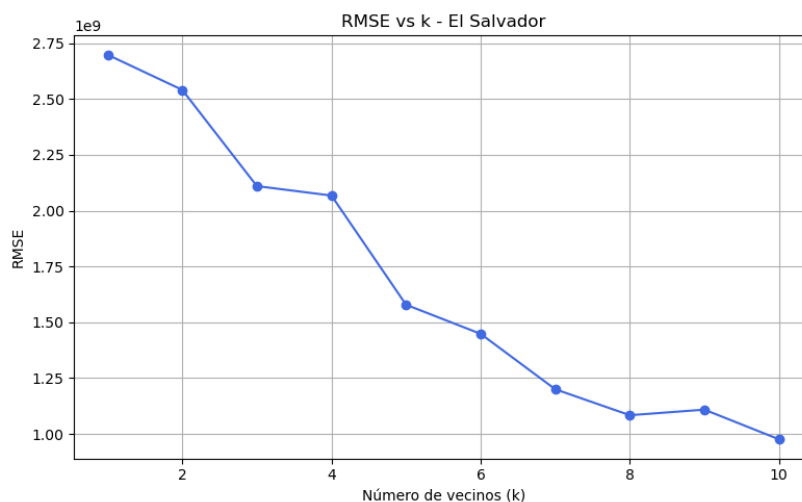


Figura 10: Evolución del error cuadrático medio (RMSE) según el número de vecinos (k) para El Salvador en la imputación con KNNImputer.

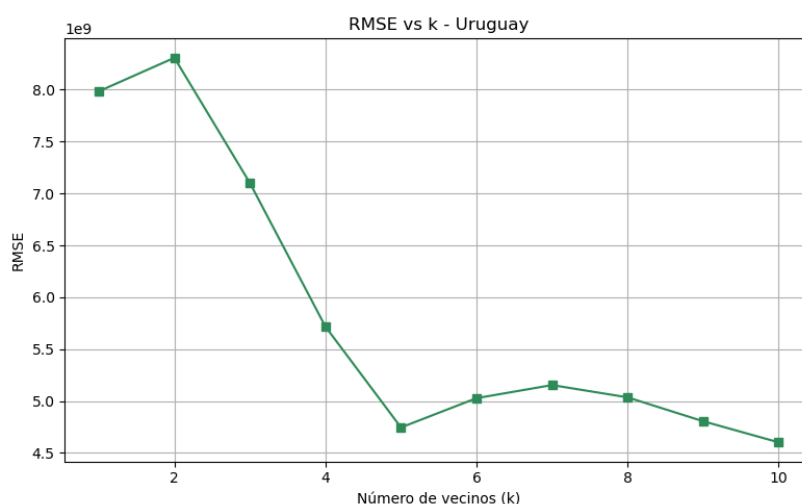


Figura 11: Evolución del error cuadrático medio (RMSE) según el número de vecinos (k) para Uruguay en la imputación con KNNImputer.

Con los valores óptimos del hiperparámetro k , se aplicó la imputación final para cada país utilizando KNNImputer de forma separada. Como resultado, se obtuvieron tres conjuntos de datos, uno por país, sin valores nulos.

Esta limpieza ha permitido **preservar la mayor cantidad posible de información**, garantizando al mismo tiempo la calidad y completitud del conjunto de datos final. Finalmente, los datos imputados se exportaron en archivos Excel independientes para cada país con el fin de facilitar su posterior análisis.

5.4. Estadísticas descriptivas

Con el objetivo de caracterizar el comportamiento de las variables clave en cada uno de los países analizados (Venezuela, El Salvador y Uruguay), se ha elaborado una tabla de estadísticas descriptivas a partir del conjunto de datos limpio tras la imputación de valores faltantes mediante el algoritmo KNNImputer. Para cada país se han calculado la media, mediana, desviación estándar, así como los valores mínimo y máximo de todas las variables disponibles.

Los resultados permiten identificar diferencias estructurales relevantes entre Venezuela, El Salvador y Uruguay en términos de corrupción, gobernanza y desempeño económico. A continuación, se destacan algunas de las principales diferencias observadas.

En la Tabla 1 se recogen los estadísticos relativos a las variables de **corrupción** (*Bribery Incidence* y *Bribery Depth*), donde se aprecia:

- **Venezuela** muestra una media de **7.05** en *Bribery Incidence* (con un máximo de **10.30**), la más alta entre los tres países. Esto refleja una mayor frecuencia de sobornos en las operaciones empresariales. Asimismo, la variable *Bribery Depth* alcanza una media de **4.30**, más del triple que Uruguay (**1.35**).
- **Uruguay**, por el contrario, se posiciona como el país con menor incidencia de soborno, con una media de **2.23**, y una desviación estándar especialmente baja (**0.048**). Este dato sugiere un entorno institucional sólido y condiciones más estables y transparentes para la actividad empresarial.
- **El Salvador** se sitúa en una posición intermedia, con una media de **4.38** en *Bribery Incidence* y **1.72** en *Bribery Depth*, reflejando una situación menos crítica que la venezolana, pero aún con margen de mejora frente al modelo uruguayo.

	Bribery Incidence					Bribery Depth				
	min	max	mean	median	std	min	max	mean	median	std
El Salvador	3,1000	4,7000	4,3843	4,5000	0,2773	1,5000	2,3000	1,7176	1,6500	0,1353
Uruguay	2,2000	2,4000	2,2259	2,2000	0,0483	1,2000	2,0000	1,3528	1,2500	0,1902
Venezuela, RB	3,8000	10,3000	7,0500	7,0500	0,6313	3,1000	5,5000	4,3000	4,3000	0,2331

Tabla 1: Estadísticas descriptivas de variables de soborno y corrupción.

La Tabla 2 muestra las estadísticas descriptivas de los indicadores de gobernanza, revelando las siguientes diferencias:

- **Uruguay** se posiciona consistentemente como el país con mejores valores medios en todos los indicadores institucionales. Por ejemplo, su media en *Control of Corruption* es de **1.13**, mientras que en *Voice and Accountability* alcanza **1.05**, y en *Government Effectiveness* **0.52**. Estas cifras reflejan un entorno institucional sólido, con una administración eficaz, marcos regulatorios funcionales y altos niveles de participación democrática.
- **Venezuela**, en contraste, presenta los peores registros institucionales. Su media en *Rule of Law* es de **-1.96**, y en *Control of Corruption*, de **-1.39**, evidenciando una débil capacidad estatal para hacer cumplir normas, baja confianza en las instituciones y una administración pública ineficaz (*Government Effectiveness*: **-1.37**). La alta desviación estándar en varias de estas variables también sugiere una gran inestabilidad institucional a lo largo del tiempo.
- **El Salvador** se encuentra en una posición intermedia, aunque sus valores tienden a estar más cerca de los de Venezuela que de Uruguay. Sus indicadores de gobernanza se sitúan en pequeños valores negativos, por ejemplo, su media en *Political Stability* es de **-0.04**, mientras que en *Regulatory Quality* y *Control of Corruption* presenta **-0.04** y **-0.70**, respectivamente, reflejando un entorno normativo aún en desarrollo y desafíos en la lucha contra la corrupción.

	Control of Corruption					Rule of Law					Political Stability				
	min	max	mean	median	std	min	max	mean	median	std	min	max	mean	median	std
El Salvador	-0,8907	-0,2302	-0,7049	-0,8780	0,2194	-0,9303	-0,3694	-0,7425	-0,7805	0,1076	-0,3892	0,3292	-0,0407	-0,0623	0,1332
Uruguay	0,8941	1,6105	1,1340	1,0497	0,1715	0,4564	0,7868	0,6378	0,6654	0,0711	0,5910	1,0858	0,8293	0,7748	0,1133
Venezuela, RB	-1,6900	-0,6427	-1,3948	-1,5378	0,2362	-2,3319	-0,7065	-1,9621	-2,2054	0,4280	-1,5518	-0,5377	-1,1938	-1,2206	0,1889

	Regulatory Quality					Voice and Accountability					Government Effectiveness				
	min	max	mean	median	std	min	max	mean	median	std	min	max	mean	median	std
El Salvador	-0,4627	0,4039	-0,0365	-0,0573	0,1541	-0,4744	0,1886	-0,0057	-0,0298	0,1069	-0,6915	0,0504	-0,4963	-0,6517	0,2182
Uruguay	0,3000	0,9065	0,5211	0,4340	0,1350	0,8298	1,3068	1,0492	1,0234	0,1015	0,3032	0,8502	0,5244	0,5228	0,1022
Venezuela, RB	-2,3867	-0,3087	-1,7809	-2,0055	0,5169	-1,6177	0,1150	-1,1442	-1,3746	0,4120	-1,8894	-0,5407	-1,3705	-1,5134	0,2996

Tabla 2: Estadísticas descriptivas de variables de gobernanza.

En la Tabla 3 se presentan las estadísticas descriptivas de cuatro indicadores económicos: *Domestic Credit to Private Sector (% of GDP)*, *Gross Domestic Savings (% of GDP)*, *Gross Capital Formation (% of GDP)* e *Inflation (consumer prices, annual %)*.

A partir de los datos se pueden destacar las siguientes observaciones relevantes:

- **El Salvador** es el país con mayor acceso al crédito interno en términos relativos, con una media del **49.09%** del PIB destinado al sector privado, muy por encima

de Uruguay (**31.36%**) y Venezuela (**19.14%**), lo que sugiere una mayor capacidad de financiamiento empresarial.

- En cuanto al ahorro interno (*Gross Domestic Savings*), **Venezuela** presenta el valor medio más alto (**28.75%** del PIB), pero con una desviación estándar significativa (**7.11**), lo que refleja alta inestabilidad macroeconómica. **El Salvador**, por el contrario, muestra la tasa de ahorro más baja (**4.56%**), lo que puede comprometer la capacidad de inversión sin recurrir al endeudamiento externo o a la inversión extranjera.
- La formación bruta de capital (*Gross Capital Formation*) también es mayor en Venezuela (**25.77%**), aunque nuevamente con alta volatilidad. **Uruguay** y **El Salvador** presentan medias similares (**17.27%** y **17.29%**, respectivamente), con menor dispersión, reflejando trayectorias de inversión más estables.
- El indicador de *Inflación* muestra grandes contrastes entre los países: **Venezuela** presenta una situación crítica, con una media anual del **156.21%** y un máximo de más del **250%**, lo que confirma su estado de hiperinflación crónica. En contraste, **El Salvador** muestra una inflación controlada (**8.27%**), mientras que **Uruguay** mantiene un nivel medio (**34.36%**), con menor dispersión que Venezuela, reforzando su imagen de estabilidad macroeconómica.

	Domestic credit to private sector (% GDP)					Gross domestic savings (% of GDP)				
	min	max	mean	median	std	min	max	mean	median	std
El Salvador	45,5605	65,3332	49,0857	45,7367	5,2651	-7,0121	24,1076	4,5612	2,5185	7,0426
Uruguay	12,1736	72,3823	31,3623	26,1720	13,0560	2,0755	35,0206	17,8581	18,0209	4,3858
Venezuela, RB	8,1266	33,4597	19,1420	18,3183	7,3955	7,5924	42,4046	28,7575	29,6051	7,1189

	Gross capital formation (% GDP)					Inflation, consumer prices (annual %)				
	min	max	mean	median	std	min	max	mean	median	std
El Salvador	10,8458	26,0745	17,2973	17,4853	3,3275	-0,7314	31,9350	8,2692	4,9099	7,8105
Uruguay	3,3962	34,1083	17,2741	17,2061	4,3904	4,3593	112,5259	34,3645	19,1861	31,1561
Venezuela, RB	9,8897	51,8588	25,7708	24,7853	8,8450	21,0690	254,9485	156,2080	188,3433	62,0379

Tabla 3: Estadísticas descriptivas de factores económicos.

En conjunto, los resultados del análisis descriptivo han permitido establecer una primera comparación entre los tres países:

- **Uruguay** se presenta como el país con el entorno más favorable, caracterizado por bajos niveles de soborno, instituciones robustas y una economía relativamente estable.

- **Venezuela**, en cambio, presenta los mayores niveles de corrupción, debilidad institucional e indicadores económicos marcados por la volatilidad y la hiperinflación.
- **El Salvador** ocupa una posición intermedia, con algunos indicadores positivos, como el acceso al crédito privado, pero también con carencias importantes en ahorro interno, calidad regulatoria y control de la corrupción.

Estas diferencias evidencian la existencia de **relaciones potenciales** entre los distintos bloques de variables.

5.5. Correlaciones entre variables

Con el objetivo de explorar las posibles relaciones lineales entre las variables clave, se ha calculado una **matriz de correlación de Pearson** entre todas las variables numéricas del conjunto de datos. Los resultados se visualizan en forma de **mapa de calor** (Figura 12), lo que permite identificar visualmente tanto asociaciones positivas como negativas, así como su intensidad relativa.

El resultado revela varios patrones consistentes con lo esperado desde un punto de vista teórico, contribuyendo a comprender cómo los factores de gobernanza, desarrollo económico y bienestar social se asocian con la incidencia del soborno y corrupción.

Variables institucionales: Las variables como *Control of Corruption*, *Rule of Law*, *Political Stability*, *Regulatory Quality*, *Voice and Accountability* y *Government Effectiveness* presentan correlaciones positivas fuertes entre sí, evidenciando la consistencia interna del bloque institucional. Asimismo, estas variables de gobernanza presentan correlaciones positivas con indicadores de desarrollo como: *FDI*, *Literacy rate*, *Primary school enrollment*, *Access to electricity* y *Health expenditure*. Este patrón sugiere que un entorno institucional sólido se asocia con mayores niveles de desarrollo económico y social.

Sobornos y gobernanza: Las variables *Bribery Incidence* y *Bribery Depth* presentan correlaciones fuertemente negativas con las 6 variables de gobernanza. Esto indica que los países con instituciones más robustas tienden a tener niveles más bajos de soborno y corrupción.

Corrupción y desarrollo social: La variable *Bribery Incidence* también se correlaciona negativamente con indicadores sociales como: *Access to electricity*, *Literacy rate*, *Primary school enrollment*, *Female labor force participation* y *Health expenditure*. Con ello, se sugiere que la corrupción representa un obstáculo para el progreso social.

Desarrollo económico: Las variables *GDP*, *FDI* y *Domestic credit to private sector* muestran correlaciones positivas con *Health expenditure*, *Primary school enrollment* y *Female labor force participation*, lo que refleja que un mayor nivel de actividad económica se traduce en desarrollo social.

Uso de servicios financieros: Las variables *Firms using banks to finance investment / working capital* presentan correlaciones negativas con los indicadores institucionales, lo que sugiere que en el conjunto de países analizados, un mayor uso del financiamiento bancario para inversión se da en contextos con menor calidad institucional, posiblemente como una estrategia alternativa ante la falta de otras fuentes de capital.

Pobreza y gobernanza: La variable *Poverty headcount ratio* presenta una correlación negativa con *Control of Corruption*, *GDP*, *Access to electricity* y *Primary school enrollment*. Esto refuerza la idea de que la pobreza está vinculada con un entorno institucional, y económico menos desarrollado. Por otro lado, la correlación de la tasa de pobreza con el índice de Gini es altamente positiva. Lo que reitera que, a mayor desigualdad en la distribución del ingreso, mayor es el porcentaje de población que queda excluida de los beneficios del desarrollo económico. Esta relación sugiere que no basta con que una economía crezca, sino que es crucial garantizar una distribución más equitativa para reducir los niveles de pobreza.

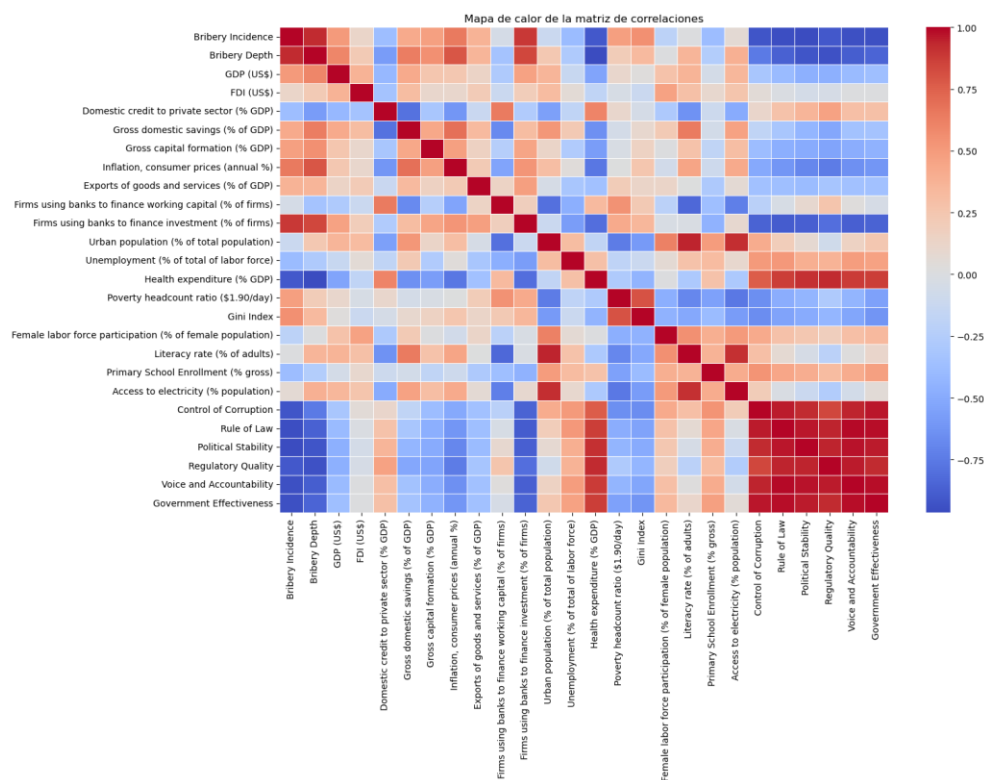


Figura 12: Mapa de calor de la matriz de correlaciones entre las variables clave.

En adición, en las figuras: Figura 13, Figura 14 y Figura 15 se muestran el mapa de calor de las matrices de correlaciones para cada país. Este análisis individual revela diferencias estructurales en la relación entre las variables institucionales, económicas y sociales.

En el caso de **Venezuela**. Se observa una débil relación entre la calidad institucional y la incidencia del soborno. Esto sugiere un entorno institucional deteriorado, donde las prácticas corruptas están más normalizadas y desconectadas del desempeño social y económico.

Por el contrario, **El Salvador** presenta un patrón de correlaciones más coherente: la incidencia de sobornos muestra relaciones negativas claras con los indicadores de gobernanza. Además, las variables institucionales también se asocian positivamente con variables de desarrollo como la educación, el acceso a servicios básicos o el uso del sistema financiero.

Finalmente, **Uruguay** destaca por mostrar el perfil más estructurado: las variables institucionales presentan correlaciones muy elevadas entre sí, y se asocian tanto con menores niveles de corrupción como con mejores indicadores sociales y económicos. Este

patrón sugiere una sólida coherencia institucional y una mayor capacidad del entorno gubernamental para promover el desarrollo sostenible y limitar las prácticas corruptas.

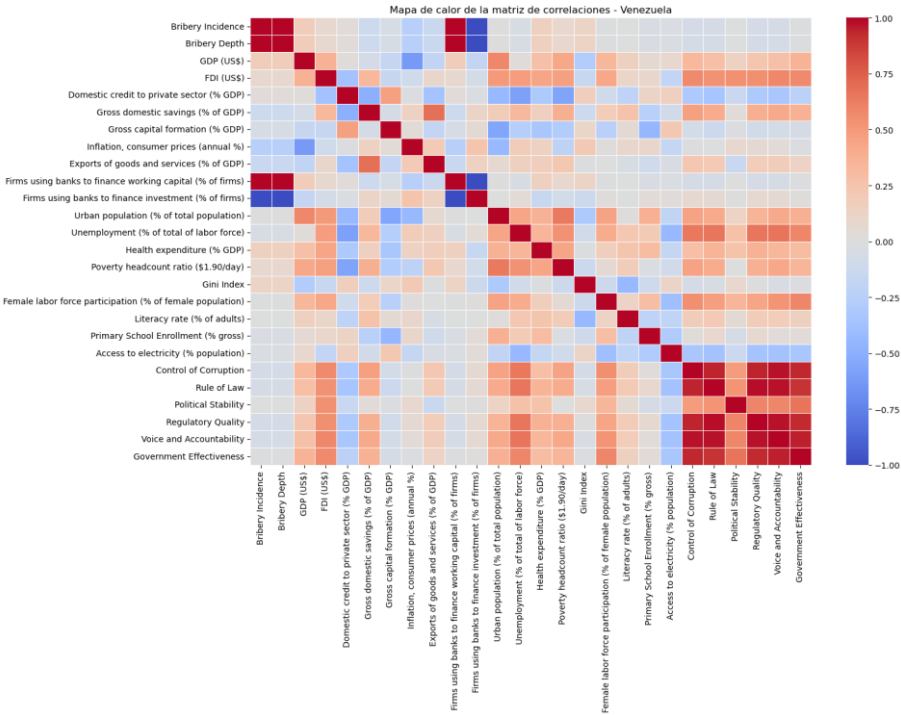


Figura 13: Mapa de calor de la matriz de correlaciones entre las variables clave para Venezuela.

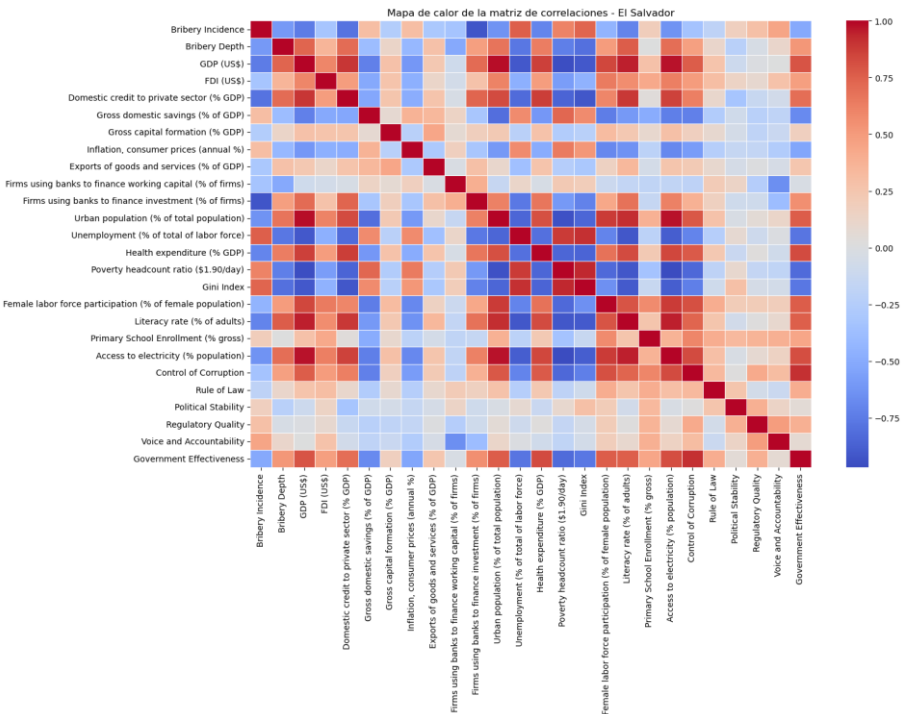


Figura 14: Mapa de calor de la matriz de correlaciones entre las variables clave para El Salvador.

