



Facultad de Ciencias Económicas y Empresariales (ICADE)

Impacto de la Formalidad, la Localización y el Tamaño sobre la Digitalización de Micro y Miniempresas en República Dominicana.

Autor: Alejandro Ruiz Miguel
Director: Raúl González Fabre

RESUMEN

En la República Dominicana, las micro y pequeñas empresas desempeñan un papel crucial en la economía, contribuyendo aproximadamente con el 39% al Producto Interno Bruto (PIB) y destacándose como las principales generadoras de empleo. La digitalización es una tendencia que está transformando la forma en que estas empresas operan, ayudando a mejorar su productividad, eficiencia y competitividad. A pesar de los avances, la adopción tecnológica en este tipo de empresas dominicanas sigue siendo limitada debido a una serie de barreras que son objeto de estudio en este trabajo. La presente investigación analiza una serie de variables obtenidas de una encuesta realizada a miles de empresarios de micro y pequeñas empresas dominicanas para verificar si existe una relación directa y positiva entre estas variables y la digitalización de las empresas. A través de la implementación de diversas técnicas de análisis, se obtienen resultados que confirman la existencia de las relaciones mencionadas anteriormente.

Palabras Clave: República Dominicana, Informalidad, Digitalización, Transformación digital, MiPymes, Formalidad y tecnología, Localización y tecnología, Tamaño de empresa y tecnología, Micro y miniempresas

ABSTRACT

In the Dominican Republic, micro and small enterprises play a crucial role in the economy, contributing approximately 39% to the Gross Domestic Product (GDP) and standing out as the main job creators. Digitization is a trend that is transforming the way these enterprises operate, helping to improve their productivity, efficiency, and competitiveness. Despite advancements, technological adoption in these types of Dominican enterprises remains limited due to various barriers that are the subject of study in this work. This research examines a series of variables obtained from a survey conducted with thousands of micro and small business entrepreneurs in the Dominican Republic to verify if there is a direct and positive relationship between these variables and the digitization of the enterprises. Through the implementation of various analytical techniques, results are obtained that confirm the existence of the aforementioned relationships.

Keywords: Dominican Republic, Informality, Digitalization, Digital transformation, SMEs, Formality and technology, Location and technology, Company size and technology, Micro and small businesses

Índice

1. Introducción	6
1.1. Contexto Económico en República Dominicana	6
1.2. Informalidad vs Digitalización	7
1.2.1. Informalidad en República Dominicana	7
1.2.2. Digitalización en República Dominicana	7
1.3. Hipótesis	8
1.3.1 Primera Hipótesis	8
1.3.2. Segunda Hipótesis	9
1.3.3 Tercera Hipótesis	9
1.4. Objetivos del trabajo	9
1.5. Metodología de la investigación	10
1.6. Estructura del trabajo	10
2. Marco Teórico	11
2.1. Formalización	11
2.1.1. Formalización en las empresas.....	11
2.1.2. Ventajas de la Formalización	12
2.1.3. Desafíos de la formalización	13
2.1.4. Formalización en MiPymes de República Dominicana	14
2.2. Digitalización en Micro y Miniempresas	14
2.2.1. Ventajas y oportunidades de la Digitalización	15
2.2.2. Desafíos de la Digitalización	16
3. Datos	17
3.1. Extracción de Datos	17
3.1.2. Tratamiento de los datos	21
4. Metodología	23
4.1. Metodología de la Primera Hipótesis	23
4.1.1. Contraste Primera Hipótesis.....	25
4.2. Metodología de la Segunda Hipótesis	26
4.2.1. Contraste Segunda Hipótesis	29
4.3. Metodología de la Tercera Hipótesis	30
4.3.1. Contraste Tercera Hipótesis	32
4.4. Creación de Modelo de Regresión Logística con las variables independientes	33
4.4.1. Resultados del Modelo de Regresión Logística	34
5. Conclusión	37
6. Bibliografía	40
7. Anexos	44
Anexo I: Código Python Cálculo Chi-Cuadrado	44
Anexo II: Código Python Modelo de Regresión Lineal	46
Anexo III: Código Python Proporción Empresas Digitalizadas por Localidad	47

Índice de Figuras

Figura N°1: Tabla de Contingencia Primera Hipótesis.....	24
Figura N°2: Tabla de Frecuencias Esperadas Primera Hipótesis.....	24
Figura N°3: Cálculo de Chi-cuadrado Primera Hipótesis.....	24
Figura N°4: Tabla de Contingencia Segunda Hipótesis.....	27
Figura N°5: Tabla de Frecuencias Esperadas Segunda Hipótesis.....	27
Figura N°6: Cálculo de Chi-cuadrado Segunda Hipótesis.....	28
Figura N°7: Tabla de Contingencia Tercera Hipótesis.....	30
Figura N°8: Tabla de Frecuencias Esperadas Tercera Hipótesis.....	31
Figura N°9: Cálculo de Chi-cuadrado Tercera Hipótesis.....	31
Figura N°10: Resultado Rendimiento del Modelo.....	35
Figura N°11: Resultados por Variable del Modelo.....	36
Figura N°12: Proporción Empresas Digitalizadas por Localidad.....	38

1.Introducción

1.1. Contexto Económico en República Dominicana

República Dominicana es un país que cuenta con una economía en desarrollo y con gran dinamismo que presenta un crecimiento de 5.6% en los últimos diez años, experimentando uno de los crecimientos económicos más rápidos de su región. Este crecimiento se ha debido a una combinación de reformas estructurales orientadas al mercado, implementadas a principios de la década de los 90, y a condiciones externas favorables. Sin embargo, el país enfrenta desafíos relacionados con el bajo crecimiento de la productividad, la necesidad de mejorar el capital humano para satisfacer las demandas del sector empresarial, y los desafíos impuestos por el cambio climático y distorsiones en mercados claves (Banco Mundial, 2023).

En la República Dominicana, existen aproximadamente 1.5 millones de Pymes, que aportan alrededor del 39% al Producto Interior Bruto (PIB). Este segmento es considerado como una de las principales fuerzas de desarrollo económico y es un articulador vital del tejido empresarial, especialmente en economías emergentes. Además, las Pymes desempeñan un papel clave en la generación de empleos y en la producción y oferta de bienes y servicios (Acosta, 2021).

Según estadísticas del Informe Caracterización de las MiPymes módulo ENHOGAR 2022 (2023), del total de micro, pequeñas y medianas empresas, se refleja que la mayor proporción de estas se encuentra compartida con la vivienda, con 46.2%. Por otro lado, un 15% están ubicadas en plazas comerciales, y el sector agropecuario también representa un significativo 10.8%, localizadas en fincas y parcelas. Las empresas con menor participación son aquellas ubicadas en locales ambulantes (4.9%) y puestos fijos en mercados (2.0%). Adicionalmente, más del 83.3% de las MiPymes están concentradas en la zona urbana y tan solo 16.7% se localiza en la zona rural. En términos de sectores económicos, el 46.7% de estas empresas pertenecen al sector comercio, seguido por el sector servicios con un 38.4% y el sector de industrias con un 14.9% (Programa de las Naciones Unidas para el Desarrollo, 2020).

Este elevado nivel de participación destaca la gran vitalidad del tejido empresarial dominicano y la capacidad de las pymes de actuar como motores en el crecimiento económico y la creación de empleo. Esto se debe en parte a su flexibilidad, capacidad de adaptación y cercanía con las comunidades locales, lo que les permite responder de manera ágil a las necesidades del mercado y a las demandas específicas de los consumidores.

1.2. Informalidad vs Digitalización

1.2.1. Informalidad en República Dominicana

La informalidad es un problema importante en la República Dominicana y puede tener un impacto negativo en la productividad de las empresas, ya que les dificulta acceder a los recursos y servicios que necesitan para crecer y desarrollarse.

El alto nivel de informalidad en la economía de la República Dominicana, especialmente en las microempresas, es un tema de gran relevancia. Según la Confederación Dominicana de la Pequeña y Mediana Empresa (Codopyme), alrededor del 80% de las microempresas en el país son informales, lo que representa un desafío para su acceso a créditos y servicios necesarios para su crecimiento y desarrollo (Forbes, 2023).

Además, la tasa de empleo informal en 2021 fue del 59%, superior al promedio de América Latina y el Caribe, lo que impacta negativamente en la recaudación de impuestos y el crecimiento de la productividad (Diario Libre, 2023).

1.2.2. Digitalización en República Dominicana

La digitalización es una tendencia global que está transformando la forma en que operan las empresas. La digitalización puede ayudar a las empresas a mejorar su productividad, eficiencia y competitividad.

Sin embargo, la adopción de tecnología en las pymes dominicanas sigue siendo baja, en parte debido a los desafíos que enfrentan las empresas informales para integrarse en la economía formal y digital (Franklin, 2018).

La transformación digital en las MiPymes, micro, pequeñas o medianas empresas, de la República Dominicana ha avanzado considerablemente, especialmente impulsada por la pandemia de COVID-19. Según un estudio de Microsoft realizado por Edelman, aproximadamente el 85% de las MiPymes dominicanas han acelerado su proceso de transformación digital durante la pandemia. Más del 60% de estas empresas invirtieron en tecnologías clave, y el 87% indicó que toman decisiones basadas en datos, reflejando una creciente orientación hacia una cultura orientada a datos (Microsoft News Center Latinoamérica, 2023).

Sin embargo, las MiPymes en la República Dominicana enfrentan desafíos como el volumen de ventas, la conexión o adquisición de clientes, la adaptación al trabajo remoto/híbrido y la ciberseguridad, con un 40% de las MiPymes reportando problemas de ciberseguridad (CDN, 2023).

La digitalización es vista como un factor clave para la reactivación económica y la generación de empleo, y se enfatiza la necesidad de formación en competencias digitales y una visión humanista en el proceso de digitalización (López Valerio, 2023).

1.3. Hipótesis

1.3.1. Primera Hipótesis

H1: La formalidad afecta positivamente al uso de tecnología en las micro y miniempresas dominicanas.

