



Facultad de Ciencias Económicas y Empresariales

ICADE

**DESARROLLO DEL MODELO ANALÍTICO DE LA
STARTUP: DATABRIDGE**

Autor: Gonzalo Leopoldo Puig de Marcos

Director: Carlos Bellón Núñez-Mera

MADRID | Junio 2026

RESUMEN EJECUTIVO

Las pequeñas y medianas empresas son el núcleo del tejido empresarial español, pero conviven desde hace años con un acceso a la financiación bancaria que, en muchos casos, no funciona bien. En su raíz, esa dificultad es un problema de información: el banco evalúa a cada pyme con datos contables agregados y desfasados, mientras que su comportamiento financiero cotidiano, que es la información verdaderamente relevante, permanece fuera de su alcance. En este contexto se sitúa DataBridge, una plataforma digital desarrollada en el marco del programa Comillas Emprende, con el apoyo de Elevatorfy y en colaboración con AsFin, que convierte los datos transaccionales de las pymes, obtenidos con su consentimiento a través del open banking, en un diagnóstico financiero para la empresa y en un lead cualificado para el banco.

Frente al Trabajo de Fin de Grado de Administración y Dirección de Empresas, que desarrolló el modelo de negocio, este trabajo aborda el problema de datos que subyace a la propuesta. Siguiendo la metodología CRISP-DM, genera un conjunto sintético de transacciones calibrado con estadística pública española, automatiza su categorización y deriva indicadores de comportamiento financiero. Sobre ellos construye un perfil de solvencia transparente, lo traduce a un diagnóstico claro y demuestra cómo se convierte en un lead cualificado, cuantificando el valor económico que ese análisis aporta al banco frente a una recomendación sin él. Los resultados confirman que el problema es abordable y que su solución genera valor para ambas partes, dentro de los límites propios de un trabajo apoyado en datos sintéticos.

Palabras clave: pymes, financiación empresarial, asimetría de información, open banking, datos transaccionales, credit scoring, CRISP-DM

ABSTRACT

Small and medium-sized enterprises are the backbone of the Spanish business fabric, yet they have long struggled with access to bank financing that, in many cases, does not work well. At its root, this difficulty is an information problem: banks assess each SME using aggregated and outdated accounting data, while the company's day-to-day financial behaviour, which is the truly relevant information, remains beyond their reach. It is in this context that DataBridge emerges, a digital platform developed within the Comillas Emprende programme, with the support of Elevatorfy and in collaboration with AsFin, which turns SMEs' transactional data, obtained with their consent through open banking, into a financial diagnosis for the company and a qualified lead for the bank.

Unlike the Business Administration final degree project, which developed the business model, this work addresses the underlying data problem. Following the CRISP-DM methodology, it generates a synthetic dataset of transactions calibrated against Spanish public statistics, automates its categorisation and derives financial-behaviour indicators. On these it builds a transparent solvency profile, translates it into a clear diagnosis and shows how it becomes a qualified lead, quantifying the economic value this analysis provides to the bank over a recommendation made without it. The results confirm that the problem is tractable and that its solution creates value for both parties, within the limits inherent to a work based on synthetic data.

Palabras clave: SMEs, business financing, information asymmetry, open banking, transactional data, credit scoring, CRISP-DM.

ÍNDICE

1.	INTRODUCCIÓN.....	6
1.1.	Justificación e interés del tema desde la perspectiva analítica	6
1.2.	Pregunta de investigación e hipótesis analíticas.....	7
1.3.	Objetivos del trabajo.....	8
1.4.	Metodología: CRISP-DM como marco de referencia	8
1.5.	Estructura del trabajo.....	10
2.	LA PYME, LA REGULACIÓN Y EL DATO.....	11
2.1.	La pyme española: peso en el tejido empresarial y baja profesionalización de su gestión financiera	11
2.2.	El acceso a la financiación y la asimetría de información como problema de datos	12
2.3.	El marco regulatorio del open banking: la directiva PSD2	13
2.4.	El ecosistema fintech y los agregadores de datos transaccionales	14
2.5.	El dato transaccional como nueva fuente de información financiera.....	15
3.	MODELO DE NEGOCIO DE DATABRIDGE	17
3.1.	Origen del proyecto y problema que resuelve.....	17
3.2.	La propuesta: del dato transaccional al diagnóstico y al lead cualificado.....	18
3.3.	Modelo de negocio en síntesis.....	21
4.	ENTRADA Y ESTRUCTURA DE LOS DATOS EN LA PLATAFORMA	23
4.1.	Flujo de datos: del consentimiento de la pyme a la conexión vía open banking	23
4.2.	Modelo de datos: estructura de una transacción y esquema de la plataforma	24
4.3.	Generación de un conjunto de datos sintético calibrado con estadística pública española	26
4.4.	Pipeline de categorización automática de movimientos.....	29
4.5.	De transacciones a indicadores de comportamiento financiero.....	31
5.	DEL DATO AL DIAGNÓSTICO: PERFIL FINANCIERO DE LA PYME.....	34
5.1.	Definición del perfil de solvencia a partir de los indicadores	34
5.2.	Cálculo del score y traducción a un lenguaje claro para la pyme.....	35
5.3.	Lectura e interpretación del diagnóstico.....	37
6.	DEL DIAGNÓSTICO A LA RECOMENDACIÓN PARA EL BANCO Y SU VALOR.....	40

6.1.	De la puntuación al lead cualificado: cómo el diagnóstico se traduce en una recomendación para el banco	40
6.2.	La capacidad de discriminación del score en términos de negocio	41
6.3.	Cuantificación del valor para el banco frente a la recomendación sin análisis	43
7.	CONCLUSIONES.....	46
8.	BIBLIOGRAFÍA	48
9.	APÉNDICES DE CÓDIGO	50
9.1.	Generación dataset.....	50
9.2.	Categorización automática de movimientos.....	56
9.3.	Generación de indicadores de comportamiento.....	57
9.4.	La capacidad de discriminación del score en términos de negocio	60
9.5.	Cuantificación del valor para el banco frente a la recomendación sin análisis	62

1. INTRODUCCIÓN

1.1. Justificación e interés del tema desde la perspectiva analítica

Las pequeñas y medianas empresas constituyen el núcleo del tejido empresarial español. Representan el 99,8 % del total de empresas del país y dan empleo a más de once millones de personas, lo que las convierte en el principal soporte de la actividad económica (Ministerio de Industria y Turismo, 2025). Sin embargo, su gestión financiera está, en muchos casos, poco profesionalizada. La mayoría de las pymes no cuenta con un departamento financiero estructurado ni con herramientas que le permitan analizar de forma sistemática su propia información, y buena parte de sus decisiones sobre financiación se siguen tomando por intuición o por la relación con el banco de siempre. Esta situación resulta llamativa desde una perspectiva analítica: las pymes generan, en su operativa diaria, un volumen considerable de información financiera que rara vez se explota.

Esa falta de aprovechamiento del dato tiene consecuencias directas sobre el acceso a la financiación. Cuando una pyme solicita un préstamo, el banco evalúa su riesgo con una información escasa y, a menudo, desfasada: cuentas anuales depositadas con meses de retraso, ratios contables poco representativos de la situación real y un historial limitado. Este desajuste entre lo que el banco necesita saber y lo que realmente puede observar es una manifestación de la asimetría de información que la literatura económica describe desde hace décadas. Stiglitz y Weiss (1981) mostraron que, cuando el prestamista no puede distinguir con precisión el riesgo de cada solicitante, el mercado de crédito no se ajusta únicamente a través del precio, sino que tiende al racionamiento: hay empresas solventes que no obtienen financiación simplemente porque el banco no dispone de información suficiente para evaluarlas. En el segmento de las pymes, este problema es especialmente acusado.

En los últimos años, dos circunstancias han modificado las condiciones de partida. La primera es de naturaleza regulatoria: el marco europeo de servicios de pago ha abierto la posibilidad de que un tercero autorizado acceda, con el consentimiento del titular, a sus datos bancarios. Esta apertura, conocida como open banking, ha puesto en circulación una infraestructura de datos transaccionales sobre la que es posible construir servicios que antes no eran viables. La segunda es de naturaleza empírica: la investigación reciente ha demostrado que este tipo de información de comportamiento tiene un valor predictivo

real en la evaluación del crédito. Berg, Burg, Gombović y Puri (2020) comprobaron que variables sencillas derivadas de la huella digital de un usuario alcanzan una capacidad de predicción del impago comparable a la de las puntuaciones de las agencias de crédito tradicionales, y que además aportan información complementaria a la que estas ya recogen. Si esto se cumple con datos relativamente indirectos, el potencial de los datos transaccionales, que reflejan de forma directa cómo se mueve el dinero en una empresa, es aún mayor.

Es en este punto donde se sitúa el interés analítico del presente trabajo. El proyecto DataBridge, desarrollado como plan de negocio en el Trabajo de Fin de Grado de Administración y Dirección de Empresas, propone aprovechar precisamente esa información para conectar a las pymes con la financiación adecuada. Este trabajo se ha desarrollado en el marco del programa Comillas Emprende, con el apoyo de la plataforma Elevatorfy y en colaboración con AsFin como socio estratégico, y abordó la oportunidad de mercado y el modelo de negocio. Este se centra en el problema que subyace a la propuesta y que constituye su verdadero núcleo: cómo convertir un conjunto de transacciones en bruto en un diagnóstico financiero útil para la pyme y en un lead cualificado para el banco. Se trata, en esencia, de un problema de datos, y abordarlo exige definir cómo entra y se estructura la información en la plataforma, qué indicadores de comportamiento pueden derivarse de ella y cómo se traduce todo ello en valor para ambas partes.

1.2. Pregunta de investigación e hipótesis analíticas

El planteamiento anterior conduce a la pregunta que guía este trabajo: ¿es posible transformar los datos transaccionales de una pyme en un diagnóstico financiero comprensible para la propia empresa y en un perfil de solvencia que permita cualificarla como lead ante una entidad financiera? Responderla exige recorrer el proceso completo que va del dato en bruto al resultado interpretable, y ese recorrido se apoya en tres hipótesis que el trabajo pone a prueba en los capítulos siguientes.

La primera es que, a partir de los movimientos bancarios de una pyme, pueden derivarse de forma automatizada indicadores de comportamiento financiero (liquidez, estabilidad de los ingresos, estacionalidad, carga de deuda o puntualidad en los pagos) que resumen su situación con mayor cercanía a la realidad operativa que la información contable tradicional. La segunda es que esos indicadores permiten construir un perfil de solvencia,

expresado como una puntuación, que ordena a las pymes según su aptitud para acceder a financiación y que puede traducirse a un lenguaje claro para la empresa. La tercera es que ese mismo perfil, presentado a través de una interfaz sencilla, sirve a los dos lados del mercado al mismo tiempo: ofrece a la pyme un diagnóstico útil y proporciona al banco un lead ya cualificado, reduciendo la asimetría de información descrita en el apartado anterior.

1.3. Objetivos del trabajo

El objetivo general de este Trabajo de Fin de Grado es demostrar, desde una perspectiva analítica, el proceso mediante el cual DataBridge convierte los datos transaccionales de una pyme en un diagnóstico financiero útil para la propia empresa y en un perfil de solvencia que permite cualificarla como lead ante una entidad bancaria. Para alcanzarlo se definen los siguientes objetivos específicos:

- Situar la dificultad de acceso a la financiación de las pymes como un problema de información, e identificar el marco del open banking y el dato transaccional como la base que hace viable una solución analítica.
- Sintetizar el modelo de negocio de DataBridge y traducir su propuesta de valor a un problema de datos concreto y abordable.
- Definir el flujo y la estructura de entrada de la información en la plataforma y construir un conjunto de datos sintético, calibrado con estadística pública española, que permita ilustrar el proceso ante la ausencia de datos reales de pymes.
- Desarrollar un procedimiento automatizado que categorice los movimientos bancarios y derive de ellos un conjunto de indicadores de comportamiento financiero.
- Construir un perfil de solvencia a partir de esos indicadores, expresado como una puntuación, y traducirlo a un diagnóstico comprensible para la pyme.
- Demostrar cómo el perfil de solvencia se traduce en una recomendación cualificada para el banco y cuantificar el valor que ese análisis aporta frente a una recomendación sin él.

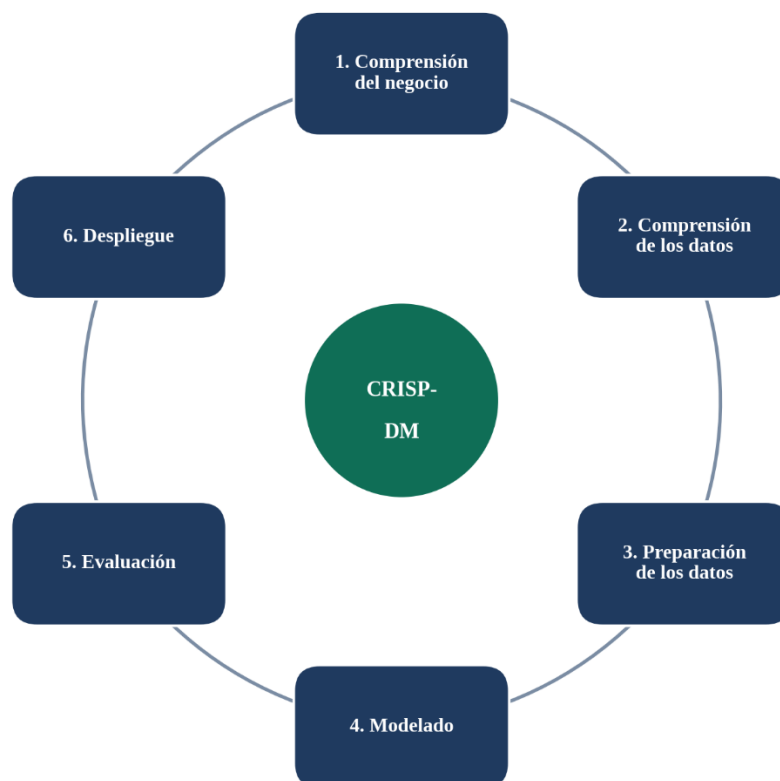
1.4. Metodología: CRISP-DM como marco de referencia

El trabajo se estructura siguiendo CRISP-DM (Cross-Industry Standard Process for Data Mining), un marco metodológico estándar para proyectos de análisis de datos que organiza el proceso en seis fases: comprensión del negocio, comprensión de los datos,

preparación de los datos, modelado, evaluación y despliegue (Chapman et al., 2000). Su utilidad reside en que ordena el recorrido completo que va del problema de negocio al resultado analítico, garantizando que cada decisión técnica responde a una necesidad previamente definida.

Las fases de CRISP-DM se corresponden con los capítulos de este trabajo. La comprensión del negocio se aborda en los capítulos 2 y 3, que sitúan el problema de las pymes y sintetizan el modelo de DataBridge. La comprensión de los datos y su preparación se desarrollan en el capítulo 4, donde se define el flujo de entrada de la información, se genera el conjunto de datos sintético y se transforma en indicadores de comportamiento financiero. El modelado y la evaluación corresponden al capítulo 5, dedicado a construir el perfil de solvencia y a interpretar el diagnóstico resultante. Por último, la fase de despliegue se aproxima en el capítulo 6, que muestra cómo el diagnóstico se convierte en una recomendación cualificada para el banco y cuantifica el valor que ese análisis aporta.