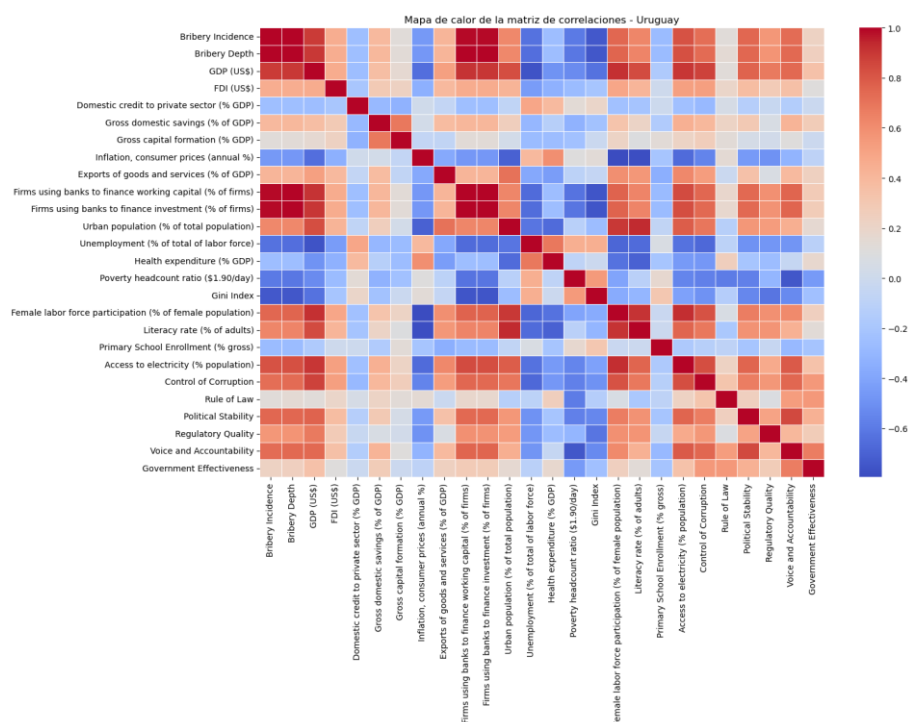


Figura 15: Mapa de calor de la matriz de correlaciones entre las variables clave para Uruguay.

Este análisis preliminar basado en correlaciones respalda la existencia de relaciones estructurales entre corrupción, instituciones, economía y bienestar social, proporcionando una base empírica sólida para el desarrollo de modelos explicativos en los siguientes apartados, en los que se explorarán estas relaciones con mayor profundidad mediante técnicas de aprendizaje automático.

5.6. Distribuciones

Con el objetivo de comprender mejor la naturaleza estadística de las variables empleadas en este estudio, se han generado histogramas con curvas de densidad para cada grupo de indicadores. Este análisis permite identificar la forma de las distribuciones, detectar posibles valores atípicos, asimetrías, y evaluar la adecuación de las variables para su inclusión en modelos de Machine Learning.

5.6.1. Variables de corrupción

En la Figura 16 se muestran los histogramas agregados de las variables de soborno. En el caso de *Bribery Incidence*, se observan picos diferenciados, lo que sugiere que algunos países presentan niveles bajos mientras que otros se agrupan en valores medios-altos. Esta

distribución multimodal refleja la heterogeneidad en la prevalencia del soborno entre los países considerados.

Por su parte, la variable *Bribery Depth* se concentra mayoritariamente en valores bajos (alrededor de 1), pero presenta una segunda acumulación significativa en torno a valores más elevados (alrededor de 4). Esta doble agrupación indica que existen contextos donde el soborno no solo es más frecuente, sino también más profundo.

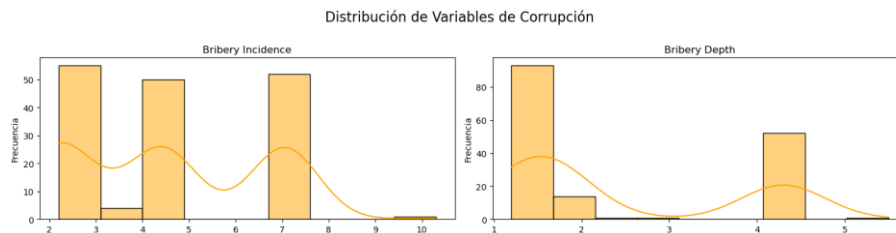


Figura 16: Histogramas de las variables de corrupción.

Adicionalmente, la Figura 17 permite desagregar estos patrones por país. En ella se aprecia claramente cómo **Uruguay** presenta valores muy bajos y concentrados para ambas variables, confirmando su reputación como país con baja incidencia de prácticas corruptas. **El Salvador** muestra valores intermedios, con una distribución de *Bribery Incidence* centrada en torno a 4, y de *Bribery Depth* centrada entre 1.5 y 2, lo que sugiere una experiencia moderada con prácticas de soborno. **Venezuela**, por el contrario, destaca por una distribución claramente desplazada hacia valores altos, lo que evidencia un entorno institucional más frágil y propenso a prácticas de corrupción sistemáticas.

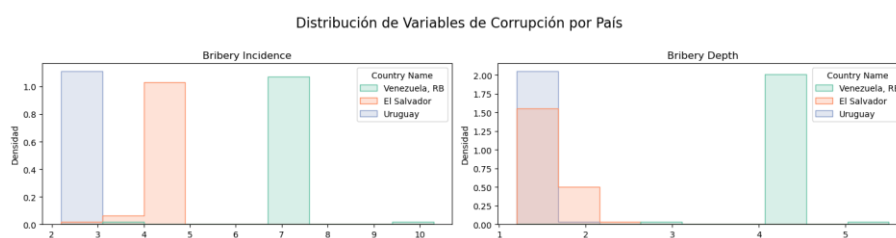


Figura 17: Histogramas de las variables de corrupción por país.

5.6.2. Variables económicas

En la Figura 18 se representan los histogramas de las variables económicas, que presentan distribuciones notablemente heterogéneas.

Las variables *GDP (US\$)* y *FDI (US\$)* presentan distribuciones **sesgadas a la derecha**, con una alta concentración de observaciones en valores bajos, y la presencia de valores extremos muy elevados. Esta asimetría resulta esperable al comparar países con niveles dispares de desarrollo.

En contraste, indicadores como *Gross capital formation (% GDP)* y *Gross domestic savings (% of GDP)* presentan **distribuciones más simétricas**, lo que sugiere una variabilidad más equilibrada entre las economías. De forma similar, la variable *Exports of goods and services (% of GDP)* muestra una distribución aproximadamente **normal**, lo que refleja cierta homogeneidad en el grado de apertura comercial entre los países analizados.

La variable *Inflation, consumer prices (annual %)* presenta una distribución con **cola alargada hacia la derecha**, indicando la presencia de situaciones de hiperinflación en economías inestables, como es el caso de Venezuela.

Por otro lado, los indicadores *Firms using banks to finance working capital (% of firms)* y *Firms using banks to finance investment (% of firms)* presentan distribuciones **bimodal** y **trimodal**, respectivamente. Esta forma sugiere la existencia de **múltiples perfiles diferenciados en el acceso al financiamiento empresarial**, lo cual puede estar reflejando diferencias estructurales en el desarrollo del sistema financiero o limitaciones en el acceso al crédito.

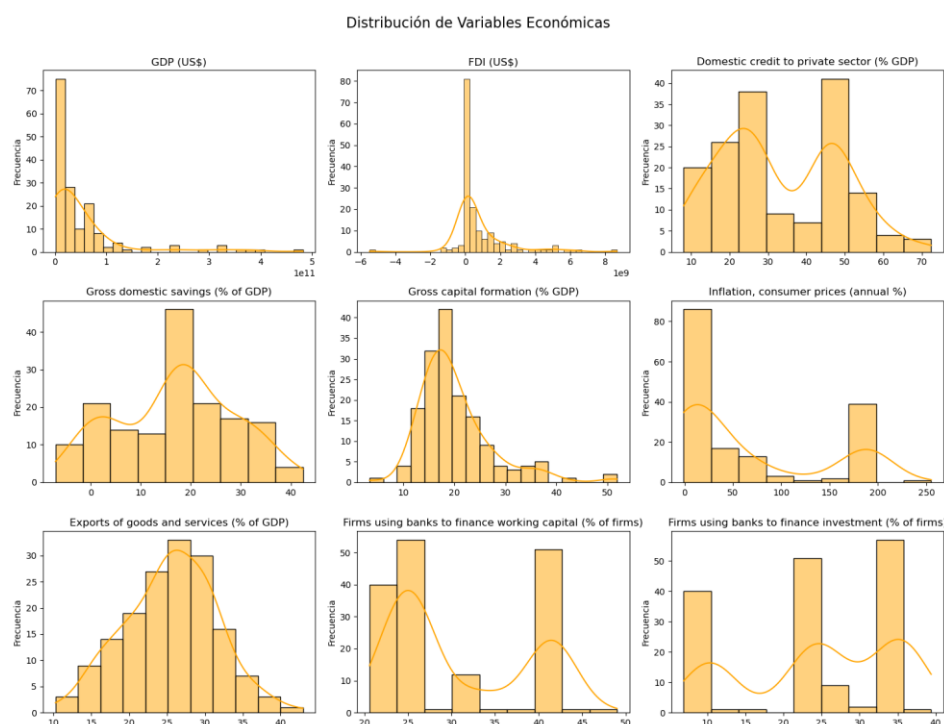


Figura 18: Histogramas de las variables económicas.

Adicionalmente, la Figura 19 complementa este análisis al mostrar las distribuciones desagregadas por país.

En el caso de *GDP (US\$)* y *FDI (US\$)*, se observa que **Uruguay** concentra los valores más elevados, reflejo de una mayor estabilidad macroeconómica y un entorno más favorable para la atracción de capital extranjero. Por otro lado, **Venezuela** domina los valores extremos en *Inflation, consumer prices (annual %)*, lo cual es consistente con los episodios de hiperinflación que experimenta.

En lo relativo al acceso al financiamiento, la **trimodalidad** observada en *Firms using banks to finance investment (% of firms)* puede explicarse por una segmentación estructural entre países con mayor desarrollo financiero (como Uruguay) y aquellos donde el acceso al crédito es más limitado (como Venezuela). Por su parte, **El Salvador** se posiciona en un punto intermedio.

Cabe mencionar que, en el caso de *Firms using banks to finance working capital (% of firms)*, la aparente **bimodalidad** observada resulta algo **atípica**, ya que **El Salvador** concentra la mayoría de los valores elevados, en torno al 40%, mientras que **Uruguay** y **Venezuela** se sitúan en rangos inferiores, cercanos al 22-27%.

Distribución de Variables Económicas por País

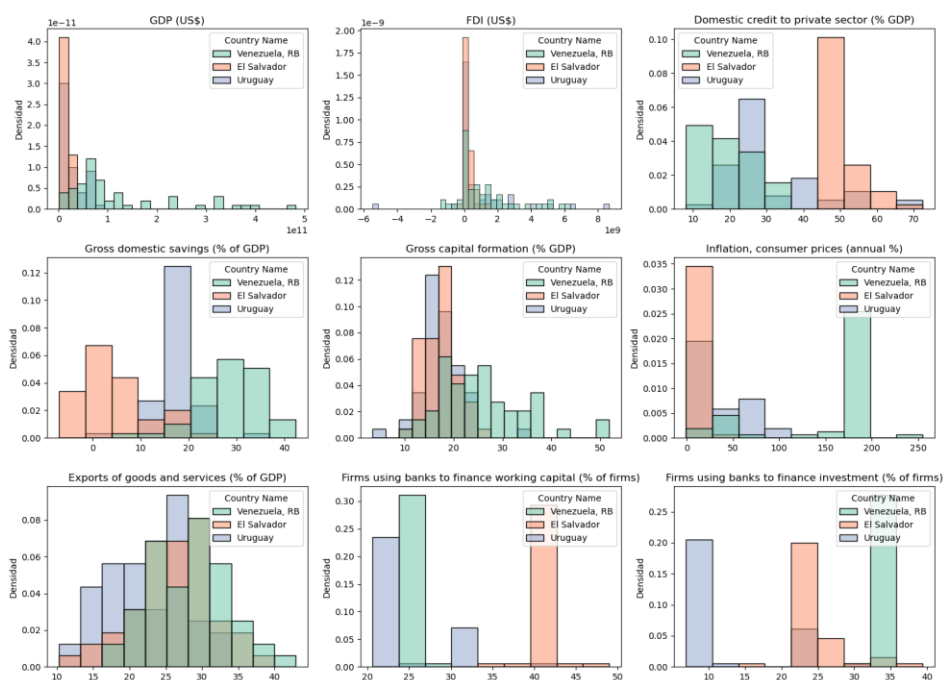


Figura 19: Histogramas de las variables económicas por país.

5.6.3. Variables sociales

En la Figura 20 podemos ver que los indicadores sociales presentan una alta dispersión, reflejando diferencias sustanciales en el bienestar y el desarrollo humano entre los países analizados:

La variable *Access to electricity (% population)* se concentra mayoritariamente en valores altos, pero muestra cierta variabilidad hacia la izquierda, lo que evidencia desniveles en el desarrollo de infraestructuras básicas entre países.

Tanto *Literacy rate (% of adults)* como *Primary School Enrollment (% gross)* muestran distribuciones **asimétricas hacia la derecha**, con muchas observaciones cercanas al 100%, pero también con valores más bajos que indican brechas educativas en algunas regiones.

El *Gini Index*, que mide la desigualdad en la distribución del ingreso, presenta una distribución **aproximadamente simétrica**, con la mayoría de los valores agrupados entre 40 y 55. Esta forma sugiere que, a lo largo del tiempo, los países analizados se han mantenido en torno a un **nivel moderado-alto de desigualdad** de forma relativamente estable. En contraste, la variable *Poverty headcount ratio (\$1.90/day)* muestra una

distribución **sesgada a la derecha**, lo que indica que en la mayoría de los años se registran **niveles bajos de pobreza extrema**.

Por último, las variables *Unemployment (% of total of labor force)* y *Female labor force participation (% of female population)* exhiben distribuciones **irregulares y multimodales**, lo que indica que los mercados laborales y la participación femenina en el empleo han experimentado variaciones significativas a lo largo del tiempo y entre los países considerados. Estas formas no uniformes sugieren que las dinámicas laborales son sensibles al contexto institucional, económico y político de cada periodo.

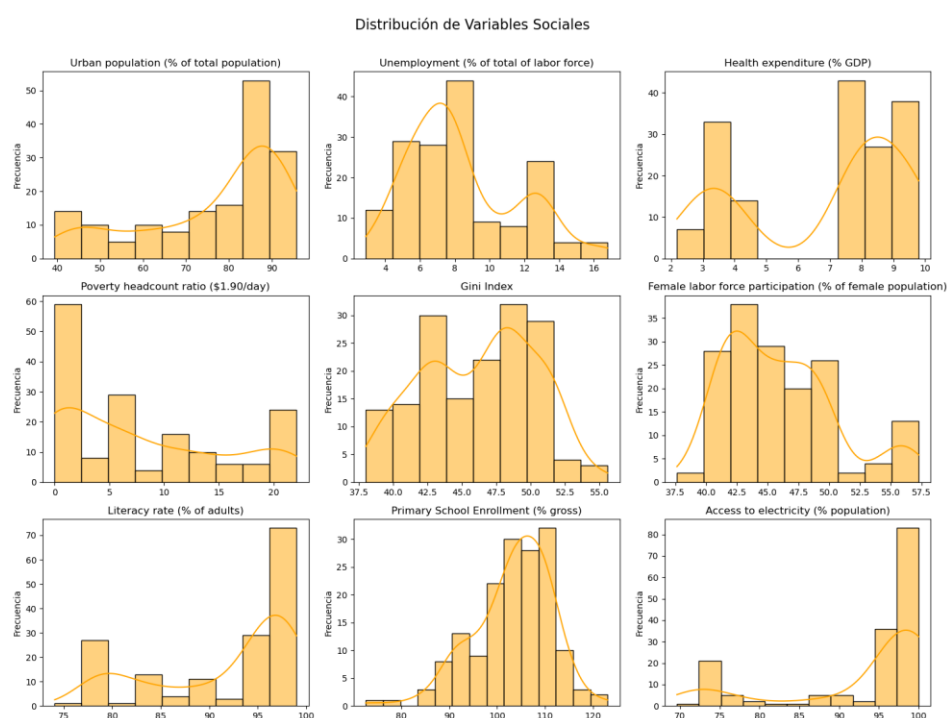


Figura 20: Histogramas de las variables sociales.

La Figura 21 permite analizar estas distribuciones de forma desagregada por país. En ella se evidencia, por ejemplo, que **Uruguay** presenta valores más altos y concentrados en *Literacy rate*, *Access to electricity* y *Primary school enrollment*, mientras que **Venezuela** tiende a agrupar los valores más extremos en *Poverty headcount ratio* y *Unemployment*. **El Salvador**, por su parte, muestra distribuciones intermedias, aunque en algunas variables como *Female labor force participation* presenta una mayor dispersión.

Distribución de Variables Sociales por País

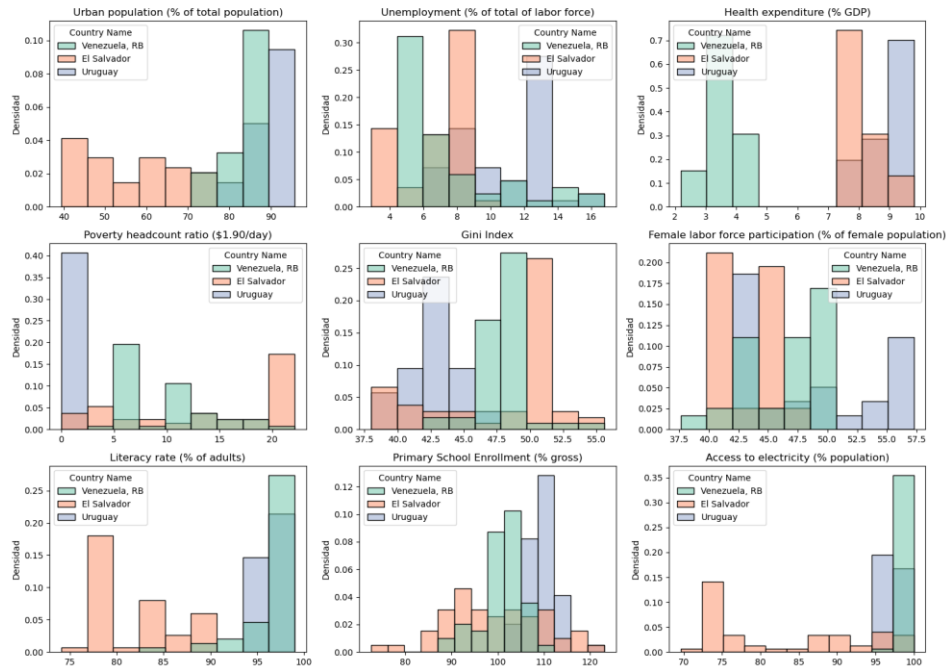


Figura 21: Histogramas de las variables sociales por país.

5.6.4. Variables de gobernanza

En la Figura 22 se representan los histogramas de las variables institucionales, las cuales se expresan en una escala que oscila entre -2.5 y +2.5. En general, las distribuciones de estos indicadores presentan formas **multimodales**, lo que sugiere la existencia de tres grupos bien diferenciados: países con instituciones sólidas (valores positivos), países con instituciones debilitadas (valores negativos) y países con instituciones moderadas (valores centrales).

Distribución de Variables Institucionales

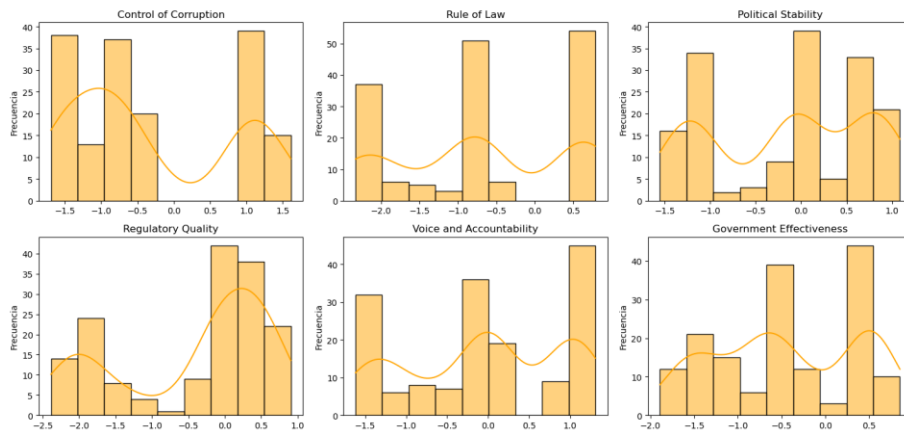


Figura 22: Histograma de las variables institucionales.

La Figura 23 permite analizar estas diferencias comparando por país. Se observa que **Uruguay** concentra sistemáticamente valores más altos en casi todas las variables institucionales, lo que respalda su posición como país con un entorno institucional más robusto. En cambio, **Venezuela** agrupa los valores más bajos, evidenciando un entorno marcado por una gobernanza limitada. **El Salvador**, por su parte, presenta valores más dispersos, situándose en posiciones intermedias.

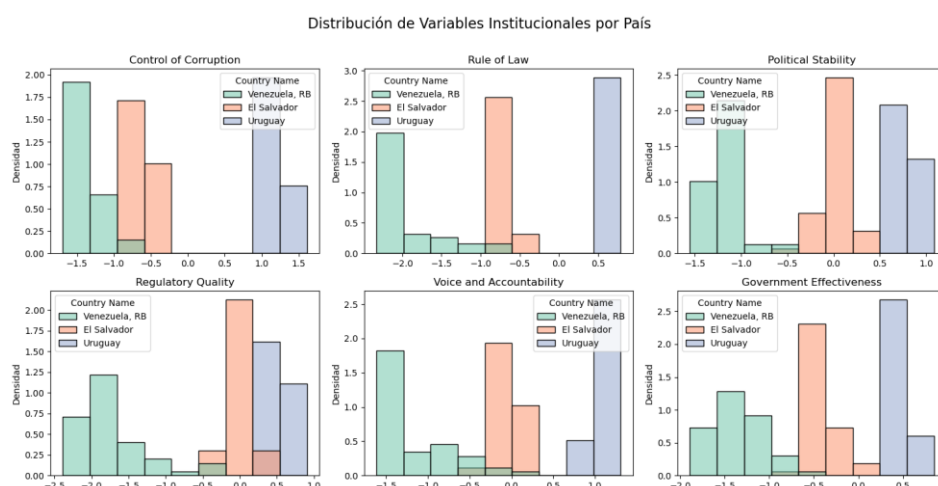


Figura 23: Histograma de las variables institucionales por país.