H0: La formalidad no afecta positivamente al uso de tecnología en las micro y miniempresas dominicanas.

1.3.2. Segunda Hipótesis

H1: La localización de las empresas afecta directamente al uso de tecnología en las micro y miniempresas dominicanas.

H0: La localización de las empresas no afecta directamente al uso de tecnología en las micro y miniempresas dominicanas.

1.3.3. Tercera Hipótesis

H1: El tamaño de las empresas afecta positivamente al uso de tecnología en las micro y miniempresas dominicanas.

H0: El tamaño de las empresas no afecta positivamente al uso de tecnología en las micro y miniempresas dominicanas.

1.4. Objetivos del trabajo

El objetivo principal de este trabajo es investigar si la formalidad, la localización y el tamaño afecta al uso de tecnología en MiPymes dominicanas, a través de los datos obtenidos de una encuesta realizada a micro y miniempresas localizadas en diferentes localidades de República Dominicana. Para la consecución de este objetivo, voy a llevar a cabo una serie de objetivos secundarios que van a ser de gran utilidad:

- Discutir sobre la formalización de micro y miniempresas, sus ventajas y desafíos, y la situación en República Dominicana al respecto.
- Presentar los elementos fundamentales de la digitalización junto con sus oportunidades y desafíos.
- Estudiar diferentes técnicas de análisis de estadísticos para interpretar los resultados de la encuesta.

1.5. Metodología de la investigación

La sección teórica del trabajo se basará en recabar de información para la justificación de contenido que se va a transmitir a lo largo del estudio. Las herramientas utilizadas serán:

- Google Académico
- Artículos web, blogs e informes encontrados en Internet, debido a la carencia de contenido académico relacionado con la formalización y digitalización de micro y miniempresas en República Dominicana.

La sección práctica del trabajo se basará en la realización de un análisis mediante el estadístico Chi-cuadrado. Será fundamental para determinar si hay una asociación significativa entre variables categóricas. Esta prueba estadística será utilizada de forma similar para comprobar las tres hipótesis del estudio.

Posterior al análisis del Chi-cuadrado, y para justificar de mayor forma los resultados obtenidos, se realizará un modelo de regresión logística para descubrir cómo afectan cada una de las variables en la probabilidad de que una empresa esté digitalizada. Estudiándose el efecto y la importancia de cada variable independiente en la probabilidad de que una empresa esté digitalizada.

Por último, se valorarán los resultados de ambos análisis con la finalidad de verificar o rechazar las hipótesis planteadas previamente.

1.6. Estructura del trabajo

El trabajo está desarrollado en dos grandes bloques. El bloque teórico se centra en el análisis del concepto de formalización empresarial, estudiando primero sus ventajas y desafíos, y luego abordando este concepto en el entorno de las MiPymes dominicanas. Y continúa con el estudio de la digitalización de micro y miniempresas, centrándose nuevamente en las oportunidades y desafíos que enfrentan estas empresas.

En el segundo bloque, el práctico, se lleva a cabo la explicación de la obtención de los datos para el estudio, su tratamiento y toda la metodología relacionada con el cálculo de los estadísticos y del modelo de regresión logística. También se incluyen interpretaciones de los resultados de cada análisis, finalizando en la conclusión de la investigación.

Por último, se incluye una última sección de anexos en la que están reflejados los scripts de Python utilizados para el análisis del trabajo.

2. Marco Teórico

2.1. Formalización

2.1.1. Formalización en las empresas

En primer lugar, es necesario comprender qué implica la formalización en las empresas y cuáles son las ventajas que conlleva la incorporación a la economía formal, así como identificar los principales desafíos que surgen en este proceso.

La formalización en las empresas se refiere al proceso que llevan a cabo para incorporarse a la economía formal. Implica cumplir con las regulaciones y requisitos legales, como registros empresariales, normativas laborales y fiscales. La formalización no solo implica costes, sino que también debe proporcionar beneficios, como acceso a financiación, participación en licitaciones públicas, mayor seguridad jurídica y mejora de la legitimidad del negocio.

En muchos casos, la informalidad permite operar con una estructura de costos más baja en comparación con la economía formal, pero ello se logra evadiendo impuestos, incumpliendo normas y traspasando costos a los trabajadores (Chacaltana e Infante, 2014).

2.1.2. Ventajas de la Formalización

La formalización presenta a las empresas acceso a diversas oportunidades y recursos que de otra forma podrían estar fuera de su alcance. Un estudio llevado a cabo por Laboratorio de Acción contra la Pobreza Abdul Latif Jameel (2022) sugiere que la formalización puede permitir a las empresas crecer expandiendo su base de clientes, accediendo a más contratos gubernamentales, servicios financieros y préstamos comerciales.

Una de las utilidades más destacadas es la oportunidad de acceso a financiación formal, ya que las instituciones financieras suelen requerir documentos y registros legales para avalar las solicitudes de préstamos. Esta facilidad para obtener recursos financieros va a afectar de manera positiva al crecimiento y la expansión de la empresa (MIICM, 2019).

Además, la formalización abre las puertas para que las empresas participen en licitaciones públicas, facilitando su implicación en proyectos gubernamentales y contrataciones. Este acceso a oportunidades de contratación pública puede ser un motor fundamental para el desarrollo de la empresa, generando nuevas fuentes de ingresos y ampliando su cartera de clientes (MICM, 2019).

Otro componente y ventaja de la formalización es la ganancia de seguridad jurídica. Cumpliendo con las diferentes normativas y regulaciones, se establece un marco legal que protege tanto a las empresas como a sus clientes y socios. De esta forma las empresas van a conseguir crear una reputación sólida y un clima de confianza en el entorno empresarial (MICM, 2019).

Lo comentado anteriormente también ayuda a mejorar la legitimidad del negocio sobre todo a nivel local. Si estas empresas operan dentro de lo legal y cumplen los estándares establecidos, van a ser vistas como confiables y éticas. Así van a poder ganarse una mejor imagen que va a ayudar a fortalecer la relación con sus clientes y socios.

2.1.3. Desafíos de la formalización

En el proceso de formalización empresarial, las empresas pueden encontrar varias dificultades que impidan o retrasen el proceso. Estos desafíos pueden tener carácter institucional, cultural o económico.

Los factores institucionales incluyen la dificultad para entender y seguir el paso del constante cambio en los marcos regulatorios que las empresas deben conocer para adoptar la formalización. Este proceso puede ser complicado y a menudo requiere de un asesoramiento especializado para garantizar el cumplimiento de todas las leyes. La adaptación a estos marcos regulatorios puede ser especialmente difícil para las pequeñas y medianas empresas que pueden no tener los recursos o el conocimiento especializado necesario para manejar estos desafíos de manera eficiente (Galiani y Meléndez, 2013) (Feal y Ventura, 2023).

Los factores culturales también son un desafío a tener en cuenta en el proceso de formalización empresarial. Puede haber alguna preferencia o pensamiento que prefiera mantener las operaciones de manera informal, principalmente porque la burocracia y la regulación se perciben como barreras importantes. Esta barrera cultural puede influir en la decisión de una empresa de formalizarse, ya que puede ver el proceso como un impedimento más que como una oportunidad para crecer y expandirse. (Galiani y Meléndez, 2013) (Feal y Ventura, 2023).

Por último, los factores económicos son desafíos asociados al coste monetario que sufre el empresario para adoptar el proceso de formalización. Entre estos se incluyen tanto los gastos directos como tarifas y pagos de impuestos, como el tiempo y los recursos dedicados a cumplir con los requisitos de formalización. Estos costes pueden ser especialmente gravosos para las MiPymes, ya que podrían ser difíciles de asumir sin verse afectadas su rentabilidad y operatividad regular. (Galiani y Meléndez, 2013) (Feal y Ventura, 2023).

2.1.4. Formalización en MiPymes de República Dominicana

Hemos visto los beneficios, pero también los desafíos que implica la formalización para las MiPymes. En la República Dominicana es un tema de interés ya que aportan más del 38% del PIB y generan más del 54% de empleos en todo el país.

Un informe del Consejo Nacional de Promoción y Apoyo a la Micro, Pequeña y Mediana Empresa (Promipyme) identificó problemas como la necesidad de acceso a recursos financieros, asistencia técnica, educación financiera, y la existencia de una burocracia excesiva para obtener créditos. Estos problemas se agravan debido a las dificultades para ofrecer garantías y la percepción de impuestos injustos para las MiPymes. Además, enfrentan obstáculos como la inseguridad ciudadana y el alto coste de la energía eléctrica (García, 2021).

El gobierno dominicano, reconociendo la importancia de las MiPymes, ha implementado políticas públicas para promover su desarrollo. Esto incluye el acceso a financiamiento, la implementación de soluciones tecnológicas y la creación de espacios para incentivar la productividad y el crecimiento de estos negocios.

Desde la página web “Formálizate” del Ministerio de Industria, se ofrece gran variedad de información y recursos para aquellas empresas que buscan iniciar el proceso de formalizarse.

También se ha sugerido desde el gobierno que una reforma fiscal podría ayudar a romper el pánico de las MiPymes a formalizarse. Desde el ministerio de MiPymes se propone un sistema de seguimiento, control y asistencia técnica y educativa para las empresas formales. Además, se plantea la necesidad de una escala impositiva que permita a las MiPymes pagar sus impuestos de manera satisfactoria y proporcional a su tamaño y capacidad económica (Severino, 2021).

2.2. Digitalización en Micro y Miniempresas

La digitalización empresarial consiste en proveer de un proceso de sustitución tecnológica, representada por la computación en la nube, big data, inteligencia artificial, aprendizaje automático e Internet de las cosas, que transforma los modelos

de negocio; altera las cadenas de valor de los productos; y desvanece las fronteras que separan a las industrias entre sí (Armas, 2018).