Ilustración 1. Fases de la metodología CRISP-DM.



Fuente: elaboración propia a partir Chapman et al. (2000).

Conviene matizar dos adaptaciones del marco. La evaluación no se realiza sobre datos reales, sino sobre el conjunto sintético, por lo que su alcance es ilustrativo y no de validación predictiva. Ambas limitaciones derivan de la ausencia de datos reales de pymes y se asumen de forma explícita.

1.5. Estructura del trabajo

El trabajo se organiza en seis capítulos, además de esta introducción y de las conclusiones. El capítulo 2 sitúa el contexto: el peso de la pyme en el tejido empresarial español, su dificultad de acceso a la financiación entendida como un problema de información, y el marco del open banking y el dato transaccional que hacen posible una respuesta analítica. El capítulo 3 presenta DataBridge, sintetizando el problema que resuelve, su propuesta de valor y su modelo de negocio, y sirve de puente entre el contexto y la parte técnica.

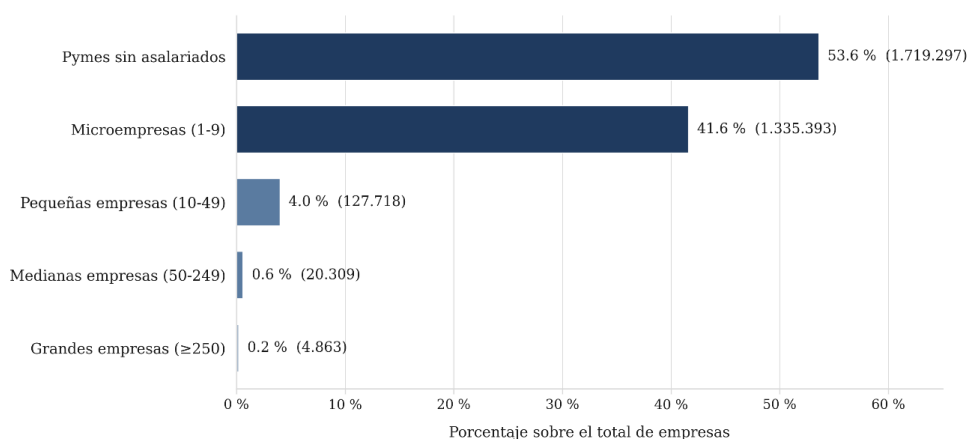
A partir de ahí comienza el núcleo analítico del trabajo. El capítulo 4 define cómo entra y se estructura la información en la plataforma, genera el conjunto de datos sintético con el que se ilustra el proceso y lo transforma en indicadores de comportamiento financiero. El capítulo 5 construye, a partir de esos indicadores, el perfil de solvencia de la pyme, lo expresa como una puntuación y lo traduce a un diagnóstico comprensible. El capítulo 6 muestra cómo ese diagnóstico se traduce en una recomendación cualificada para el banco y cuantifica, sobre los propios datos, el valor que el análisis aporta frente a una recomendación sin él.

2. LA PYME, LA REGULACIÓN Y EL DATO

2.1. La pyme española: peso en el tejido empresarial y baja profesionalización de su gestión financiera

Las pymes constituyen la práctica totalidad del tejido empresarial español. Representan el 99,8 % de las empresas del país (Ministerio de Industria y Turismo, 2025), pero ese dato agregado oculta una característica más relevante para este trabajo: su reducido tamaño. Más de la mitad de las empresas españolas no tiene ningún asalariado, y la inmensa mayoría de las que sí lo tienen son microempresas de menos de diez trabajadores (INE, 2025). El tejido productivo está formado, por tanto, por un número muy elevado de unidades de dimensión muy pequeña. Esa reducida dimensión se refleja también en su facturación: según el Retrato de la PYME, el 96,7 % de las empresas españolas ingresan menos de dos millones de euros al año (Ministerio de Industria y Turismo, 2024).

Ilustración 2. Estructura del tejido empresarial español por tamaño.



Fuente: elaboración propia a partir de Ministerio de Industria y Turismo (2024).

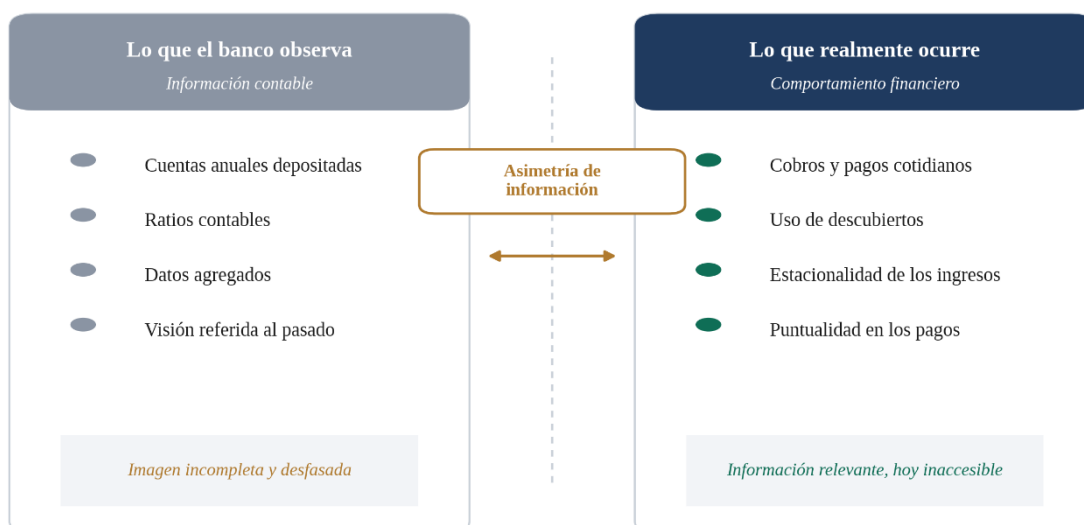
Esta estructura condiciona de forma directa la manera en que las pymes gestionan sus finanzas. Una empresa de tamaño reducido rara vez dispone de un departamento financiero propio o de personal especializado, de modo que las decisiones sobre tesorería, inversión y financiación recaen habitualmente sobre la persona que dirige el negocio, que las compatibiliza con el resto de sus funciones. El resultado es una gestión financiera poco sistematizada, apoyada más en la experiencia y en la relación con la entidad bancaria de referencia que en el análisis de la propia información económica. Esta baja profesionalización no responde a una falta de interés, sino a una limitación de medios inherente al tamaño, y explica buena parte de las dificultades que estas empresas encuentran cuando necesitan demostrar su solvencia ante un tercero.

2.2. El acceso a la financiación y la asimetría de información como problema de datos

La financiación ajena de las pymes españolas descansa, en lo esencial, sobre el crédito bancario. A diferencia de las grandes empresas, que pueden recurrir a los mercados de capitales o a la emisión de deuda, la pequeña empresa dispone de pocas alternativas reales y acude al banco como vía principal cuando necesita financiación (Banco de España, 2025). Esto sitúa a la entidad en una posición central: es ella quien decide si concede el crédito y en qué condiciones, y para hacerlo necesita evaluar el riesgo de cada solicitante.

El problema está en la información sobre la que se apoya esa evaluación. El banco juzga la solvencia de una pyme, fundamentalmente, a partir de sus cuentas anuales y de los ratios contables que de ellas se derivan. Esta información es agregada, se refiere al pasado y, además, llega con retraso, ya que las cuentas de un ejercicio no están disponibles hasta bien entrado el siguiente. El comportamiento financiero real de la empresa, cómo entran y salen los ingresos, con qué regularidad cobra, si recurre con frecuencia a descubiertos o si atiende sus pagos con puntualidad, no se refleja en esos documentos o lo hace de forma muy parcial. Como consecuencia, el banco evalúa con una imagen incompleta y desfasada, y la asimetría de información descrita por Stiglitz y Weiss (1981) se manifiesta aquí en su forma más concreta: existe información relevante sobre la solvencia de la pyme, pero no está al alcance del banco en un formato utilizable.

Ilustración 3. La asimetría de información en la evaluación del crédito a la pyme.



Fuente: elaboración propia.

Visto así, la dificultad de acceso a la financiación de las pymes no es únicamente un problema económico, sino, en su raíz, un problema de datos. La información que permitiría evaluar mejor a estas empresas existe y se genera cada día en su operativa bancaria, pero permanece dispersa y fuera del alcance de quien debe tomar la decisión de crédito. Reducir la asimetría no exige, por tanto, generar información nueva, sino hacer accesible y legible la que ya existe. Es precisamente esa posibilidad la que abre el marco del open banking, objeto del siguiente apartado.

2.3. El marco regulatorio del open banking: la directiva PSD2

El acceso a esa información dispersa fue, durante mucho tiempo, técnicamente posible pero jurídicamente vedado: los datos pertenecían a la relación entre el cliente y su banco, y ningún tercero podía acceder a ellos. La segunda Directiva de servicios de pago, la Directiva (UE) 2015/2366, conocida como PSD2, modificó esa situación. Su aportación esencial fue obligar a las entidades a permitir que terceros autorizados accedan a la información de las cuentas de pago de un cliente, siempre que este lo consienta de forma expresa, y a hacerlo a través de interfaces técnicas normalizadas. Esta apertura es lo que se conoce como open banking. En España, la directiva se transpuso mediante el Real Decreto-ley 19/2018, de 23 de noviembre, de servicios de pago, que atribuye al Banco de España la supervisión de las nuevas entidades.

La PSD2 reconoce dos figuras nuevas de proveedores de servicios. Los proveedores de servicios de iniciación de pagos (PISP) pueden ordenar pagos en nombre del cliente, y los proveedores de servicios de información sobre cuentas (AISP) pueden acceder a la información de sus cuentas y agregarla. Es esta segunda figura la relevante para una plataforma como DataBridge, cuyo funcionamiento se apoya en el acceso a los movimientos de la pyme. El marco rodea ese acceso de garantías estrictas: exige el consentimiento explícito del titular, lo limita a la información expresamente autorizada y obliga a emplear mecanismos de autenticación reforzada de cliente (Reglamento Delegado (UE) 2018/389), de modo que la apertura de los datos no se produce a costa de su seguridad.

La relevancia de este marco para el presente trabajo es directa. La PSD2 convierte una información que antes estaba encerrada en los sistemas del banco en un recurso al que un tercero puede acceder, con el consentimiento del cliente y en un formato estructurado y legible por máquina. Resuelve así el obstáculo identificado en el apartado anterior: la

información que permite evaluar mejor a una pyme deja de ser inaccesible y pasa a poder tratarse de forma sistemática. Sin esta apertura regulatoria, la propuesta de DataBridge no sería viable.

El marco, además, se encuentra en plena evolución hacia una mayor apertura. En 2023 la Comisión Europea presentó un paquete para sustituir la PSD2 por una tercera directiva (PSD3) acompañada de un reglamento de servicios de pago (PSR), sobre el que el Parlamento Europeo y el Consejo alcanzaron un acuerdo político en noviembre de 2025, pendiente todavía de adopción formal. Junto a ellos avanza la propuesta de un marco de acceso a los datos financieros (FIDA), que pretende extender los principios del open banking más allá de las cuentas de pago hacia un open finance de alcance más amplio. Esta evolución refuerza la base sobre la que se apoya el trabajo, al consolidar y ampliar el acceso regulado a los datos financieros.

2.4. El ecosistema fintech y los agregadores de datos transaccionales

La apertura regulatoria descrita en el apartado anterior no habría tenido efecto sin un conjunto de empresas capaces de aprovecharla. Ese papel lo desempeña el ecosistema fintech, formado por compañías que aplican la tecnología a la prestación de servicios financieros. En España es ya un sector consolidado: la Asociación Española de Fintech e Insurtech (AEFI), su principal organización de referencia, agrupa a más de un centenar de empresas asociadas, repartidas en distintas verticales que van desde los pagos digitales y los neobancos hasta la financiación alternativa (AEFI, 2026). Es en esta última vertical, la que conecta a las empresas con financiación al margen de los canales tradicionales, donde se sitúa una propuesta como la de DataBridge.

Dentro de este ecosistema, los agregadores de datos transaccionales ocupan una posición central para este trabajo. Cada banco expone su información a través de su propia interfaz, con formatos y estructuras distintas, de modo que conectarse de forma individual con cada entidad resultaría costoso e inviable para la mayoría de las empresas. Los agregadores resuelven ese problema: se conectan a las API de las distintas entidades, normalizan la información en un formato común y la ponen a disposición de terceros mediante una única integración. Para hacerlo legalmente, actúan como proveedores de servicios de información sobre cuentas (AISP) autorizados y supervisados por el Banco de España, lo que les habilita para acceder, en modo de solo lectura y con el consentimiento del cliente, a los saldos y movimientos de sus cuentas. Empresas como

Tink, adquirida por Visa en 2022, o servicios de ámbito nacional como Afterbanks operan bajo este modelo.

Esta infraestructura es la que permite que una plataforma como DataBridge funcione sin necesidad de desarrollar y mantener integraciones con cada banco, ni de asumir por sí misma la condición de entidad regulada. Apoyándose en un agregador o en un socio que actúe como tal, DataBridge puede acceder a los datos transaccionales de la pyme, previo consentimiento, y concentrarse en lo que constituye su verdadero valor: el tratamiento analítico de esa información. El ecosistema fintech aporta, así, tanto el contexto competitivo en el que se inserta la propuesta como la infraestructura técnica sobre la que se construye.

2.5. El dato transaccional como nueva fuente de información financiera

De todo lo anterior se desprende que el dato transaccional, los movimientos de entrada y salida de las cuentas de una empresa, constituye una fuente de información financiera con características propias que lo distinguen de la información contable tradicional. Es, en primer lugar, granular, ya que recoge cada operación de forma individual y no como un agregado anual. Es continuo y está actualizado, porque refleja la actividad de la empresa prácticamente en tiempo real, frente al desfase de las cuentas depositadas. Y es, sobre todo, de naturaleza conductual: no describe lo que la empresa declara, sino lo que realmente hace, cómo y cuándo cobra, cómo afronta sus pagos o con qué frecuencia tensiona su tesorería. Esta última característica le confiere, además, una mayor resistencia a la manipulación que la información autodeclarada.

La investigación académica ha confirmado el valor de esta fuente para la evaluación del riesgo. Tobback y Martens (2019) construyeron modelos de credit scoring a partir de los datos de pago de las cuentas de los clientes de un banco y demostraron que su uso mejora de forma significativa la capacidad de predicción del impago, y que lo hace manteniendo la interpretabilidad del modelo, un requisito clave en un ámbito sometido a exigencias regulatorias de explicación. En la misma línea, y como ya se señaló, Berg et al. (2020) habían mostrado el poder predictivo de datos de comportamiento incluso más indirectos. El dato transaccional no sustituye a la información contable, sino que la complementa con una imagen más cercana y oportuna de la situación real de la empresa.