En conjunto, el análisis de las distribuciones ha permitido identificar patrones clave en la configuración de cada grupo de variables. Las diferencias en la forma de las distribuciones ya sean unimodales, multimodales o asimétricas, reflejan **la diversidad estructural existente entre los países y a lo largo del tiempo**, en términos económicos, sociales, institucionales y de corrupción.

6. Desarrollo de modelos

En este estudio se ha optado por transformar la variable continua Bribery Incidence en una variable binaria Bribery Class, con el objetivo de clasificar los contextos en situaciones de alta o baja corrupción. Esta decisión metodológica se fundamenta en la naturaleza comparativa del trabajo, centrado en identificar patrones y factores asociados a entornos significativamente corruptos. El enfoque de clasificación permite simplificar la interpretación de resultados y facilita el análisis de métricas como la precisión, el recall o el F1-score, especialmente relevantes en contextos con desbalance de clases. Asimismo, técnicas como SMOTE resultan más efectivas bajo este esquema, permitiendo corregir la escasez de observaciones de alta corrupción. Dado que el interés del estudio no reside en predecir con exactitud la magnitud del soborno, sino en comprender las condiciones que caracterizan su presencia, la clasificación binaria ofrece una aproximación más adecuada y alineada con los objetivos analíticos y explicativos del trabajo.

6.1. Modelo de Random Forest

El algoritmo **Random Forest**, disponible en la biblioteca *scikit-learn* de Python, es una técnica de aprendizaje supervisado basada en ensamblado de árboles de decisión. En este análisis se han usado dos versiones: **Random Forest Classifier** (*RandomForestClassifier*, 2025) para clasificación, y **Random Forest Regressor** (*RandomForestRegressor*, 2025) para regresión.

Estas dos versiones de Random Forest difieren en la naturaleza de la variable objetivo, pero ambos comparten la técnica de construcción del modelo de ensamblado, conocida como *bagging* (Bootstrap Aggregating) (Fernández Casal et al., 2024). Bagging consiste en entrenar múltiples árboles de decisión sobre subconjuntos aleatorios del conjunto de entrenamiento, generados mediante muestreo con reemplazo, y luego combinar sus predicciones para mejorar la precisión y reducir la varianza del modelo.

Random Forest Classifier se emplea cuando la variable dependiente es **categorica**. El modelo genera un conjunto de árboles de decisión entrenados sobre subconjuntos aleatorios del conjunto de datos (método *bagging*) y realiza la predicción final a través de una votación mayoritaria entre todos los árboles. Esta estrategia proporciona modelos con

una alta precisión y robustez frente al sobreajuste, especialmente útil cuando existen relaciones no lineales o datos ruidosos.

En el contexto de este trabajo, el clasificador se ha utilizado para **predecir la presencia o ausencia de que una observación presente prácticas de soborno** según la variable *Bribery Incidence*, a partir de los indicadores institucionales, económicos y sociales. Este enfoque permite analizar patrones comunes y diferenciales en los tres países estudiados.

Por otro lado, **Random Forest Regressor** se aplica cuando la variable dependiente es **continua**. En este caso, cada árbol predice un valor numérico, y la predicción final se obtiene como el promedio de todas las predicciones individuales. Esta técnica es permite modelar relaciones no lineales y complejas sin necesidad de supuestos sobre la distribución de los datos.

En este proyecto, el modelo de regresión se ha utilizado para **estimar niveles de corrupción o sobornos** expresados de forma continua, facilitando una comparación cuantitativa entre los países seleccionados.

En definitiva, ambas versiones del modelo Random Forest comparten las siguientes ventajas clave:

- Aplican bagging, lo que reduce la varianza y mitiga el sobreajuste, siendo eficaces con datos ruidosos o con valores faltantes.
- Capturan **relaciones no lineales** entre variables sin necesidad de transformaciones complejas.
- Permiten calcular la **importancia relativa de cada predictor**, lo cual es fundamental en este caso para interpretar los resultados.

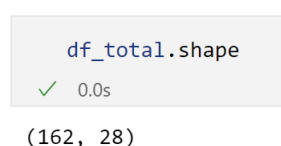
6.1.1. Preparación de los datos

Antes de aplicar los modelos de Machine Learning, se llevó a cabo un proceso de preparación y transformación del conjunto de datos con el objetivo de garantizar su idoneidad para el análisis predictivo. A continuación, se detallan las principales etapas realizadas:

1- Carga y revisión inicial del dataset:

Se trabajó con el conjunto de datos *df_total*, en formato Excel, construido a partir del proceso descrito en el capítulo 5. Limpieza y análisis exploratorio de datos. Este archivo integra información de indicadores institucionales, económicos y sociales, así como variables relacionadas con la corrupción y los sobornos, para los tres países analizados: Venezuela, El Salvador y Uruguay.

En la Figura 24 se muestran las dimensiones del conjunto de datos utilizado como base para el desarrollo de los modelos, el cual cuenta con 28 columnas o variables y, 162 filas u observaciones.



```
df_total.shape
✓ 0.0s
(162, 28)
```

Figura 24: Dimensiones del conjunto de datos total del que se parte en el desarrollo de los modelos.

Una vez cargado el archivo, se procedió a verificar la existencia de valores nulos. El resultado de esta comprobación indicó que no era necesario aplicar técnicas de imputación en esta fase del análisis, ya que el conjunto se encontraba **completo** gracias a los pasos previos.

2- Creación de la variable objetivo Bribery Class:

Dado que uno de los objetivos del presente estudio es **identificar patrones asociados a contextos de mayor o menor corrupción**, inicialmente se abordó el problema como una tarea de clasificación supervisada.

Para poder aplicar el modelo de clasificación de Random Forest, fue necesario crear una nueva variable binaria, denominada *Bribery Class*, a partir de la variable continua *Bribery Incidence* – que representa el porcentaje de empresas que reportaron al menos una solicitud de soborno. De esta forma se pasó a tener un total de 29 variables. Esto permitió transformar el problema de regresión en uno de clasificación binaria.

Como punto de corte, se utilizó la mediana de la variable *Bribery Incidence*, cuyo valor es 4.5, tal como se muestra en la Figura 25. A partir de este umbral, se definieron dos clases:

- **Bribery Class = 1.** Observaciones con valores mayores o iguales a 4.5 fueron clasificadas como 1, lo que implica una alta incidencia de sobornos.
- **Bribery Class = 0.** Observaciones con valores menores a 4.5 fueron clasificadas como 0, lo que indica baja incidencia de sobornos.

```
df_total['Bribery Incidence'].median()
✓ 0.0s
4.5
```

Figura 25: Mediana de la variable Bribery Incidence.

Esta recodificación facilitó el entrenamiento y la interpretación de modelos clasificatorios capaces de revelar qué variables institucionales, económicas o sociales están más asociadas a contextos de alta o baja corrupción.

3- Codificación de variables categóricas:

La variable *Country Name* fue transformada mediante la técnica de codificación *one-hot* para representar cada país como una variable dummy, sin eliminar ninguna categoría con *drop_first=False*. Con este paso, el número de variables del estudio pasa a ser 31, ya que se elimina automáticamente la variable *Country Name* y en su lugar, se pasa a tener 3 variables dummies binarias (Figura 26), una para cada país. Esto permite preservar toda la información de los países durante el entrenamiento del modelo.

Country Name_El Salvador	Country Name_Uruguay	Country Name_Venezuela, RB
False	False	True
False	False	True
False	False	True
False	False	True
False	False	True
...	...	...
False	True	False
False	True	False

Figura 26: Nuevas variables dummies para los países.

4- Separación de variables predictoras y variable objetivo:

Una vez creada la variable objetivo *Bribery Class*, se procedió a dividir el dataset en dos subconjuntos diferenciados:

- **X:** conjunto de variables explicativas, excluyendo *Bribery Class*, *Bribery Incidence* y *Year*. De esta forma, este subconjunto contiene 28 columnas y 162 observaciones.
- **y:** variable objetivo binaria *Bribery Class*, con 1 columna y 162 observaciones.

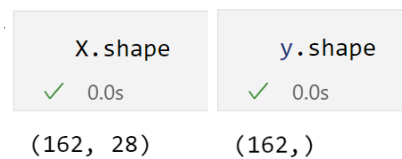


Figura 27: Dimensiones de los subconjuntos *X* e *y*.

Esta separación es un paso previo fundamental para aplicar modelos supervisados, permitiendo entrenar algoritmos que aprendan a predecir la clase de soborno en función de las variables explicativas.

5- División en entrenamiento y prueba:

Con el objetivo de evaluar adecuadamente el rendimiento del modelo, se procedió a dividir el dataset, utilizando la función *train_test_split* (*Train_test_split*, 2025) de la biblioteca *scikit-learn*, en dos subconjuntos: **entrenamiento (80%)** y **prueba (20%)**. Esto resultó en un total de 129 muestras para el entrenamiento y 33 muestras para la prueba.

Para asegurar que la distribución de clases en la variable objetivo *Bribery Class* se mantuviera proporcional en ambos subconjuntos, se utilizó la opción *stratify=y*. Este muestreo **estratificado** es especialmente importante en problemas de clasificación binaria para evitar que alguna clase quede subrepresentada en el conjunto de entrenamiento o test.

6- Balanceo de clases con SMOTE:

Dado que se observó un leve desbalanceo en la variable objetivo *Bribery Class* – con 74 observaciones de clase 1 (alta corrupción) frente a 55 observaciones de clase 0 (baja

corrupción) – se optó por aplicar la técnica *oversampling* **SMOTE (Synthetic Minority Oversampling Technique)** (SMOTE, 2024) sobre el conjunto de entrenamiento.

Se optó por una técnica de *oversampling* como SMOTE en lugar de eliminar observaciones de la clase mayoritaria (*undersampling*) con el fin de preservar al máximo la información disponible en el conjunto de datos, especialmente en contextos como este donde el número total de observaciones no es elevado. Esto permite evitar la pérdida de información y mantener la riqueza del conjunto original para el entrenamiento del modelo.

Esta técnica genera nuevas observaciones sintéticas de la clase minoritaria, utilizando interpolación entre vecinos cercanos, con el objetivo de **equilibrar la distribución de clases** y mejorar así la capacidad predictiva del modelo. El balanceo previo al entrenamiento es especialmente importante en problemas de clasificación binaria, ya que un desequilibrio puede sesgar el modelo hacia la clase mayoritaria.

Tras aplicar SMOTE, se obtuvo una distribución perfectamente balanceada, con 74 observaciones para cada clase, resultando en un total de 148 muestras en el conjunto de entrenamiento. En la Figura 28 se muestra el cambio en la distribución de clases antes y después de aplicar la técnica, observándose un incremento de 19 muestras sintéticas para la clase 0, logrando así la paridad.

```
Distribución original:
Bribery Class
1      74
0      55
Name: count, dtype: int64

Distribución tras SMOTE:
Bribery Class
0      74
1      74
Name: count, dtype: int64
```

Figura 28: Resultado de aplicar la técnica SMOTE sobre el conjunto de entrenamiento.

6.1.2. Ajuste de hiperparámetros

Con el objetivo de **optimizar el rendimiento** del modelo de clasificación basado en Random Forest, se llevó a cabo un proceso de ajuste de hiperparámetros mediante la técnica de **búsqueda en malla** (*GridSearchCV*, 2025) combinada con **validación**

cruzada estratificada (*StratifiedKFold*, 2025), ambas implementadas en la biblioteca *scikit-learn*.

El procedimiento consistió en definir un conjunto de combinaciones posibles de hiperparámetros relevantes del modelo:

- `n_estimators`: número de árboles en el bosque ([50, 100, 200]).
- `max_depth`: profundidad máxima de los árboles ([None, 5, 10]).
- `min_samples_split`: número mínimo de muestras necesarias para dividir un nodo ([2, 5]).
- `min_samples_leaf`: número mínimo de muestras en una hoja ([1, 2]).
- `max_features`: número de variables consideradas al dividir un nodo (['sqrt', 'log2']).

La búsqueda de hiperparámetros se realizó utilizando la clase *GridSearchCV*, combinada con un esquema de validación cruzada estratificada de 10 particiones, implementada mediante *StratifiedKFold*. Esta técnica divide el conjunto de entrenamiento en 10 subconjuntos (*folds*), de forma que en cada iteración se entrena el modelo con 9 de ellos y se valida con el restante. Se activó el parámetro `shuffle=True` para mezclar aleatoriamente los datos antes de dividirlos, garantizando que cada partición sea representativa y no dependa del orden original del dataset.

Como criterio de optimización se empleó **F1-Score**, especialmente adecuado para problemas de clasificación binaria. Aunque el conjunto de entrenamiento fue previamente balanceado con SMOTE, es importante utilizar una métrica que no dependa únicamente de la proporción de clases, sino que evalúe con precisión el rendimiento en ambas. Esto se debe a que, incluso tras el balanceo, pueden persistir **desequilibrios latentes**: por ejemplo, el conjunto de prueba sigue reflejando la distribución original, o una clase puede resultar sistemáticamente más difícil de predecir que la otra.

El F1-Score combina de forma armónica dos métricas clave:

- **Precision** (precisión): proporción de verdaderos positivos entre todas las predicciones positivas. Es decir, mide cuántas de las observaciones clasificadas como positivas lo eran realmente.

- **Recall** (exhaustividad): proporción de verdaderos positivos entre todas las observaciones positivas reales. Es decir, mide cuántas de las observaciones positivas reales fueron correctamente identificadas.

El F1-Score es útil en este estudio, ya que se busca **detectar correctamente tanto los contextos de alta como de baja corrupción**, sin favorecer a ninguna clase en particular. Además, evita que el modelo se optimice únicamente en función de la precisión global (*accuracy*), métrica que podría resultar engañosa en presencia de desequilibrios latentes.

El ajuste se aplicó sobre el conjunto de entrenamiento balanceado. En total, se evaluaron **72 combinaciones** de hiperparámetros, lo que implicó **720 iteraciones** de entrenamiento y validación.

En la Figura 29 podemos ver los parámetros óptimos encontrados para la creación del modelo de Random Forest Classifier. Este conjunto de valores proporciona **un F1-Score medio de validación de 0.9933**, lo que evidencia un alto rendimiento del modelo.

Mejores parámetros: {'max_depth': None, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50}
 Mejor puntuación F1: 0.9933333333333334

Figura 29: Mejores hiperparámetros para el modelo de Random Forest Classifier.

En la Figura 30 se presenta la creación final del modelo de *Random Forest Classifier* con los hiperparámetros óptimos, el cual será utilizado en las fases posteriores de evaluación y análisis de importancia de variables.

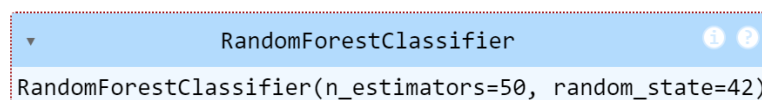


Figura 30: Creación del modelo de Random Forest Classifier con los hiperparámetros óptimos.

6.1.3. Evaluación del modelo

Una vez entrenado el modelo de *Random Forest Classifier* con los hiperparámetros óptimos, se procedió a su evaluación sobre el conjunto de prueba. Para ello, se utilizaron diversas métricas estándar de clasificación: precisión (*precision*), exhaustividad (*recall*), F1-Score (media armónica entre la precisión y la exhaustividad) y exactitud (*accuracy*), además de la matriz de confusión y la curva ROC con su área bajo la curva (AUC).

Respecto a las **métricas de clasificación**, los resultados obtenidos se resumen en la Tabla 4. El modelo muestra un rendimiento equilibrado en ambas clases:

- Para la clase Baja Corrupción (clase 0), el F1-Score fue de 0.9630.
- Para la clase Alta Corrupción (clase 1), el F1-Score fue de 0.9744.

Además, la exactitud global (accuracy) fue de 0.9697, lo que significa que el 96.97% de las observaciones fueron correctamente clasificadas. Se presentan también dos promedios globales que permiten valorar el rendimiento general del modelo:

- **Macro average:** calcula la media aritmética simple de cada métrica, asignando igual peso a ambas clases. Es útil para evaluar si el modelo es justo con todas las clases, independientemente de su tamaño.
- **Weighted average:** calcula una media ponderada por el número de observaciones de cada clase, lo que proporciona una visión más realista del rendimiento global cuando existe un ligero desbalance de clases.

Ambos promedios de F1-Score fueron superiores a 0.96, lo que indica una **alta capacidad de generalización** sin favorecer a una clase en particular.

	Precisión	Recall	F1-Score	Soporte
0 (Baja Corrupción)	1,0000	0,9285	0,9630	14
1 (Alta Corrupción)	0,9500	1,0000	0,9744	19
Accuracy global	–	–	0,9697	33
Macro avg	0.9750	0,9643	0,9687	33
Weighted avg	0.9712	0,9697	0,9695	33

Tabla 4: Informe de clasificación sobre el conjunto de prueba del modelo de Random Forest Classifier con todos los países.

En la Figura 31 se presenta la **matriz de confusión** del modelo. Esta muestra que el modelo clasificó correctamente:

- 13 de las 14 observaciones con baja corrupción (clase 0).
- 19 de las 19 observaciones con alta corrupción (clase 1).

Solo se produjo **un falso positivo**, es decir, un caso de baja corrupción clasificado erróneamente como alta corrupción.

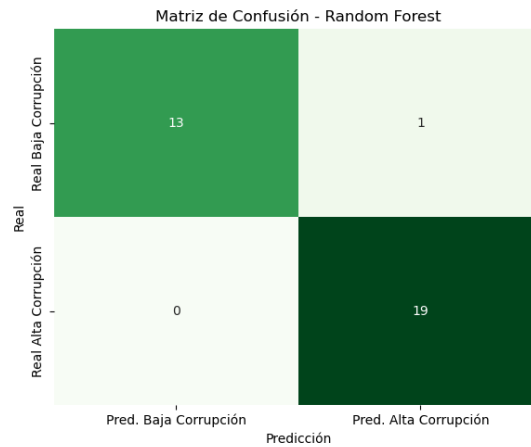


Figura 31: Matriz de confusión del modelo de Random Forest Classifier.

En la Figura 32 se muestra la curva ROC del modelo, generada a partir de las probabilidades de predicción del modelo. El área bajo la curva (AUC) es igual a 1.00, lo que indica una capacidad de discriminación perfecta entre las clases para este conjunto de prueba. Este resultado respalda las métricas de clasificación obtenidas y confirma que el modelo es altamente eficaz al predecir tanto escenarios de alta como de baja corrupción.

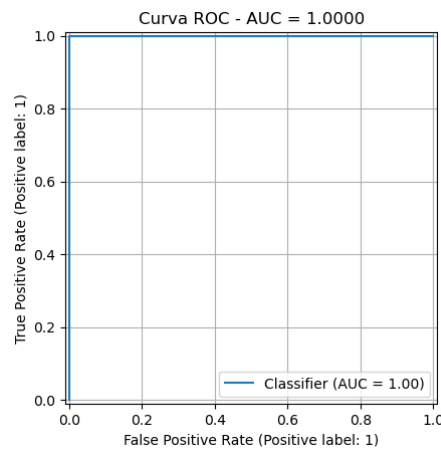


Figura 32: Curva ROC del modelo de Random Forest Classifier.

6.1.4. Validación cruzada

Con el objetivo de verificar la robustez y estabilidad del modelo de **Random Forest Classifier**, se aplicó una validación cruzada estratificada de 10 particiones con *StratifiedKfold* sobre el conjunto completo de datos. Esta técnica permite estimar el rendimiento medio del modelo al entrenarlo y validarlo en diferentes subconjuntos, manteniendo la proporción de clases en cada fold.

Se evaluaron las métricas de Accuracy, F1-score y AUC-ROC en cada partición. Las medias obtenidas fueron consistentes con los resultados en el conjunto de prueba, reforzando la validez del modelo construido.

En la Figura 33 se muestran los resultados de cada una de las 10 iteraciones, así como la media y la desviación estándar de cada métrica. El modelo alcanzó un accuracy promedio de 0.9813, un F1-score promedio de 0.9836 y un ROC AUC promedio de 0.9952. Estos valores reflejan un desempeño consistente y alto a lo largo de las diferentes particiones, sin variaciones significativas entre los *folds*.

	Accuracy	F1-score	ROC AUC
0	1.000000	1.000000	1.000000
1	1.000000	1.000000	1.000000
2	1.000000	1.000000	1.000000
3	1.000000	1.000000	1.000000
4	1.000000	1.000000	1.000000
5	0.937500	0.941176	1.000000
6	0.937500	0.947368	1.000000
7	0.937500	0.947368	0.952381
8	1.000000	1.000000	1.000000
9	1.000000	1.000000	1.000000
Media	0.981250	0.983591	0.995238
Desviación estándar	0.028641	0.025116	0.014286

Figura 33: Resultados de la validación cruzada para Random Forest Classifier.

Adicionalmente, se generó un **informe de clasificación agregado** (Figura 34) a partir de las predicciones cruzadas. El modelo mantuvo un F1-score de 0.9781 para la clase Baja Corrupción y de 0.9839 para la clase Alta Corrupción, con una exactitud general del 98.15%. Los valores de macro y weighted average superaron el 0.98 en todas las métricas.

	precision	recall	f1-score	support
0	0.985294	0.971014	0.978102	69.000000
1	0.978723	0.989247	0.983957	93.000000
accuracy	0.981481	0.981481	0.981481	0.981481
macro avg	0.982009	0.980131	0.981030	162.000000
weighted avg	0.981522	0.981481	0.981463	162.000000

Figura 34: Informe de clasificación con validación cruzada para Random Forest Classifier.

Por otro lado, la **matriz de confusión combinada** de la Figura 35 muestra que el modelo clasificó correctamente:

- 67 de 69 observaciones de Baja Corrupción (2 errores de clase 0).
- 92 de 93 observaciones de Alta Corrupción (1 error de clase 1).

Lo cual se traduce en una tasa de error muy baja del 1.85%, con únicamente tres observaciones clasificadas incorrectamente en todo el conjunto.

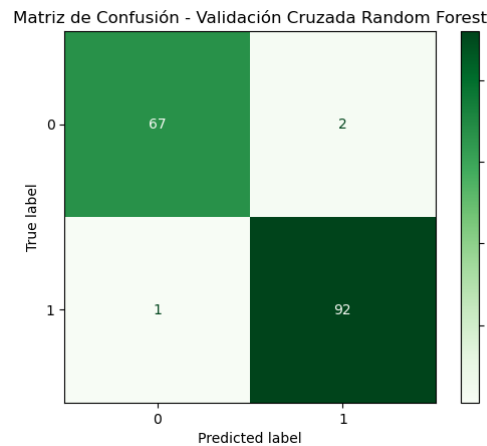


Figura 35: Matriz de confusión para validación cruzada en Random Forest Classifier.