Mediante la digitalización, se pueden acceder a diversos beneficios como la reducción de costes en áreas de gestión de documentos y logística o mejoras de eficiencia y agilidad en la comunicación entre *stakeholders*. También va a contribuir a una mejora en la experiencia de los clientes a través de la digitalización de los canales y automatización de procesos, creando una mayor rapidez y satisfacción en todo el proceso.

Es evidente que la digitalización tiene un impacto importante en las micro, pequeñas y medianas empresas en forma de oportunidad, pero también hay que ser consecuente y entender los desafíos que puede provocar iniciar este proceso.

2.2.1. Ventajas y oportunidades de la Digitalización

En primer lugar, está la disminución de los gastos en infraestructura. Pero también, la aceleración en la entrada al mercado, la creación de modelos de negocio innovadores y la expansión de la gama de productos existentes son tres ventajas adicionales significativas. Otra ventaja de la digitalización incluye la colaboración y la formación de identidad, la flexibilidad y capacidad de ampliación, y la distribución eficiente de datos (Armas, 2018).

Una de las utilidades más relevantes es la ampliación del alcance de las MiPymes. Pueden acceder a nuevos mercados mediante la implementación de plataformas digitales y su presencia en línea (Plaza, 2020). Esto también ayuda a generar una capacidad de adaptación a las nuevas tendencias que les va a permitir operar de forma más flexible y competitiva.

Otra de las mayores ventajas de la digitalización es la automatización de procesos. Esto implica una reducción significativa de los costes operativos (Plaza, 2020). Áreas como la gestión de inventarios y la comunicación interna pueden beneficiarse enormemente de la automatización, lo que a su vez puede conducir a una mayor productividad y eficiencia.

Aunque los beneficios de la digitalización no están solo centrados en la reducción de costes y mejora de la productividad, aprovechar la digitalización para ser más innovador y diferenciarse de los competidores en clave en el entorno actual (Rivera-Torres y Salas-Fumás, 2023).

En resumen, la digitalización en las Mipymes se traduce en una revolución positiva al ofrecer oportunidades para optimizar operaciones, alcanzar nuevos horizontes de mercado y adaptarse proactivamente a las dinámicas cambiantes del entorno empresarial. A pesar de los desafíos asociados, la adopción inteligente de tecnologías digitales promete posicionar a las Mipymes en un camino hacia la sostenibilidad y el crecimiento a largo plazo.

2.2.2. Desafíos de la Digitalización

La digitalización en las empresas pequeñas y medianas presenta desafíos especiales. Es importante manejar estos desafíos con cuidado y un buen plan para sacar el mayor provecho posible.

Uno de los obstáculos más significativos es la resistencia al cambio por parte de los empleados (Moglia, 2023). Éstos temen que sus habilidades y capacidades actuales ya no sean suficientes y que tengan que cambiar sus hábitos a la hora de trabajar. Pueden sentirse inseguros ante las nuevas tecnologías, pensando que estas podrían tomar su lugar o hacer que su rol en la empresa sea menos importante. Para que un nuevo proyecto o plan de implementación funcione correctamente es importante adoptar una sólida cultura de gestión del cambio. Para ello las organizaciones deben proporcionar una formación integral y un apoyo constante al rendimiento de los empleados (Calzadilla, 2022).

Denning (2023), sugiere que adopción de tecnologías "low code/no-code" se está convirtiendo rápidamente en una prioridad para la construcción de capacidades digitales en todas las organizaciones. Este tipo de tecnología es muy popular entre MiPymes ya que posibilita a los usuarios a crear aplicaciones y procesos digitales sin

tener grandes conocimientos sobre la materia. Son fáciles de usar y permiten personalización y flexibilidad en su desarrollo.

3. Datos

3.1. Extracción de Datos

Para intentar comprobar los objetivos e hipótesis previamente planteados, utilizaré una base de datos facilitada por mi tutor. Esta base de datos contiene información obtenida de una serie de respuestas a una encuesta con una amplia variedad de preguntas, realizada a los propietarios de micro y miniempresas en la República Dominicana. Las preguntas se dividen en diferentes tipos dependiendo de la clase de información que se quiera obtener.

En primer lugar, me voy a centrar en analizar las respuestas a las preguntas que aparecen en el apartado de Tecnologías, donde los entrevistados fueron preguntados por los diferentes recursos tecnológicos que utilizan para operar su negocio. Así voy a poder observar cual es el nivel de conocimiento y adaptación digital que tienen los empresarios sobre sus micro y miniempresas.

Estas preguntas solo se pueden contestar con sí o no, por tanto, las respuestas van a ser dicotómicas 1 o 0. Esto me va a ser útil para eliminar rápidamente aquella información que considere ineficaz.

Dentro del apartado de Tecnologías, me enfocaré en obtener información, transformando en variables las siguientes preguntas.

De los siguientes, ¿de cuáles equipos dispone su negocio?

- Computadoras fijas
- Laptops, tablets
- Celulares
- Servidor central
- Caja registradora conectada (TPV)
- Robots: máquinas computarizadas con movimiento

- Usamos cibercafés de pago
- Usamos infocentros u otros locales gratuitos con computadoras

Entre estas opciones, considero más relevante que un negocio disponga de una computadora fija, laptops, celulares y una caja registradora TPV. Aunque la presencia de alguno de estos recursos tecnológicos es un indicativo de digitalización, no necesariamente implica que el negocio esté completamente digitalizado. Es posible que se cuente con una computadora, pero que no se le esté dando el uso más adecuado para facilitar la transición digital del negocio. Por ejemplo, puede que el equipo se utilice únicamente para tareas administrativas básicas, como el procesamiento de texto o la gestión de correos electrónicos, sin explotar su potencial para optimizar procesos, analizar datos o interactuar con clientes de manera digital.

Creo que si se cuenta con alguno o varios de estos dispositivos es un indicio de que el negocio está digitalizado o en proceso de hacerlo. Sin embargo, la digitalización implica una integración y utilización efectiva de estas herramientas en todas las áreas del negocio.

Otra pregunta que me va a dar información sobre la situación tecnológica de cada negocio está relacionada con las herramientas digitales que utilizan en el día a día. De esta forma voy a poder comprobar la efectividad de la contribución de los recursos tecnológicos que he mencionado anteriormente.

¿Cuenta con alguna de estas herramientas en el negocio?

- Página web propia
- App propia
- Cuentas de la empresa en redes sociales
- Integración en plataformas de venta online (MercadoLibre, Amazon, Glovo)
- Conexión fija a proveedores
- Conexión fija a clientes
- Sistema integrado para el manejo de clientes (CRM)
- Sistema integrado de empresa (ERP)

Entre las diferentes opciones, encuentro realmente útiles las cuatro primeras para evaluar el nivel de digitalización de las MiPymes. Si éstas han conseguido expandir sus canales de venta y comunicación a través de webs y apps propias o la gestión activa de cuentas en redes sociales y plataformas de venta online, entonces considero que están digitalizadas, o al menos han iniciado el proceso de forma significativa.

En paralelo al análisis de la digitalización de las micro y miniempresas, me gustaría analizar si su localización geográfica es un factor relevante a la hora de confirmar con la segunda hipótesis. Es probable que la ubicación de estas empresas influya significativamente en su proceso de digitalización, ya que factores como el acceso a la infraestructura tecnológica, la disponibilidad de recursos digitales y la cultura empresarial pueden variar considerablemente de una región a otra.

Lugar donde opera el negocio:

- Dajabón (1)
- Los Cacaos (2)
- Santo Domingo Savio (3)
- San Cristóbal (4)
- Santo Domingo (ciudad) (5)

Esta pregunta señala la zona de operación de los negocios encuestados. Para comprender mejor los resultados que obtendré, creo que es necesario conocer un poco sobre el contexto de estas localidades y sus ubicaciones dentro del país.

Dajabón: Situada en el noroeste de la República Dominicana, en la frontera con Haití, Dajabón es una ciudad pequeña con una población de aproximadamente 25,000-30,000 habitantes. Esta ciudad es conocida por su vital mercado fronterizo, donde se comercian diversos productos con Haití. La economía local depende significativamente de este intercambio, aunque la región no se cuenta entre las más ricas del país y enfrenta varios retos en términos de desarrollo (Agüera y Cuadra, 2013).

Los Cacaos: Ubicado en la provincia de San Cristóbal, en el sur del país, Los Cacaos es un municipio pequeño con una población que ronda los pocos miles. La economía

de esta área es principalmente agrícola, reflejando un nivel de desarrollo más modesto. La comunidad, rodeada de un paisaje natural, se centra en actividades rurales y agrícolas, sin destacar por su riqueza económica (Ayuntamiento Municipal Los Cacaos, 2023).

Santo Domingo Savio: Este barrio de Santo Domingo, la capital de la República Dominicana se caracteriza por ser una zona densamente poblada. Santo Domingo Savio muestra una mezcla de clases sociales, con áreas de desarrollo variado, aunque es predominantemente una zona de ingreso bajo (URBE).

San Cristóbal: Como capital de la provincia de San Cristóbal y situada al sur de Santo Domingo, cuenta con una población de aproximadamente 200,000-250,000 habitantes. Esta ciudad histórica es un importante centro de industria y comercio, con una economía diversificada que incluye actividades industriales, comerciales y culturales (VisitaRepublicaDominicana).

Santo Domingo (ciudad): La capital de la República Dominicana, ubicada en la costa sur, es la ciudad más grande del país, con una población que supera los 2 millones en su área metropolitana. Santo Domingo es el centro económico, cultural y político del país, destacando por un PIB per cápita más alto en comparación con otras regiones, y alberga importantes centros de negocios, turismo y administración gubernamental (Sagawe, 1985).

También me gustaría añadir el tamaño de la empresa como factor clave para determinar si una empresa con más empleados tiende a estar más digitalizada, o, por el contrario, si el tamaño de la empresa no tiene una relación directa con la adopción de la tecnología.