Es precisamente esta fuente la que DataBridge se propone aprovechar. Una vez establecido por qué el dato transaccional es valioso y cómo el marco regulatorio y el

ecosistema fintech permiten acceder a él, el resto del trabajo se ocupa de la cuestión analítica que constituye su núcleo: cómo transformar ese flujo de movimientos en un diagnóstico útil para la pyme y en un perfil de solvencia que la cualifique ante una entidad financiera. A ese proceso se dedican los capítulos siguientes.

3. MODELO DE NEGOCIO DE DATABRIDGE

3.1. Origen del proyecto y problema que resuelve

DataBridge es una plataforma digital que parte de una observación concreta sobre el mercado de la financiación a las pymes en España. Dos partes que se necesitan mutuamente, las pymes y los bancos, mantienen desde hace años una relación poco eficiente que perjudica a ambas. Por el lado de la pyme, el problema es de información y de criterio: el responsable del negocio rara vez sabe con seguridad qué producto financiero se ajusta mejor a su situación, desconoce qué entidad se lo ofrecería en mejores condiciones y, ante esa incertidumbre, tiende a recurrir por defecto a su banco habitual sin comparar alternativas. No es una cuestión de desinterés, sino de falta de tiempo y de medios para evaluar un mercado opaco. Por el lado del banco, el problema es de identificación: con la información de que dispone en la fase inicial del análisis, una entidad no consigue distinguir bien a qué pymes puede servir mejor ni anticipar cuáles encajan realmente con su oferta, de modo que dedica esfuerzo comercial a operaciones que no se ajustan y deja pasar otras que sí lo harían.

Ilustración 4. Logo de DataBridge.



Fuente: elaboración propia.

El resultado de ese doble desconocimiento es un mercado con fricción en los dos sentidos. Se quedan sin cerrar operaciones que tendrían sentido para las dos partes, se contratan productos que no encajan del todo con la necesidad real de la empresa, y tanto la búsqueda por parte de la pyme como la originación por parte del banco resultan más lentas y más caras de lo que deberían. La hipótesis de partida del proyecto es que buena parte de ese desajuste no responde a una incompatibilidad de fondo entre la oferta y la demanda, sino a una falta de información útil en el momento adecuado, y que, por tanto, puede reducirse de forma significativa si se utilizan los datos transaccionales reales de la pyme para

diagnosticar su situación financiera y conectarla, ya cualificada, con la oferta bancaria que mejor se ajusta a su perfil.

El proyecto surgió en el marco del programa Comillas Emprende, organizado por la Universidad Pontificia Comillas, y se desarrolló a través de la plataforma Elevatorfy, que aporta una metodología guiada para estructurar la idea desde su definición inicial hasta la validación con clientes potenciales. El recorrido por estas herramientas permitió pasar de una intuición de partida a una propuesta articulada, con los clientes definidos, el valor diferencial identificado y el modelo de monetización planteado. A ese trabajo se sumó, desde el origen, la colaboración con AsFin como socio estratégico, que ha resultado determinante en dos sentidos. En la fase de diseño, AsFin ha aportado conocimiento del sector y experiencia operativa con pymes, que han servido para orientar el enfoque y validar los supuestos iniciales. En la fase de lanzamiento, puede facilitar el acceso a un volumen significativo de pymes de su cartera que autoricen incorporarse al servicio, lo que proporciona al proyecto una masa inicial de usuarios reales con la que poner el modelo a prueba en condiciones realistas.

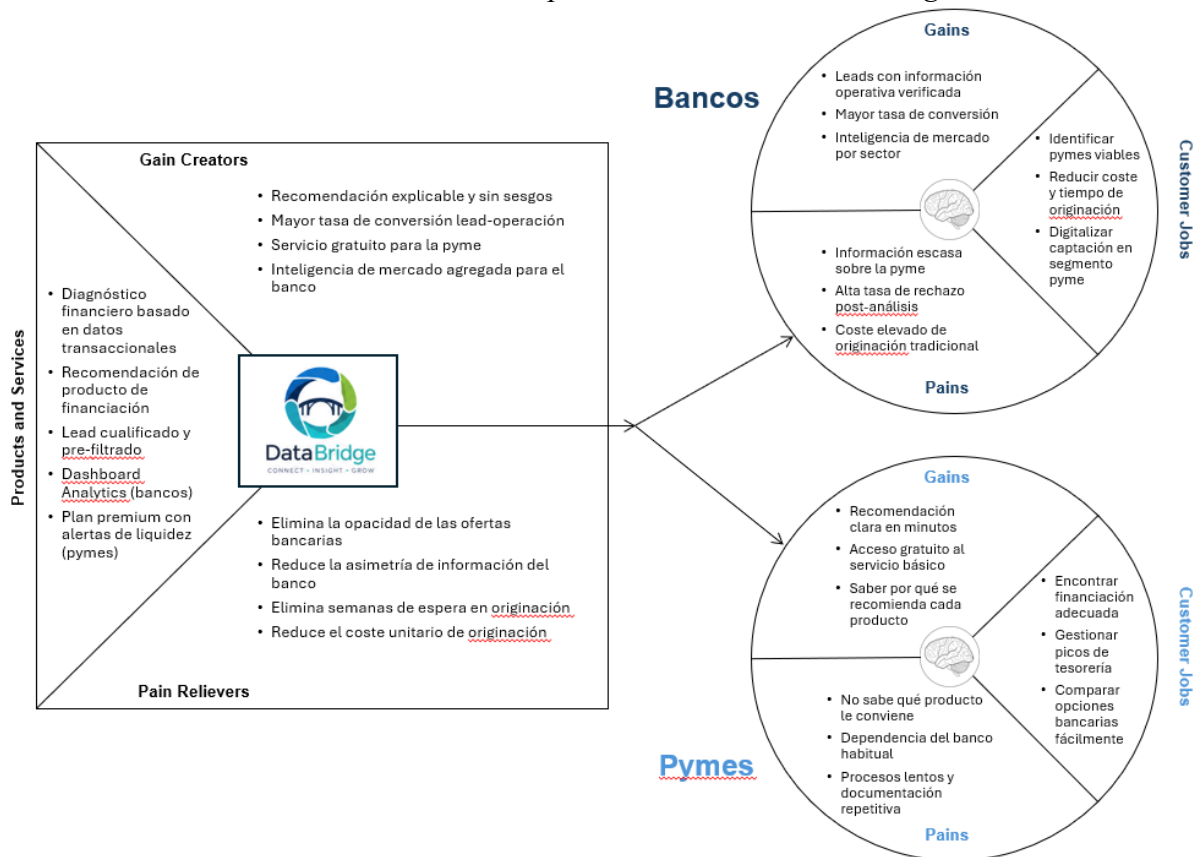
De esa manera de entender el proyecto procede su propio nombre. DataBridge no se concibe como una herramienta dirigida a uno solo de los dos lados, sino como un puente entre dos clientes que se necesitan pero que hoy se relacionan con fricción. En términos operativos, la plataforma recibe los datos transaccionales de cada pyme, con su consentimiento explícito y a través del marco del open banking, los analiza para diagnosticar su situación financiera y segmentarla según su comportamiento, y devuelve un resultado distinto a cada lado del mercado. A la pyme le ofrece una recomendación clara sobre qué producto de financiación encaja mejor con su perfil, en qué tipo de entidad podría conseguirlo en buenas condiciones y por qué se le recomienda esa opción y no otra. Al banco le entrega un lead cualificado y prefiltrado, con la información operativa relevante ya estructurada y listo para incorporarse a su proceso de originación. Esa doble salida, construida sobre un mismo insumo, es la idea que desarrollan los apartados siguientes.

3.2. La propuesta: del dato transaccional al diagnóstico y al lead cualificado

Como corresponde a un modelo de plataforma, DataBridge no tiene una única propuesta de valor, sino dos, una para cada lado del mercado al que se dirige. Esta es una característica propia de los modelos de negocio de dos lados, en los que el servicio solo

funciona si las dos partes encuentran una razón clara para participar. Lo relevante en el caso de DataBridge es que esas dos propuestas no son independientes ni parten de procesos distintos: ambas se construyen sobre el mismo insumo, el análisis de los datos transaccionales de la pyme. El diagnóstico que recibe la empresa y el lead que recibe el banco son, en realidad, dos resultados del mismo tratamiento analítico, presentados de forma distinta según a quién se dirijan. Articular bien las dos propuestas, y entender que se sostienen la una a la otra, es lo que hace que el modelo funcione.

Ilustración 5. Value Proposition Canvas de DataBridge.



Fuente: Osterwalder et al. (2014).

Para la pyme, la propuesta de valor es una recomendación de financiación adaptada a su comportamiento real, obtenida en minutos y sin complicaciones. El recorrido es sencillo desde el punto de vista del usuario: la pyme conecta sus datos, DataBridge analiza cómo se mueve el dinero en la empresa y devuelve una respuesta concreta sobre qué tipo de producto financiero encaja mejor con su perfil y por qué motivo. La utilidad de esa respuesta se entiende mejor frente a las alternativas de las que dispone hoy una pyme. Puede preguntar a su banco habitual, que le ofrecerá sus propios productos sin una visión comparada del mercado; puede comparar manualmente entre distintas webs, una tarea lenta y poco fiable; o puede recurrir a una asesoría externa, con el coste y la demora que

eso implica. Frente a todas ellas, DataBridge le ahorra tiempo, reduce el riesgo de contratar un producto inadecuado y elimina la opacidad habitual de las ofertas tradicionales en cuanto a tipos de interés, comisiones y condiciones de permanencia. El servicio es gratuito en su versión básica, lo que elimina la fricción de entrada y se alinea con uno de los valores fundacionales del proyecto. Para las pymes con necesidades más avanzadas existe, además, un plan premium opcional que incorpora alertas de liquidez, un panel de tesorería continuo e integración multibanco.










Para el banco, la propuesta de valor es un flujo de leads cualificados y prefiltrados, listos para entrar en su proceso de originación. La diferencia respecto a los leads que una entidad puede adquirir hoy a un comparador tradicional es sustancial y reside en la información que los acompaña. Mientras que un lead convencional se limita, en la práctica, a un contacto potencialmente interesado, el de DataBridge llega acompañado de información operativa real sobre la pyme: su comportamiento de cobros y pagos, su estacionalidad, la concentración de sus clientes o su uso del circulante. Esa información reduce de forma apreciable la asimetría que el banco arrastra en la fase inicial del análisis, que es precisamente el cuello de botella descrito en el capítulo segundo, y le permite priorizar mejor su esfuerzo comercial. El efecto es doble y medible: una mejora en la tasa de conversión de lead a operación cerrada y una reducción del coste unitario de originación, dado que el banco deja de dedicar recursos a contactos que no encajan.

A estos dos beneficios se añade un tercero que afecta a las dos partes a la vez y que conviene explicitar, porque es el que en última instancia justifica el modelo: el ahorro de tiempo y la reducción del riesgo de un emparejamiento incorrecto entre necesidad y producto. En la situación actual, una pyme puede tardar semanas en recibir una oferta concreta, y un banco puede dedicar recursos a operaciones que finalmente no se ajustan. DataBridge comprime ese plazo de manera notable y aumenta la probabilidad de que el match entre la necesidad de la empresa y el producto contratado sea acertado. Esa reducción de fricción es lo que explica que las dos propuestas se sostengan mutuamente. Sin pymes que conecten sus datos y generen diagnósticos no hay leads que ofrecer al banco, y sin bancos dispuestos a pagar por esos leads no es posible mantener el servicio gratuito para la pyme. Esta interdependencia, más que una característica accesorio, es el mecanismo central sobre el que se construye el modelo de negocio, y determina directamente su esquema de monetización, que se sintetiza en el apartado siguiente.

3.3. Modelo de negocio en síntesis

El modelo de negocio de DataBridge puede representarse de forma sintética a través del Business Model Canvas (Osterwalder y Pigneur, 2010), la herramienta más extendida para mostrar en un único lienzo la lógica con la que una empresa crea, entrega y captura valor. Su utilidad radica en que permite ver de un vistazo los nueve bloques que componen un modelo de negocio y comprobar la coherencia entre ellos, algo especialmente valioso en proyectos en fase temprana como este, donde conviene validar que las piezas encajan antes de entrar en el detalle cuantitativo. El lienzo que se presenta a continuación es el resultado del trabajo realizado en el programa Elevatorfy y resume la lógica que los apartados anteriores han ido desgranando.

Ilustración 6. Business Model Canvas de DataBridge.

<p>Alianzas clave </p> <ul style="list-style-type: none"> AsFin (socio estratégico, base inicial de pymes y conocimiento sectorial) Bancos <u>partners</u> (clientes y fuente de datos vía PSD2) ERPs y gestorías (canal de distribución indirecto) Proveedores de infraestructura <u>cloud</u> 	<p>Actividades clave </p> <ul style="list-style-type: none"> Desarrollo y mantenimiento del motor de análisis y segmentación Captación y onboarding de pymes Venta y gestión de relación con bancos <u>partners</u> Cumplimiento regulatorio (RGPD, PSD2) <p>Recursos claves </p> <ul style="list-style-type: none"> Motor de <u>scoring</u> y recomendación financiera Volumen de datos transaccionales de pymes (open <u>banking</u>) Alianza con AsFin Equipo técnico y comercial B2B 	<p>Propuesta de valor </p> <ul style="list-style-type: none"> Pyme: diagnóstico financiero basado en datos reales, segmentación por comportamiento y recomendación de producto de financiación adecuado (póliza de crédito, <u>factoring</u>, <u>confirming</u>, <u>circulante</u>), en minutos y sin jerga Banco: leads cualificados y <u>pre-filtrados</u> con información operativa real de la pyme, listos para entrar en el flujo de <u>originación</u>. Reducción de fricción y tiempo para ambos lados 	<p>Relación con clientes </p> <ul style="list-style-type: none"> Pyme: autoservicio con asistencia ligera vía chat y email Banco: cuenta dedicada con gestor asignado <p>Canales </p> <ul style="list-style-type: none"> Pyme: autoservicio digital, base inicial de AsFin, marketing de contenidos, asociaciones empresariales, gestorías, alianzas con ERPs Banco: venta directa B2B con <u>inside sales</u> y demos personalizadas, eventos <u>fintech</u> (AEFI) 	<p>Segmentos de clientes </p> <ul style="list-style-type: none"> Pymes de 1 a 50 empleados con gestión financiera básica y picos recurrentes de tesorería Bancos y <u>neobancos</u> con foco en financiación a pymes y <u>originación digital</u>
<p>Estructura de costes </p> <ul style="list-style-type: none"> Equipo (principal partida desde año 1) Desarrollo inicial de la plataforma Infraestructura <u>cloud</u> escalable Legal y <u>compliance</u> (RGPD, PSD2) Marketing y comercial <u>Backoffice</u> y administración 		<p>Fuentes de ingresos </p> <ul style="list-style-type: none"> Fee por lead cualificado: 150 €/lead Comisión a éxito: 0,3% sobre volumen financiado Suscripción <u>Analytics</u> opcional: 1.500 €/mes por banco Plan premium pyme opcional: 29 €/mes 		

Fuente: Osterwalder y Pigneur (2010).