Por último, la **curva ROC agregada** en la Figura 36 confirma el excelente comportamiento del modelo, con un AUC de 0.9969, reflejando una capacidad de discriminación casi perfecta entre contextos de baja y alta corrupción, evidenciando su utilidad como herramienta analítica para el estudio propuesto.

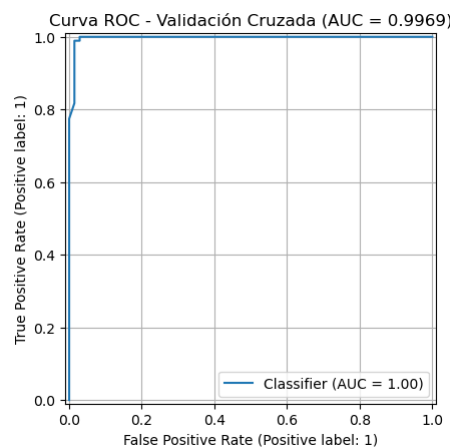


Figura 36: Curva ROC del modelo de Random Forest Classifier con validación cruzada.

En conjunto, los resultados obtenidos a través de la validación cruzada confirman que el modelo de **Random Forest Classifier** presenta un **rendimiento altamente consistente y robusto**. La baja variabilidad entre particiones, junto con las altas puntuaciones en F1-score, precisión y AUC, evidencian que el modelo **generaliza adecuadamente y es fiable para realizar análisis comparativos entre países**. Esto lo convierte en una herramienta válida para identificar patrones y diferencias en los niveles de corrupción empresarial en contextos institucionales y económicos diversos.

6.1.5. Importancia de las variables

Uno de los beneficios clave del modelo Random Forest es su capacidad para estimar la **importancia relativa** de cada variable predictora, lo cual permite identificar qué factores están más asociados a contextos de alta o baja corrupción, tal como se refleja en la variable *Bribery Class*. Esta información es esencial para comprender el fenómeno de los sobornos desde una perspectiva multivariable.

En la Figura 37 se muestran todas las variables ordenadas por su importancia en el modelo, y en la Figura 38 se destacan las 10 variables más influyentes.

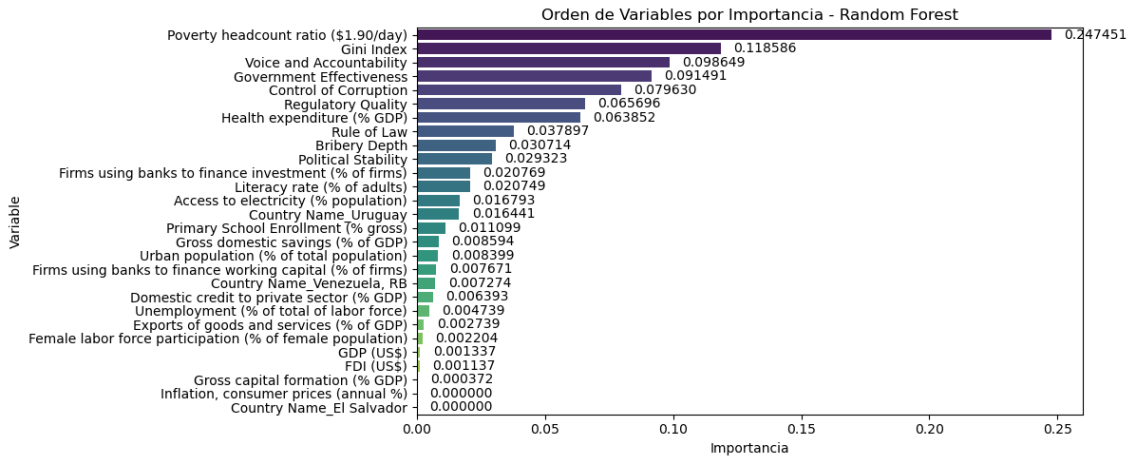


Figura 37: Orden de las variables por importancia en el modelo de Random Forest Classifier.

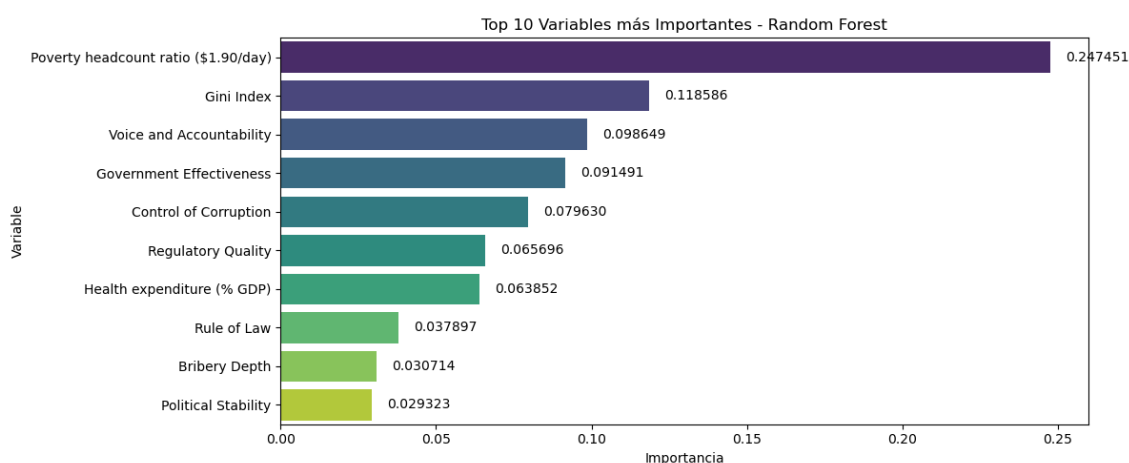


Figura 38: Top 10 variables más importantes según el modelo de Random Forest Classifier.

El modelo destaca a factores sociales y económicos. Encabezando la lista, la variable más relevante con diferencia fue el *Poverty Headcount Ratio (\$1.90/day)*, con un valor de importancia del 24.75%. Esta variable mide el porcentaje de la población que vive en situación de pobreza extrema, y su fuerte peso en el modelo sugiere una **relación directa entre altos niveles de pobreza y mayor propensión a prácticas de soborno**. Esto puede interpretarse desde una doble perspectiva: por un lado, la vulnerabilidad económica puede aumentar la disposición a aceptar o solicitar sobornos; por otro, la falta de recursos básicos puede generar presión sobre instituciones públicas, facilitando conductas corruptas.

En segundo lugar, se encuentra el *Gini Index*, indicador de desigualdad económica. Su alta importancia de 11.86%, refuerza la hipótesis de que **la desigualdad estructural contribuye a la corrupción**, ya sea porque reduce la confianza institucional o porque concentra el poder económico y político en manos de pocos actores con capacidad de influencia.

La variable *Health Expenditure (% of GDP)* también aparece como relevante con 6.39% de importancia. Este resultado podría interpretarse como un indicador del compromiso del Estado con el bienestar público: **mayores niveles de gasto en salud podrían estar asociados a una menor necesidad de prácticas informales** o pagos extraoficiales para acceder a servicios básicos.

Por otro lado, el indicador *Bribery Depth* también figura dentro de las 10 más importantes, validando que el porcentaje de transacciones públicas en las que se solicita un soborno

tiene un peso significativo para distinguir contextos de alta y baja corrupción, reforzando la utilidad de los datos de Enterprise Surveys para este tipo de análisis.

En cuanto a calidad institucional, cinco de las diez variables más importantes están directamente relacionadas con la gobernanza: *Voice and Accountability*, *Government Effectiveness*, *Control of Corruption*, *Regulatory Quality* y *Rule of Law*.

Estos resultados reflejan claramente que **la calidad institucional es un determinante clave de la corrupción**. Países con mayor capacidad administrativa, sistemas regulatorios sólidos y mecanismos de rendición de cuentas más robustos tienden a presentar menores niveles de soborno. En particular, *Voice and Accountability* y *Government Effectiveness* destacan como variables especialmente sensibles, lo cual sugiere que la libertad de expresión, la participación ciudadana y el buen funcionamiento del aparato público son factores protectores frente a la corrupción.

En conjunto, los resultados muestran que **la corrupción no es un fenómeno aislado**, sino que está profundamente conectado con condiciones estructurales de pobreza, desigualdad e institucionalidad. El modelo Random Forest Classifier ha permitido identificar un patrón claro: los países con **mayores niveles de pobreza y menor calidad institucional presentan mayor incidencia de sobornos**, mientras que aquellos con estructuras sociales más equitativas y sistemas de gobernanza más sólidos tienden a registrar menores niveles de corrupción empresarial.

Este análisis respalda la idea de que las políticas públicas orientadas a **fortalecer las instituciones, reducir la desigualdad y mejorar el acceso a servicios básicos** pueden ser herramientas efectivas para disminuir la prevalencia del soborno en el ámbito empresarial.

6.1.6. Estudio por países con Random Forest

Con el objetivo de analizar la capacidad del modelo para identificar correctamente contextos de corrupción específicos en cada país, se examinó la **distribución de clases reales y predichas** de la variable *Bribery Class* segmentada por país: El Salvador, Uruguay y Venezuela. Este análisis permite verificar si el modelo reproduce patrones coherentes con la realidad observada y evaluar el peso relativo del país en la predicción.

En la Figura 39, se observa la distribución real de clases por país. Uruguay destaca por registrar el 100% de sus observaciones en la clase Baja Corrupción (0). En el caso de Venezuela, el 98% de las observaciones corresponden a la clase Alta Corrupción (1), mientras que El Salvador muestra una distribución más mixta, aunque también dominada por la clase de alta corrupción con el 74%.

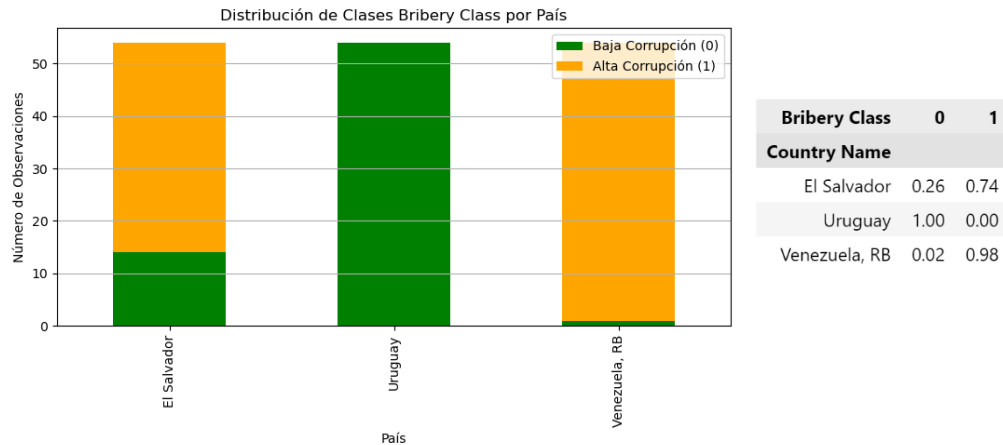


Figura 39: Distribución real de Bribery Class por país.

Al aplicar el modelo Random Forest Classifier, la **distribución de predicciones** por país que podemos ver en la Figura 40 resulta altamente coherente con la observada en los datos reales. El modelo predice: Uruguay como baja corrupción en el 100% de los casos, Venezuela como alta corrupción en el 98% de los casos, y El Salvador con una distribución muy similar a la real con el 76% como alta corrupción. Este alineamiento entre clases reales y predichas respalda la **capacidad del modelo para capturar las diferencias estructurales entre países**.

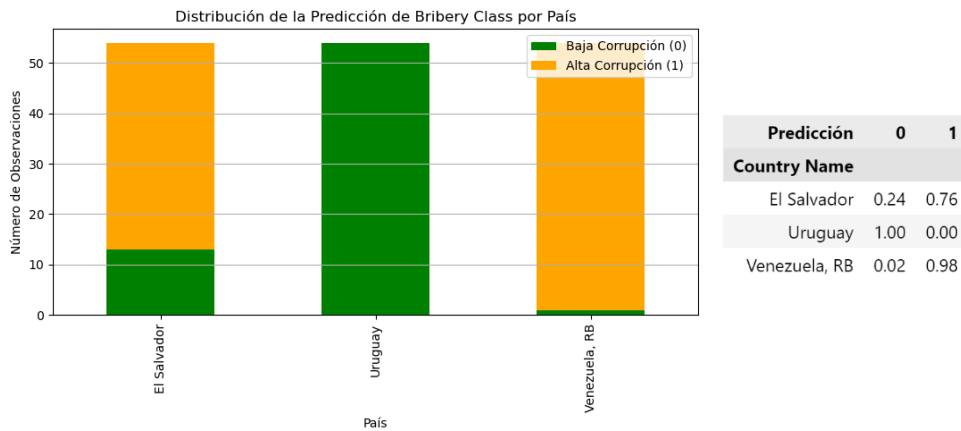


Figura 40: Proporción de predicciones de Bribery Class por país con Random Forest Classifier.

En la Figura 41 se muestra la **importancia relativa** de las variables *dummy* asociadas al país. Los resultados indican que estas variables tienen una importancia muy baja en el modelo, teniendo incluso la variable *Country Name_El Salvador* importancia igual a 0.

Esto significa que **el modelo no depende directamente del país de origen para predecir si hay alta o baja corrupción**. En cambio, toma sus decisiones principalmente en función de otras variables más explicativas, como pobreza, desigualdad, calidad institucional o gasto en salud.

Este resultado es positivo desde el punto de vista analítico: muestra que el modelo **no se limita a identificar la corrupción simplemente por el país**, sino que **aprende patrones más generales y objetivos**, aplicables a cualquier contexto. En otras palabras, el modelo predice la clase de corrupción basándose en las condiciones estructurales que caracterizan a cada país.

	Variable	Importancia
26	Country Name_Uruguay	0.016441
27	Country Name_Venezuela, RB	0.007274
25	Country Name_El Salvador	0.000000

Figura 41: Importancia relativa de las variables de país en el modelo de Random Forest Classifier.

Modelo Random Forest Regressor por país

Dado que en Uruguay no se registran observaciones clasificadas como alta corrupción, la aplicación de un modelo de clasificación binaria por país resulta no es viable.

Por este motivo, se optó por implementar un modelo básico de **Random Forest con regresión** en cada país, utilizando como variable dependiente el valor original de *Bribery Incidence*. Este enfoque permite analizar **qué factores explican mejor los niveles de sobornos en cada país**.

En la Figura 42 se presentan tres métricas para cada país obtenidas como resultado del estudio.

- **RMSE (Root Mean Squared Error):** mide el error medio cuadrático de las predicciones.
- **MAE (Mean Absolute Error):** mide el error medio absoluto.

- **R² (Coeficiente de determinación):** indica qué proporción de la variabilidad de *Bribery Incidence* es explicada por el modelo.

Los tres modelos alcanzan un **alto nivel de ajuste**, especialmente en El Salvador y Uruguay, donde el R² supera el 0.96, lo que indica que las variables independientes seleccionadas explican muy bien la variabilidad en los niveles de corrupción.

	País	RMSE	MAE	R ²
0	Venezuela, RB	0.203924	0.054167	0.893700
1	El Salvador	0.049174	0.018130	0.967949
2	Uruguay	0.005707	0.002019	0.985781

Figura 42: Métricas del modelo Random Forest con regresión por país.

Las figuras: Figura 43, Figura 44 y Figura 45 muestran las 15 variables más influyentes en la predicción del valor de *Bribery Incidence* en cada país.

En la Figura 43 vemos que en el caso de Venezuela destacan variables financieras y laborales como:

Firms using banks to finance working capital, Female labor force participation, Bribery Depth, Firms using banks to finance investment y GDP (US\$).

En conjunto, estas variables reflejan un entorno de **precariedad económica, informalidad y exclusión**, donde el soborno aparece como mecanismo informal de acceso a recursos y oportunidades.

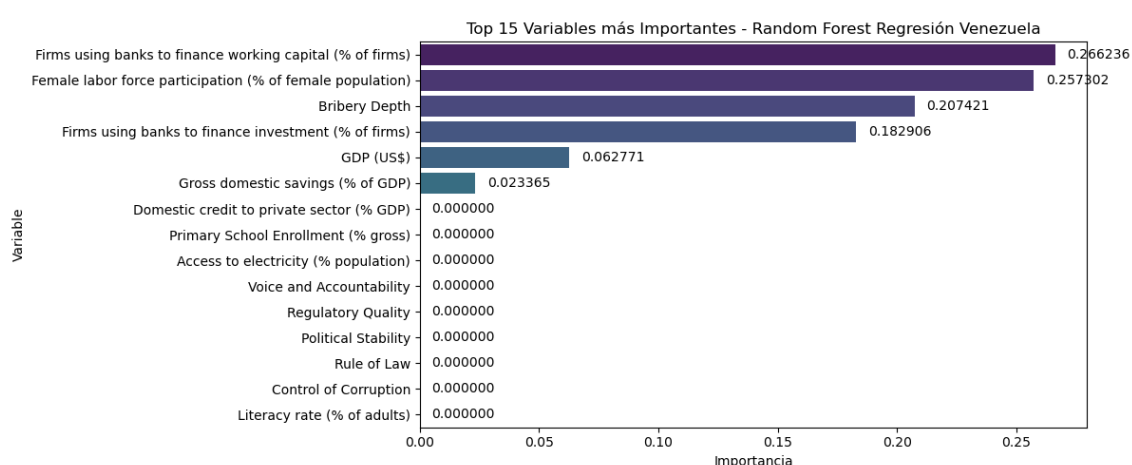


Figura 43: Top 15 Variables más importantes para Venezuela según el modelo de Regresión de Random Forest.

En la Figura 44 vemos que en El Salvador destacan factores vinculados al desarrollo:

Firms using banks to finance investment, GDP (US\$), Urban population, Access to electricity y Gini Index.

El Salvador cuenta con niveles de corrupción moderados. En este sentido, el modelo sugiere un **proceso de transición y mejora**, donde el fortalecimiento de la infraestructura y el sistema económico está relacionado con una reducción gradual de la corrupción.

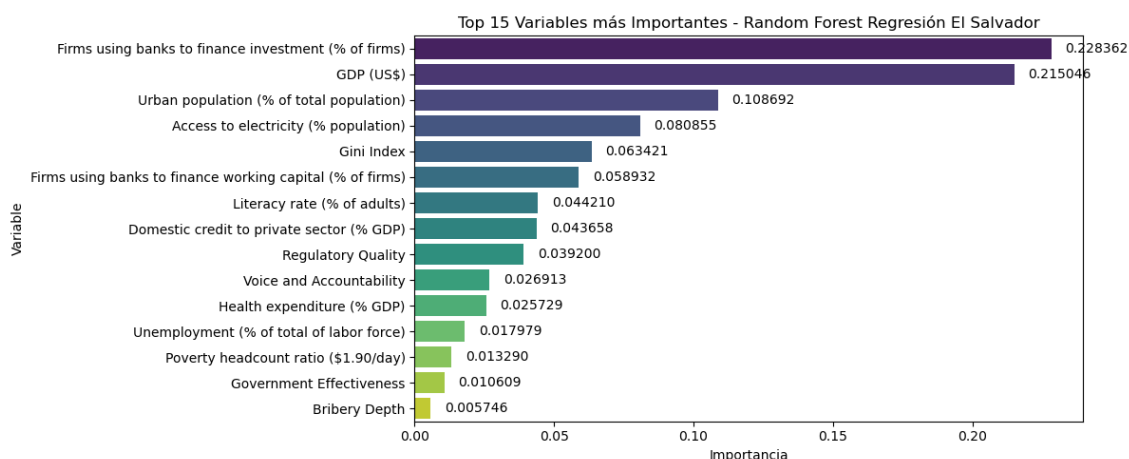


Figura 44: Top 15 Variables más importantes para El Salvador según el modelo de Regresión de Random Forest.

En la Figura 45 se analiza el caso de Uruguay, país que presenta niveles muy bajos de corrupción. El modelo identifica como variables clave:

Urban population, Firms using banks to finance investment, GDP (US\$), Literacy Rate y Bribery Depth.

Con el modelo desarrollado para Uruguay se busca explicar **qué condiciones permiten mantener la corrupción en niveles bajos** y estables. Las variables más importantes reflejan una creciente urbanización, un entorno económico sólido, una alta inclusión social y un acceso amplio a servicios básicos. Por lo que, estos elementos están relacionados con un nivel bajo de corrupción.

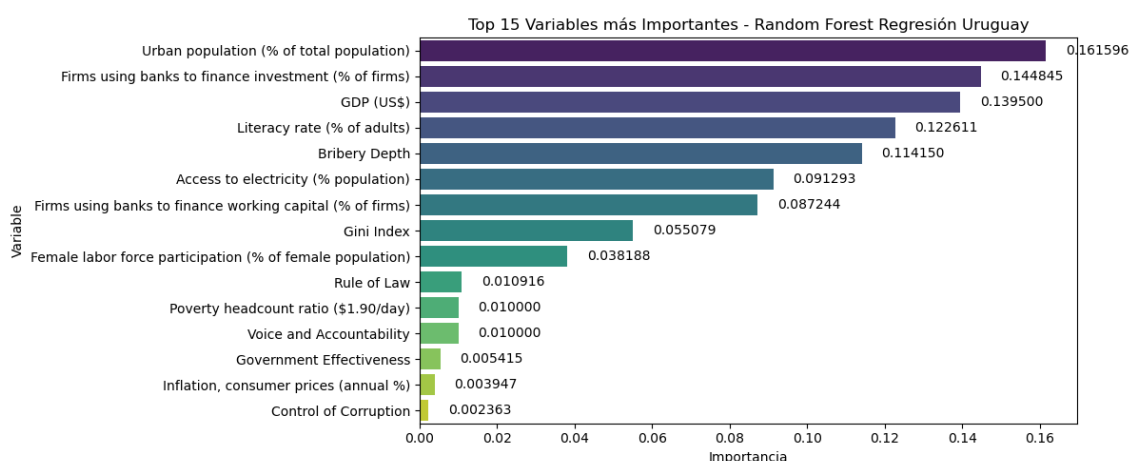


Figura 45: Top 15 Variables más importantes para Uruguay según el modelo de Regresión de Random Forest.

La Figura 46 resume comparativamente las cinco variables más influyentes por país, lo que permite observar que, aunque algunos factores como el acceso a financiamiento o el PIB son comunes, el peso y el impacto de cada uno varía según el nivel de desarrollo y contexto institucional del país.