Número total de trabajadores, incluido el dueño

- 1-2 (1)
- 3-5 (2)
- 6-10 (3)
- 11-20 (4)
- 21-50 (5)

Por último, uno de los datos más importantes para llevar a cabo mi investigación es saber si el negocio está formalizado o no.

¿Tiene que hacer cada año la declaración jurada ante la DGII (Dirección General de Impuestos Internos)?

Si se tiene que hacer la declaración jurada, implica que la empresa está registrada y, por lo tanto, es formal; de lo contrario, sería informal. Para investigar si la formalidad afecta al uso de la tecnología en las empresas la información de esta pregunta va a ser de gran utilidad, ya que me va a servir para comparar los diferentes análisis que lleve a cabo y va a determinar el devenir de la primera hipótesis.

3.1.2. Tratamiento de los datos

A continuación, se ha utilizado simultáneamente tanto Excel como Python para el tratamiento de datos y la elaboración de análisis, estadísticos y modelos. Del total de preguntas o variables que conforman la base de datos creada a través de la encuesta que proporcionó el tutor del trabajo, se han seleccionado cuáles iban a ser más útiles a la hora de generar unos outputs que conformarían un alto nivel de significación para poder comprobar las hipótesis.

En primer lugar, se ha seleccionado la pregunta que ayudaría a identificar aquellas MiPymes que podrían ser consideradas como digitalizadas o en proceso de digitalización. Dentro de las opciones de la pregunta, se puede encontrar la opción de que el encuestado tuviera una computadora fija, un *laptop* o *tablet*, un celular, un servidor de red o una máquina TPV. En la base de datos, si la respuesta del encuestado es afirmativa, está marcada como un 1; en el caso de ser negativa, como un 0. Esto ha facilitado el análisis a la hora de contabilizar qué empresas estaban digitalizadas. Para esta pregunta se ha requerido que los entrevistados tuvieran al menos dos de estos aparatos en su empresa para ser considerados un negocio digitalizado.

La segunda pregunta de la encuesta que se ha tenido en cuenta para determinar si estos negocios son digitalizados o no, era la que preguntaba sobre una serie de herramientas o plataformas a la hora de llevar a cabo la operativa del negocio. Estos recursos son: tener una página web, utilizar alguna aplicación de comercio para la venta o conexión con clientes, la utilización de redes sociales y la creación de una cuenta del negocio en alguna de ellas o la creación de una plataforma digital propia del negocio. Como en la anterior pregunta, la contabilización en las respuestas de las bases de datos es igual, por tanto, el proceso de filtración para determinar aquellas empresas que considero digitalizadas ha sido el mismo.

Pero no se ha tratado estas dos preguntas de forma independiente, sino que se han entendido como complementarias y, por tanto, si un negocio no cumplía con una pregunta, se ha quedado fuera del subgrupo al que yo he considerado negocios digitalizados. En resumen, si un negocio cuenta con dos o más de los aparatos tecnológicos previamente mencionados y, además, cuenta con alguna de las plataformas o recursos digitales mencionados, entonces se ha considerado que esa empresa o ese negocio sí está digitalizado o al menos en un proceso de digitalización avanzado.

Una vez creados los dos subgrupos, negocios digitalizados y no digitalizados, se ha clasificado su formalidad. De igual manera que en las preguntas anteriores, si los encuestados respondieron afirmativamente a su situación formal, en este caso, la respuesta es un 1; en el caso contrario, la respuesta se contabiliza como 0. Este primer tratamiento de datos ha servido para analizar y testear la primera hipótesis.

La tercera pregunta de la encuesta que he analizado y cuyos datos he tratado es relativa a la localidad donde operan los negocios en cuestión. La encuesta se realizó en 5 puntos diferentes de República Dominicana, los cuales han sido explicados en detalle previamente. Por tanto, las respuestas podían ir enumeradas del 1 al 5, siendo Dajabón, Los Cacaos, Santo Domingo Savio, San Cristóbal y Santo Domingo (ciudad), respectivamente.

Por último, la pregunta que se refiere al tamaño de la empresa según el número de empleados diferencia entre 5 tipos de respuesta enumeradas del 1 al 5, siendo 1 para

1-2 empleados, 2 para 3-5 empleados, 3 para 6-10 empleados, 4 para 11-20 empleados, y 5 para 21-50 empleados.

De esta manera he podido filtrar aquellas empresas digitalizadas y no digitalizadas que había conseguido previamente, según su área y localización y tamaño. Esto último es lo que va a ser de ayuda para determinar el análisis de la segunda y tercera hipótesis.

4. Metodología

4.1. Metodología de la Primera Hipótesis

Para la comprobación de la primera hipótesis planteada, se ha realizado un análisis mediante la prueba estadística Chi-cuadrado. Ésta mide la discrepancia entre las frecuencias observadas y las esperadas en una tabla de contingencia. Se usa para determinar si hay una asociación significativa entre dos variables categóricas. Si las variables son independientes, las frecuencias observadas deberían ser similares a las esperadas; grandes diferencias sugieren dependencia (Stewart, 2023).

Paso 1: Identificación de frecuencias observadas y esperadas

La tabla de contingencia muestra las frecuencias observadas de dos categorías: "Digitalizada" y "No Digitalizada", en relación con "Total Formal" y "Total No Formal". También se presenta la tabla de frecuencias esperadas, que se calculan asumiendo que las variables son independientes. Para calcular estas frecuencias esperadas, se utiliza la fórmula: $f_{e_{ij}} = \frac{(f_{i \text{ fila}}) \times (f_{j \text{ columna}})}{n}$, donde $f_{i \text{ fila}}$ y $f_{j \text{ columna}}$ son las sumas marginales y n es el total de observaciones.

Figura N°1: Tabla de Contingencia Primera Hipótesis

Tabla de contingencia			
	TOTAL FORMAL	TOTAL NO FORMAL	
DIGITALIZADA	409	85	494
NO DIGITALIZADA	263	365	628
	672	450	1122

Fuente: Elaboración Propia.

Figura N°2: Tabla de Frecuencias Esperadas Primera Hipótesis

Tabla de frecuencias esperadas			
	TOTAL FORMAL	TOTAL NO FORMAL	
DIGITALIZADA	296	198	494
NO DIGITALIZADA	376	252	628
	672	450	1122

Fuente: Elaboración Propia.

Se han identificado las frecuencias observadas en la tabla de contingencia y calculado las frecuencias esperadas bajo la suposición de independencia.

Paso 2: Cálculo del estadístico Chi-cuadrado (χ^2)

El estadístico Chi-cuadrado mide la discrepancia entre las frecuencias observadas y las esperadas. Se calcula usando la fórmula $\chi^2 = \sum \frac{(f_{o_{ij}} - f_{e_{ij}})^2}{f_{e_{ij}}}$, donde $f_{o_{ij}}$ son las frecuencias observadas y $f_{e_{ij}}$ las frecuencias esperadas para cada celda de la tabla.

Figura N°3: Cálculo de Chi-cuadrado Primera Hipótesis

Calculo Chi-cuadrado			
LOCALIDAD	DIGITALIZADAS	NO DIGITALIZADAS	
1	43.26	64.59	
2	34.03	50.81	
X ² =	77.28	115.41	192.69

Fuente: Elaboración Propia.

Paso 3: Determinación de los grados de libertad

Los grados de libertad para la distribución Chi-cuadrado se calculan como $(I - 1)(J - 1)$, donde I es el número de filas y J es el número de columnas en la tabla de contingencia. En este caso, tenemos 2 filas y 2 columnas, resultando en $(2 - 1)(2 - 1) = 1$ grado de libertad.

Paso 4: Comparación con la distribución Chi-cuadrado

Con el valor de χ^2 calculado y los grados de libertad, comparamos con la distribución Chi-cuadrado para determinar si la discrepancia observada es estadísticamente significativa. Si el valor calculado es mayor que el valor crítico de la distribución para un nivel de significancia dado (usualmente 0.05), rechazamos la hipótesis nula de independencia entre las variables.

Paso 5: Cálculo de C y Cmax

El coeficiente C de Contingencia permite calcular la intensidad de la asociación entre las variables de la tabla de contingencia. La fórmula es la siguiente: $C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$. C_{max} es el máximo valor que puede tomar C, dado por $C_{max} = \sqrt{\frac{k-1}{k}}$ donde k es el menor entre el número de filas y columnas. La relación entre C y C_{max} se expresa como un porcentaje y nos da una idea de cuán fuerte es la asociación en relación con la máxima posible.

4.1.1. Contraste Primera Hipótesis

Los resultados obtenidos del cálculo del estadístico Chi-cuadrado son bastante reveladores:

El valor calculado de Chi-cuadrado es 192.69. Este es un valor alto, lo cual sugiere que hay una diferencia notable entre las frecuencias observadas y las esperadas si las variables fueran independientes. Con solo 1 grado de libertad, la comparación de este valor se hará contra el valor crítico de la distribución Chi-cuadrado correspondiente a ese grado de libertad. Por lo general, para un nivel de significancia de 0.05, el valor crítico es 3.841. Dado que 192.69 es mucho mayor que 3.841,

podemos rechazar la hipótesis nula con un alto grado de confianza. Esto significa que hay suficiente evidencia estadística para sugerir que la formalidad sí afecta al uso de tecnología en las micro y miniempresas dominicanas.

El coeficiente C , que mide la fuerza de la asociación entre las variables, es 0.383. Este valor es una proporción del valor máximo que podría tomar, que es C_{max} (0.707 en este caso).

La proporción de C respecto a C_{max} es 0.541, lo que significa que el coeficiente de contingencia alcanzado es aproximadamente el 54.1% de su valor máximo posible para esa tabla de contingencia. Esto nos ayuda a entender que hay una asociación relevante entre las variables, aunque no es una asociación perfecta o máxima.

Los resultados indican que existe una asociación significativa entre la formalidad de las empresas y la adopción de tecnología digital. La hipótesis nula, que postula que la formalidad no afecta positivamente al uso de tecnología, se rechaza. Esto sugiere que la formalidad de las empresas podría tener un efecto positivo en la adopción de tecnología digital en las micro y miniempresas dominicanas.