DataBridge dirige su propuesta a dos segmentos de clientes claramente diferenciados, cada uno con sus propias necesidades y su propia mecánica de relación. El primero son las pymes de entre 1 y 50 empleados, con una gestión financiera básica y picos recurrentes de tesorería. Es el grupo donde el problema resulta más agudo, porque son las empresas con menos recursos internos para evaluar opciones de financiación y las que más dependen de su banco habitual, y es también el grupo donde el comportamiento financiero

es lo bastante diverso como para que el análisis de datos transaccionales produzca segmentaciones útiles y recomendaciones diferenciadas. El segundo segmento son los bancos y neobancos con foco en la financiación a pymes y en la originación digital, que constituyen el cliente que paga y, por tanto, el que sostiene económicamente el modelo. No todas las entidades son clientes potenciales por igual: las más relevantes son las que tienen mayor orientación a la financiación de pymes y mayor inversión en digitalización comercial, que son las que más valor obtienen de un flujo de leads cualificados.

La plataforma llega a cada segmento por canales distintos, porque la lógica de descubrimiento y conversión de uno y otro no tiene nada que ver. Para la pyme, el canal es de autoservicio con un onboarding digital rápido: la empresa descubre el servicio, se registra, conecta sus datos y obtiene su recomendación sin necesidad de pasar por un comercial. Para el banco, el canal es de venta directa B2B con un gestor dedicado, dado que una entidad no contrata un servicio de este tipo a través de una página web, sino que requiere conversación previa, demostración del producto y validación de seguridad y cumplimiento. Para operar, DataBridge se apoya en dos recursos críticos. El primero es un volumen suficiente de datos transaccionales de calidad, obtenidos vía open banking con el consentimiento de cada pyme y complementados en la fase de lanzamiento por la base aportada por AsFin. El segundo es un motor de análisis y recomendación capaz de convertir esos datos en un diagnóstico para la pyme y en un lead cualificado para el banco. Estos dos recursos son los que dan sustento a la parte analítica que desarrollan los capítulos siguientes, y por eso el resto del trabajo se concentra precisamente en ellos.

La lógica de monetización es asimétrica y se carga deliberadamente en el lado del banco. Para la pyme, el servicio es gratuito en su versión básica, con un plan premium opcional para quienes requieren funcionalidades adicionales, de modo que la empresa que aporta el dato no soporta ningún coste de entrada. Los ingresos provienen del banco a través de un modelo principalmente transaccional, con un pago por lead cualificado y una comisión a éxito sobre las operaciones que llegan a cerrarse, al que se añade una suscripción Analytics opcional que da acceso a una plataforma de inteligencia de mercado con información agregada de comportamiento financiero por sectores. Esta estructura, gratuita para quien genera el dato y de pago para quien obtiene el lead, no es una decisión arbitraria, sino la consecuencia directa de la interdependencia entre los dos segmentos descrita en el apartado anterior: situar el coste en el lado que captura el valor económico de la operación es lo que permite eliminar la fricción de entrada en el lado que la alimenta

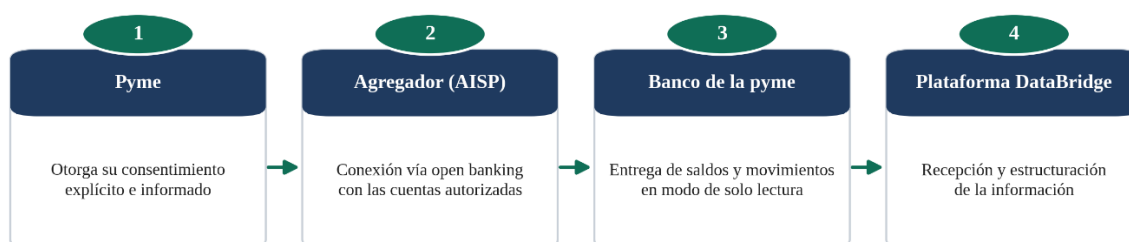
con datos. Con esto queda definido el modelo de negocio en sus líneas esenciales, lo que permite pasar al núcleo analítico del trabajo, que comienza por la forma en que esos datos entran y se estructuran en la plataforma.

4. ENTRADA Y ESTRUCTURA DE LOS DATOS EN LA PLATAFORMA

4.1. Flujo de datos: del consentimiento de la pyme a la conexión vía open banking

El núcleo analítico del trabajo comienza, en rigor, antes de procesar ningún dato. El primer eslabón del flujo no es técnico, sino el consentimiento de la pyme, que es la condición que habilita todo lo demás. Sin una autorización explícita de la empresa titular de la cuenta, DataBridge no puede acceder a ninguna información, y el proceso ni siquiera se inicia. Esta exigencia no es una elección de diseño, sino un requisito que imponen tanto el marco del open banking descrito en el capítulo segundo como la normativa de protección de datos.

Ilustración 7. Flujo de entrada de datos en DataBridge.



Fuente: elaboración propia.

El recorrido completo, representado en la ilustración anterior, se compone de cuatro etapas. En la primera, la pyme accede a la plataforma y otorga su consentimiento para que DataBridge acceda a la información de sus cuentas. Este consentimiento es explícito, informado y revocable en cualquier momento, en línea con las condiciones que para el tratamiento de datos personales establece el Reglamento general de protección de datos (Reglamento (UE) 2016/679). En la segunda etapa, una vez autorizado, el acceso se canaliza a través de la infraestructura de open banking. DataBridge no se conecta directamente a cada banco, sino que se apoya en un agregador que actúa como proveedor de servicios de información sobre cuentas (AISP) autorizado, o en AsFin desempeñando esa función, tal como se explicó en el apartado 2.4. En la tercera etapa, el agregador recupera de la entidad o entidades de la pyme la información de sus cuentas, en concreto

los saldos y el detalle de los movimientos, mediante un acceso de solo lectura que en ningún caso permite operar sobre la cuenta. En la cuarta y última etapa, esa información llega a la plataforma de DataBridge, donde queda disponible para su estructuración y posterior análisis.

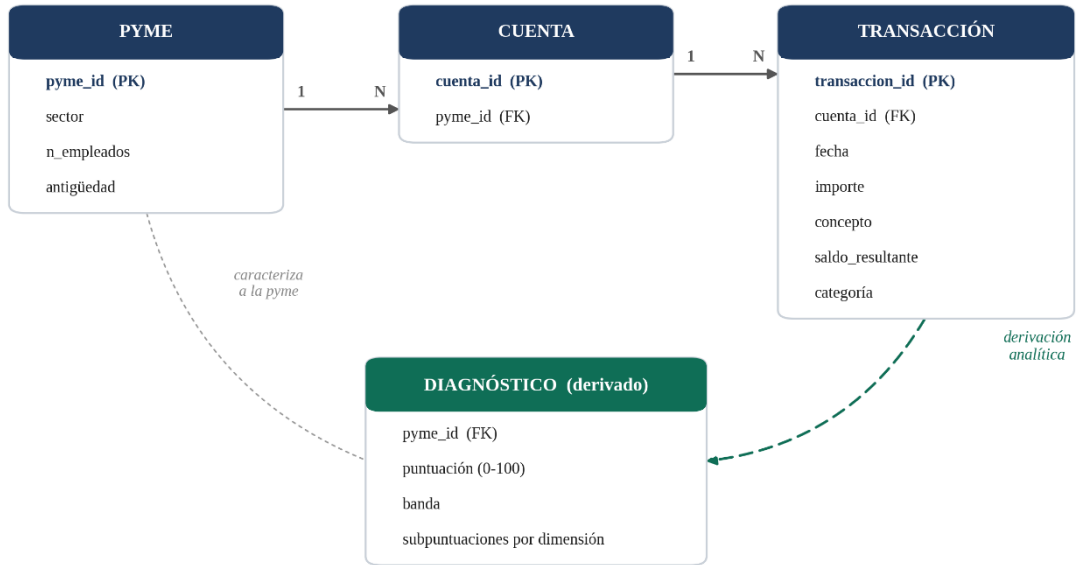
Todo el flujo está sujeto a las garantías propias de este marco. El acceso se limita estrictamente a la información que la pyme ha autorizado, se realiza bajo los mecanismos de autenticación reforzada de cliente exigidos por la normativa (Reglamento Delegado (UE) 2018/389) y se rige por el principio de minimización del dato, de modo que solo se trata la información necesaria para prestar el servicio. Estas garantías conectan directamente con el valor de protección del dato que el proyecto asume como propio, y que en el plano normativo se concreta, además de en el RGPD, en la Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales.

Conviene subrayar que, a partir de este punto, todo el tratamiento posterior opera sobre la información ya recibida. La calidad, la cobertura temporal y el nivel de detalle de esos movimientos condicionan todo lo que viene después, desde la categorización hasta el diagnóstico final. Por ello, antes de procesarlos, es necesario definir cómo se estructura esa información dentro de la plataforma, que es el objeto del apartado siguiente.

4.2. Modelo de datos: estructura de una transacción y esquema de la plataforma

Antes de procesar la información recibida es necesario fijar cómo se organiza dentro de la plataforma. Definir un modelo de datos explícito cumple dos funciones. Por un lado, hace el proceso reproducible, ya que establece con precisión qué campos contiene cada elemento y cómo se relacionan entre sí. Por otro, permite que todo el tratamiento posterior, desde la categorización hasta el cálculo de indicadores, opere sobre una estructura estable y conocida.

Ilustración 8. Esquema de datos de la plataforma.



Fuente: elaboración propia.

La unidad mínima de información es la transacción, es decir, cada movimiento individual de entrada o salida en una cuenta. Una transacción queda descrita por un conjunto reducido de campos que recogen lo esencial de cada movimiento: un identificador, la fecha en que se produce, el importe, el signo que distingue un cobro de un pago, el saldo resultante en la cuenta tras la operación, una descripción o concepto, y la cuenta a la que pertenece. La tabla siguiente recoge ese diccionario de datos con el detalle de cada campo, su tipo y su contenido.

Ilustración 9. Diccionario de datos de una transacción.

Campo	Tipo	Descripción
<code>transaccion_id</code>	Texto	Identificador único del movimiento (clave primaria)
<code>cuenta_id</code>	Texto	Cuenta a la que pertenece el movimiento (clave ajena)
<code>fecha</code>	Fecha	Fecha en que se produce el movimiento
<code>importe</code>	Numérico	Cuántía del movimiento; el signo distingue cobro (+) de pago (-)
<code>concepto</code>	Texto	Descripción textual del movimiento
<code>saldo_resultante</code>	Numérico	Saldo de la cuenta tras el movimiento
<code>categoría</code>	Texto	Categoría asignada en el procesamiento (ventas, nóminas, etc.)

Fuente: elaboración propia.

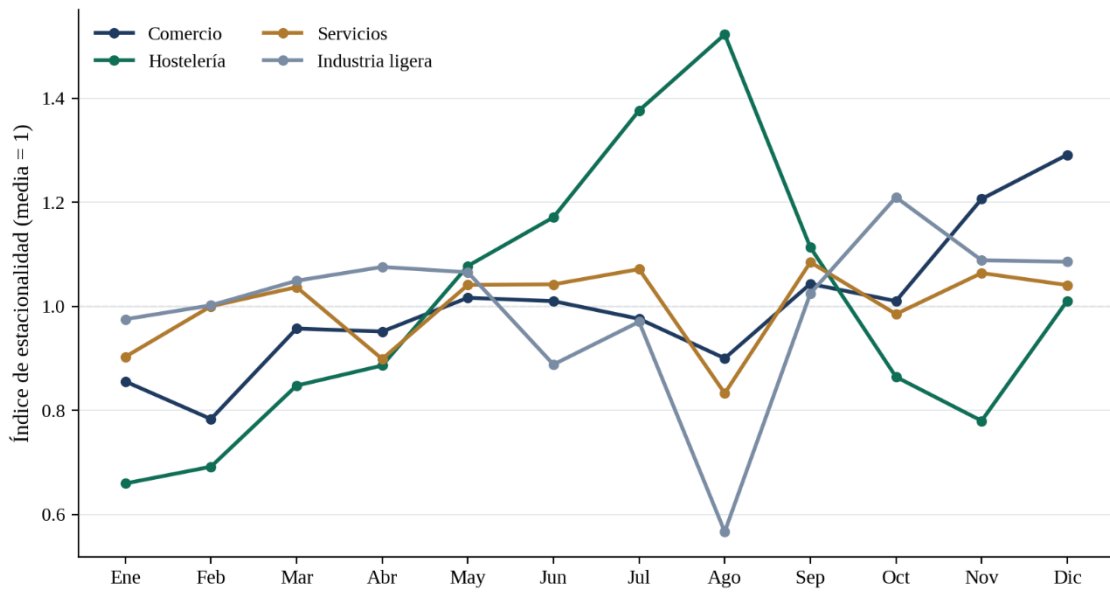
A partir de esa unidad se construye un esquema relacional sencillo, formado por tres entidades principales. La entidad pyme representa a cada empresa usuaria e incorpora los atributos que la describen, como su sector de actividad o su antigüedad. La entidad cuenta recoge cada una de las cuentas bancarias que la pyme ha autorizado, de modo que una misma pyme puede tener varias cuentas asociadas. La entidad transacción contiene todos los movimientos, cada uno vinculado a la cuenta en la que se ha producido. La relación entre ellas es jerárquica: una pyme tiene una o varias cuentas, y cada cuenta acumula numerosas transacciones.

Sobre esta estructura se asienta una cuarta entidad, de naturaleza distinta, que no proviene de los datos en bruto sino que se deriva de ellos: el diagnóstico o perfil financiero de la pyme. Mientras que las tres primeras entidades describen lo que ha ocurrido en las cuentas, el diagnóstico es el resultado del análisis y se calcula a partir de los indicadores que se obtienen del conjunto de transacciones de cada empresa. Esta separación entre el dato observado y el dato derivado es relevante, porque ordena el resto del trabajo: los apartados 4.3 a 4.5 se ocupan de construir y preparar las tres primeras entidades, y el capítulo 5 de obtener la cuarta. Definido el modelo, el paso siguiente es disponer de un conjunto de datos sobre el que aplicarlo, lo que, ante la ausencia de datos reales, exige generarlo de forma sintética.

4.3. Generación de un conjunto de datos sintético calibrado con estadística pública española

Como se anticipó en la introducción, este trabajo no dispone de datos transaccionales reales de pymes, ya que ninguna entidad cede esa información para un proyecto académico. Esta limitación, lejos de invalidar el ejercicio, es habitual en trabajos de esta naturaleza y admite una solución metodológicamente sólida: generar un conjunto de datos sintético que reproduzca de forma verosímil el comportamiento financiero de un grupo de pymes. La condición para que ese conjunto sea defendible es que sus parámetros no se fijen de forma arbitraria, sino que se calibren con rangos procedentes de estadística pública contrastada y que esa calibración se documente de manera transparente. Bajo esa condición, el dataset sirve a su propósito, que no es validar un modelo predictivo sobre datos reales, sino ilustrar el proceso completo que va del movimiento bancario al diagnóstico.