	Venezuela	El Salvador	Uruguay
Top 1	Firms using banks to finance working capital	Firms using banks to finance investment	Urban population
Top 2	Female labor force participation	GDP (US\$)	Firms using banks to finance investment
Top 3	Bribery Depth	Urban population	GDP (US\$)
Top 4	Firms using banks to finance investment	Access to electricity	Literacy rate
Top 5	GDP (US\$)	Gini Index	Bribery Depth

Figura 46: Comparación de las 5 variables más importantes por país según el modelo de regresión de Random Forest.

El análisis demuestra que los distintos **niveles de corrupción** observados en cada país se explican por conjuntos de **factores estructurales diferentes**, vinculados a su realidad social, económica e institucional.

- En **Venezuela**, la corrupción está vinculada a **limitaciones estructurales**, como la falta de acceso al crédito formal y la baja inclusión laboral.
- En **El Salvador**, influyen variables que reflejan **desarrollo económico y acceso a servicios**, en un contexto de mejora progresiva del nivel de corrupción.
- En **Uruguay**, se confirma que **un entorno social, educativo e institucional sólido** es clave para sostener bajos niveles de soborno.

Por tanto, el uso de Random Forest Regressor no solo permite predecir la corrupción, sino también comprender sus causas profundas. Así, se demuestra que la corrupción no es solo un fenómeno legal o ético, sino un **síntoma complejo de desigualdades, condiciones socioeconómicas y calidad institucional** que pueden ser detectadas y analizadas.

6.2. Modelos de Regresión Logística y Regresión Lineal

En este análisis también se han empleado dos modelos clásicos de aprendizaje supervisado: **Regresión Logística** (*LogisticRegression*, 2025) y **Regresión Lineal** (*LinearRegression*, 2025). Ambas técnicas pertenecen a la familia de los modelos paramétricos, y se diferencian en función del tipo de variable objetivo que pretenden predecir.

La **Regresión Logística** se utiliza cuando la variable dependiente es categórica binaria. En este caso, el modelo estima la **probabilidad de pertenencia a una clase**, utilizando como función de enlace la sigmoide o logística. A partir de esa probabilidad, se asigna la clase correspondiente. Este modelo es especialmente útil cuando se desea interpretar cómo cada variable influye en la **probabilidad de ocurrencia de un evento**, lo que en este trabajo se traduce en estimar la probabilidad de que exista corrupción (clase 1) frente a ausencia de corrupción (clase 0), según la variable *Bribery Class*.

El modelo de regresión logística se ha aplicado **al conjunto total de observaciones**, permitiendo identificar qué variables aumentan o reducen la probabilidad de soborno en los tres países estudiados de forma conjunta.

Por su parte, la **Regresión Lineal** se emplea cuando la variable dependiente es continua. El modelo asume una relación lineal entre las variables independientes y la variable objetivo, y estima los coeficientes que mejor ajustan esa relación minimizando el error cuadrático medio. Esta técnica permite interpretar la **contribución de cada variable al nivel esperado de sobornos**, medido mediante la variable *Bribery Incidence*, facilitando un análisis cuantitativo de la corrupción en términos absolutos.

En este trabajo, la regresión lineal se ha utilizado **de forma individual en cada país** (Venezuela, El Salvador y Uruguay), con el objetivo de identificar qué factores explican mejor la incidencia de sobornos en cada contexto nacional.

Ambos modelos comparten una ventaja importante: su **interpretabilidad**. A diferencia de técnicas más complejas como Random Forest, los coeficientes obtenidos permiten analizar de forma directa y transparente el **sentido (positivo o negativo)** y la **intensidad de la relación** entre cada variable explicativa y la variable dependiente. Esto resulta especialmente valioso en estudios aplicados como este, donde el objetivo no es solo predecir, sino también **comprender las causas subyacentes de la corrupción**.

6.2.1. Preparación de los datos

Para el análisis con modelos de regresión logística, se realizó una preparación similar del conjunto de datos *df_total* a la descrita en el punto 6.1.1 Preparación de los datos. En primer lugar, se recodificó la variable continua *Bribery Incidence* en una variable binaria denominada *Bribery Class*, asignando el valor 1 a aquellas observaciones con valores iguales o superiores a la mediana de 4.5, y 0 en caso contrario. Esta transformación permitió convertir el problema en una tarea de clasificación binaria.

Posteriormente, se codificó la variable categórica *Country Name* mediante variables *dummy*, y se eliminaron del conjunto de predictores las variables *Bribery Incidence*, *Bribery Class* (variable objetivo) y *Year*, dejando únicamente las variables explicativas.

A continuación, se dividió el dataset en conjuntos de entrenamiento y prueba, utilizando una partición estratificada del 80%-20% para garantizar la misma proporción de clases en ambas muestras. Finalmente, se aplicó la técnica **SMOTE** al conjunto de entrenamiento, con el fin de corregir el desbalance entre clases y mejorar la capacidad del modelo para detectar observaciones de alta corrupción (clase 1).

6.2.2. Ajuste de hiperparámetros

Para optimizar el rendimiento del modelo de Regresión Logística, se llevó a cabo un proceso de ajuste de hiperparámetros mediante **búsqueda en malla** (*GridSearchCV*, 2025), utilizando validación cruzada estratificada de 10 particiones. El procedimiento se aplicó sobre el conjunto de entrenamiento balanceado con SMOTE, y se integró en un pipeline que incluyó la **normalización de variables** con *Standard Scaler* (*StandardScaler*, 2025), necesaria para garantizar una correcta aplicación de la regularización.

El ajuste se centró en los siguientes dos hiperparámetros clave:

- **penalty**: determina el tipo de regularización que se aplica al modelo de Regresión Logística. La regularización es una técnica que introduce una penalización sobre los coeficientes del modelo con el objetivo de **evitar el sobreajuste** y mejorar la generalización. En este caso, se evaluaron dos tipos:
 - Regularización **L1** (Lasso): aplica una penalización proporcional a la suma de los valores absolutos de los coeficientes. Este tipo de regularización tiende a reducir algunos coeficientes exactamente a cero, lo que facilita la **selección automática de variables** y promueve modelos más simples.
 - Regularización **L2** (Ridge): aplica una penalización proporcional a la suma de los cuadrados de los coeficientes. En lugar de anularlos, esta técnica reduce su magnitud sin hacerlos exactamente cero. Es especialmente útil cuando existe **multicolinealidad entre variables independientes**, ya que distribuye el peso entre variables correlacionadas y mejora la estabilidad del modelo.
- **C**: controla la intensidad de la regularización aplicada al modelo. Un valor más bajo de C implica una regularización más fuerte, lo que restringe los coeficientes y reduce el riesgo de sobreajuste, aunque podría disminuir la capacidad del modelo para capturar relaciones complejas. Por el contrario, valores más altos de C reducen la penalización, permitiendo coeficientes más grandes y un mayor ajuste al conjunto de entrenamiento. En este análisis, se exploraron los valores [0.001, 0.01, 0.1, 1, 10] para encontrar el equilibrio óptimo entre **simplicidad del modelo y rendimiento predictivo**.

Como métrica de optimización se utilizó, igualmente que con Random Forest, el **F1-score**, dada la naturaleza binaria del problema y el interés en lograr un equilibrio entre precisión y recall en la detección de corrupción. Esto permite evaluar el rendimiento del modelo de forma equilibrada, evitando favorecer modelos que acierten en una clase a costa de fallar en la otra.

En la Figura 47 podemos ver que, tras 100 combinaciones evaluadas mediante validación cruzada, el mejor modelo correspondió a un *penalty* de tipo *Ridge* y un valor de *C* igual a 1.

Mejores hiperparámetros: {'logisticregression__C': 1, 'logisticregression__penalty': 'l2'}

Figura 47: Mejores hiperparámetros para el modelo de Regresión Logística.

En la Figura 48 se presenta la creación final del modelo de *Regresión Logística* con los hiperparámetros óptimos, el cual será utilizado en las fases posteriores de evaluación y análisis de importancia de variables.

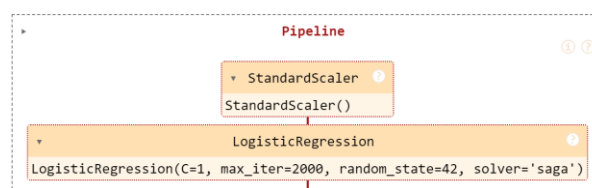


Figura 48: Creación del modelo de Regresión Logística con los hiperparámetros óptimos.

6.2.3. Evaluación del modelo

Una vez entrenado el modelo de **Regresión Logística** con los hiperparámetros óptimos obtenidos mediante *GridSearchCV*, se procedió a su evaluación sobre el conjunto de prueba. Para ello, se utilizaron diversas métricas estándar de clasificación: precisión (*precision*), exhaustividad (*recall*), F1-Score (media armónica entre precisión y recall) y exactitud (*accuracy*), además de la matriz de confusión y la curva ROC con su área bajo la curva (AUC).

Respecto a las métricas de clasificación, los resultados obtenidos se resumen en la Tabla 5. El modelo presenta un rendimiento perfecto en ambas clases:

- Para la clase Baja Corrupción (clase 0), el F1-Score fue de 1.
- Para la clase Alta Corrupción (clase 1), el F1-Score también fue de 1.

La exactitud global (*accuracy*) alcanzó el 100%, lo que significa que **todas las observaciones del conjunto de prueba fueron clasificadas correctamente**. También se calcularon dos promedios globales:

- **Macro average:** calcula la media simple de cada clase, útil para comprobar si el modelo trata las clases por igual.
- **Weighted average:** pondera el desempeño por el tamaño de cada clase, ofreciendo una evaluación más representativa del rendimiento general.

En este caso, ambos promedios también alcanzaron el valor máximo de 1, lo que indica un ajuste perfecto sin sesgos hacia ninguna clase.

	Precisión	Recall	F1-Score	Soporte
0 (Baja Corrupción)	1,0000	1,0000	1,0000	14
1 (Alta Corrupción)	1,0000	1,0000	1,0000	19
Accuracy global	–	–	1,0000	33
Macro avg	1,0000	1,0000	1,0000	33
Weighted avg	1,0000	1,0000	1,0000	33

Tabla 5: Informe de clasificación sobre el conjunto de prueba del modelo de Regresión Logística con todos los países.

En la Figura 49 se muestra la **matriz de confusión** del modelo. Esta confirma que el modelo clasificó correctamente: 14 de 14 observaciones con baja corrupción (clase 0) y 19 de 19 observaciones con alta corrupción (clase 1), sin ningún error de clasificación.

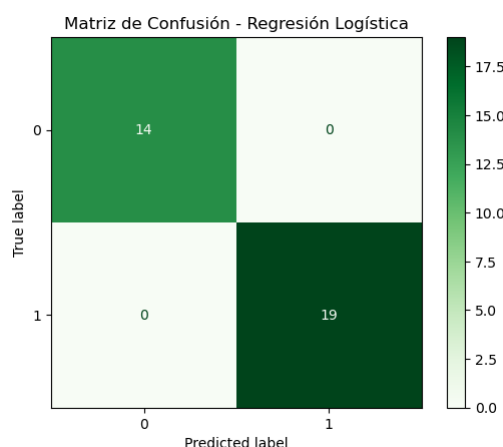


Figura 49: Matriz de confusión del modelo de Regresión Logística.

En la Figura 50 se presenta la **curva ROC** del modelo, generada a partir de las probabilidades de clase predichas. El área bajo la curva ($AUC = 1.00$) indica una **capacidad de discriminación perfecta entre ambas clases**, lo cual respalda los excelentes resultados obtenidos en las métricas de evaluación.

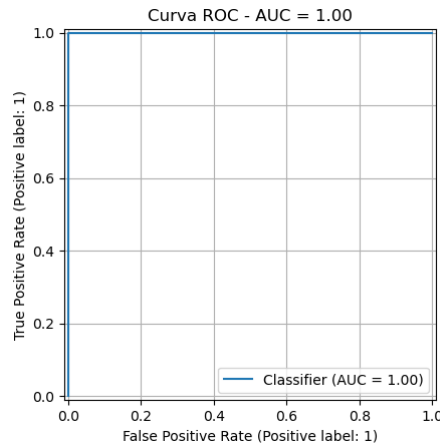


Figura 50: Curva ROC del modelo de Regresión Logística.

6.2.4. Validación cruzada

Tras obtener un rendimiento perfecto en el conjunto de prueba, se planteó la necesidad de **verificar la robustez del modelo de regresión logística**. Un desempeño del 100% puede indicar que el modelo está bien ajustado, pero también puede deberse a un leve sobreajuste, incluso a pesar de haber aplicado técnicas como SMOTE para balancear las clases y regularización L2 (Ridge) para controlar la complejidad del modelo.

Por este motivo, se aplicó un esquema de **validación cruzada estratificada de 10 particiones**, utilizando el conjunto total de observaciones disponibles. Este procedimiento permite evaluar **la estabilidad del rendimiento** del modelo a lo largo de múltiples subconjuntos de los datos, y comprobar si los resultados obtenidos son realmente consistentes y **generalizables**.

Además, esta validación es fundamental antes de interpretar la importancia de las variables explicativas, ya que garantiza que las relaciones detectadas por el modelo no dependen de una sola división entre entrenamiento y test, sino que son **consistentes a lo largo del conjunto de datos completo**.

Los resultados de las 10 iteraciones se presentan en la Figura 51, donde se muestran las métricas de accuracy, F1-score y ROC AUC para cada *fold*, junto con su media y desviación estándar.

El modelo alcanzó una precisión media del 99.38%, con un F1-score promedio de 0.9947 y un área bajo la curva ROC (AUC) media de 0.9984, lo que indica una **capacidad casi**

perfecta para distinguir entre casos de alta y baja corrupción. La baja variabilidad entre particiones, con desviaciones estándar inferiores al 2% en todas las métricas, confirma que el modelo mantiene un rendimiento estable, robusto y generalizable.

	Accuracy	F1-score	ROC AUC
0	1.00000	1.000000	1.000000
1	1.00000	1.000000	1.000000
2	1.00000	1.000000	1.000000
3	1.00000	1.000000	1.000000
4	1.00000	1.000000	1.000000
5	1.00000	1.000000	1.000000
6	1.00000	1.000000	1.000000
7	0.93750	0.947368	0.984127
8	1.00000	1.000000	1.000000
9	1.00000	1.000000	1.000000
Media	0.99375	0.994737	0.998413
Desviación estándar	0.01875	0.015789	0.004762

Figura 51: Resultados de la validación cruzada con Regresión Logística.

En la Figura 52 se muestra el informe de clasificación global obtenido con predicciones cruzadas. El modelo mantiene su excelente rendimiento:

- Clase 0 (baja corrupción): F1-score de 0.9927.
- Clase 1 (alta corrupción): F1-score de 0.9947.
- La *Accuracy* global es de 0.9938.

	precision	recall	f1-score	support
0	1.000000	0.985507	0.992701	69.000000
1	0.989362	1.000000	0.994652	93.000000
accuracy	0.993827	0.993827	0.993827	0.993827
macro avg	0.994681	0.992754	0.993677	162.000000
weighted avg	0.993893	0.993827	0.993821	162.000000

Figura 52: Informe de clasificación con validación cruzada para Regresión Logística.

La Figura 53 muestra la **matriz de confusión agregada**, que revela solo **un error de clasificación en 162 observaciones**, reforzando la alta capacidad discriminativa del modelo.

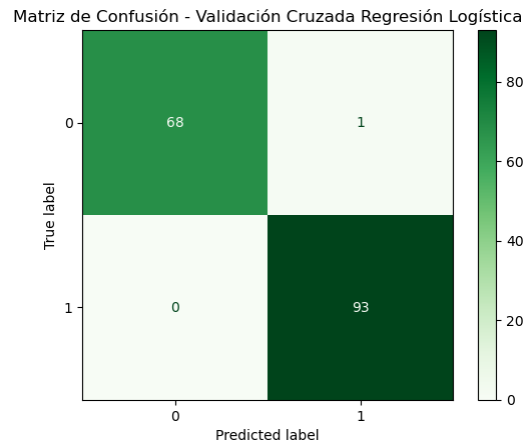


Figura 53: Matriz de confusión para validación cruzada en Regresión Logística.

Por último, la Figura 54 presenta la **curva ROC** con su área bajo la curva (AUC = 0.9984), lo que confirma una **capacidad casi perfecta de separación entre clases** incluso en validación cruzada.

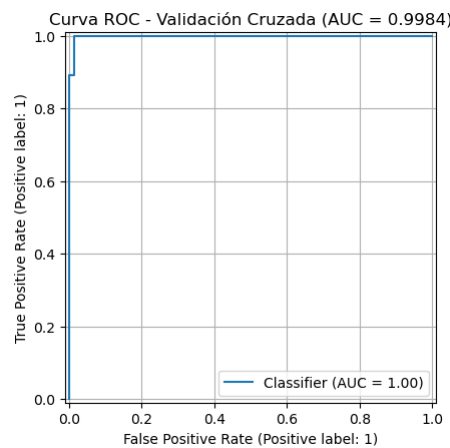


Figura 54: Curva ROC con la validación cruzada para Regresión Logística.

Los resultados obtenidos en la validación cruzada confirman que el modelo de Regresión Logística **mantiene su eficacia** de forma consistente a lo largo de diferentes subconjuntos del conjunto de datos. Su elevada precisión, la baja variabilidad entre iteraciones y el AUC cercano a 1.0 respaldan la **robustez y capacidad de generalización** del modelo. Esto refuerza la fiabilidad de los coeficientes estimados y, por tanto, la validez del análisis de importancia de variables.

6.2.5. Importancia de las variables

Al igual que Random Forest, el modelo de Regresión Logística permite identificar qué variables tienen mayor influencia sobre la variable objetivo, en este caso, la probabilidad de corrupción (*Bribery Class* = 1). La medida de importancia en este modelo se basa en el **valor absoluto de los coeficientes estimados**: cuanto mayor es el valor absoluto, mayor es el efecto que esa variable ejerce sobre la decisión del modelo. Este enfoque permite, por un lado, **ordenar las variables según su relevancia**, y por otro, interpretar **el sentido del efecto**.

En la Figura 55 se muestran todas las variables ordenadas por la magnitud de sus coeficientes.

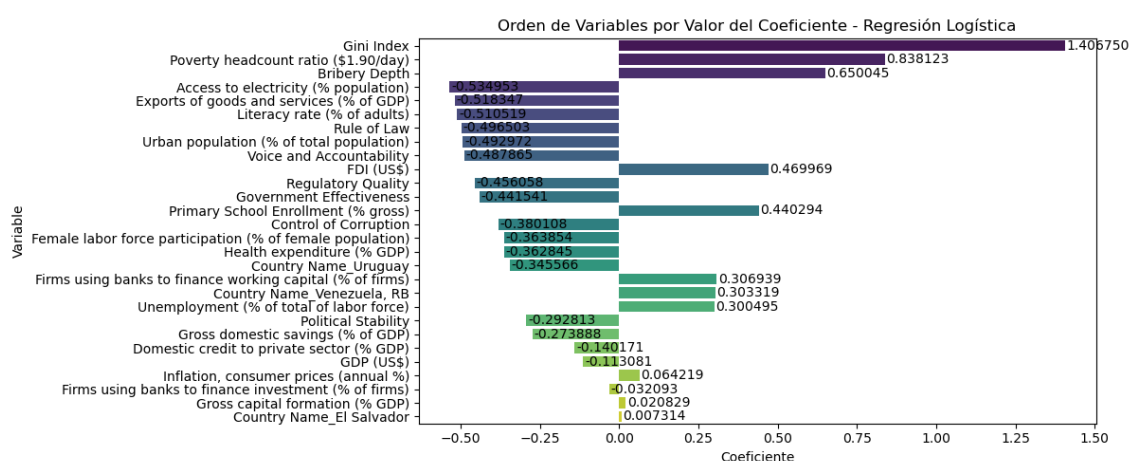


Figura 55: Orden de las variables por importancia absoluta en el modelo de Regresión Logística.

Cabe destacar que, una de las principales fortalezas del modelo de Regresión Logística es que no solo identifica las variables más influyentes, sino que permite interpretar el sentido del efecto que ejercen sobre la probabilidad de corrupción. Para facilitar esta interpretación, se han representado por separado las variables con **coeficiente positivo** (el aumento de esa variable incrementa la probabilidad de soborno) y las de **coeficiente negativo** (el aumento de esa variable actúa disminuye la probabilidad de soborno).

En la Figura 56 se presentan las 11 variables con coeficiente positivo, es decir, aquellas cuyo aumento está asociado con una **mayor probabilidad de corrupción**. Encabezan la lista factores estructurales críticos como el *Gini Index*, la pobreza extrema (*Poverty Headcount Ratio*) y la profundidad del soborno (*Bribery Depth*). Estos resultados

refuerzan la idea de que los niveles elevados de desigualdad, la precariedad económica y las prácticas corruptas previas alimentan entornos propensos al soborno.

También se observan variables como el desempleo y la inflación, cuyo efecto positivo sobre la corrupción puede interpretarse como reflejo de que, en situaciones de incertidumbre laboral o pérdida de poder adquisitivo, el soborno puede emerger como un mecanismo compensatorio o de supervivencia.

Resulta interesante la aparición de variables como *Primary School Enrollment* y *FDI* (inversión extranjera directa), tradicionalmente asociadas al desarrollo. Sin embargo, en este análisis presentan un coeficiente positivo, lo que sugiere que **su mera presencia no es suficiente para reducir la corrupción** si no va acompañada de mejoras institucionales efectivas.

En el caso de la matrícula escolar primaria, un mayor acceso a la educación no garantiza un sistema educativo de calidad ni libre de corrupción. Así, un sistema con alta cobertura, pero mal gestionado, capturado por intereses políticos o marcado por ineficiencias, puede no traducirse en un entorno más ético reproduciendo entornos corruptos.

Por otro lado, la inversión extranjera puede incentivar prácticas informales cuando se da en contextos con poca regulación, priorizando la rentabilidad sobre el cumplimiento ético o laboral.

Por último, destaca el coeficiente positivo de *Country Name_Venezuela, RB*, lo que sugiere que, manteniendo constantes el resto de las variables, pertenecer al caso venezolano incrementa significativamente la probabilidad de corrupción según el modelo. En contraste, el valor del país El Salvador es prácticamente nulo, lo que refleja un contexto de corrupción moderada o intermedia.