4.2. Metodología de la Segunda Hipótesis

Para realizar la comprobación de la segunda hipótesis, voy a volver a realizar un estudio basado en la obtención de la prueba estadística Chi-cuadrado. Esta vez, la tabla de contingencia obtenida será diferente a la de la primera hipótesis, ya que cuenta con 5 filas, referentes a las 5 localidades posibles

Paso 1: Entender la Tabla de Contingencia

La tabla de contingencia muestra la distribución de frecuencias de dos variables categóricas: digitalización, “Digitalizadas” y “No Digitalizadas” y localidades, Dajabón, Los Cacaos, Santo Domingo Savio, San Cristóbal, Santo Domingo (ciudad). Las frecuencias observadas se muestran en la siguiente tabla.

Figura N°4: Tabla de Contingencia Segunda Hipótesis

Tabla de contingencia			
LOCALIDAD	DIGITALIZADAS	NO DIGITALIZADAS	
1	142	160	302
2	15	140	155
3	19	153	172
4	55	57	112
5	263	118	381
	494	628	1122

Fuente: Elaboración Propia.

Paso 2: Calcular Frecuencias Esperadas

Las frecuencias esperadas se calculan bajo la hipótesis de que las variables son independientes. Para calcular estas frecuencias esperadas, se utiliza la fórmula:

$$f_{e_{ij}} = \frac{(f_{i \text{ fila}}) \times (f_{j \text{ columna}})}{n}, \text{ donde } f_{i \text{ fila}} \text{ y } f_{j \text{ columna}} \text{ son las sumas marginales y } n \text{ es el}$$

total de observaciones. general. Por ejemplo, para Dajabón y Digitalizadas: $\frac{(302 \times 494)}{1122}$.

Figura N°5: Tabla de Frecuencias Esperadas Segunda Hipótesis

Tabla de frecuencias esperadas			
LOCALIDAD	DIGITALIZADAS	NO DIGITALIZADAS	
1	132.97	169.03	302
2	68.24	86.76	155
3	75.73	96.27	172
4	49.31	62.69	112
5	167.75	213.25	381
	494	628	1122

Fuente: Elaboración Propia.

Paso 3: Realizar el Cálculo Chi-cuadrado

Como he realizado en la prueba anterior, voy a volver a calcular el Chi-cuadrado para medir la discrepancia entre lo observado y lo esperado y poder observar si las variables fueran independientes.

Figura N°6: Cálculo de Chi-cuadrado Segunda Hipótesis

Calculo Chi-cuadrado			
LOCALIDAD	DIGITALIZADAS	NO DIGITALIZADAS	
1	0.61	0.48	
2	41.54	32.68	
3	42.50	33.43	
4	0.66	0.52	
5	54.09	42.55	
X ² =	139.39	109.65	249.04

Fuente: Elaboración Propia.

Paso 4: Determinación de los grados de libertad

Los grados de libertad para la distribución Chi-cuadrado se calculan como $(I - 1)(J - 1)$, donde I es el número de filas y J es el número de columnas en la tabla de contingencia. En este caso, tenemos 5 filas y 2 columnas, resultando en $(5 - 1)(2 - 1) = 4$ grados de libertad.

Paso 5: Interpretación del valor p

El valor p obtenido del cálculo de Chi-cuadrado nos indica la probabilidad de obtener un resultado al menos tan extremo como el observado, bajo la suposición de que la hipótesis nula es cierta. Un valor p muy bajo sugiere que es improbable observar tales diferencias si la hipótesis nula fuera cierta, llevándonos a rechazarla.

Paso 6: Cálculo de C y Cmax

Como he explicado previamente, el coeficiente C ofrece una medida de la fuerza de la asociación entre las variables y Cmax se refiere al valor máximo que C puede alcanzar, dado el tamaño de la tabla de contingencia. También es útil volver a calcular la relación entre ambos para obtener una mejor idea de la fuerza de la asociación relativa al máximo posible para la tabla dada.

4.2.1. Contraste Segunda Hipótesis

En el contexto de este análisis, la prueba se ha utilizado para determinar si la localización geográfica de las micro y miniempresas dominicanas tiene un efecto en su estado de digitalización.

Los resultados obtenidos del test de Chi-cuadrado muestran un valor de aproximadamente 249.04, un número muy alto, lo cual indica que la probabilidad de obtener la distribución observada de empresas digitalizadas y no digitalizadas por localidad, asumiendo que la localización no tiene ningún efecto, es prácticamente nula. Con 4 grados de libertad, la comparación de este valor se hará contra el valor crítico de la distribución Chi-cuadrado correspondiente a estos grados de libertad. Por lo general, para un nivel de significancia de 0.05, el valor crítico es 9.488. Al ser 249.04 mucho mayor que 9.488, se sugiere fuertemente que la hipótesis nula, que postula la independencia entre la localización y la digitalización, debe ser rechazada.

El coeficiente de contingencia calculado es 0.426. Este valor, que oscila entre 0 y 1, es una medida de la fuerza de la asociación. Aunque nunca alcanza 1 en tablas de contingencia mayores a 2x2, un valor de aproximadamente 0.43 indica una asociación moderada. Significa que hay una relación notable entre la localización de las empresas y su digitalización, pero esta relación no es extremadamente fuerte.

El valor de CMax es 0.707, que es el valor teórico máximo que C puede alcanzar para una tabla de contingencia con el número mínimo de categorías igual a 2. En la práctica, es difícil alcanzar este valor máximo, especialmente en tablas de contingencia con muchas categorías. La relación entre el coeficiente de contingencia y su máximo es 0.60 Este valor relativo nos dice cuán cercano está el coeficiente de contingencia observado al valor máximo posible. En este caso, la relación es mayor al 60%, lo cual refuerza la idea de que existe una asociación significativa entre las variables analizadas.

En resumen, podemos rechazar la hipótesis nula con un alto grado de confianza. Los resultados indican que la localización de las micro y miniempresas en la República Dominicana tiene un efecto sobre su digitalización. Esto puede deberse a factores

como la disponibilidad de infraestructura tecnológica, el acceso a la formación en TIC, las políticas locales de promoción empresarial, entre otros factores que pueden variar de una localidad a otra.

4.3. Metodología de la Tercera Hipótesis

Para realizar la comprobación de esta tercera hipótesis voy a obtener el estadístico Chi-cuadrado de nuevo. Esta vez, la tabla de contingencia obtenida será igual que la de la hipótesis anterior ya que cuenta con 5 filas, referentes a los 5 tipos de tamaños posibles.

Paso 1: Entender la Tabla de Contingencia

La tabla de contingencia muestra la distribución de frecuencias de dos variables categóricas. Para este análisis, las filas representan los diferentes tamaños de empresa, y las columnas representan si están digitalizadas o no.

Figura N°7: Tabla de Contingencia Tercera Hipótesis

Tabla de contingencia			
TAMAÑO	DIGITALIZADAS	NO DIGITALIZADAS	
1	100	357	457
2	164	180	344
3	83	43	126
4	94	33	127
5	53	15	68
	494	628	1122

Fuente: Elaboración Propia.

Paso 2: Calcular Frecuencias Esperadas

Para este paso, he realizado exactamente el mismo procedimiento que en la hipótesis anterior.

Figura N°8: Tabla de Frecuencias Esperadas Tercera Hipótesis

Tabla de frecuencias esperadas			
TAMAÑO	DIGITALIZADAS	NO DIGITALIZADAS	
1	201.21	255.79	457
2	151.46	192.54	344
3	55.48	70.52	126
4	55.92	71.08	127
5	29.94	38.06	68
	494	628	1122

Fuente: Elaboración Propia.

Paso 3: Realizar el Cálculo Chi-cuadrado

Como he realizado en las anteriores pruebas, voy a volver a calcular el Chi-cuadrado para evaluar la discrepancia entre lo observado y lo esperado, y determinar la posible independencia de las variables.

Figura N°9: Cálculo de Chi-cuadrado Tercera Hipótesis

Calculo Chi-cuadrado			
TAMAÑO	DIGITALIZADAS	NO DIGITALIZADAS	
1	50.91	40.05	
2	1.04	0.82	
3	13.66	10.74	
4	25.94	20.40	
5	17.76	13.97	
X ² =	109.30	85.98	195.29

Fuente: Elaboración Propia.

Paso 4: Determinación de los grados de libertad

Para interpretar el valor de chi-cuadrado, necesitamos saber los grados de libertad, que dependen del número de categorías en cada variable. En este caso, tenemos 5

tamaños de empresas y 2 estados de digitalización, resultando en $(5 - 1)(2 - 1) = 4$ grados de libertad.

Paso 5: Interpretación del valor p

Como he mencionado anteriormente, el valor p nos dice la probabilidad de obtener un valor de chi-cuadrado tan grande como el calculado o mayor, bajo la suposición de que la hipótesis nula es cierta. Un valor p muy bajo sugiere que los datos observados son muy improbables bajo esta suposición, llevando a rechazar la hipótesis nula.

Paso 6: Cálculo de C y Cmax

También es útil volver a calcular la relación entre ambos para obtener una mejor idea de la fuerza de la asociación relativa al máximo posible para la tabla dada.

4.3.1. Contraste Tercera Hipótesis

El valor de chi-cuadrado que se ha obtenido, 195.29, es bastante alto, lo que nos señala que las diferencias entre las observaciones y las expectativas no son pequeñas, sugiriendo que la variabilidad que observamos probablemente no es debido al azar.

En este caso, el valor p es prácticamente cero ($3.874146E-41$), menor que el umbral típico de 0.05. Esto significa que la probabilidad de que las diferencias observadas sean aleatorias es muy baja, por lo tanto, rechazamos la hipótesis nula. En términos prácticos, es casi seguro que el tamaño de la empresa tiene algún efecto en si una empresa está digitalizada o no.