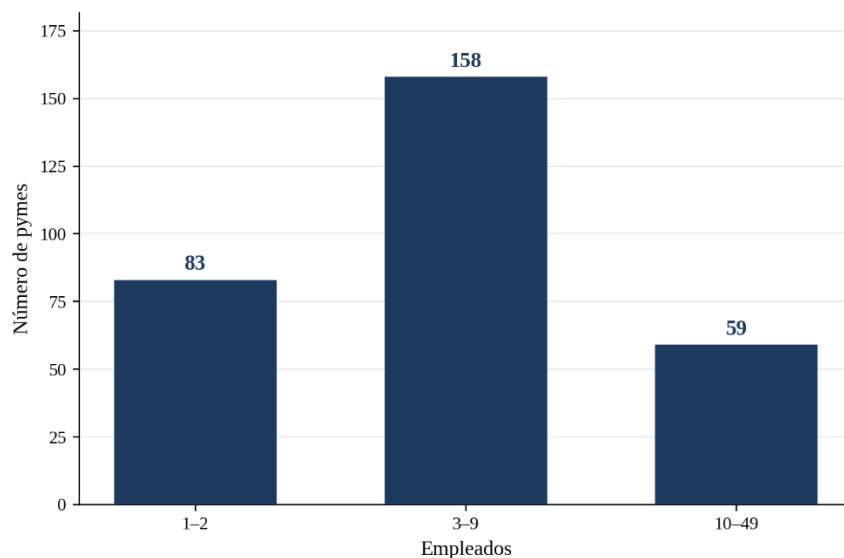
Ilustración 10. Estacionalidad mensual de los ingresos por sector.



Fuente: elaboración propia.

El conjunto generado está formado por 300 pymes repartidas en cuatro sectores (comercio, hostelería, servicios e industria ligera), cada una con dieciocho meses de movimientos bancarios, lo que da lugar a algo más de 56.000 transacciones. La construcción se realiza en dos niveles. En el primero se generan las empresas con sus atributos, y en el segundo se genera, para cada una de ellas, la secuencia completa de sus movimientos a lo largo del periodo.

Ilustración 11. Distribución de las pymes por tamaño.



Fuente: elaboración propia.

Cada pyme queda definida por un conjunto reducido de atributos. El sector se asigna de forma aleatoria entre los cuatro considerados. El número de empleados se extrae de una distribución que favorece deliberadamente los tamaños pequeños, de modo que la mayoría de las empresas tienen entre uno y nueve trabajadores y solo una minoría alcanza la franja de diez a cuarenta y nueve, en coherencia con el predominio de las microempresas en el tejido real (INE, 2025). La facturación anual no se fija de forma independiente, sino que se deriva del tamaño y del sector: se calcula a partir del número de empleados, multiplicado por una facturación media por empleado propia de cada sector y modulada por un factor de variación aleatoria. El resultado es una distribución de facturación sesgada hacia valores bajos, en la que la mayoría de las pymes facturan cantidades modestas, como ocurre en la realidad (CESGAR, 2025). A la antigüedad de la empresa se le asigna un valor de entre uno y veinticinco años.

Sobre esos atributos se genera, mes a mes, la actividad de cada empresa, distinguiendo varios tipos de movimiento que reproducen la operativa habitual de una pyme. Los ingresos por ventas se reparten cada mes en varios cobros de clientes, en lugar de en un único apunte, para imitar la entrada escalonada de la facturación. Su importe se modula por un factor de estacionalidad propio de cada sector, que recoge el pico navideño del comercio, la concentración estival de la hostelería, el parón de agosto en la industria ligera o el perfil más plano de los servicios. Frente a esos ingresos se generan los gastos característicos: las nóminas, proporcionales al número de empleados; el alquiler y los suministros, como gastos fijos mensuales; los pagos a proveedores, calculados como una fracción de las ventas y ejecutados con un retraso que reproduce el periodo medio de pago en España, situado en torno a los ochenta días, por encima del límite legal de sesenta (CEPYME, 2025); las liquidaciones trimestrales de IVA, siguiendo el calendario fiscal; y, en las empresas que tienen deuda, una cuota mensual de préstamo. Por último, el saldo de la cuenta se calcula de forma encadenada, sumando cada movimiento al saldo anterior a partir de un colchón inicial de tesorería, y cuando ese saldo queda en negativo se genera una comisión por descubierto, lo que reproduce las consecuencias reales de una tensión de liquidez.

Un elemento del diseño merece una explicación específica, porque es el que dota al conjunto de la variabilidad necesaria para que el análisis posterior tenga sentido. A cada pyme se le asigna, además de sus atributos visibles, un perfil de salud financiera latente, un valor entre cero y uno que orienta de forma transversal la generación de sus

movimientos. Las empresas con un perfil más tensionado presentan ingresos más volátiles, pagan a sus proveedores con mayor retraso, soportan con más probabilidad una carga de deuda y, al partir de un colchón de tesorería menor, incurren con más frecuencia en descubiertos. Las empresas saneadas muestran el comportamiento opuesto. Esta variable reparte así a las pymes a lo largo de un continuo que va de la solidez a la fragilidad financiera. Es importante subrayar que ese perfil latente no se utiliza en ningún momento del análisis posterior, sino que constituye una verdad oculta que permitirá, en la fase de validación, comprobar si el diagnóstico derivado de los datos coincide con la situación real de cada empresa. La distinción es esencial para la honestidad del ejercicio: el diagnóstico se construye exclusivamente a partir de los movimientos observados, igual que sucedería con datos reales, y nunca a partir de esa variable.

Todo el proceso se ejecuta con una semilla aleatoria fija, lo que garantiza que el conjunto de datos sea reproducible y que cualquier persona que ejecute el código obtenga exactamente el mismo resultado. El código completo se recoge en el anexo. Conviene insistir, para cerrar el apartado, en el alcance de lo que sigue: los resultados que se obtengan a partir de este conjunto tienen valor ilustrativo y demostrativo del procedimiento, no de validación predictiva, y así deben interpretarse a lo largo de todo el trabajo.

4.4. Pipeline de categorización automática de movimientos

Los movimientos que llegan a la plataforma son, en bruto, una sucesión de importes con un concepto en texto y una fecha. Para que esa información resulte útil es necesario agruparla en categorías con significado financiero, de modo que pueda distinguirse lo que la empresa ingresa por ventas de lo que destina a nóminas, a proveedores, a impuestos o al pago de su deuda. Esta categorización es el primer paso de transformación del dato y la condición previa para calcular cualquier indicador, ya que los indicadores de comportamiento del apartado siguiente se construyen precisamente sobre esas categorías.

Ilustración 12. Reglas de categorización.

Categoría	Palabras clave en el concepto
Ventas	VENTA, FACTURA CLIENTE, TPV, ABONO
Nóminas	NOMINA
Alquiler	ALQUILER
Suministros	SUMINISTRO, TELEFONIA
Proveedores	PROVEEDOR
Impuestos	IMPUESTO, AEAT, IVA, IRPF
Deuda	PRESTAMO, CUOTA, HIPOTECA
Comisiones	COMISION, DESCUBIERTO
Sin coincidencia (respaldo)	Ingreso → Ventas; gasto → Otros gastos

Fuente: elaboración propia.

El pipeline desarrollado clasifica cada movimiento mediante un conjunto de reglas basadas en el concepto. A cada categoría se le asocia una lista de palabras clave, de manera que un movimiento cuyo concepto contiene el término "NOMINA" se clasifica como nómina, uno que contiene "PROVEEDOR" como pago a proveedores, y así sucesivamente. Las reglas se evalúan empezando por las categorías más específicas, y cuando ninguna encaja se aplica un criterio de respaldo basado en el signo del importe, según el cual un ingreso no identificado se atribuye a ventas y un gasto no identificado se etiqueta como gasto sin clasificar. Es un enfoque deliberadamente sencillo y transparente, en coherencia con la prioridad que el trabajo concede a la interpretabilidad: cada clasificación puede justificarse señalando la regla que la ha producido.

Para comprobar su funcionamiento, la salida del pipeline se contrasta con la categoría real que el generador había guardado como verdad oculta. Sobre el conjunto sintético, el categorizador acierta el 100 % de los movimientos. Este resultado debe interpretarse con cautela y no como una medida del rendimiento que cabría esperar en un entorno real. El acierto es total porque los conceptos del conjunto sintético son limpios y unívocos, de manera que cada uno se corresponde sin ambigüedad con una categoría. En los datos reales, los conceptos bancarios son mucho más irregulares: contienen abreviaturas,

errores, referencias variables y formatos distintos según la entidad, lo que haría descender ese acierto y obligaría a enriquecer el pipeline con técnicas más robustas, como la correspondencia aproximada de cadenas o un clasificador entrenado. Conviene además señalar que, en la práctica, los propios agregadores suelen ofrecer la información ya categorizada, por lo que este paso ilustra una lógica que en un entorno productivo estaría en parte delegada.

Con los movimientos ya clasificados, la información está preparada para el paso que da sentido a todo lo anterior: convertir ese conjunto de transacciones categorizadas en un grupo reducido de indicadores que resuman el comportamiento financiero de cada pyme. A ello se dedica el apartado siguiente.

4.5. De transacciones a indicadores de comportamiento financiero

Una sucesión de movimientos bancarios, por muy bien categorizada que esté, no es todavía información sobre la que pueda emitirse un juicio. Para diagnosticar la situación de una pyme es necesario resumir ese conjunto de transacciones en un grupo reducido de magnitudes que capturen los rasgos relevantes de su comportamiento financiero. Esta transformación, que en la terminología analítica se conoce como ingeniería de variables, es el paso que convierte el dato en información útil y constituye el verdadero núcleo del proceso, ya que de la calidad de estos indicadores depende la del diagnóstico posterior.

Ilustración 13. Indicadores de comportamiento financiero.

Indicador	Definición	Interpretación favorable
Escala y estabilidad		
Ingresos medios mensuales	Media de las ventas mensuales	Mayor (más escala)
Volatilidad de los ingresos	Coefficiente de variación de las ventas mensuales	Menor
Índice de estacionalidad	Cociente entre el mes de mayor y el de menor venta	Menor (matizar por sector)
Sostenibilidad		
Ratio gastos / ingresos	Gastos totales sobre ingresos totales	Menor
Flujo de caja neto medio mensual	Ingresos menos gastos, en media mensual	Mayor
Liquidez		
Días de tesorería	Saldo medio dividido entre el gasto medio diario	Mayor
Frecuencia de descubiertos	Proporción de meses con saldo en negativo	Menor
Endeudamiento		
Ratio de servicio de la deuda	Cuotas de préstamo sobre ingresos	Menor
Peso de los gastos fijos	Nóminas, alquiler y suministros sobre ingresos	Menor

Fuente: elaboración propia.

A partir de las transacciones de cada pyme se calculan nueve indicadores, agrupados en cuatro dimensiones del comportamiento financiero. La primera es la escala y la estabilidad del negocio, recogida por los ingresos medios mensuales, por la volatilidad de los ingresos, que mide cuánto fluctúan las ventas de un mes a otro, y por un índice de estacionalidad que compara el mes de mayor facturación con el de menor. La segunda dimensión es la sostenibilidad, capturada por el ratio entre gastos e ingresos y por el flujo de caja neto medio mensual, que indican si la empresa vive dentro de sus posibilidades o consume recursos de forma sistemática. La tercera es la liquidez, medida por los días de tesorería, que estiman cuántos días de gasto puede cubrir la empresa con su saldo medio, y por la frecuencia de descubiertos, que registra con qué asiduidad la cuenta entra en negativo. La cuarta es el endeudamiento, recogido por el ratio de servicio de la deuda, que relaciona las cuotas de préstamo con los ingresos, y por el peso de los gastos fijos sobre la facturación.

Cada uno de estos indicadores tiene un significado financiero claro y un sentido de interpretación definido. Una volatilidad alta, una frecuencia de descubiertos elevada o un ratio de servicio de la deuda exigente apuntan a una situación de fragilidad, mientras que unos días de tesorería holgados o un flujo de caja neto positivo apuntan a solidez. Esta transparencia no es casual, sino una decisión de diseño coherente con la prioridad que el trabajo concede a la interpretabilidad: a diferencia de las variables opacas de un modelo de caja negra, cada uno de estos indicadores puede explicarse a la pyme en términos que entienda, lo que más adelante permitirá traducir el diagnóstico a un lenguaje claro.

Para comprobar que los indicadores capturan realmente la situación financiera de cada empresa, y no ruido, se contrasta su valor con el perfil de salud latente que el generador había guardado como verdad oculta. La correlación va en la dirección esperada en todos los casos. Los indicadores más informativos resultan ser la volatilidad de los ingresos y la estacionalidad, seguidos de la liquidez, la frecuencia de descubiertos y la carga de deuda. Esta comprobación es relevante porque confirma que el proceso de transformación funciona: a partir de un conjunto de movimientos, y sin recurrir en ningún momento a la variable oculta, se obtienen indicadores que reflejan la solidez o la fragilidad real de la empresa. Conviene matizar, no obstante, un punto sobre la estacionalidad, ya que esta no depende solo de la salud de la empresa sino también del sector al que pertenece, de modo que en el capítulo siguiente deberá tratarse con cuidado para no penalizar a negocios que son estacionales por su propia naturaleza, como los de hostelería.

El resultado de este apartado es una tabla con una fila por pyme y una columna por indicador, que sintetiza en unas pocas magnitudes lo que antes eran decenas de miles de movimientos. Esa tabla es el punto de partida del capítulo siguiente, en el que los indicadores se combinan para construir el perfil de solvencia de cada empresa.

5. DEL DATO AL DIAGNÓSTICO: PERFIL FINANCIERO DE LA PYME

5.1. Definición del perfil de solvencia a partir de los indicadores

El capítulo anterior terminó con una tabla que resume el comportamiento de cada pyme en nueve indicadores. El objetivo de este capítulo es combinar esos indicadores en un único perfil de solvencia que ordene a las pymes según su aptitud para acceder a financiación y que pueda traducirse a un diagnóstico comprensible. Antes de calcularlo conviene definir con precisión qué se entiende por ese perfil y con qué criterio se construye, porque de esa definición dependen tanto su validez como su defensa.

Ilustración 14. Construcción del perfil de solvencia.



Fuente: elaboración propia.