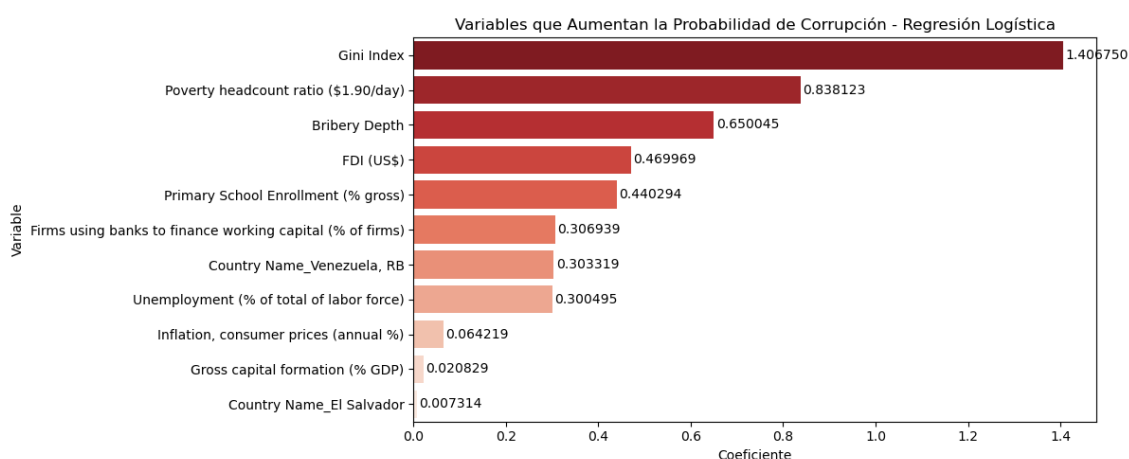


Figura 56: Orden por importancia de las variables que aumentan la probabilidad de corrupción en el modelo de Regresión Logística.

La Figura 57 muestra las 17 variables con coeficientes negativos en el modelo de Regresión Logística, es decir, aquellas cuyo aumento se asocia con una **menor probabilidad de corrupción**. Estas variables representan factores estructurales, sociales e institucionales que actúan como **barreras frente a las prácticas corruptas**.

En primer lugar, destacan variables de infraestructura y acceso equitativo a servicios básicos, como el acceso a la electricidad y el nivel de exportaciones. Su efecto negativo sugiere que una mayor cobertura energética y una mayor apertura comercial tienden a asociarse con contextos más institucionalizados, donde la corrupción es menos necesaria o tolerada para acceder a bienes y servicios.

La alfabetización, junto a Rule of Law y Voice and Accountability, refuerzan la idea de que el fortalecimiento institucional y la participación ciudadana son fundamentales para reducir la corrupción. Estos factores tienden a favorecer la transparencia, la rendición de cuentas y la cultura cívica. La alfabetización, en concreto, puede reflejar una mayor capacidad de los ciudadanos para conocer sus derechos y exigirlos sin recurrir a sobornos.

Del mismo modo, la calidad regulatoria, la eficacia gubernamental y el control de la corrupción aparecen como pilares centrales. Su presencia dentro del grupo protector de corrupción respalda la hipótesis de que **la corrupción no es solo un problema ético o individual, sino un fallo sistémico en la calidad de las instituciones**.

El modelo también reconoce como factores protectores la urbanización, la participación femenina en el mercado laboral y el gasto en salud, los cuales apuntan a una mayor

inclusión social y cohesión estructural. Estas variables contribuyen a disminuir la vulnerabilidad de la población ante pagos informales, al ampliar el acceso equitativo a derechos y oportunidades.

Entre las variables económicas, destacan el ahorro interno bruto, el crédito al sector privado y el PIB per cápita (*GDP*), todas ellas asociadas a entornos de más formales y de menor dependencia de canales informales. Estos elementos fomentan una mayor transparencia económica y menor necesidad de sobornos para operar o acceder a recursos.

Por último, podemos ver que Uruguay aparece con un coeficiente negativo, lo que refuerza su perfil de país con instituciones sólidas y bajos niveles de corrupción dentro del modelo.

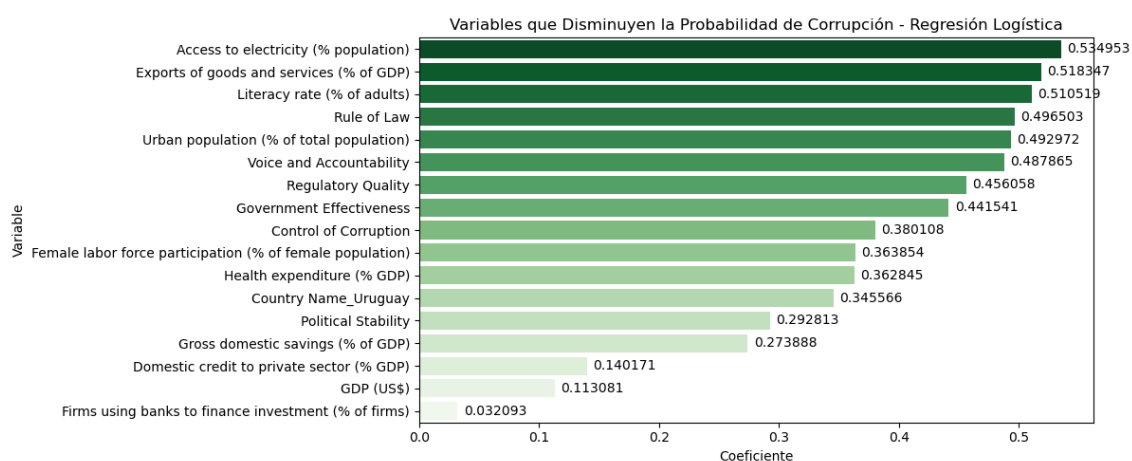


Figura 57: Orden por importancia de las variables que disminuyen la probabilidad de corrupción en el modelo de Regresión Logística.

En conjunto, el modelo de Regresión Logística no solo permite identificar las variables más influyentes en la probabilidad de corrupción, sino también **comprender con claridad el sentido del efecto que cada una ejerce**.

Esta capacidad interpretativa ha permitido distinguir entre factores que agravan la corrupción, como la desigualdad, la pobreza extrema, la inflación o la precariedad laboral, y aquellos que la atenúan, como la inclusión social, la calidad institucional o el acceso a servicios básicos. La coincidencia de varias variables clave con los resultados obtenidos en Random Forest refuerza la **validez y robustez de los patrones detectados**.

En definitiva, este análisis confirma que la corrupción no es un fenómeno aislado, sino el resultado de desequilibrios estructurales que pueden ser explicados, medidos y enfrentados mediante evidencia empírica. La combinación de técnicas explicativas y predictivas permite no solo anticipar contextos de riesgo, sino también fundamentar políticas públicas orientadas a fortalecer la integridad institucional y reducir las condiciones que permiten el soborno.

6.2.6. Estudio por países con Regresión Lineal

Con el objetivo de identificar los factores que explican los distintos niveles de corrupción observados en Venezuela, El Salvador y Uruguay, se aplicó un modelo de **Regresión Lineal Ridge** de forma independiente para cada país. Esta variante de la regresión lineal incorpora una penalización sobre los coeficientes del modelo, lo que permite **reducir el riesgo de sobreajuste y mitigar problemas de multicolinealidad** entre las variables predictoras, especialmente útil en conjuntos de datos con muchas variables o alta correlación entre ellas.

Este enfoque permite examinar cómo varía la influencia de diferentes variables económicas, sociales e institucionales según el contexto nacional, tomando como variable dependiente la incidencia de sobornos (*Bribery Incidence*).

La Figura 58 muestra las métricas de ajuste para cada modelo nacional. El coeficiente de determinación R^2 indica que los modelos explican una proporción muy elevada de la variabilidad en la incidencia de sobornos, con un valor para Uruguay de 0.99 y El Salvador 0.94. En el caso de Venezuela, el modelo alcanza un ajuste perfecto, lo que refleja la fuerte capacidad explicativa de las variables seleccionadas en este contexto concreto.

	País	RMSE	MAE	R^2
0	Venezuela, RB	0.000068	0.000039	1.000000
1	El Salvador	0.067482	0.042216	0.939641
2	Uruguay	0.005205	0.003179	0.988176

Figura 58: Métricas del modelo Regresión Lineal por país.

Las figuras: Figura 59, Figura 60 y Figura 61 presentan las 15 variables más relevantes en cada país, según los coeficientes estimados por el modelo. Estas variables permiten

interpretar cómo determinados factores están asociados a un mayor o menor nivel de corrupción percibida en el ámbito empresarial. Cabe mencionar que, un coeficiente positivo (+) indica que el aumento en esa variable está asociado a un **mayor nivel de corrupción** (medido como *Bribery Incidence*), mientras que un coeficiente negativo (-) implica que un mayor valor de esa variable está asociado a una **reducción en la corrupción**.

En la Figura 59 podemos ver que, en el caso de **Venezuela**, las variables con mayor peso están directamente asociadas al contexto financiero e institucional del país:

- *Firms using banks to finance working capital* (+) presenta un coeficiente positivo, lo que sugiere que el uso de servicios bancarios para financiar el capital de trabajo está relacionado con mayores niveles de soborno, posiblemente debido corrupción en procesos administrativos.
- *Firms using banks to finance investment* (-) tiene un coeficiente negativo, lo que indica que una mayor proporción de empresas que acceden a financiamiento bancario para inversión está asociada con una reducción de la corrupción. Esto puede interpretarse como un efecto positivo del acceso formal a crédito productivo sobre la transparencia empresarial.
- Variables como *Regulatory Quality* (-) y *Urban population* (-) presentan coeficientes negativos, lo que sugiere que una mejor calidad institucional y una mayor urbanización podrían estar asociadas a una reducción del soborno.
- Además, cabe destacar que, aunque conceptualmente la **inflación** suele asociarse a inestabilidad institucional y corrupción, en el caso de Venezuela el modelo de regresión lineal resulta en un **coeficiente negativo**, lo cual podría explicarse por efectos estructurales propios de economías en hiperinflación: reducción de la actividad económica formal, desplazamiento de prácticas corruptas fuera del sistema encuestado o colinealidad con otras variables más directamente vinculadas a la corrupción.

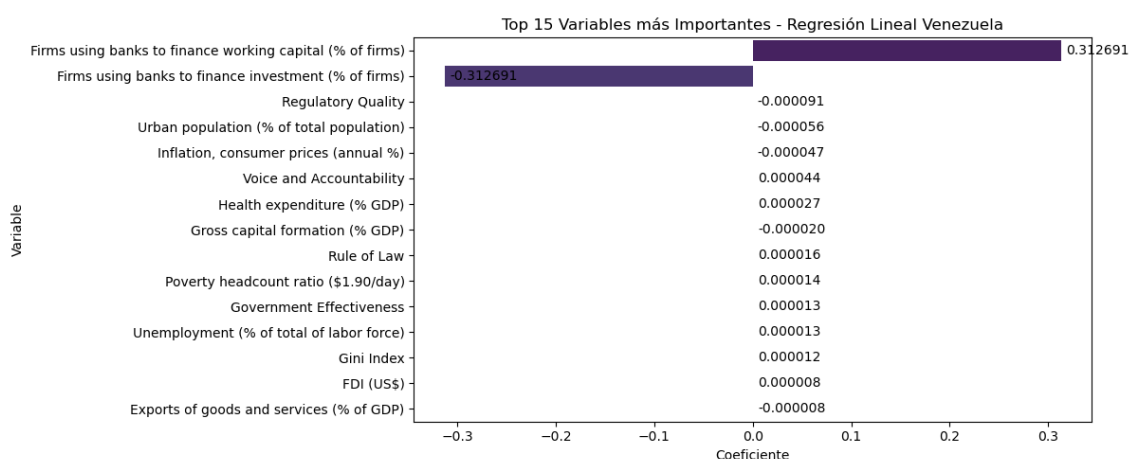


Figura 59: Top 15 Variables más importantes para Venezuela según el modelo de Regresión Lineal.

En la Figura 60 podemos observar que, en **El Salvador** los resultados reflejan un contexto mixto, donde ciertas mejoras sociales coexisten con vulnerabilidades:

- El uso de la banca para financiar inversiones (*Firms using banks to finance investment* (-)) es la variable que más contribuye a reducir la corrupción.
- En cambio, variables como *Primary School Enrollment* (+), *Unemployment* (+) y *Voice and Accountability* (+) se asocian con un aumento en los niveles de corrupción. Esto puede interpretarse como una señal de tensiones en la estructura institucional: aunque existan mejoras educativas o participación ciudadana, estas aún no se traducen en menor corrupción.
- La variable *Gini Index* (+) también tiene un impacto positivo sobre la corrupción, indicando que una mayor desigualdad se asocia a una mayor incidencia de sobornos.

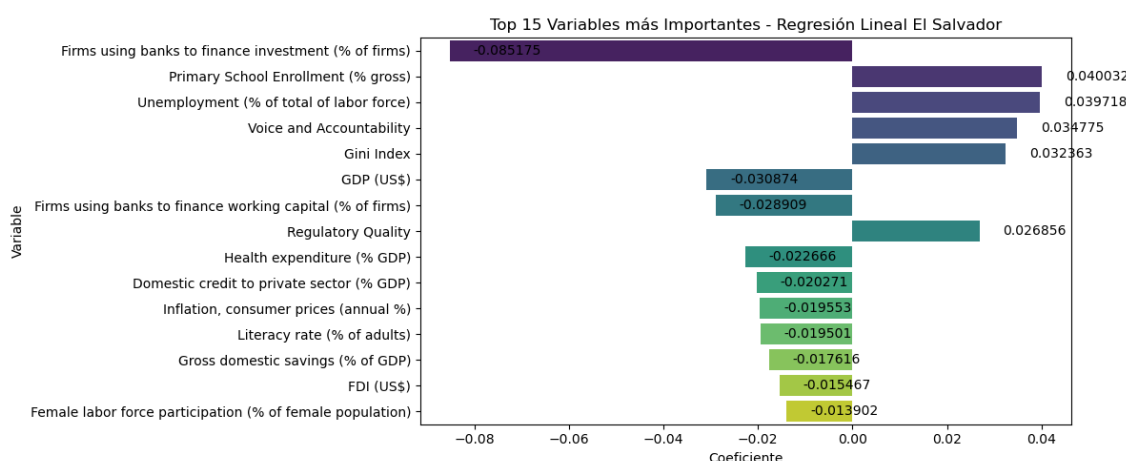


Figura 60: Top 15 Variables más importantes para El Salvador según el modelo de Regresión Lineal.

En la Figura 61 se observa que en el caso de **Uruguay**, el modelo de regresión lineal refleja una combinación de factores económicos, institucionales y de acceso a servicios como los principales determinantes de los niveles de corrupción empresarial. Aunque el país presenta bajos niveles de soborno en general, el modelo permite identificar qué variables contribuyen positiva o negativamente a explicar su variabilidad interna.

- *Firms using banks to finance investment* (+) y *Firms using banks to finance working capital* (+) presentan los coeficientes más altos y positivos del modelo. Esto sugiere que, en Uruguay, un mayor uso del financiamiento bancario está asociado a un **leve aumento en los niveles de corrupción percibida**. Esta relación podría indicar que algunos procesos administrativos asociados al acceso al crédito siguen siendo susceptibles a prácticas irregulares.
- Entre los factores que **reducen la corrupción**, destacan *Government Effectiveness*, *Access to electricity*, *Rule of Law* y *Urban population*. Estos resultados refuerzan la idea de que una gobernanza eficaz, el acceso amplio a servicios públicos y un entorno urbano consolidado siguen actuando como elementos estabilizadores que limitan la aparición de sobornos.
- Llama la atención que las variables *Political Stability* (+) y *Control of Corruption* (+), pese a su definición institucional como indicadores de buena gobernanza, aparecen con un coeficiente positivo en el modelo aplicado a Uruguay.
 - Este resultado puede explicarse por **efectos estadísticos**, como la **multicolinealidad** con otras variables institucionales altamente

correlacionadas, lo que puede distorsionar la interpretación directa del signo.

- Además, desde un punto de vista **contextual**, es posible que este índice no actúe como un factor preventivo, sino como una **respuesta institucional ante el aumento de prácticas corruptas**. Es decir, a mayor nivel de soborno, mayor esfuerzo del Estado por controlarlo, generando así una **correlación positiva no causal** entre ambas variables.

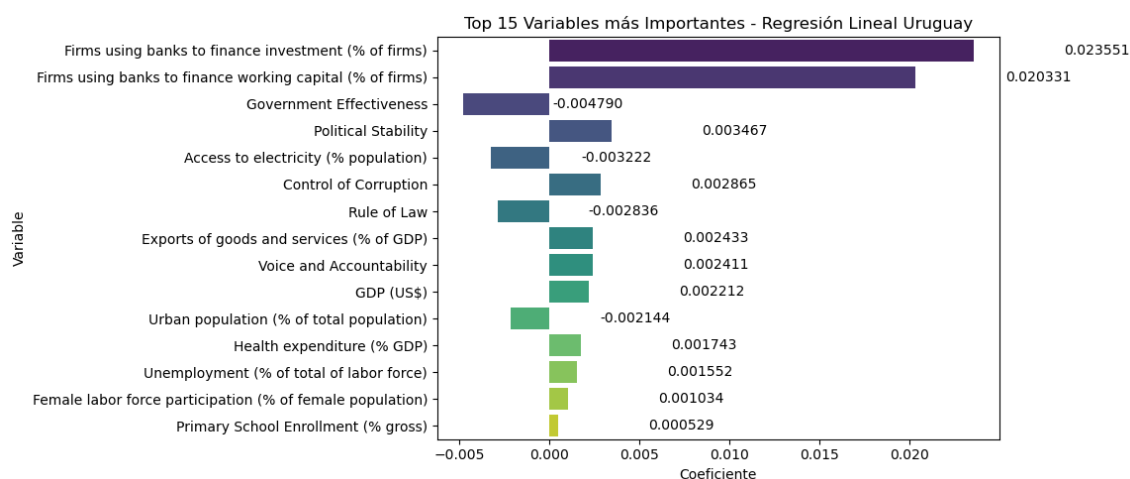


Figura 61: Top 15 Variables más importantes para Uruguay según el modelo de Regresión Lineal.

La Figura 62 resume comparativamente las cinco variables más influyentes en la explicación de la corrupción empresarial por país, distinguiendo si su efecto es positivo (aumenta la corrupción) o negativo (la reduce).

	Venezuela	El Salvador	Uruguay
Top 1	Bribery Depth (+)	Firms using banks to finance investment (-)	Firms using banks to finance investment (+)
Top 2	Firms using banks to finance investment (-)	Primary School Enrollment (+)	Firms using banks to finance working capital (+)
Top 3	Firms using banks to finance working capital (+)	Unemployment (+)	Government Effectiveness (-)
Top 4	Regulatory Quality (-)	Voice and Accountability (+)	Political Stability (+)
Top 5	Urban population (-)	Gini Index (+)	Access to electricity (-)

Figura 62: Comparación de las 5 variables más importantes por país según el modelo de Regresión Lineal. El símbolo (+) indica que la variable contribuye a aumentar la corrupción; (-) indica que está asociada a una menor corrupción.

El análisis comparativo evidencia que **los factores que explican la corrupción varían significativamente entre países**, reflejando sus respectivos niveles de desarrollo institucional, inclusión financiera y contexto social.

- En **Venezuela**, el soborno se asocia a restricciones financieras estructurales y fragilidad institucional.
- En **El Salvador**, hay una mezcla de variables con signo positivo y negativo. Es decir, se observa un escenario de transición, donde los avances en desarrollo humano conviven con desigualdades estructurales que continúan facilitando la aparición de sobornos.
- En **Uruguay**, el entorno institucional sólido actúa como protección frente al soborno.

En conjunto, estos resultados confirman que la corrupción no es un fenómeno homogéneo, sino que está determinada por **dinámicas específicas a cada país**, y que **su reducción efectiva requiere políticas integradas** que combinen gobernanza, acceso equitativo al crédito y fortalecimiento de los servicios públicos.

7. Discusión de los resultados

Este apartado analiza de forma integrada los resultados obtenidos por los cuatro modelos desarrollados: dos globales (Random Forest Classifier y Regresión Logística) y dos aplicados por país (Random Forest Regressor y Regresión Lineal Ridge). En conjunto, los modelos muestran un rendimiento altamente robusto, lo que otorga solidez a los análisis posteriores.

En el caso de los modelos de clasificación global, el Random Forest Classifier alcanzó una precisión del 98.15%, mientras que la Regresión Logística obtuvo un valor aún superior, con 99.38% de precisión en validación cruzada. En cuanto a los modelos por país, el Random Forest Regressor mostró un coeficiente de determinación R^2 de 0.8937 para Venezuela, 0.9679 para El Salvador y 0.9858 para Uruguay. Por su parte, la Regresión Lineal Ridge alcanzó un R^2 igual a 1.0000 en Venezuela, 0.9396 en El Salvador y 0.9882 en Uruguay.

Estos resultados indican que los modelos han capturado de forma adecuada la estructura subyacente de los datos y han logrado explicar una proporción muy elevada de la variabilidad en los niveles de corrupción reportados. Por tanto, las conclusiones derivadas del análisis de la importancia de las variables en cada uno de ellos pueden considerarse válidas y representativas de las dinámicas reales de la corrupción.

7.1. Random Forest Classifier

Por otro lado, un aspecto destacado es que el Random Forest Classifier no se apoya fuertemente en la variable "país" en sí misma para predecir el nivel de corrupción, sino en características estructurales subyacentes. En la Figura 41 se presenta la importancia relativa de las variables dummy asociadas a cada país dentro del modelo, evidenciando que su contribución es prácticamente nula (incluso la dummy *Country Name_El Salvador* tiene importancia 0 en el modelo). Es decir, el algoritmo **no “aprende” simplemente a etiquetar la corrupción por país**, sino que basa su decisión en otras variables más explicativas como indicadores socioeconómicos y de gobernanza.

Esto es un resultado positivo desde el punto de vista analítico: sugiere que el modelo está capturando patrones generales de corrupción vinculados a condiciones estructurales (pobreza, desigualdad, calidad institucional, gasto social, etc.) aplicables a cualquier país,

en lugar de limitarse a reconocer etiquetas nacionales. En otras palabras, **la pertenencia a un país específico no determina por sí sola la predicción del modelo**, sino que son los factores socioeconómicos e institucionales los que realmente explican si un caso se clasifica como de alta o baja corrupción.

Si bien la importancia de cada variable individual en el Random Forest global se reparte entre muchos predictores, el análisis sugiere que entre las variables más influyentes para explicar la corrupción a nivel global figuran aquellas relacionadas con **desigualdad y pobreza**, medidas de **incidencia de soborno y calidad institucional**.

Por ejemplo, indicadores como el *Gini Index* (desigualdad de ingresos), la incidencia de pobreza extrema (*Poverty Headcount Ratio*), el gasto en salud (*Health expenditure*) o la frecuencia con que las empresas pagan sobornos (*Bribery Depth*) destacan como factores asociados a mayor corrupción en distintos contextos. Asimismo, las seis variables de gobernanza (*Voice and Accountability*, *Government Effectiveness*, *Control of Corruption*, *Regulatory Quality*, *Rule of Law* y *Political Stability*), junto con indicadores de desarrollo socioeconómico (por ejemplo, tasa de alfabetización, acceso a electricidad, uso de servicios bancarios, PIB per cápita), aparecen vinculados a entornos de menor corrupción. Estas tendencias cualitativas se confirmarán con mayor claridad al contrastar con el modelo logístico, que permite identificar explícitamente el signo del efecto de cada variable.