Aquí, el valor de C es aproximadamente 0.385, lo que indica una fuerza de asociación moderada. No es tan cercano a 0 como para ser insignificante, ni es tan cercano a 1 como para sugerir una asociación perfecta. Sin embargo, es lo suficientemente fuerte como para ser notable. El valor de CMax es el máximo posible que C podría ser dado el número de filas y columnas en la tabla. Aquí es aproximadamente 0.707, lo que nos dice que, dado el tamaño de nuestra tabla, nuestro C es un poco más de la mitad del máximo posible, lo cual es bastante significativo. La relación entre los coeficientes,

aproximadamente 0.544, es otra manera de expresar la fuerza de la asociación relativa al máximo posible. La asociación entre el tamaño de la empresa y la digitalización es más de la mitad de fuerte que la asociación más fuerte posible, reforzando la conclusión de que hay una asociación notable entre ambas.

En resumen, el análisis de chi-cuadrado proporciona evidencia estadística fuerte de que el tamaño de una empresa influye en su probabilidad de estar digitalizada en el contexto de las micro y miniempresas dominicanas, así que podemos rechazar H_0 . Esta conclusión es importante para las políticas y estrategias empresariales, especialmente para aquellos que buscan promover la digitalización en este sector.

4.4. Creación de Modelo de Regresión Logística con las variables independientes

Tras haber realizado el estudio de la prueba Chi-cuadrado he podido comprobar que existe una relación fuerte o media entre la digitalización de las empresas y las otras 3 variables independientes. Pero para ir un poco más en detalle y descubrir cómo afectan cada una de estas variables en la probabilidad de que una empresa esté digitalizada, voy a crear un modelo de regresión logística.

La regresión logística nos permitiría modelar la probabilidad de que una empresa esté digitalizada como función de una o más variables independientes, para comprender mejor los factores que influyen en la digitalización. Y además voy interpretar los coeficientes del modelo para entender el efecto y la importancia de cada variable independiente en la probabilidad de que una empresa esté digitalizada.

Paso 1: Definición del Problema

Hay que diferenciar entre la variable dependiente que voy a predecir, la Digitalización de las Empresas, y las variables independientes, Lugar, Numero de Trabajadores, y Formalidad.

Paso 2: Recopilación y Preparación de Datos

He creado un archivo Excel diferente al inicial para poder limpiar correctamente los datos y asegurarme de que no haya valores errores.

Paso 3: Transformación de Variables

A partir de este paso he utilizado la herramienta Python para todo el tratamiento de los datos. He convertido las variables categóricas en variables dummy y me he asegurado de que la variable dependiente sea binaria (1 para digitalizadas, 0 para no digitalizadas).

Paso 4: División de Datos

He dividido los datos en un conjunto de entrenamiento y un conjunto de prueba para poder validar la capacidad predictiva del modelo.

Paso 5: Ajuste del Modelo

Para ajustar el modelo de regresión logística al conjunto de entrenamiento, he utilizado la librería scikit-learn.

Paso 6: Validación del Modelo

He evaluado la precisión del modelo utilizando el conjunto de prueba y métricas como la matriz de confusión, precisión, recall y F1-score.

Paso 7: Interpretación de Coeficientes

Una vez que el modelo es ajustado, he examinado los coeficientes para entender la dirección y magnitud del efecto de cada variable independiente sobre la probabilidad de que la empresa esté digitalizada.

Paso 8: Evaluación de Significancia

Hay que observar los valores p asociados con cada coeficiente para determinar si las variables son estadísticamente significativas.

4.4.1. Resultados del Modelo de Regresión Logística

Primero he evaluado la precisión del modelo para cerciorarme que los resultados posteriores que voy a obtener son relevantes. Para ello hay que fijarse en las siguientes métricas y comprobar si mi modelo es apto:

Figura N°10: Resultado Rendimiento del Modelo

	precision	recall	f1-score	support
NO	0.76	0.81	0.78	118
SI	0.78	0.71	0.74	107
accuracy			0.76	225
macro avg	0.77	0.76	0.76	225
weighted avg	0.77	0.76	0.76	225
[[96 22]				
[31 76]]				

Fuente: Elaboración Propia.

La precisión es una medida que indica cuántos de los elementos identificados como pertenecientes a una clase son realmente de esa clase. Para la clase "NO" (no digitalizada), el modelo tiene una precisión del 76%, lo que significa que el 76% de las veces que el modelo predijo la clase "NO", estaba en lo correcto. Para la clase "SI" (digitalizada), tiene una precisión del 78%.

El recall es la sensibilidad, e indica cuántos de los elementos verdaderamente pertenecientes a una clase fueron identificados correctamente. Para la clase "NO", el modelo tiene un recall del 81%, lo que significa que logró identificar el 81% de todas las instancias reales de la clase "NO". Para la clase "SI", el recall es del 71%.

El F1-score es una medida que combina la precisión y el recall en un solo número mediante su media armónica. Un F1-score alto indica un buen equilibrio entre precisión y recall. En este caso, los F1-scores son 0.78 para "NO" y 0.74 para "SI".

La accuracy del modelo es del 76%, lo que indica una capacidad adecuada para clasificar correctamente si las empresas están digitalizadas o no, basándose en las variables proporcionadas. Por tanto, tengo un modelo consistente con un desempeño general sólido.

Por último, he analizado los coeficientes y evaluado su significancia a partir de los siguientes resultados:

Figura N°11: Resultados por Variable del Modelo

Logit Regression Results						
Dep. Variable:	DIG/NODIG	No. Observations:	1122			
Model:	Logit	Df Residuals:	1119			
Method:	MLE	Df Model:	2			
Date:	Mon, 15 Apr 2024	Pseudo R-squ.:	0.03773			
Time:	13:31:48	Log-Likelihood:	-740.65			
converged:	True	LL-Null:	-769.69			
Covariance Type:	nonrobust	LLR p-value:	2.438e-13			
	coef	std err	z	P> z	[0.025	0.975]
Lugar	-0.1985	0.033	-6.005	0.000	-0.263	-0.134
NumTrabajadores	0.0633	0.048	1.305	0.192	-0.032	0.158
Formalidad	0.9936	0.144	6.887	0.000	0.711	1.276
Lugar	1.918827e-09					
NumTrabajadores	1.918441e-01					
Formalidad	5.700034e-12					
dtype: float64						

Fuente: Elaboración Propia.

El coeficiente en un modelo de regresión logística representa el cambio en el logaritmo de probabilidad (log-odds) de la variable dependiente por cada unidad de cambio en la variable independiente, manteniendo todas las otras variables independientes constantes. Mientras que log-odds es el logaritmo natural del cociente de probabilidades (odds) de que ocurra un evento. En este caso, sería el logaritmo de la probabilidad de que una empresa esté digitalizada dividido por la probabilidad de que no lo esté.

Los coeficientes de las variables independientes son:

- **Lugar:** El coeficiente de 0.1910 sugiere que hay una relación positiva entre la variable 'Lugar' y la probabilidad de que una empresa esté digitalizada. Por cada unidad que aumenta 'Lugar', el logaritmo de la probabilidad de estar digitalizado aumenta en 0.1910.
- **NumTrabajadores:** Un coeficiente de 0.5451 indica una relación positiva entre el número de trabajadores y la probabilidad de digitalización. Esto sugiere que a medida que aumenta el número de trabajadores, también lo hace la probabilidad de digitalización de la empresa.
- **Formalidad:** Con un coeficiente de 1.3233, esta variable tiene un impacto positivo fuerte en la probabilidad de que una empresa esté digitalizada,

sugiriendo que las empresas formales tienen mucha más probabilidad de estar digitalizadas que las informales.

Los errores estándar (std err) asociados con cada coeficiente son relativamente bajos, lo que sugiere una precisión razonable en las estimaciones de los coeficientes. Todos los valores p ($P > |z|$) son 0.000 (realmente son números muy cercanos a 0, como se ha visto previamente en el cálculo de las Chi-cuadrado), lo que indica que todas las variables independientes son estadísticamente significativas en el modelo al nivel de confianza estándar. Esto significa que hay una probabilidad muy baja de que estos efectos sean debidos al azar.

La precisión del modelo ha demostrado ser consistente con un desempeño general sólido que respalda la relevancia de los resultados obtenidos. Este modelo ha permitido evaluar la influencia de cada variable independiente en la probabilidad de que una empresa esté digitalizada. Y los coeficientes indican que las variables utilizadas tienen efectos positivos y estadísticamente significativos sobre la digitalización.

5. Conclusión

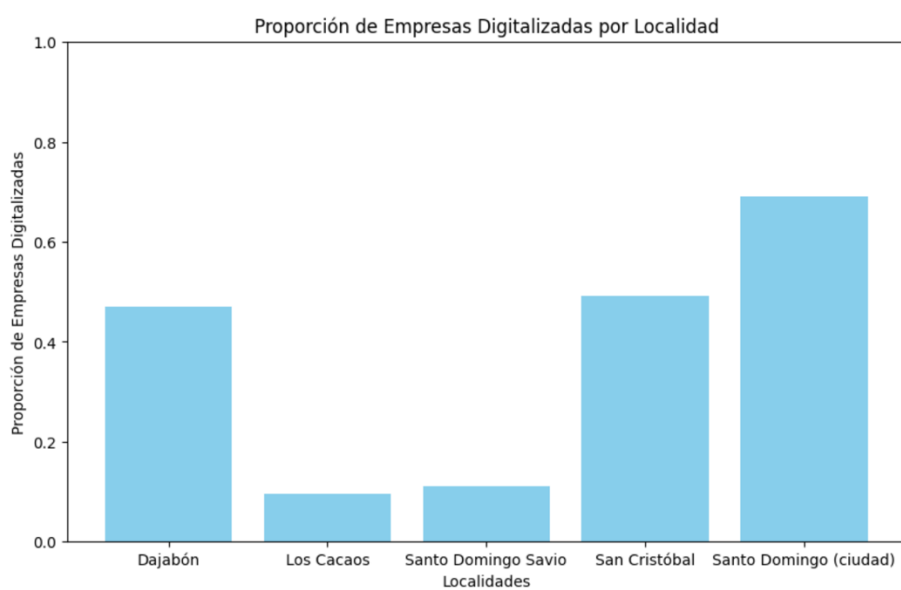
Los objetivos de este trabajo eran demostrar cómo la formalidad, la localización y el tamaño de las empresas influyen la adopción de tecnologías digitales en las micro y miniempresas dominicanas. A través de un análisis utilizando el test de Chi-cuadrado y un modelo de regresión logística, se ha podido demostrar con evidencia estadística significativa que estos factores no solo están asociados con la digitalización, sino que son determinantes críticos en la transformación digital de este segmento empresarial.