La primera decisión, y la más importante, es la del tipo de modelo. En la literatura de credit scoring conviven dos grandes enfoques: los modelos estadísticos o de aprendizaje automático, que aprenden a predecir un resultado, típicamente el impago, a partir de un conjunto de casos etiquetados, y los modelos de puntuación basados en reglas explícitas, que combinan un conjunto de indicadores con pesos definidos de antemano. Este trabajo adopta el segundo enfoque, por dos razones. La primera es de fondo: no se dispone de un resultado real de impago sobre el que entrenar un modelo predictivo. La única variable de resultado disponible es la salud financiera latente del conjunto sintético, y entrenar un modelo sobre ella sería circular, ya que el modelo se limitaría a recuperar lo que se introdujo en la generación. La segunda razón es de principio: un perfil destinado a explicarse a la pyme y a justificarse ante un banco debe ser interpretable, de modo que cada puntuación pueda descomponerse y argumentarse, algo que un modelo de caja negra dificulta y que la propia normativa de protección de datos respalda al reconocer el derecho a obtener una explicación de las decisiones automatizadas (Reglamento (UE) 2016/679). Esta prioridad por la interpretabilidad es, además, coherente con la evidencia de que un modelo transparente puede ofrecer un buen rendimiento sin renunciar a la explicación (Tobback y Martens, 2019).

Conviene precisar también qué representa el perfil y qué no representa. El perfil de solvencia que aquí se construye no es una probabilidad de impago, sino una medida ordinal de aptitud crediticia, es decir, una puntuación que sitúa a cada pyme en una escala que va de menor a mayor solidez financiera según su comportamiento observado. No sustituye al análisis de riesgo que el banco realiza con sus propios modelos, sino que lo precede: su función es cualificar y ordenar, ofreciendo a la pyme un diagnóstico y al banco una primera señal del encaje de cada empresa. Esta distinción es importante para no atribuir al perfil una capacidad predictiva que no tiene.

A partir de ahí, la construcción del perfil sigue una lógica sencilla y transparente. De los nueve indicadores se seleccionan los que reflejan solvencia, dejando fuera los ingresos medios mensuales, que miden el tamaño del negocio más que su solidez, y el índice de estacionalidad en su forma bruta, ya que la estacionalidad es un rasgo propio del sector y no un defecto de solvencia, como se advirtió en el apartado 4.5. Se excluye también el flujo de caja neto medio mensual, por recoger en buena medida la misma información de sostenibilidad que el ratio entre gastos e ingresos, con el que resultaría redundante. Cada indicador seleccionado se orienta de manera que un valor más alto signifique siempre mayor solidez, invirtiendo los que se interpretan en sentido contrario, y se normaliza comparando a cada pyme con las demás empresas de su mismo sector, lo que evita penalizar a un negocio por características intrínsecas a su actividad. Por último, los indicadores normalizados se agrupan por dimensión (estabilidad, sostenibilidad, liquidez y endeudamiento) y se combinan con unos pesos que reflejan su importancia relativa para la aptitud crediticia, dando lugar a una puntuación final en una escala de 0 a 100. El cálculo concreto de esa puntuación, junto con los pesos empleados, se desarrolla en el apartado siguiente.

5.2. Cálculo del score y traducción a un lenguaje claro para la pyme

Definida la lógica en el apartado anterior, el cálculo del perfil de solvencia se realiza en cuatro pasos. En primer lugar, de los nueve indicadores se seleccionan los seis que reflejan solvencia: la volatilidad de los ingresos, el ratio entre gastos e ingresos, los días de tesorería, la frecuencia de descubiertos, el ratio de servicio de la deuda y el peso de los gastos fijos. En segundo lugar, cada indicador se normaliza calculando la posición relativa de la pyme dentro de su sector, de manera que una empresa se compara siempre con otras de su misma actividad y no con el conjunto. Esa posición se orienta, además, para que un valor más alto signifique siempre mayor solidez, invirtiendo los indicadores que se

interpretan en sentido contrario. En tercer lugar, los indicadores normalizados se agrupan en sus cuatro dimensiones y se combinan con unos pesos que reflejan la importancia relativa de cada una. En cuarto y último lugar, el resultado se expresa en una escala de 0 a 100.

Ilustración 15. Pesos asignados a cada dimensión en el cálculo del perfil de solvencia.

Dimensión	Peso	Indicadores que la componen
Liquidez	35 %	Días de tesorería; frecuencia de descubiertos
Estabilidad	25 %	Volatilidad de los ingresos
Sostenibilidad	20 %	Ratio gastos / ingresos
Endeudamiento	20 %	Ratio de servicio de la deuda; peso de los gastos fijos

Fuente: elaboración propia.

Los pesos asignados conceden mayor importancia a la liquidez (35 %), por ser la dimensión que con más inmediatez condiciona la solvencia de una empresa pequeña, seguida de la estabilidad de los ingresos (25 %), y reparten el resto entre la sostenibilidad y el endeudamiento (20 % cada una). Esta asignación responde a un criterio financiero razonado y no a un ajuste posterior sobre los datos, lo que preserva la independencia entre la construcción del score y su validación. Aplicado al conjunto, el cálculo produce una puntuación media en torno a 50 puntos, con un recorrido que va de 5 a 88, lo que indica que el score discrimina y no concentra a todas las empresas en un valor central.

Sobre esa puntuación se construye el diagnóstico, que es lo que finalmente recibe la pyme. La escala se divide en cuatro bandas, que traducen un número a una valoración comprensible: por debajo de 40 el perfil se considera frágil, entre 40 y 60 ajustado, entre 60 y 80 sólido, y por encima de 80 muy sólido. A esa banda se añade un mensaje que identifica, para cada empresa, su dimensión más fuerte y la que conviene vigilar, obtenidas a partir de las subpuntuaciones por dimensión. El resultado es un diagnóstico como el siguiente, generado para una pyme del sector comercio con una puntuación de 73,6: «Tu perfil financiero es sólido y te sitúa en buena posición para acceder a financiación. Tu punto más fuerte es la dimensión de sostenibilidad y el aspecto a vigilar es la dimensión de estabilidad».

A modo de ilustración, la tabla siguiente recoge el resultado del proceso para una muestra de cuatro pymes. Para cada empresa se muestra su puntuación final, su banda y la

descomposición en las cuatro dimensiones, lo que permite apreciar cómo una misma puntuación se sostiene sobre perfiles distintos: la pyme de industria ligera, por ejemplo, presenta buena estabilidad, pero queda lastrada por su sostenibilidad y su liquidez, mientras que la empresa de comercio en banda sólida destaca por su sostenibilidad y su endeudamiento contenido.

Ilustración 16. Ejemplo de resultado del perfil de solvencia para una muestra de pymes.

Pyme	Sector	Puntuación	Banda	Estabilidad	Sostenib.	Liquidez	Endeud.
P0080	Servicios	80.7	Muy sólido	90	93	72	72
P0006	Comercio	73.6	Sólido	51	93	73	84
P0002	Industria ligera	44.0	Ajustado	81	23	38	29
P0001	Comercio	39.9	Frágil	28	38	49	41

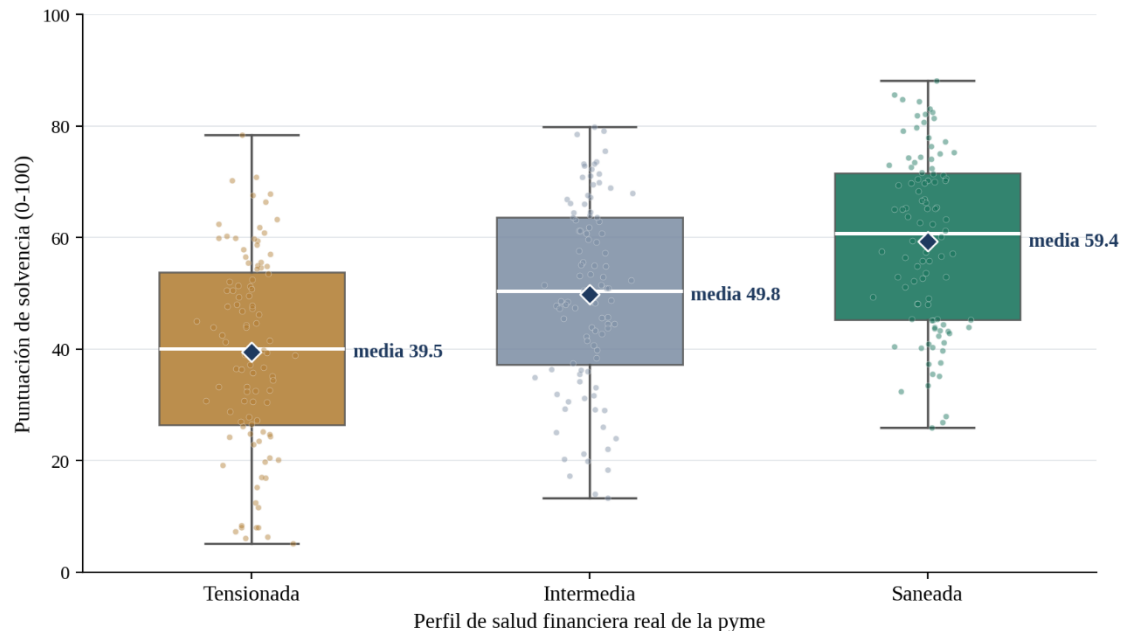
Fuente: elaboración propia.

Esta traducción es la que materializa la propuesta de valor para la pyme descrita en el capítulo 3. La empresa no recibe un número aislado ni una caja negra, sino una valoración de su situación expresada en términos que entiende y acompañada de los aspectos concretos sobre los que puede actuar. Además, y por la forma en que se ha construido, cada diagnóstico es enteramente trazable: una puntuación puede descomponerse en sus dimensiones, y cada dimensión en los indicadores que la forman, lo que permite justificar ante la pyme, o ante el banco, por qué su perfil es el que es. La interpretación detallada de estos resultados y su validación se abordan en el apartado siguiente.

5.3. Lectura e interpretación del diagnóstico

Para comprobar si el perfil de solvencia refleja realmente la situación de cada empresa, se contrasta la puntuación obtenida con el perfil de salud financiera latente que el generador había guardado como verdad oculta y que, conviene recordar, no ha intervenido en ningún momento del cálculo. La relación entre ambos es positiva y clara: las empresas que el generador había definido como más tensionadas obtienen una puntuación media de 39,5 puntos, las intermedias de 49,8 y las saneadas de 59,4. La ordenación es, por tanto, la esperada, y el score separa de forma consistente a las pymes según su solidez real, lo que confirma que todo el proceso, desde la transformación de las transacciones hasta la ponderación final, funciona en su conjunto.

Ilustración 17. Distribución de la puntuación según la salud financiera real.



Fuente: elaboración propia.

Esta correspondencia no es perfecta, y no debe serlo. Existe solapamiento entre los grupos, ya que hay empresas tensionadas que alcanzan puntuaciones razonables y empresas saneadas que quedan por debajo de la media de su grupo. Lejos de ser un defecto, esa imperfección es una señal de que el ejercicio es honesto. Un score que recuperase exactamente la variable latente delataría un proceso circular, en el que el resultado se habría filtrado de algún modo en el cálculo. Lo que se observa, en cambio, es lo propio de un modelo construido a partir de información parcial: capta la señal dominante del comportamiento financiero sin reproducir el dato de origen, igual que ocurriría al trabajar con datos reales, donde ninguna puntuación explica por completo la realidad de una empresa.

Más allá de esta validación agregada, el valor del diagnóstico reside en su lectura individual. Cada pyme no recibe únicamente una puntuación, sino también una banda y una descomposición en las cuatro dimensiones que permite entender de dónde procede esa valoración. Esa descomposición es lo que convierte un número en información accionable: dos empresas con una puntuación similar pueden tener perfiles muy distintos, una sólida en liquidez pero frágil en endeudamiento y otra a la inversa, y el diagnóstico

lo refleja señalando en cada caso el punto fuerte y el aspecto a vigilar. Para la pyme, esta lectura supone un punto de partida concreto sobre el que actuar. Para el banco, la misma información, estructurada y comparable entre empresas, es la que permite ordenar y priorizar, y constituye el contenido del lead cualificado descrito en el capítulo tercero.

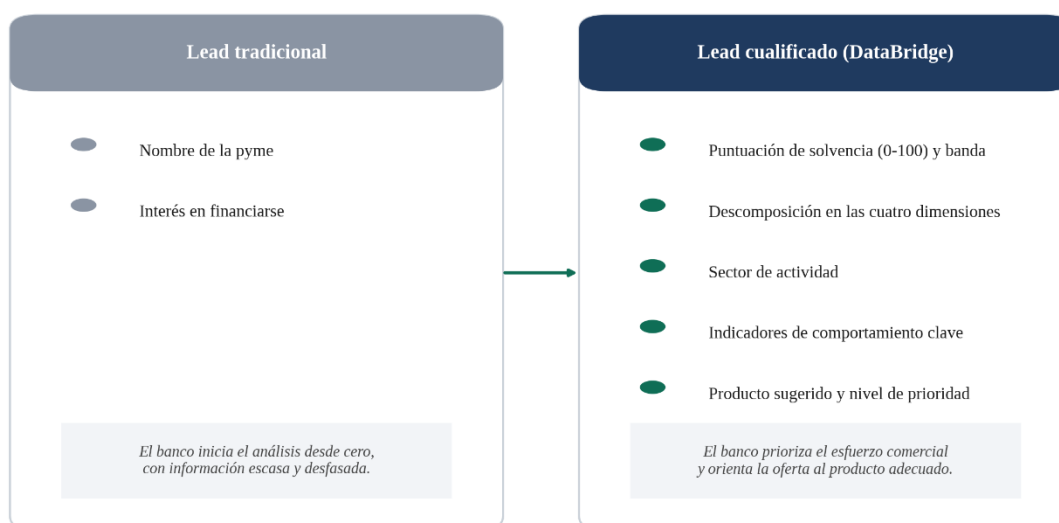
Conviene, por último, situar la interpretación de estos resultados dentro de los límites del ejercicio. El perfil se ha validado contra una variable de salud financiera, no contra un resultado real de impago, por lo que mide aptitud crediticia y no probabilidad de default. Los pesos que combinan las dimensiones responden a un criterio financiero razonado y no a un proceso de aprendizaje sobre datos, de modo que un trabajo posterior que dispusiera de resultados reales podría calibrarlos de forma empírica. Y todo el análisis se apoya en un conjunto sintético, con el alcance ilustrativo que ya se ha señalado. Dentro de esos límites, el resultado es consistente: a partir de los movimientos de una cuenta, y mediante un procedimiento enteramente transparente, se obtiene un diagnóstico que ordena correctamente a las empresas y que puede explicarse a cada una de ellas. El último paso es mostrar cómo ese diagnóstico se traduce en una recomendación cualificada para el banco y qué valor aporta, objeto del capítulo siguiente.

6. DEL DIAGNÓSTICO A LA RECOMENDACIÓN PARA EL BANCO Y SU VALOR

6.1. De la puntuación al lead cualificado: cómo el diagnóstico se traduce en una recomendación para el banco

El capítulo anterior terminó con un diagnóstico dirigido a la pyme. Ese mismo diagnóstico, sin embargo, es también el insumo a partir del cual se construye el valor para el otro lado del mercado. Como se explicó en el capítulo tercero, el análisis de los datos transaccionales produce dos resultados a partir de un único proceso: la recomendación que recibe la pyme y el lead cualificado que recibe el banco. Este apartado describe cómo se realiza esa segunda traducción, de la puntuación al lead.