En síntesis, la coherencia entre la Figura 39 (clases reales) y la Figura 40 (clases predichas) avala la capacidad predictiva del modelo Random Forest Classifier global. Además, su independencia de la variable país, refuerza la idea de que existen determinantes universales detrás de la corrupción que el modelo ha sabido aprovechar.

7.2. Regresión Logística

El modelo de Regresión Logística global ofrece no solo una buena capacidad predictiva sino también una ventana a la interpretación de los factores que aumentan o disminuyen la probabilidad de corrupción. En la Figura 55 se muestran todas las variables ordenadas por la magnitud absoluta de sus coeficientes en la regresión logística. Esta visualización permite identificar cuáles son los predictores más influyentes en la clasificación de

corrupción a nivel global. Además, al separar los coeficientes positivos y negativos, el modelo logístico aporta información valiosa sobre **el sentido del efecto** de cada variable.

Factores que aumentan la probabilidad de corrupción: coeficientes positivos

La Figura 56 destaca las variables cuyo aumento se asocia con una mayor probabilidad de pertenecer a la clase de Alta Corrupción. Encabezando esta lista se encuentran **tres variables estructurales críticas**: el Índice de Gini, la proporción de población en pobreza extrema (*Poverty Headcount Ratio*) y la Profundidad del soborno (*Bribery Depth*). Estos resultados refuerzan la idea de que **mayores niveles de desigualdad, precariedad económica y prácticas corruptas arraigadas** alimentan un entorno proclive al soborno. Junto a ellas aparecen otras variables con coeficientes positivos significativos, como la tasa de desempleo y la inflación. Un aumento en el desempleo o en la inestabilidad de precios suele traducirse en mayor vulnerabilidad económica, lo que puede empujar a individuos y empresas a recurrir al soborno como mecanismo compensatorio o de supervivencia en contextos de incertidumbre.

Llama la atención la presencia de algunos indicadores tradicionalmente asociados al desarrollo que, en este análisis, presentan coeficientes positivos. Por ejemplo, una mayor tasa de matriculación en primaria (*Primary School Enrollment*) se correlaciona con más corrupción, al igual que un incremento en la Inversión Extranjera Directa (*FDI*). A primera vista, esto podría parecer contradictorio, pues la educación y la inversión externa suelen considerarse fuerzas modernizadoras que reducen la corrupción. Sin embargo, el modelo sugiere que su mera existencia no garantiza mejoras en integridad. Es posible que una alta cobertura educativa no se traduzca en un sistema de calidad o ético – por ejemplo, si el sistema escolar está mal gestionado o infiltrado por intereses particulares, una mayor matrícula por sí sola no disminuye las prácticas corruptas. De modo similar, la entrada de inversión extranjera en entornos con débil regulación puede incentivar prácticas informales (sobornos, elusión de normas), con lo cual mayores flujos de *FDI* podrían asociarse a más corrupción en ausencia de controles adecuados.

Por último, dentro de los factores que elevan el riesgo, aparece significativa la dummy de país para Venezuela. El coeficiente positivo y elevado de *Country Name_Venezuela* sugiere que, manteniéndose el resto de los indicadores estables, pertenecer a Venezuela incrementa significativamente la probabilidad de corrupción. Esto indica la existencia de

características no capturadas por las demás variables que hacen de Venezuela un entorno inherentemente más corrupto en el periodo analizado.

Factores que disminuyen la probabilidad de corrupción: coeficientes negativos

La Figura 57 reúne las variables cuyo aumento se asocia con una menor probabilidad de soborno, es decir, factores “protectores” frente a la corrupción. En primer lugar, destacan variables de infraestructura y apertura económica como el acceso a la electricidad y el nivel de exportaciones. Su efecto negativo indica que una mayor cobertura de servicios básicos y una mayor integración comercial tienden a acompañar entornos más institucionalizados, donde el soborno es menos necesario o tolerado para acceder a bienes y servicios.

Asimismo, les siguen varios indicadores de calidad institucional y capital humano. La tasa de alfabetización, junto con los índices de Estado de Derecho (*Rule of Law*) y Voz y Rendición de Cuentas (*Voice and Accountability*), aparecen entre las variables protectoras más importantes. Esto respalda la hipótesis de que un fortalecimiento institucional (mejores leyes, mayor accountability gubernamental) y una ciudadanía educada y empoderada crean barreras efectivas contra las prácticas corruptas. Sociedades más alfabetizadas, donde los ciudadanos conocen y exigen sus derechos, tienden a mostrar menor tolerancia al soborno.

Del mismo modo, indicadores de buen gobierno como la Calidad Regulatoria, la Eficacia Gubernamental y el propio índice de Control de la Corrupción se incluyen en este grupo: su influencia negativa y destacada refuerza la idea de que la corrupción no es solo un asunto moral individual, sino un síntoma de fallas sistémicas en la calidad institucional. Los resultados del modelo logístico confirman que allí donde las instituciones son sólidas y eficientes, las oportunidades y necesidades de incurrir en sobornos disminuyen drásticamente.

Otros factores socioeconómicos complementan el perfil de protección. Por ejemplo, una mayor tasa de urbanización, una mayor participación femenina en el mercado laboral, y un incremento en el gasto en salud aparecen con coeficientes negativos. Todos ellos apuntan a **mayor inclusión social y cohesión estructural**: poblaciones mayoritariamente urbanas (con mejor acceso a servicios públicos), economías que incorporan a la mujer de

forma amplia en el trabajo remunerado, y Estados que invierten más en salud, tienden a generar ambientes menos propensos a la informalidad y, por tanto, al soborno.

En el plano macroeconómico, variables como el ahorro interno bruto, el crédito al sector privado y el PIB per cápita también se asocian a entornos más transparentes y menos dependientes de canales informales: una mayor profundidad financiera y prosperidad económica reducen la motivación para recurrir a la corrupción como atajo.

Cabe notar que incluso la variable *dummy* de Uruguay se incluye con coeficiente negativo en esta lista, subrayando cuantitativamente el estatus de Uruguay como caso de instituciones fuertes y bajos niveles de soborno, consistente con las observaciones empíricas.

En conjunto, la Regresión Logística proporciona evidencia de que los **principales impulsores de la corrupción a nivel global son factores estructurales**: desigualdad, pobreza, informalidad económica (medida por sobornos previos), desempleo e inestabilidad macroeconómica. Paralelamente, **los factores que frenan la corrupción** son la inclusión social (educación básica ampliamente disponible, participación femenina), el desarrollo económico formal (ahorro, crédito, urbanización) y, sobre todo, la **calidad institucional** (buen gobierno, estado de derecho, voces ciudadanas activas). La coincidencia de muchas de estas variables clave con las identificadas por el Random Forest Classifier refuerza la robustez de los patrones detectados. En definitiva, ambos enfoques globales coinciden en pintar un panorama donde la corrupción surge no aleatoriamente, sino como **resultado de desequilibrios sociales, económicos e institucionales** identificables mediante datos.

7.3. Random Forest Regressor – Análisis por país

Debido a las peculiaridades de los datos por país – en particular, el hecho de que Uruguay no presentó ningún caso de alta corrupción en la muestra – se optó por un enfoque de modelado específico para cada país usando Random Forest en versión regresión. En lugar de predecir una clase binaria, estos modelos predicen directamente el valor continuo de la incidencia de sobornos (*Bribery Incidence*) en cada país, permitiendo evaluar qué factores explican mejor los niveles observados de corrupción empresarial en contextos nacionales particulares.

Venezuela (Figura 43): El modelo resalta principalmente variables de índole **financiera y laboral**. Entre las más importantes se encuentran: la proporción de empresas que usan bancos para financiar el capital de trabajo, la participación femenina en la fuerza laboral, la profundidad de sobornos (*Bribery Depth*), el porcentaje de empresas que usan financiamiento bancario para inversiones, y el PIB del país. Este conjunto de variables sugiere un entorno de precariedad económica e informalidad: un acceso limitado al financiamiento formal y una baja inclusión laboral (especialmente de las mujeres) pintan un escenario donde las empresas operan con restricciones, y el soborno emerge como un mecanismo informal para acceder a recursos u oportunidades. La fuerte presencia de variables financieras (uso de la banca) indica que en Venezuela la falta de crédito formal accesible podría estar vinculada a mayores niveles de soborno: las empresas que no logran financiación por canales normales pueden recurrir a “facilitar” trámites mediante pagos indebidos. Asimismo, la alta importancia de *Bribery Depth* sugiere que donde hay soborno, este tiende a ser profundo y generalizado, perpetuando un círculo vicioso. En suma, el perfil de importancia en Venezuela apunta a que su elevada corrupción es un **fenómeno sistémico**, alimentado por la combinación de restricciones estructurales (financieras, laborales) y la consiguiente normalización del soborno como parte del funcionamiento empresarial.

El Salvador (Figura 44): Para El Salvador, país con niveles moderados de corrupción, las variables más influyentes según el Random Forest se vinculan en buena medida con el **desarrollo socioeconómico**. Destacan factores como el porcentaje de empresas que utilizan la banca para financiar sus inversiones, el PIB, el porcentaje de población urbana, el acceso a la electricidad y el Índice de Gini. A diferencia de Venezuela, aquí aparecen con fuerza variables de infraestructura y modernización económica: una mayor urbanización y electrificación suelen acompañar mejoras institucionales y de control, mientras que un mayor uso del crédito bancario para inversión señala una formalización creciente de la economía. De hecho, el modelo sugiere que en **El Salvador el fortalecimiento de la infraestructura y del sistema financiero formal se asocia con reducciones graduales de la corrupción**. Sin embargo, la presencia simultánea del Índice de Gini como variable importante recuerda que la **desigualdad** sigue siendo un factor de riesgo: altas brechas socioeconómicas pueden estar frenando una mejora más rápida en los niveles de integridad. En resumen, el patrón salvadoreño es **mixto**: por un

lado, indicadores positivos de desarrollo económico tiran hacia una menor corrupción; por otro, persisten vulnerabilidades (desigualdad, posiblemente ciertos nichos de pobreza o informalidad) que mantienen el soborno presente.

Uruguay (Figura 45): Uruguay presenta niveles muy bajos de corrupción, y el modelo por país busca explicar qué condiciones sustentan ese estado. Las variables clave señaladas por el Random Forest para Uruguay incluyen la población urbana, el porcentaje de empresas que financian inversiones con la banca, el PIB, la tasa de alfabetización y la profundidad de soborno (*Bribery Depth*). Este resultado confirma que Uruguay se distingue por un **entorno social, educativo e institucional sólido**: una población altamente urbanizada y alfabetizada, una economía formal robusta y con amplio acceso a servicios financieros, y en general una baja prevalencia de sobornos (la aparición de la variable *Bribery Depth* se explica con que los pocos casos de soborno que ocurren podrían estar asociados a situaciones particulares detectables). Las primeras posiciones las ocupan variables que reflejan **inclusión y desarrollo**: prácticamente toda la población uruguaya vive en zonas urbanas con acceso a infraestructura, y el crédito bancario para inversiones es común, lo que sugiere transparencia en el financiamiento empresarial. La alta alfabetización indica un capital humano capaz de ejercer control social y demandar rendición de cuentas, coherente con un entorno de baja corrupción. Cabe destacar que **los niveles de corrupción en Uruguay son tan bajos que la variabilidad interna es reducida**; aun así, el modelo logra identificar qué factores explican las pequeñas variaciones observables. Pero en general, el perfil de Uruguay sugiere que su inmunidad relativa frente a la corrupción se apoya en **altos niveles de desarrollo humano e institucional**.

Finalmente, la Figura 46 proporciona una comparación directa de las cinco variables más importantes en los modelos de Venezuela, El Salvador y Uruguay. Este resumen visual evidencia que, si bien hay **factores comunes** que se repiten como determinantes en los tres países (por ejemplo, el acceso al financiamiento bancario y el nivel de PIB aparecen en todos), **el peso y la influencia de cada uno varían considerablemente según el nivel de desarrollo y el contexto institucional de cada nación**. Esta comparación refuerza la noción de que **la corrupción en cada país está explicada por combinaciones particulares de variables**, aunque dentro de un conjunto de factores estructurales compartidos.

7.4. Regresión Lineal Ridge – Análisis por país

En el estudio de **Regresión Lineal Ridge** que se llevó a cabo de forma independiente para cada país (Venezuela, El Salvador y Uruguay), al igual que con el Random Forest por país, la variable dependiente es la incidencia de sobornos (*Bribery Incidence*) y el objetivo es comprender cómo varía la influencia de distintos factores económicos, sociales e institucionales según el contexto nacional. A diferencia del Random Forest, la regresión lineal permite identificar **el signo del efecto** (positivo o negativo) de cada variable en cada país, aportando un matiz interpretativo adicional.

Venezuela (Figura 59): El modelo de regresión lineal confirma en gran medida los resultados obtenidos con Random Forest, aunque brinda información sobre la dirección de las relaciones. Se confirma que el entorno venezolano está fuertemente influenciado por factores financieros e institucionales.

La variable “Empresas que usan bancos para financiar capital de trabajo” aparece con uno de los coeficientes más altos y positivos, lo que indica que un mayor uso del financiamiento bancario de corto plazo está asociado a un aumento en la corrupción. Esto podría parecer contraintuitivo, pues cabría esperar que el acceso a crédito formal redujera la necesidad de sobornos, pero en el contexto venezolano sugiere que obtener capital de trabajo formal podría implicar mayores contactos con funcionarios o sistemas débiles, exponiendo a las empresas a más solicitudes de pagos informales.

Contrastando con lo anterior, el uso de la banca para financiar inversiones de capital fijo tiene un coeficiente negativo importante. Una posible interpretación es que la inversión de largo plazo en entornos formales conlleva mayor planificación, reduciendo la improvisación y las oportunidades para la corrupción en las operaciones empresariales. Dicho de otro modo, cuando las empresas pueden financiar su crecimiento por vías legales y planificadas, dependen menos de atajos corruptos para funcionar.

Entre los demás factores significativos en Venezuela, la calidad regulatoria y la población urbana presentan coeficientes negativos, reforzando la noción de que mejores instituciones y mayor modernización (urbanización) van de la mano con menos sobornos.

Un resultado inesperado del modelo de Venezuela es el coeficiente negativo de la inflación, que sugiere que a mayor inflación, menor incidencia de sobornos. Esta relación,

contraria a la tendencia global, puede explicarse por las **condiciones extremas de hiperinflación** en Venezuela, que llevaron al colapso de la economía formal y a la reducción de interacciones empresa-Estado, desplazando muchas prácticas corruptas al ámbito informal y fuera del alcance de las encuestas empresariales sobre las que se basa la variable *Bribery Incidence*. Por tanto, con esta hipótesis **la corrupción no desaparece**, sino que se dificulta su medición.

Como se observa en la Figura 13, correspondiente a la matriz de correlaciones para Venezuela, la inflación presenta correlaciones negativas con indicadores como el PIB, el acceso al crédito bancario, la formación bruta de capital o el crédito al sector privado, lo cual refleja el colapso económico estructural que acompaña a la hiperinflación. A su vez, guarda correlación positiva con la pobreza extrema (*Poverty Headcount Ratio*), evidenciando su impacto regresivo. En este contexto de multicolinealidad, el modelo Ridge podría haber asignado a la inflación un coeficiente negativo **no por efecto causal**, sino como ajuste estadístico tras haber atribuido el impacto real a otras variables correlacionadas.

En conjunto, estos resultados reflejan un entorno de corrupción sistémica, donde la precariedad económica, la debilidad institucional y la informalidad marcan las dinámicas del soborno en Venezuela.

El Salvador (Figura 60): El modelo de Regresión Lineal revela un **escenario híbrido** en el que coexisten indicios de mejora con algunas vulnerabilidades.

Por el lado positivo, el acceso al crédito bancario para inversiones destaca como el principal factor asociado a una menor corrupción, lo que refuerza el papel positivo de la formalización financiera en la transparencia empresarial. Este factor aparece como el principal “protector” en El Salvador, sugiriendo que políticas que fomenten el crédito productivo formal pueden seguir debilitando las redes de corrupción.

Sin embargo, variables como la matriculación en primaria y *Voice and Accountability* presentan coeficientes positivos, lo que sugiere que, pese a ciertos avances sociales y democráticos, estos aún **no se traducen en una reducción efectiva del soborno**, probablemente por rezagos institucionales. Es posible que los aumentos en la participación ciudadana ocurran como respuesta al descontento ante la corrupción, más que a un control real de la misma.

Además, la desigualdad (*Gini Index*) y el **desempleo** también se asocian a mayores niveles de soborno, en línea con el patrón observado en otros países. En conjunto, el caso de El Salvador muestra que **el fortalecimiento económico va en la dirección correcta**, pero **la eficacia institucional aún debe consolidarse** para que los avances en desarrollo social y cívico se reflejen en menor corrupción.

Uruguay (Figura 61): Aunque presenta niveles muy bajos de corrupción, el modelo de Regresión Lineal revela que existen factores que explican la pequeña variabilidad interna en los niveles de soborno percibidos.

Sorprendentemente, las variables con mayor peso son el uso del crédito bancario para inversión y capital de trabajo, con coeficientes positivos, lo que sugiere que **el soborno en Uruguay se vincula a sectores con mayor dinamismo económico**, donde existen más trámites y contratos, y, por tanto, puede que surjan oportunidades de corrupción. La relación positiva, sugiere entonces que **las áreas de expansión económica en Uruguay requieren atención para mantenerlas libres de corrupción**.

En cuanto a los factores protectores, se corresponden a indicadores de calidad institucional y acceso equitativo a servicio: *Government Effectiveness*, *Access to electricity*, *Control of Corruption*, *Rule of Law* y *Urban population* destacan como elementos que ayudan a contener la corrupción. A diferencia de otros países analizados, en Uruguay no sobresalen la pobreza o la desigualdad como determinantes, lo que refuerza la idea de que su baja corrupción se debe más a la **solidez institucional y social**, que a la simple ausencia de carencias económicas.

En resumen, la Regresión Lineal por país complementa al modelo de Random Forest al aportar claridad sobre el **sentido del efecto de cada variable**. Los resultados muestran que **el impacto de los factores varía según el contexto nacional**: variables como el crédito bancario, la participación ciudadana o la educación pueden tener efectos positivos, neutros o incluso adversos, según el nivel institucional. Esto subraya la importancia de diseñar **estrategias anticorrupción adaptadas a las condiciones específicas de cada país**.

7.5. Comparación de resultados entre modelos

Considerando los cuatro modelos en conjunto – dos modelos globales (Random Forest Classifier y Regresión Logística) y dos enfoques por país (Random Forest Regressor y Regresión Lineal Ridge) – emergen varios **patrones coherentes**, así como algunas **contradicciones interesantes**:

Comparación entre Random Forest Classifier y la Regresión Logística:

Ambos modelos globales identifican conjuntos muy similares de variables clave, lo cual refuerza la robustez de las conclusiones. Las variables asociadas a **mayor corrupción** tales como la desigualdad de ingresos (*Gini Index*), pobreza extrema y prácticas corruptas previas (*Bribery Depth*), aparecen como importantes en el Random Forest (Figura 38) y cuentan con coeficientes positivos significativos en la Regresión Logística (Figura 56).

Del mismo modo, los factores vinculados a **menor corrupción** como la alta calidad institucional (*Voice and Accountability*, *Rule of Law*, *Regulatory Quality*, *Government Effectiveness*, *Control of Corruption* y *Political Stability*), y el gasto en salud (*Health Expenditure*) aparecen en ambos enfoques (Figura 38 y Figura 57) entre los indicadores más destacados.

Esta convergencia de resultados ofrece una base sólida para afirmar que la corrupción **tiende a incrementarse en contextos de fragilidad socioeconómica e institucional**, mientras que retrocede en entornos caracterizados por un **Estado fuerte, servicios públicos accesibles y mayor equidad social**.

Comparación entre Random Forest Regressor y Regresión Lineal Ridge:

A nivel de análisis por país, tanto el Random Forest Regressor como la Regresión Lineal Ridge identificaron conjuntos de variables muy similares en cada contexto, lo que refuerza la validez de los hallazgos.

En **Venezuela** (Figura 43 y Figura 59), ambos modelos destacaron de forma clara el papel central del acceso al crédito de corto plazo (financiamiento de capital de trabajo) como motor de la corrupción, reflejando la precariedad institucional y las restricciones del sistema financiero

En **El Salvador** (Figura 44 y Figura 60), ambos enfoques coincidieron en que el uso del crédito bancario para inversión se asocia a menores niveles de corrupción, mientras que

factores como la desigualdad (*Gini Index*) y el desempleo se vinculan con un mayor riesgo. Además, se destacó la relevancia del desarrollo socioeconómico (PIB, gasto en salud y tasa de alfabetización) como variables que contrarrestan la corrupción.

En **Uruguay** (Figura 45 y Figura 61), la convergencia entre modelos también fue notable. Tanto el Random Forest como la regresión lineal señalaron la **urbanización**, la **alfabetización** y el uso del crédito como determinantes. Si bien algunos coeficientes del modelo lineal apuntan a relaciones contraintuitivas (como la correlación positiva del crédito con el soborno), estos resultados deben interpretarse dentro de un contexto de corrupción estructuralmente baja, donde pequeñas variaciones pueden estar asociadas a dinámicas muy específicas.

En **Uruguay**, tanto el Random Forest como la Regresión Lineal señalaron el uso del financiamiento formal como potenciador de la corrupción, y la urbanización y el acceso a la electricidad como protector de la corrupción.

Esta consonancia entre algoritmos tan distintos (uno no paramétrico, otro lineal penalizado) refuerza la credibilidad de los hallazgos y sugiere que los patrones detectados no dependen del tipo de modelo ni de sus limitaciones, por lo que reflejan tendencias reales del fenómeno estudiado: la corrupción.

Comparación de patrones contextuales y globales:

La comparación revela que **la relevancia y el signo de ciertos factores varía según el contexto nacional**. Por ejemplo, *Voice and Accountability* actúa como freno a la corrupción a nivel global, pero en El Salvador muestra un efecto positivo, quizás porque la participación ciudadana aún no se traduce en control efectivo, posiblemente porque las instituciones no logran canalizar ese empoderamiento cívico.

Del mismo modo, el **uso del financiamiento bancario** reduce la corrupción en El Salvador, pero se asocia a leves aumentos en Uruguay, posiblemente por su concentración en sectores más expuestos. Incluso la **inflación**, generalmente un factor de riesgo presenta una relación inversa en Venezuela debido a su colapso económico.

Estos casos muestran que **los determinantes de la corrupción no son universales**, y que es esencial considerar el contexto nacional para interpretar adecuadamente su efecto.