El análisis comenzó con la hipótesis de que la formalidad de las empresas afecta positivamente al uso de tecnología. Los resultados obtenidos han sido reveladores, mostrando un valor de Chi-cuadrado extremadamente alto que permite rechazar la hipótesis nula con un alto grado de confianza. El coeficiente de contingencia obtenido sugiere una asociación relevante, aunque no máxima, entre la formalidad y la digitalización, lo que implica que las empresas formales tienen mayores

probabilidades de incorporar tecnologías digitales en sus operaciones. Además, para reafirmar más este punto, tras analizar los resultados del modelo, la formalidad mostró el impacto más fuerte entre las demás variables, destacando que las empresas formales son mucho más propensas a estar digitalizadas que las informales.

En lo que respecta a la localización, los resultados también han sido relevantes. La significancia del Chi-cuadrado sugiere que la localización geográfica de las empresas es un factor de peso en su digitalización. Y aunque se rechaza la hipótesis nula, el grado de asociación, aunque es moderado no es tan fuerte como podría ser. Esto implica que, aunque la localización tiene un efecto, hay otros factores que también podrían estar influyendo mayormente en el nivel de digitalización de las empresas. A parte del análisis del estadístico, creo que también es importante visualizar las proporciones de empresas digitalizadas en las cinco localidades diferentes:

Figura N°12: Proporción Empresas Digitalizadas por Localidad



Fuente: Elaboración Propia.

Este análisis proporciona una vista clara de la disparidad en la digitalización entre distintas localidades. Podría ser útil para las autoridades enfocar esfuerzos en localidades con baja digitalización y entender las barreras que enfrentan estas empresas. De este resultado se puede destacar cómo el entorno regional, incluyendo la infraestructura disponible y las políticas de apoyo local, pueden afectar notoriamente la capacidad de una empresa para adoptar nuevas tecnologías.

Por último, el tamaño de la empresa también ha mostrado ser un predictor significativo de la digitalización. El valor de Chi-cuadrado obtenido en este caso refuerza la idea de que el tamaño de una empresa afecta directamente su probabilidad de estar digitalizada. Tras observar los resultados del modelo también llegamos a esta conclusión, el número de trabajadores mostró una relación positiva considerable, lo que implica que a medida que el tamaño de la empresa aumenta, también lo hace la probabilidad de estar digitalizada. Las empresas más grandes tienden a tener más recursos, lo que probablemente les facilita invertir en tecnologías digitales. Este hallazgo es crucial para entender que las políticas y estrategias de digitalización necesitan ser adaptadas según el tamaño de las empresas, promoviendo soluciones escalables que puedan beneficiar a empresas de todos los tamaños.

Conjuntamente, estos resultados no solo confirman y verifican las hipótesis planteadas, sino que también proporcionan una perspectiva clara de los factores que deben ser considerados al desarrollar estrategias para incrementar la adopción de tecnología en las micro y miniempresas dominicanas. Empezando por la formalización, que se posiciona como un proceso fundamental, sugiriendo que las políticas que promueven la formalidad podrían ser especialmente efectivas para fomentar la digitalización. También, el impacto de la localización implica que debe prestarse atención especial a mejorar la infraestructura tecnológica y ofrecer capacitaciones en áreas menos desarrolladas para asegurar que la digitalización no se concentre únicamente en zonas urbanas o desarrolladas, como se puede apreciar en el último gráfico.

En conclusión, el estudio resalta la importancia de enfoques personalizados y específicos que no solo se centren en promover la digitalización en sí misma, sino que también consideren las condiciones particulares y necesidades de las micro y miniempresas. Al diseñar estrategias de digitalización que sean inclusivas y efectivas, las autoridades y líderes políticos pueden garantizar que el progreso en esta adaptación se convierta en un crecimiento sostenible y equitativo para el sector empresarial en la República Dominicana.

Sin embargo, a pesar de los hallazgos significativos y las implicaciones prácticas del estudio, existen varias limitaciones que deben ser consideradas. Primero, la generalización de los resultados puede estar limitada por la selección y el tamaño de la muestra, que se centró exclusivamente en micro y miniempresas en la República Dominicana de 5 localidades, sin incluir más puntos del país. Además, se tienen en cuenta solamente 3 variables independientes para la estimación del modelo, lo que puede provocar un sesgo de especificación, pudiendo llevar a generar estimaciones incorrectas. El estudio tampoco explora variables externas, como el entorno macroeconómico, cambios regulatorios recientes, o crisis globales (como la pandemia de COVID-19) podrían influir en la adopción de tecnologías digitales. Estas variables podrían alterar o moderar las relaciones observadas entre la formalidad, la localización, el tamaño de la empresa y su nivel de digitalización.

Para futuras investigaciones sería importante considerar estos aspectos, dado que representa oportunidades de mejorar de este estudio. Esto facilitaría el desarrollo de un modelo más generalizado e imparcial, que podría aplicarse en una variedad de contextos empresariales.

6. Bibliografía

Acosta, C. (2021-07-03). *En el país existe 1.5 millones de Pymes que aportan el 39% del PIB*. Listín Diario. <https://listindiario.com/economia/2021/07/03/677669/en-el-pais-existe-1-5-millones-de-pymes-que-aportan-el-39-del-pib.html>

Agüera, F. y Cuadra, S. (2013). *El impacto de la responsabilidad social corporativa en las empresas de servicios de la ciudad de Dajabon, República Dominicana*. <https://www.eumed.net/cursecon/ecolat/la/13/empresas-servicios.html>

Amat, J. (2023-02-23). *Machine learning con Python y Scikit-learn*. Ciencia de datos. https://cienciadedatos.net/documentos/py06_machine_learning_python_scikitlearn

Armas, J. A. (2018). *Digitalización empresarial: una nueva adopción tecnológica*. Review of Global Management, 4(1), 17-17.

López Valerio, A. (2020-09-30). *Retos de la digitalización de las empresas en la República Dominicana*. <https://www.arturolopezvalerio.com/retos-de-la-digitalizacion-de-las-empresas-en-la-republica-dominicana/>

Ayuntamiento Municipal Los Cacaos. (2023) *Historia*. Recuperado de <https://ayuntamientoloscacaos.gob.do/historia/>

Banco Mundial. (2023). *República Dominicana: panorama general*. Banco Mundial. <https://www.bancomundial.org/es/country/dominicanrepublic/overview>

Calzadilla, C. (2022). *9 desafíos de la transformación digital a superar en 2022*. *Mundo Posgrado*. Recuperado de <https://www.mundoposgrado.com/desafios-de-la-transformacion-digital-en-2022/>

CDN. (2023-03-29). 88% de PyMEs dominicanas está en transformación digital. CDN. <https://www.cdn.com.do/2023/88-de-pymes-dominicanas-esta-en-transformacion-digital/>

Denning, S. (2023), "*Recognizing and outmaneuvering the resistance to digital transformation*", *Strategy & Leadership*, Vol. 51 No. 2, pp. 10-16.

Diario Libre. (2022). *Informalidad laboral en país era de 45.5 % en 2021*. Diario Libre. Recuperado de <https://www.diariolibre.com/actualidad/economia/informalidad-laboral-en-pais-era-de-45-5-en-2021-YN39549931>

Feal, F. y Ventura, J. (2023). *El desafío de la formalización empresarial en Paraguay: Causas, motivaciones y propuestas de política pública*. Banco Interamericano de Desarrollo. <https://publications.iadb.org/es/el-desafio-de-la-formalizacion-empresarial-en-paraguay-causas-motivaciones-y-propuestas-de-politica>

Forbes Staff. (2017-11-23). *El 80% de las microempresas en RD son informales*. Forbes México. <https://www.forbes.com.mx/el-80-de-las-microempresas-en-rd-son-informales/>

Franklin, V. (2018-11-29). *La esencia de la informalidad de la economía dominicana*. El Dinero. <https://eldinero.com.do/66942/la-esencia-de-la-informalidad-de-la-economia-dominicana/>

Galiani, S. y Meléndez, M. (2013). *Lecciones a partir de experimentos de formalización empresarial*. Banco Interamericano de Desarrollo. <https://publications.iadb.org/es/publicacion/15870/lecciones-partir-de-experimentos-de-formalizacion-empresarial>

García, P. (2021-09-11). *Mipymes dominicanas limitadas por las trabas para acceso al crédito y a la formalidad*. Diario Libre. <https://www.diariolibre.com/economia/mipymes-dominicanas-limitadas-por-las-trabas-para-acceso-al-credito-y-a-la-formalidad-FE28689873>

J-PAL Policy Insight. (2022). *“Encouraging micro and small enterprises to formalize.”* Cambridge, MA: Abdul Latif Jameel Poverty Action Lab. <https://www.povertyactionlab.org/policy-insight/encouraging-micro-and-small-enterprises-formalize>

Li, S. (2017-09-29). *Building A Logistic Regression in Python, Step by Step*. Towards Data Science. <https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8>

Microsoft News Center Latinoamérica. (2022-02-11). *85% de las MiPyMEs dominicanas aceleraron su proceso de transformación digital durante la pandemia*. <https://news.microsoft.com/es-xl/85-de-las-mipymes-dominicanas-aceleraron-su-proceso-de-transformacion-digital-durante-la-pandemia/>

Ministerio de Industria, Comercio y Mipymes (MICM) & Instituto Tecnológico de Santo Domingo (INTEC). (2019). *Formalizarse es un Buen Negocio: Apoyando a las micro, pequeñas y medianas empresas de la República Dominicana*.