Ilustración 18. Contenido de un lead cualificado frente a uno tradicional.



Fuente: elaboración propia.

La diferencia esencial entre un lead cualificado y un contacto convencional está en la información que lo acompaña. Un lead tradicional, como el que proporciona un comparador, se reduce en la práctica a una pyme que ha manifestado interés en financiarse, sin más contexto. El banco recibe un nombre y debe iniciar desde cero el análisis de la empresa, con la misma información escasa y desfasada que se describió en el capítulo segundo. El lead de DataBridge, en cambio, llega acompañado del perfil que el análisis ha generado: la puntuación de solvencia, su descomposición en las cuatro dimensiones, el sector de la empresa y los indicadores de comportamiento más relevantes,

todo ello estructurado y comparable entre empresas. El banco no recibe solo un interesado, sino un interesado del que ya conoce cómo se comporta financieramente.

Esa información es la que convierte el lead en una recomendación. La puntuación permite al banco ordenar y priorizar las pymes según su aptitud crediticia, decidiendo a cuáles dedicar antes su esfuerzo comercial. La descomposición por dimensiones añade el matiz necesario para orientar la oferta: una empresa sólida en ingresos pero con tensiones de liquidez encaja con una línea de circulante, mientras que otra con un perfil estable y poco endeudada es candidata a un préstamo a plazo en buenas condiciones. De este modo, el lead no se limita a indicar que una pyme quiere financiación, sino que sugiere qué tipo de producto se ajusta mejor a su situación y con qué nivel de prioridad conviene atenderla.

El resultado es una reducción directa de la asimetría de información en la fase inicial del proceso, que es precisamente el cuello de botella identificado al comienzo del trabajo. El banco parte de una posición de conocimiento mucho mayor, lo que debería traducirse en una mayor probabilidad de cerrar la operación y en un menor esfuerzo dedicado a contactos que no encajan. Conviene precisar que toda esta información se comparte con el banco con el consentimiento de la pyme, que autoriza la cesión de su perfil como parte del servicio. Ahora bien, la magnitud de ese beneficio depende de una condición que hasta ahora se ha asumido pero no se ha demostrado: que la puntuación concentre realmente a las buenas pymes en sus tramos más altos. Comprobarlo, y medir cuánto, es el objeto del apartado siguiente.

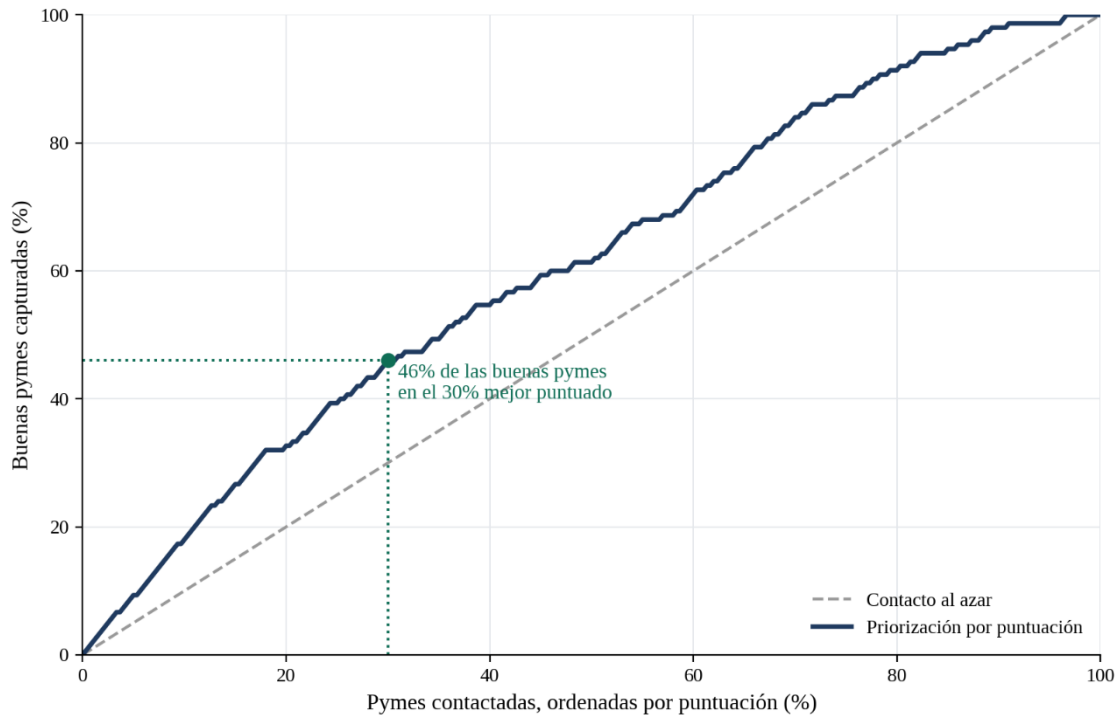
6.2. La capacidad de discriminación del score en términos de negocio

El apartado anterior asumió que la puntuación concentra a las buenas pymes en sus tramos altos, pero esa afirmación necesita comprobarse antes de poder cuantificar su valor. La forma habitual de medirlo en analítica de negocio es el análisis de ganancia acumulada, que responde a una pregunta concreta: si el banco ordena los leads por su puntuación y atiende primero a los mejor valorados, ¿cuántas buenas pymes captura en comparación con un contacto al azar?

Para realizarlo es necesario definir qué se entiende por buena pyme. Se considera como tal a la mitad de las empresas con mejor salud financiera real, es decir, la variable latente que, conviene insistir, no ha intervenido en el cálculo de la puntuación. La tasa base de buenas pymes es, por tanto, del 50 %: si el banco contactara a las empresas al azar,

encontraría una buena pyme en la mitad de los casos. La cuestión es si la priorización por puntuación mejora esa proporción.

Ilustración 19. Curva de ganancia acumulada del score.



Fuente: elaboración propia.

Los resultados confirman que sí, y de forma apreciable. Entre el diez por ciento de pymes con mejor puntuación, el 90 % son buenas pymes, frente al 50 % de base, lo que supone un *lift* de 1,8: priorizar por puntuación casi duplica la proporción de aciertos en la parte alta de la lista. El efecto se mantiene a lo largo de todo el recorrido. Contactando únicamente al 30 % de las empresas mejor puntuadas, el banco alcanza al 46 % de todas las buenas pymes, y contactando a la mitad, al 61 %. En todos los tramos, la curva de ganancia acumulada se sitúa por encima de la diagonal que representa el contacto aleatorio, lo que indica que el orden que propone el score es sistemáticamente mejor que el azar.

La lectura de negocio de este resultado es directa. Con el mismo número de contactos, el banco que prioriza por puntuación localiza más buenas pymes que el que actúa sin esa información, o, dicho de otro modo, alcanza un objetivo determinado de buenas pymes contactando a menos empresas. Esa diferencia es la base material del valor que aporta el lead cualificado: no es una mejora hipotética, sino una concentración medible sobre los

propios datos del trabajo. Sobre ella se construye, en el apartado siguiente, la estimación económica del valor que ese mejor orden representa para el banco frente a una recomendación sin análisis.

6.3. Cuantificación del valor para el banco frente a la recomendación sin análisis

Demostrada la capacidad de discriminación del score, el último paso es traducirla a términos económicos: cuánto vale para el banco recibir leads cualificados frente a operar sin ese análisis. Para estimarlo se compara una misma campaña comercial bajo dos escenarios. En el primero, sin análisis, el banco trabaja leads no cualificados, en un orden equivalente al azar. En el segundo, con DataBridge, trabaja leads cualificados y priorizados. La diferencia entre ambos se concentra en una variable: la tasa de conversión de lead a operación cerrada.

Esa mejora de la conversión no se postula, sino que se deriva del análisis de lift del apartado anterior. Dado que la priorización por puntuación eleva la proporción de buenas pymes entre las contactadas, y que una buena pyme presenta una mayor probabilidad de cerrar una operación, la conversión media aumenta. Partiendo de la tasa del 25 % que el modelo de negocio estima para los leads cualificados, y de la concentración de buenas pymes medida en el lift, el modelo sitúa la conversión sin análisis en torno al 21 %. El resto de parámetros procede del propio modelo de negocio: un ticket medio de 80.000 euros por operación, una comisión a éxito del 0,3 % sobre el volumen financiado, que equivale a 240 euros por operación cerrada, y un margen neto para el banco por operación que, a falta del dato real, se fija de forma prudente en 4.000 euros y se somete después a análisis de sensibilidad.

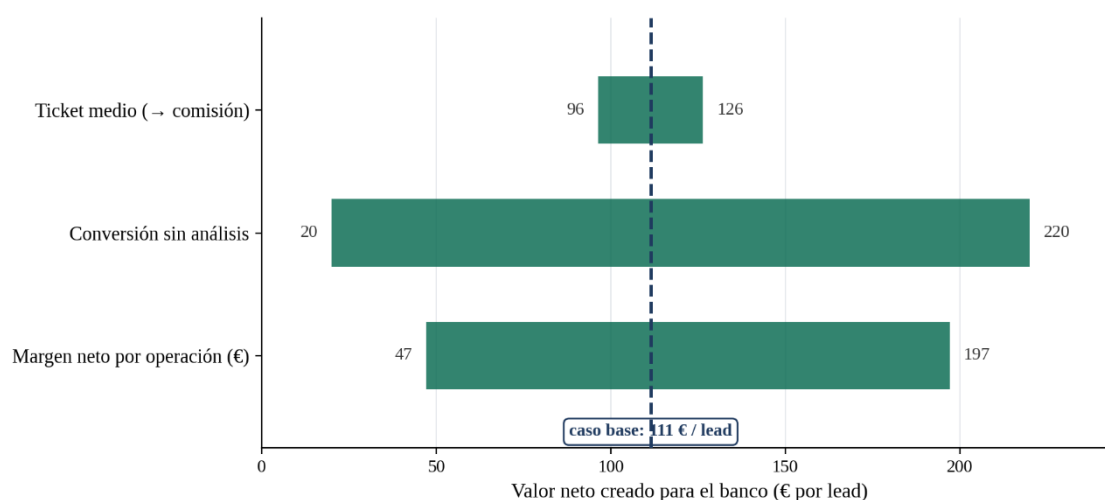
Ilustración 20. Comparación de escenarios para una campaña de 1000 leads.

Concepto	Sin análisis	Con DataBridge
Tasa de conversión	20,7 %	25,0 %
Operaciones cerradas	207	250
Margen bruto del banco	828.571 €	1.000.000 €
Comisión a DataBridge	—	60.000 €
Margen neto del banco	828.571 €	940.000 €

Fuente: elaboración propia.

Sobre una campaña de 1.000 leads, los resultados son los siguientes. Sin análisis, el banco cierra 207 operaciones; con DataBridge, 250. Esas 43 operaciones adicionales generan un margen bruto de unos 171.000 euros. De ese valor, el banco abona a DataBridge la comisión a éxito correspondiente a las operaciones cerradas, unos 60.000 euros, de modo que el valor neto que captura asciende a unos 111.000 euros por cada 1.000 leads, es decir, en torno a 111 euros por lead. Conviene subrayar la relación entre ese valor y el precio: la comisión que cobra DataBridge equivale aproximadamente a un tercio del valor bruto que su análisis genera, lo que deja los otros dos tercios en manos del banco. El servicio se paga, por tanto, con una fracción del valor que crea, que es la condición para que la propuesta sea sostenible para ambas partes.

Ilustración 21. Sensibilidad del valor neto por lead a los supuestos clave del modelo.



Fuente: elaboración propia.

El análisis de sensibilidad confirma la solidez de esta conclusión y, a la vez, delimita su alcance con honestidad. El valor creado se mantiene positivo en todos los escenarios razonables considerados, pero su magnitud depende sobre todo de dos supuestos. El primero, y más influyente, es cuánto peor convierte el banco sin el análisis: si la diferencia de conversión fuera menor de la estimada, el valor se reduciría de forma notable, mientras que si fuera mayor, crecería. El segundo es el margen neto que el banco obtiene por operación, sobre el que el trabajo no dispone de un dato real. El ticket medio, en cambio, influye poco. Esta lectura es importante, porque sitúa el resultado en su justa medida: el modelo demuestra que el valor existe y que es de un orden de magnitud relevante, pero su cifra exacta dependería de parámetros que solo una implementación real con un banco permitiría fijar.

Por último, conviene mencionar una fuente de valor adicional que el modelo no cuantifica de forma deliberada. Una recomendación sin análisis no solo convierte menos, sino que aumenta el riesgo de colocar a la pyme un producto que no encaja con su situación, con las consecuencias asociadas de insatisfacción, reclamaciones y eventual pérdida del cliente. Ese coste del mal encaje es real y refuerza el argumento, pero cuantificarlo exigiría supuestos sobre el comportamiento de la clientela que excederían el rigor del resto del análisis, por lo que se deja señalado como un factor cualitativo que se suma, sin medir, al valor estimado. Con esta cuantificación se cierra el recorrido analítico del trabajo, que ha ido del dato transaccional en bruto al valor económico que su análisis genera para las dos partes del mercado.

7. CONCLUSIONES

Al finalizar este Trabajo de Fin de Grado, se han alcanzado los objetivos específicos establecidos en el primer capítulo. El análisis, ordenado según las fases de la metodología CRISP-DM, ha permitido recorrer el camino completo que va del dato transaccional en bruto hasta el valor económico que su tratamiento genera, y responder de forma afirmativa a la pregunta de partida. A continuación se presentan las conclusiones derivadas de cada objetivo.

En relación con el primer objetivo, el trabajo ha situado la dificultad de acceso a la financiación de las pymes como un problema de información antes que económico. El banco evalúa con datos contables agregados y desfasados, mientras que el comportamiento financiero real de la empresa, que es la información verdaderamente relevante, permanece inaccesible. Se ha mostrado que el marco regulatorio del open banking y la infraestructura de los agregadores ponen esa información a disposición de un tercero autorizado, y que la literatura confirma su valor predictivo, lo que establece la viabilidad de una respuesta analítica al problema.

En relación con el segundo objetivo, se ha sintetizado el modelo de negocio de DataBridge y se ha traducido su propuesta a un problema de datos abordable. La plataforma genera, a partir de un único análisis de los datos transaccionales, un diagnóstico para la pyme y un lead cualificado para el banco, de modo que la propuesta de valor se reduce, en lo esencial, a una cuestión analítica que ha articulado el resto del trabajo.

En relación con el tercer y el cuarto objetivo, se ha definido el flujo y la estructura de entrada de la información y se ha construido el núcleo técnico del trabajo. Ante la ausencia de datos reales, se ha generado un conjunto sintético de 300 pymes con dieciocho meses de movimientos, calibrado con estadística pública contrastada, sobre el que se ha implementado la categorización automática de los movimientos y la derivación de nueve indicadores de comportamiento financiero. La comprobación de que esos indicadores reflejan la situación real de cada empresa ha confirmado la validez del proceso de transformación.