7.6. Conclusión del análisis

El análisis comparativo realizado entre Venezuela, El Salvador y Uruguay ha permitido identificar patrones diferenciados en la relación entre variables económicas, sociales e institucionales y la incidencia del soborno. Los resultados del estudio no solo confirman varias hipótesis planteadas en la literatura sobre los determinantes de la corrupción, sino que también aportan evidencia empírica que matiza o incluso contradice algunas de las teorías existentes.

Los resultados empíricos muestran que las variables institucionales, como el Control de la Corrupción, la Calidad Regulatoria, la Estabilidad Política y el Estado de Derecho, presentan una fuerte correlación negativa con la incidencia del soborno. Esta asociación fue reforzada tanto por los análisis de correlación como por los modelos de clasificación y regresión. De forma consistente, Uruguay, con instituciones sólidas, muestra los niveles más bajos de corrupción, mientras que Venezuela, con una institucionalidad debilitada, presenta los niveles más altos.

En el plano económico, se ha identificado una relación inversa entre el PIB per cápita (GDP) y la corrupción. A mayor nivel de desarrollo económico, menor incidencia de sobornos. Este hallazgo coincide con los estudios de Mauro (1995) y Desfrancois & Pastas Gutiérrez (2022), quienes también establecieron vínculos negativos entre crecimiento económico y corrupción. Igualmente, el acceso a servicios básicos (como electricidad y educación primaria) y la tasa de alfabetización se revelan como factores protectores frente a la corrupción, respaldando así la tesis de Gupta, Davoodi y Alonso-Terme (2002), quienes argumentan que la corrupción tiene efectos regresivos al reducir el acceso a servicios esenciales y profundizar las desigualdades, lo que sugiere que el desarrollo humano y social está estrechamente vinculado con entornos institucionales más transparentes.

De forma particular, el modelo Random Forest Regressor aplicado por país destaca variables como el Gini Index y la Pobreza Extrema como factores relevantes en contextos de alta corrupción, lo cual refuerza los argumentos de Vega (2025) sobre el carácter regresivo de la corrupción.

Por otro lado, un hallazgo contraintuitivo ha sido observado en El Salvador: el mayor uso del crédito bancario formal se relaciona con menores niveles de soborno. Este resultado

es inesperado dado su contexto institucional intermedio, y sugiere que la inclusión financiera podría actuar como un mecanismo de formalización económica que reduce la necesidad de prácticas corruptas. Aunque Amin y Motta (2021) argumentan que la corrupción limita el acceso al crédito formal, los resultados de este estudio complementan su enfoque al mostrar que, en sentido inverso, fortalecer el acceso al crédito también puede contribuir a reducir la corrupción. En este sentido, se aporta evidencia de que el fortalecimiento del sistema financiero y la expansión del crédito formal podrían constituir herramientas anticorrupción eficaces, incluso en contextos donde las instituciones aún no han alcanzado su consolidación plena.

La evidencia obtenida confirma la hipótesis central de Kaufmann, Kraay y Mastruzzi (2009): una buena gobernanza se asocia con menores niveles de corrupción. La experiencia de Uruguay valida esta teoría, al mostrar cómo una administración pública efectiva, con altos niveles de participación ciudadana y control institucional, actúa como barrera frente a prácticas corruptas.

Igualmente, se observa coherencia con los planteamientos de Tanzi y Davoodi (1997), quienes sostienen que la corrupción distorsiona la asignación del gasto público. En Venezuela, donde el modelo revela una fuerte asociación entre corrupción e inestabilidad macroeconómica (ejemplificada por tasas de inflación extremas), se confirma que la corrupción debilita los fundamentos económicos y deteriora el entorno inversor.

Sin embargo, los resultados también contradicen parcialmente ciertas corrientes teóricas. La hipótesis de la “mano que ayuda” o *“helping hand”*, que sugiere que un Estado activo puede reducir la corrupción si se orienta hacia la redistribución (Hopkin & Rodríguez-Pose, 2007), no encuentra pleno respaldo en este estudio. En Venezuela, por ejemplo, el gasto público no ha logrado reducir la corrupción, debido a la falta de controles eficaces y la captura institucional. Este resultado respalda la perspectiva del *“grabbing hand”* expuesta por Shleifer y Vishny (1993), al mostrar que cuando el Estado opera con amplios márgenes de discrecionalidad y sin mecanismos efectivos de control, se generan condiciones propicias para la corrupción, ya que los funcionarios públicos pueden actuar en beneficio propio sin temor a sanciones.

Del mismo modo, los datos contradicen la hipótesis de la “corrupción eficiente” (Lui, 1985), según la cual los sobornos podrían acelerar trámites y mejorar la eficiencia

económica. En todos los países, pero especialmente en Venezuela, se observa que mayores niveles de soborno se asocian con menor crecimiento económico, más informalidad y menor inversión, lo que refuta dicha teoría.

La selección de tres países con niveles de corrupción estructuralmente distintos ha permitido no solo contrastar teorías, sino también generar un marco de análisis aplicable a otros contextos. Uruguay se confirma como caso de buenas prácticas institucionales, Venezuela como advertencia de fallos sistémicos, y El Salvador como escenario de transición. Esta diversidad permite afirmar que las políticas anticorrupción más eficaces son aquellas que se adaptan a las condiciones institucionales y socioeconómicas de cada país.

En definitiva, este trabajo ha demostrado que la corrupción es un fenómeno multicausal, cuya comprensión exige el análisis integrado de variables económicas, sociales e institucionales. Además, la combinación de análisis exploratorio, técnicas de Machine Learning y contrastación teórica ha permitido validar empíricamente hipótesis clave.

8. Conclusiones y trabajos futuros

Los resultados obtenidos en este estudio permiten confirmar que la corrupción en América Latina no es un fenómeno aislado, sino profundamente vinculado con condiciones estructurales sociales, económicas e institucionales. A través de modelos de clasificación y regresión aplicados tanto a nivel global como por país, se ha evidenciado que variables como la desigualdad de ingresos, el acceso al crédito formal, la calidad institucional y el nivel educativo influyen de manera significativa en la incidencia del soborno.

En todos los modelos, la corrupción se relaciona con altos niveles de desigualdad (Gini Index), pobreza extrema, baja calidad institucional y profundidad de prácticas corruptas previas (*Bribery Depth*). Igualmente, variables como el acceso a servicios básicos (electricidad, salud), la alfabetización, la participación femenina en el mercado laboral o la urbanización suelen actuar como frenos frente al soborno.

Los modelos por país revelan que, aunque hay factores compartidos, el contexto nacional modifica su peso y dirección. Por ejemplo, el acceso al crédito bancario en El Salvador reduce la corrupción, mientras que en Uruguay puede estar vinculado con sectores más expuestos a riesgos.

Estos resultados confirman lo señalado por la literatura, que destaca la naturaleza multicausal y contextual de la corrupción. En línea con Fisman y Svensson (2007), quienes evidencian cómo la corrupción limita el crecimiento empresarial en entornos con sobornos frecuentes, y con Desfrancois y Pastas Gutiérrez (2022), que demuestran su impacto negativo sobre el PIB per cápita en América Latina, se confirma que los efectos económicos de la corrupción son sustanciales. Asimismo, Amin y Motta (2021) y Loayza (2018) refuerzan el papel central de la informalidad y el acceso desigual al sistema financiero como mecanismos a través de los cuales la corrupción afecta el desarrollo económico y laboral.

En conjunto, este estudio empírico valida la hipótesis de que la corrupción no es un problema individual, sino que es el resultado de **estructuras institucionales frágiles, desigualdades persistentes y deficiencias en la provisión de oportunidades económicas y sociales.**

La comparación entre Venezuela, El Salvador y Uruguay ha permitido identificar tres perfiles claramente diferenciados en cuanto a los niveles de corrupción y el grado de solidez institucional.

Venezuela representa un caso de alta corrupción estructural, profundamente arraigada. Con una puntuación de 13 sobre 100 en el Índice de Percepción de la Corrupción (CPI) de 2023, se ubica entre los países más corruptos del mundo ocupando el puesto 177 de 180. Esta situación extrema se asocia a la debilidad institucional crónica, a la corrupción sistemática en los niveles más altos del poder político y a la penetración del crimen organizado en estructuras clave del Estado. Todo ello configura un entorno en el que el soborno y el desvío de recursos públicos no son episodios aislados, sino manifestaciones de un fenómeno de corrupción estructural que exige reformas profundas y un reforzamiento integral del Estado de derecho.

En contraste, **El Salvador** ocupa una posición de corrupción intermedia. Con una puntuación de 31 en el CPI y situado en el puesto 126 a nivel global, presenta niveles moderados de corrupción, aunque aún preocupantes. Sin embargo, en los últimos años ha puesto en marcha reformas orientadas a combatir el fenómeno de la corrupción (Asamblea Legislativa de El Salvador, 2025). En 2021, El Salvador aprobó una nueva ley anticorrupción y reformó el código penal para endurecer las penas asociadas a delitos de corrupción, con el objetivo de reforzar la transparencia y lograr una mayor integridad en la función pública. Estas medidas reflejan una voluntad política de mejora institucional, caracterizando a El Salvador como un país en transición, en el que coexisten desafíos persistentes de corrupción con iniciativas reformistas, lo que sugiere un potencial de mejora si esos avances sostenibles se consolidan en el tiempo.

Por su parte, **Uruguay** destaca como el modelo regional de baja corrupción y elevada solidez institucional. Con una puntuación de 73 sobre 100 en el CPI 2023 y ocupando el puesto 16 de 180, se posiciona como el país más transparente de América Latina, superando incluso a democracias consolidadas como Estados Unidos.

El liderazgo de Uruguay se sustenta en una institucionalidad fuerte, basada en el respeto al Estado de derecho, en la transparencia de la gestión pública y en altos niveles de participación ciudadana. A lo largo de más de una década, ha impulsado políticas de gobierno abierto y acceso a la información pública, lo que ha contribuido a fortalecer la

confianza en las instituciones y a consolidar mecanismos efectivos de control y prevención de la corrupción (Sitio oficial de la República Oriental del Uruguay, 2025). En este sentido, Uruguay actúa como modelo de referencia de baja corrupción y buena gobernanza para otros países de la región interesados en desarrollar prácticas efectivas en la lucha contra la corrupción.

La selección intencionada de tres países con realidades institucionales, sociales y económicas tan contrastantes ha permitido construir una base de datos robusta y representativa para el análisis, capaz de capturar la diversidad de dinámicas relacionadas con la corrupción. Gracias a ello, los hallazgos obtenidos no solo han permitido comprender con mayor profundidad los factores de corrupción en América Latina, sino que también ofrecen un marco adaptable a otros entornos, facilitando la extrapolación de los resultados a países con características similares o en fases distintas de desarrollo institucional. Esta variedad comparativa refuerza la validez del estudio y amplía su utilidad como herramienta de diagnóstico y diseño de estrategias anticorrupción en distintos contextos nacionales.

Asimismo, la solidez de los resultados se sustenta en el empleo combinado de técnicas de *Business Intelligence* (BI) y algoritmos de *Machine Learning* aplicados sobre una base de datos integrada, compuesta por indicadores económicos, sociales e institucionales procedentes de diversas fuentes del Banco Mundial (World Bank Enterprise Surveys, World Development Indicators, Worldwide Governance Indicators y ESG Data). Este enfoque permitió identificar patrones, tendencias, distribuciones, correlaciones y variables clave mediante visualizaciones, análisis exploratorios y modelos predictivos como Random Forest, Regresión Logística y Regresión Lineal Ridge, tanto a nivel global como por país. De este modo, se logró contrastar hipótesis teóricas con evidencia empírica, fortaleciendo las conclusiones obtenidas.

A partir de los resultados obtenidos en este estudio – especialmente el análisis de correlaciones, distribuciones y modelos de Machine Learning – se proponen distintas **líneas de acción para combatir la corrupción**, adaptadas a los diferentes niveles de desarrollo institucional. Estas estrategias se sustentan en las variables que han demostrado ser significativas en la explicación de la incidencia del soborno, tales como la desigualdad, el acceso a servicios básicos, la calidad institucional, la inclusión financiera y la participación femenina.

En contextos de alta institucionalidad, como es el caso de Uruguay, los esfuerzos deben centrarse en consolidar los logros alcanzados. Como muestra el análisis, variables como *Rule of Law*, *Government Effectiveness* y *Control of Corruption* presentan una fuerte asociación con bajos niveles de soborno. Por ello, se recomienda profundizar en políticas de transparencia, gobierno abierto y digitalización de los servicios públicos. Asimismo, la promoción de la participación ciudadana y el fortalecimiento de la educación cívica pueden contribuir a prevenir retrocesos y mantener una cultura de integridad institucional.

En contextos de institucionalidad intermedia, como ocurre en El Salvador, los modelos muestran la relevancia de factores mixtos: aunque se detecta debilidad institucional, el mayor uso del crédito bancario formal se asocia con menores niveles de soborno, lo que sugiere que la inclusión financiera puede funcionar como mecanismo de formalización económica. Así, se recomienda fortalecer el acceso al crédito para PYMES, simplificar trámites burocráticos y consolidar reformas legales recientes que refuercen la independencia judicial y los sistemas institucionales de control y supervisión pública. También se constata la importancia de variables como *Female Labor Force Participation*, lo que respalda que la implementación de políticas orientadas a fomentar el liderazgo femenino en los ámbitos público y privado puede contribuir a fortalecer la integridad y la transparencia institucional, y, en consecuencia, a reducir los niveles de corrupción.

En contextos de baja institucionalidad, como Venezuela, los resultados evidencian que variables como la pobreza, la desigualdad (*Gini Index*) y el deterioro institucional se asocian claramente con altos niveles de soborno. Por tanto, las prioridades deben centrarse en reconstruir las capacidades básicas del Estado, restaurar el Estado de derecho y garantizar que las decisiones públicas se tomen conforme a criterios objetivos, transparentes y supervisados, evitando así que se puedan tomar decisiones importantes (como asignar recursos, aprobar licencias, o aplicar sanciones) sin controles, reglas claras o rendición de cuentas. A su vez, resulta esencial recuperar la provisión de servicios públicos básicos como salud, educación o electricidad (*Health Expenditure*, *Access to Electricity*), cuya ausencia alimenta el espacio para prácticas corruptas. Al mismo tiempo, es fundamental impulsar la participación activa de la ciudadanía en el control del poder público, fortalecer la cooperación con organismos internacionales para garantizar el seguimiento de las reformas, y fomentar la formación de líderes comprometidos con la ética y la transparencia.

Por último, este trabajo abre diversas líneas para futuras investigaciones, como la **ampliación del número de países**, lo que permitiría incluir más casos latinoamericanos o extender el análisis a otras regiones del mundo para contrastar si los patrones identificados son universales o específicos de determinados contextos; la **ampliación del número de variables**, incorporando nuevos indicadores provenientes de otras bases de datos del Banco Mundial u organismos internacionales, con el fin de captar dimensiones adicionales que enriquezcan el modelo analítico; y la **incorporación de nuevos modelos** de predicción, explorando técnicas como redes neuronales o XGBoost, que podrían detectar relaciones no lineales aún no captadas por los algoritmos empleados en este estudio.

9. Referencias bibliográficas

Amin, M., & Motta, V. (2021). The Impact of Corruption on SMEs' Access to Finance: Evidence Using Firm-Level Survey Data from Developing Countries. World Bank.

<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/935941634651569360/the-impact-of-corruption-on-smes-access-to-finance-evidence-using-firm-level-survey-data-from-developing-countries>

Asamblea Legislativa de El Salvador. (2025). Asamblea aprueba Ley Anticorrupción para reforzar la transparencia del Estado | Asamblea Legislativa de El Salvador.

<https://www.asamblea.gob.sv/node/13467>

Cárdenas Denham, S. (2012). La corrupción en sistemas educativos: Una revisión de prácticas, causas, efectos y recomendaciones. Revista electrónica de investigación educativa. https://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S1607-40412012000200005&lng=es&nrm=iso&tlng=es

DataBank | The World Bank. (2025). <https://databank.worldbank.org/>

Desfrancois, P., & Pastas Gutiérrez, E. R. (2022). Corrupción y crecimiento económico en América Latina y el Caribe. Revista de Economía del Caribe, 29.

<https://doi.org/10.14482/ecoca.29.704.937>

Enterprise Surveys | DataBank. (2025).

<https://databank.worldbank.org/source/enterprise-surveys>

Environment Social and Governance (ESG) Data | DataBank. (2025).

[https://databank.worldbank.org/source/environment-social-and-governance-\(esg\)-data](https://databank.worldbank.org/source/environment-social-and-governance-(esg)-data)

Fernández Casal, R., Costa Bouzas, J., & Oviedo de la Fuente, M. (2024). Métodos predictivos de aprendizaje estadístico. <https://doi.org/10.17979/spudc.9788497498937>

Fisman, R., & Svensson, J. (2007). Are corruption and taxation really harmful to growth? Firm level evidence. <https://doi.org/10.1016/j.jdeveco.2005.09.009>

Glossary | DataBank. (2025).

<https://databank.worldbank.org/metadataglossary/all/series?search=bribery>

- GridSearchCV. (2025). Scikit-Learn. https://scikit-learn/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
- Gupta, S., Davoodi, H., & Alonso-Terme, R. (2002). Does corruption affect income inequality and poverty? <https://doi.org/10.1007/s101010100039>
- Haberson, A. (2021). Gender and corruption: The time is now. World Bank Blogs. <https://blogs.worldbank.org/en/governance/gender-and-corruption-time-now>
- Hopkin, J., & Rodríguez-Pose, A. (2007). “Grabbing Hand” or “Helping Hand”? Corruption and the Economic Role of the State. Governance. <https://doi.org/10.1111/j.1468-0491.2007.00353.x>
- Jameel, A. L. (2023, septiembre 27). Stretching Budgets and Lifespans | Abdul Latif Jameel®. Abdul Latif Jameel. <https://alj.com/en/perspective/stretching-budgets-and-lifespans/>
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2009). Governance matters VIII: Aggregate and individual governance indicators 1996-2008. World Bank. <https://documentos.bancomundial.org/es/publication/documents-reports/documentdetail/en/598851468149673121>
- KNNImputer. (2025). Scikit-Learn. <https://scikit-learn/stable/modules/generated/sklearn.impute.KNNImputer.html>
- Lewis, M. (2006). Governance and Corruption in Public Health Care Systems. Social Science Research Network. <https://doi.org/10.2139/ssrn.984046>
- LinearRegression. (2025). Scikit-Learn. https://scikit-learn/stable/modules/generated/sklearn.linear_model.LinearRegression.html
- Loayza, N. V. (2018). Informality: Why Is It So Widespread and How Can It Be Reduced? World Bank. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/en/130391545228882358>
- LogisticRegression. (2025). Scikit-Learn. https://scikit-learn/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Los Tiempos. (2023). En América Latina, unas 18 millones de personas no tienen acceso a electricidad. <https://www.lostiempos.com/actualidad/mundo/20230903/america-latina-18-millones-personas-no-tienen-acceso-electricidad>

Lui, F. T. (1985). An Equilibrium Queuing Model of Bribery. *Journal of Political Economy*. <https://www.jstor.org/stable/1832136>

Mauro, P. (1995). Corruption and Growth. *The Quarterly Journal of Economics*. <https://doi.org/10.2307/2946696>

Özşahin, Ş., & Üçler, G. (2017). The Consequences of Corruption on Inflation in Developing Countries: Evidence from Panel Cointegration and Causality Tests. <https://doi.org/10.3390/economies5040049>

RandomForestClassifier. (2025). Scikit-Learn. <https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

RandomForestRegressor. (2025). <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

Shleifer, A., & Vishny, R. (1993). Corruption. <https://shleifer.scholars.harvard.edu/publications/corruption>

Sitio oficial de la República Oriental del Uruguay. (2025). Guía para la publicación de Datos Abiertos de Transparencia Activa. Agencia de Gobierno Electrónico y Sociedad de la Información y del Conocimiento. <https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/publicaciones/guia-para-publicacion-datos-abiertos-transparencia-activa-1>

SMOTE. (2024). https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html

StandardScaler. (2025). Scikit-Learn. <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

StratifiedKfold. (2025). Scikit-Learn. https://scikit-learn/stable/modules/generated/sklearn.model_selection.StratifiedKfold.html

Torres Fragoso, J. (2023). DESAFÍOS EN EL COMBATE A LA CORRUPCIÓN EN HISPANOAMÉRICA: LAS EXPERIENCIAS DE CHILE, COLOMBIA, COSTA RICA, GUATEMALA Y MÉXICO. <https://doi.org/10.18601/16578651.n32.07>

Train_test_split. (2025). Scikit-Learn. https://scikit-learn/stable/modules/generated/sklearn.model_selection.train_test_split.html

2023 Corruption Perceptions Index: Explore the results. (2024, enero 30).

Transparency.Org. <https://www.transparency.org/en/cpi/2023>

U4 Anti-Corruption Resource Centre. (2024). The impacts of corruption on women and their role in preventing corruption. U4 Anti-Corruption Resource Centre. <https://www.u4.no/publications/the-impacts-of-corruption-on-women-and-their-role-in-preventing-corruption>

Vega, F. J. (2025, febrero 11). ¿Qué tienen en común la corrupción y la desigualdad? Mexicanos Contra la Corrupción y la Impunidad. <https://contralacorrupcion.mx/que-tienen-en-comun-la-corrupcion-y-la-desigualdad/>

Vito Tanzi & Hamid Davoodi. (1997). Corruption, Public Investment, and Growth. IMF. <https://www.imf.org/en/Publications/WP/Issues/2016/12/30/Corruption-Public-Investment-and-Growth-2353>

Wei, S.-J. (2000). How Taxing is Corruption on International Investors? The Review of Economics and Statistics. <https://doi.org/10.1162/003465300558533>

World Bank. (2025). Combating Corruption. <https://www.worldbank.org/en/programs/anticorruption-for-development>

World Development Indicators | DataBank. (2025). <https://databank.worldbank.org/source/world-development-indicators>

Worldwide Governance Indicators | DataBank. (2025). <https://databank.worldbank.org/source/worldwide-governance-indicators>

10. ANEXO: Declaración Uso IAG

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Tatiana López Buendía, estudiante de Grado en Análisis de Negocios / Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "ESTUDIO COMPARATIVO SOBRE CORRUPCIÓN EN VENEZUELA, EL SALVADOR Y URUGUAY", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Crítico:** Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
5. **Interpretador de código:** Para realizar análisis de datos preliminares.
6. **Estudios multidisciplinares:** Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
7. **Constructor de plantillas:** Para diseñar formatos específicos para secciones del trabajo.
8. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.

9. **Generador previo de diagramas de flujo y contenido:** Para esbozar diagramas iniciales.
10. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
11. **Generador de datos sintéticos de prueba:** Para la creación de conjuntos de datos ficticios.
12. **Generador de problemas de ejemplo:** Para ilustrar conceptos y técnicas.
13. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
14. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 16/06/2025

Firma:

A handwritten signature in black ink, appearing to read 'Tatiana Lopez', with a stylized flourish at the end.