[https://micm.gob.do/transparencia/images/pdf/publicaciones/libros/boletin/Boletin Fo
rmalizarse es un Buen Negocio.pdf](https://micm.gob.do/transparencia/images/pdf/publicaciones/libros/boletin/Boletin_Fo
rmalizarse_es_un_Buen_Negocio.pdf)

Moglia, S. (2023-03-31) *Transformación digital: Desafíos y oportunidades para las empresas*. Forbes México. <https://www.forbes.com.mx/transformacion-digital-desafios-y-oportunidades-para-las-empresas/>

Oficina Nacional de Estadística (ONE). (2023). *Informe Caracterización de las Mipymes módulo ENHOGAR 2022*. República Dominicana. Recuperado de <https://www.one.gob.do/publicaciones/2023/caracterizacion-de-las-mipymes-encuesta-nacional-de-hogares-de-propositos-multiples-enhogar-2022/>

Plaza, C. (2020). *La digitalización de las PYMES*. https://crea.ujaen.es/jspui/bitstream/10953.1/12646/1/TFG_Carmen_Plaza.pdf

Presidencia de la República Dominicana. (2023). *Gobierno considera a las mipymes como el pulmón del desarrollo económico dominicano*. <https://presidencia.gob.do>

Programa de las Naciones Unidas para el Desarrollo. (2020). *Informe Encuesta Mipymes PNUD/MICM*. https://dominicanrepublic.un.org/sites/default/files/2020-07/INFORME_Encuesta_Mipymes_PNUDMICM_17_6_2020-comprimido.pdf

Programa de las Naciones Unidas para el Desarrollo. (n.d.). *Fomento a las MIPYMES / República Digital Productiva*. <https://www.undp.org/es/dominican-republic/proyectos/fomento-las-mipymes/republica-digital-productiva>

Python – Pearson’s Chi-Square Test. (2023-04-26). GeeksforGeeks. <https://www.geeksforgeeks.org/python-pearsons-chi-square-test/>

Rivera-Torres, P. y Salas-Fumás, V. (2023). *La digitalización de las empresas españolas: resumen de resultados comparados con las empresas de la UE*. Cuadernos de Información Económica, 292.

Sagawe, T. (1985). *El desarrollo industrial en República Dominicana. una perspectiva espacial*. Eme-eme: estudios dominicanos, 13(77), 3-34.

Severino, J. (2021-10-22). *La reforma fiscal debe romper el pánico de las mipymes a formalizarse*. El Dinero. <https://eldinero.com.do/176803/la-reforma-fiscal-debe-romper-el-panico-de-las-mipyme-a-formalizarse/>

Stewart, K. (2023). *Chi-squared test*. Encyclopaedia Britannica. <https://www.britannica.com/topic/chi-squared-test>

Unidad Ejecutora para la Readequación de Barrios y Entornos (URBE). (n.d.). *Nuevo Domingo Savio*. Ministerio de la Presidencia. <https://urbe.gob.do/galeria/nuevo-domingo-savio/#:~:text=El%20Barrio%20Domingo%20Savio%20se,Mar%C3%ADa%20Auxiliar%20y%20Ensanche%20Espaillat.>

VisitaRepúblicaDominicana.org. San Cristóbal. (n.d.). <https://www.visitarepublicadominicana.org/san-cristobal>

7. Anexos

Anexo I: Código Python Cálculo Chi-Cuadrado

```
import numpy as np
import scipy.stats as stats

#PRIMERA HIPOTESIS

# Copiar los datos de la tabla de contingencia
tablaC = np.array([[409, 85], [263, 365]])

# Test de chi cuadrado
chi2, p_valor, glib, F_esperadas = stats.chi2_contingency(tablaC)

# Coeficiente de contingencia
C = np.sqrt(chi2 / (np.sum(tablaC) + chi2))

# Cálculo de C_max
k = min(tablaC.shape) # Menor número entre columnas o filas
C_max = np.sqrt((k - 1) / k)
```

```

# Cálculo de la proporción de C respecto a C_max
C_prop = C / C_max

# Visualización de los datos
print("\n")
print("PRIMERA HIPÓTESIS")
print(f"Chi-cuadrado: {chi2}")
print(f"P-valor: {p_valor}")
print(f"Grados de Libertad: {glib}")
print(f"Frecuencias Esperadas: \n{F_esperadas}")
print(f"Coeficiente de Contingencia:{C}")
print(f"C_max:{C_max}")
print(f"Proporción de C respecto C_max:{C_prop}")

#SEGUNDA HIPOTESIS

# Copiar los datos de la tabla de contingencia
tablaC2 = np.array([[142, 160],
                    [15, 140],
                    [19, 153],
                    [55, 57],
                    [263, 118]])

# Test de chi cuadrado
chi2_2, p_valor2, glib2, F_esperadas2 = stats.chi2_contingency(tablaC2)

# Coeficiente de contingencia
C2 = np.sqrt(chi2_2 / (np.sum(tablaC2) + chi2_2))

# Cálculo de C_max
k2 = min(tablaC2.shape) # Menor número entre columnas o filas
C_max2 = np.sqrt((k2 - 1) / k2 )

# Cálculo de la proporción de C respecto a C_max
C_prop2 = C2 / C_max2

# Visualización de los datos
print("\n")
print("SEGUNDA HIPÓTESIS")
print(f"Chi-cuadrado: {chi2_2}")
print(f"P-valor: {p_valor2}")
print(f"Grados de Libertad: {glib2}")
print(f"Frecuencias Esperadas: \n{F_esperadas2}")
print(f"Coeficiente de Contingencia:{C2}")
print(f"C_max:{C_max2}")
print(f"Proporción de C respecto C_max:{C_prop2}")

```

```

#TERCERA HIPOTESIS

# Copiar los datos de la tabla de contingencia
tablaC3 = np.array([[100, 357],
                    [164, 180],
                    [83, 43],
                    [94, 33],
                    [53, 15]])

# Test de chi cuadrado
chi2_3, p_valor3, glib3, F_esperadas3 = stats.chi2_contingency(tablaC3)

# Coeficiente de contingencia
C3 = np.sqrt(chi2_3 / (np.sum(tablaC3) + chi2_3))

# Cálculo de C_max
k3 = min(tablaC3.shape) # Menor número entre columnas o filas
C_max3 = np.sqrt((k3 - 1) / k3 )

# Cálculo de la proporción de C respecto a C_max
C_prop3 = C3 / C_max3

# Visualización de los datos
print("\n")
print("TERCERA HIPÓTESIS")
print(f"Chi-cuadrado: {chi2_3}")
print(f"P-valor: {p_valor3}")
print(f"Grados de Libertad: {glib3}")
print(f"Frecuencias Esperadas: \n{F_esperadas3}")
print(f"Coeficiente de Contingencia:{C3}")
print(f"C_max:{C_max3}")
print(f"Proporción de C respecto C_max:{C_prop3}")

```

Anexo II: Código Python Modelo de Regresión Lineal

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.preprocessing import LabelEncoder
import statsmodels.api as sm

# Cargo datos
datos = pd.read_excel('/content/prueba_modelo_reglog.xlsx')

# Codifico la variable objetivo/dependiente en binario
le = LabelEncoder()

```

```

datos['DIG/NODIG'] = le.fit_transform(datos['DIG/NODIG'])

# Variables independientes y dependiente
X = datos[['Lugar', 'NumTrabajadores', 'Formalidad']]
y = datos['DIG/NODIG']

# Dividir los datos en training y test
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Crear y entrenar el modelo de regresión logística
logreg_modelo = LogisticRegression()
logreg_modelo.fit(X_train, y_train)

# Realizar predicciones
y_pred = logreg_modelo.predict(X_test)

# Visualización del informe de clasificación y matriz de confusión
print(classification_report(y_test, y_pred, target_names=le.classes_))
print(confusion_matrix(y_test, y_pred))

# Ajustar el modelo usando statsmodels para obtener los p-valores
modelo_logit = sm.Logit(y, X)
result = modelo_logit.fit()

# Visualización del resumen del modelo que incluye los coeficientes y p-
valores
print(result.summary())
p_valores = result.pvalues
p_valores

```

Anexo III: Código Python Proporción Empresas Digitalizadas por Localidad

```

import numpy as np
import matplotlib.pyplot as plt

# Datos observados [digitalizadas, no digitalizadas]
datos = np.array([
    [142, 160], # Dajabón
    [15, 140], # Los Cacaos
    [19, 153], # Santo Domingo Savio
    [55, 57], # San Cristobal
    [263, 118] # Santo Domingo (ciudad)
])

# Calculo de la proporción de empresas digitalizadas
proporcion = datos[:, 0] / datos.sum(axis=1)

```

```
localidades = ['Dajabón', 'Los Cacaos', 'Santo Domingo Savio', 'San
Cristóbal', 'Santo Domingo (ciudad)']

# Gráfico de barras para mostrar las proporciones
plt.figure(figsize=(10, 6))
plt.bar(localidades, proporcion, color='skyblue')
plt.xlabel('Localidades')
plt.ylabel('Proporción de Empresas Digitalizadas')
plt.title('Proporción de Empresas Digitalizadas por Localidad')
plt.ylim(0, 1)
plt.show()
```


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, Alejandro Ruiz estudiante de E2-Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado " Impacto de la Formalidad, la Localización y el Tamaño sobre la Digitalización de Micro y Miniempresas en República Dominicana.", 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. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
3. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
4. **Interpretador de código:** Para realizar análisis de datos preliminares.
5. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
6. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
7. **Generador de problemas de ejemplo:** Para ilustrar conceptos y técnicas.
8. **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: 15 de abril de 2024

Firma: Alejandro Ruiz