En relación con el quinto objetivo, se ha construido un perfil de solvencia a partir de esos indicadores. Se ha optado por un modelo de puntuación transparente, por su interpretabilidad y por la ausencia de datos reales de impago, que combina los indicadores normalizados por sector en una escala de 0 a 100 y se traduce a un diagnóstico claro para la pyme. La validación ha mostrado que el score ordena correctamente a las empresas según su solidez real, con una separación clara entre grupos y un solapamiento que confirma la honestidad del ejercicio.

En relación con el sexto objetivo, se ha demostrado cómo ese mismo perfil se convierte en un lead cualificado para el banco y se ha cuantificado su valor. Mediante un análisis de ganancia acumulada sobre los propios datos se ha verificado que la priorización por puntuación concentra a las buenas pymes en los tramos altos, y sobre esa base medida se ha estimado un valor económico para el banco positivo y de un orden relevante, equivalente a varias veces el precio del servicio.

En conjunto, el trabajo confirma que la propuesta planteada en el plan de negocio se sostiene también desde una perspectiva analítica. Sus conclusiones deben leerse, no obstante, dentro de los límites del ejercicio: todo el análisis se apoya en datos sintéticos, por lo que sus resultados tienen valor demostrativo del procedimiento y no de validación predictiva, y tanto los pesos del perfil como los parámetros del modelo de valor responden a criterios razonados que solo una implementación real permitiría fijar. De ahí se deriva la principal línea futura, que es validar el procedimiento con datos reales para calibrar los pesos de forma empírica y sustituir el sistema de puntuación por reglas por un modelo de aprendizaje automático. Con esa salvedad, el trabajo demuestra que el análisis del dato transaccional ofrece una vía concreta y defendible para reducir la asimetría de información en la financiación a las pymes.

8. BIBLIOGRAFÍA

Asociación Española de Fintech e Insurtech (AEFI). (2026). El ecosistema FinTech e InsurTech en España. AEFI.

Banco de España. (2025). Encuesta sobre el acceso a la financiación de las empresas (SAFE). Banco de España.

Berg, T., Burg, V., Gombović, A. y Puri, M. (2020). On the rise of FinTechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7), 2845–2897. <https://doi.org/10.1093/rfs/hhz099>

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. y Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.

Confederación Española de la Pequeña y Mediana Empresa (CEPYME). (2025). Observatorio de Morosidad. CEPYME.

Confederación Española de Sociedades de Garantía Recíproca (CESGAR). (2025). La financiación de la pyme en España (XV Informe). CESGAR.

Directiva (UE) 2015/2366 del Parlamento Europeo y del Consejo, de 25 de noviembre de 2015, sobre servicios de pago en el mercado interior, por la que se modifican las Directivas 2002/65/CE, 2009/110/CE y 2013/36/UE y el Reglamento (UE) n.º 1093/2010 y se deroga la Directiva 2007/64/CE. *Diario Oficial de la Unión Europea*, L 337, 23 de diciembre de 2015.

Instituto Nacional de Estadística. (2025). Directorio Central de Empresas (DIRCE) a 1 de enero de 2025 [Nota de prensa]. INE.

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. *Boletín Oficial del Estado*, núm. 294, 6 de diciembre de 2018.

Ministerio de Industria y Turismo. (2025). Cifras PYME. Ministerio de Industria y Turismo.

Ministerio de Industria y Turismo. (2024). Retrato de la PYME 2024. DIRCE a 1 de enero de 2023. Dirección General de Industria y de la Pequeña y Mediana Empresa. <https://ipyme.org/Publicaciones/Retrato%20de%20la%20PYME/Retrato-PYME-DIRCE-1-enero-2023.pdf>

Osterwalder, A. y Pigneur, Y. (2010). *Business Model Generation*. Wiley.

Osterwalder, A., Pigneur, Y., Bernarda, G. y Smith, A. (2014). *Value Proposition Design*. Wiley.

Real Decreto-ley 19/2018, de 23 de noviembre, de servicios de pago y otras medidas urgentes en materia financiera. Boletín Oficial del Estado, núm. 284, 24 de noviembre de 2018.

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (Reglamento general de protección de datos). Diario Oficial de la Unión Europea, L 119, 4 de mayo de 2016.

Reglamento Delegado (UE) 2018/389 de la Comisión, de 27 de noviembre de 2017, por el que se completa la Directiva (UE) 2015/2366 en lo relativo a las normas técnicas de regulación para la autenticación reforzada de clientes y unos estándares de comunicación abiertos, comunes y seguros. Diario Oficial de la Unión Europea, L 69, 13 de marzo de 2018.

Stiglitz, J. E. y Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3), 393–410.

Tobback, E. y Martens, D. (2019). Retail credit scoring using fine-grained payment data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(4), 1227–1246. <https://doi.org/10.1111/rssa.12469>

9. APÉNDICES DE CÓDIGO

9.1. Generación dataset

"""

Generación de un conjunto de datos sintético de pymes y transacciones bancarias.

Salida:

- pymes.csv : una fila por pyme, con sus atributos.
- transacciones.csv : una fila por movimiento bancario.

El conjunto se genera con una semilla fija para que sea reproducible.

"""

```
import numpy as np
import pandas as pd
```

```
SEMILLA = 42
```

```
rng = np.random.default_rng(SEMILLA)
```

```
# 1. PARÁMETROS GENERALES
```

```
N_PYMES = 300 # tamaño manejable, suficiente para las
gráficas
```

```
FECHA_INICIO = pd.Timestamp("2024-01-01")
```

```
N_MESES = 18 # 18 meses de historia: ene-2024 a jun-
2025
```

```
SECTORES = ["Comercio", "Hostelería", "Servicios", "Industria ligera"]
```

```
# Facturación media por empleado (€/año), aproximada por sector.
```

```
# Calibrada con órdenes de magnitud de la cifra de negocio por ocupado (INE).
```

```
FACT_POR_EMPLEADO = {
    "Comercio": 130_000,
    "Hostelería": 70_000,
    "Servicios": 90_000,
    "Industria ligera": 110_000,
}
```

```
# Factores de estacionalidad mensual (índice 1 = mes medio).
```

```
# Calibrados con patrones sectoriales conocidos: pico navideño en comercio,
```

```
# verano en hostelería, parón de agosto en industria, servicios más plano.
```

```
ESTACIONALIDAD = {
    "Comercio": [0.85, 0.80, 0.95, 1.00, 1.00, 1.00, 0.95, 0.85,
1.00, 1.05, 1.25, 1.35],
    "Hostelería": [0.70, 0.70, 0.85, 1.00, 1.10, 1.30, 1.45, 1.45,
1.10, 0.95, 0.85, 1.05],
```

```

    "Servicios": [0.95, 0.95, 1.00, 1.00, 1.05, 1.05, 1.00, 0.80,
1.05, 1.05, 1.05, 1.05],
    "Industria ligera": [1.00, 1.00, 1.05, 1.05, 1.05, 1.00, 0.95, 0.55,
1.05, 1.10, 1.10, 1.10],
}

```

2. GENERACIÓN DE LAS PYMES

```

def generar_pymes(n):
    filas = []
    for i in range(1, n + 1):
        sector = rng.choice(SECTORES)

        # Tamaño: predominio de microempresas (1-9), coherente con el peso de
        # las micro en el tejido español (INE/DIRCE). Pocas llegan a 10-49.
        tramo = rng.choice(["micro_baja", "micro_alta", "pequena"], p=[0.45,
0.35, 0.20])
        if tramo == "micro_baja":
            empleados = int(rng.integers(1, 4))
        elif tramo == "micro_alta":
            empleados = int(rng.integers(4, 10))
        else:
            empleados = int(rng.integers(10, 50))

        antiguedad = int(rng.integers(1, 26)) # 1 a 25 años

        # Facturación anual base, ligada al tamaño y al sector, con ruido.
        # La distribución resultante queda sesgada hacia valores bajos, en
        # línea con que la mayoría de las pymes facturan menos de 300.000 €
        fact_base = empleados * FACT_POR_EMPLEADO[sector] * rng.uniform(0.7,
1.3)

        # Perfil de salud financiera latente (0 = tensionada, 1 = saneada).
        # No se usa directamente en el análisis posterior: es la "verdad
oculta"
        # que permitirá, más adelante, validar el diagnóstico. Se reparte
para
        # que haya empresas saneadas, intermedias y tensionadas.
        salud = float(np.clip(rng.beta(2, 2), 0.05, 0.95))

        filas.append({
            "pyme_id": f"P{i:04d}",
            "sector": sector,
            "n_empleados": empleados,
            "antiguedad_anios": antiguedad,
            "facturacion_anual_base": round(fact_base, 2),
            "salud_latente": round(salud, 3),
        })

```

```
return pd.DataFrame(filas)
```

```
# 3. GENERACIÓN DE LAS TRANSACCIONES
```

```
def generar_transacciones(pymes):  
    movimientos = []  
    tid = 0  
  
    for _, p in pymes.iterrows():  
        sector = p["sector"]  
        salud = p["salud_latente"]  
        fact_mensual = p["facturacion_anual_base"] / 12.0  
  
        # La volatilidad de los ingresos es mayor en empresas tensionadas.  
        volatilidad = 0.10 + (1 - salud) * 0.35  
  
        # Plazo de pago a proveedores: las empresas saneadas pagan antes.  
        # Calibrado con el periodo medio de pago en España, superior a los  
        # 60 días legales en buena parte de los casos (CEPYME, morosidad).  
        plazo_pago = int(30 + (1 - salud) * 70) # 30 a 100 días  
  
        # ¿Tiene la pyme un préstamo en curso? Más probable si está  
        # tensionada.  
        tiene_prestamo = rng.random() < (0.3 + (1 - salud) * 0.4)  
        cuota_prestamo = round(fact_mensual * rng.uniform(0.04, 0.10), 2) if  
        tiene_prestamo else 0.0  
  
        cuenta_id = f"{p['pyme_id']}-C1"  
  
        for m in range(N_MESES):  
            mes_fecha = FECHA_INICIO + pd.DateOffset(months=m)  
            mes_idx = mes_fecha.month - 1  
            factor = ESTACIONALIDAD[sector][mes_idx]  
  
            # INGRESOS: ventas repartidas en varios cobros de clientes  
            ingreso_mes = fact_mensual * factor * rng.normal(1.0,  
            volatilidad)  
            ingreso_mes = max(ingreso_mes, fact_mensual * 0.2)  
            n_cobros = int(rng.integers(3, 9))  
            pesos = rng.dirichlet(np.ones(n_cobros))  
            for k in range(n_cobros):  
                dia = int(rng.integers(1, 28))  
                tid += 1  
                movimientos.append({  
                    "transaccion_id": f"T{tid:07d}",  
                    "pyme_id": p["pyme_id"],  
                    "cuenta_id": cuenta_id,
```

```

        "fecha": mes_fecha.replace(day=dia),
        "importe": round(ingreso_mes * pesos[k], 2),
        "concepto": rng.choice(
            ["TRANSF VENTA CLIENTE", "ABONO FACTURA CLIENTE",
"COBRO TPV"]),
        "categoria_real": "Ventas",
    })

# NÓMINAS: pago mensual proporcional al número de empleados
nomina = p["n_empleados"] * rng.uniform(1500, 2200)
tid += 1
movimientos.append({
    "transaccion_id": f"T{tid:07d}", "pyme_id": p["pyme_id"],
    "cuenta_id": cuenta_id, "fecha": mes_fecha.replace(day=28),
    "importe": -round(nomina, 2), "concepto": "RECIBO NOMINA
PERSONAL",
    "categoria_real": "Nóminas",
})

# ALQUILER: gasto fijo mensual
alquiler = fact_mensual * rng.uniform(0.05, 0.12)
tid += 1
movimientos.append({
    "transaccion_id": f"T{tid:07d}", "pyme_id": p["pyme_id"],
    "cuenta_id": cuenta_id, "fecha": mes_fecha.replace(day=5),
    "importe": -round(alquiler, 2), "concepto": "ADEUDO ALQUILER
LOCAL",
    "categoria_real": "Alquiler",
})

# SUMINISTROS: luz, agua, etc.
suministros = fact_mensual * rng.uniform(0.02, 0.05)
tid += 1
movimientos.append({
    "transaccion_id": f"T{tid:07d}", "pyme_id": p["pyme_id"],
    "cuenta_id": cuenta_id, "fecha": mes_fecha.replace(day=12),
    "importe": -round(suministros, 2),
    "concepto": rng.choice(["RECIBO SUMINISTRO ELECTRICIDAD",
"RECIBO SUMINISTRO AGUA", "RECIBO
TELEFONIA"]),
    "categoria_real": "Suministros",
})

# PROVEEDORES: compras, pagadas con el plazo correspondiente
compras = ingreso_mes * rng.uniform(0.35, 0.55)
fecha_pago = mes_fecha + pd.Timedelta(days=plazo_pago)
tid += 1
movimientos.append({
    "transaccion_id": f"T{tid:07d}", "pyme_id": p["pyme_id"],

```

```

        "cuenta_id": cuenta_id, "fecha": fecha_pago,
        "importe": -round(compras, 2), "concepto": "PAGO PROVEEDOR",
        "categoria_real": "Proveedores",
    })

# IMPUESTOS: liquidación trimestral de IVA
if mes_fecha.month in (1, 4, 7, 10):
    iva = fact_mensual * 3 * 0.21 * rng.uniform(0.10, 0.20)
    tid += 1
    movimientos.append({
        "transaccion_id": f"T{tid:07d}", "pyme_id": p["pyme_id"],
        "cuenta_id": cuenta_id, "fecha":
mes_fecha.replace(day=20),
        "importe": -round(iva, 2), "concepto": "IMPUESTO AEAT
IVA",
        "categoria_real": "Impuestos",
    })

# CUOTA DE PRÉSTAMO: mensual, si la pyme tiene deuda
if cuota_prestamo > 0:
    tid += 1
    movimientos.append({
        "transaccion_id": f"T{tid:07d}", "pyme_id": p["pyme_id"],
        "cuenta_id": cuenta_id, "fecha":
mes_fecha.replace(day=2),
        "importe": -cuota_prestamo, "concepto": "CUOTA PRESTAMO
BANCARIO",
        "categoria_real": "Deuda",
    })

df = pd.DataFrame(movimientos)
df = df.sort_values(["pyme_id", "cuenta_id",
"fecha"]).reset_index(drop=True)
return df

```

4. SALDO ENCADENADO Y COMISIONES POR DESCUBIERTO

```

def calcular_saldos(transacciones, pymes):
    """Calcula el saldo resultante tras cada movimiento y añade comisiones de
descubierto cuando el saldo queda en negativo (más frecuentes en empresas
tensionadas, ya que parten de un colchón de tesorería menor)."""
    salud = pymes.set_index("pyme_id")["salud_latente"].to_dict()
    fact = pymes.set_index("pyme_id")["facturacion_anual_base"].to_dict()

    nuevas_filas = []
    saldos_finales = []

```

```

for pyme_id, grupo in transacciones.groupby("pyme_id", sort=False):
    # Colchón inicial de tesorería: mayor en empresas saneadas.
    colchon = fact[pyme_id] / 12.0 * (0.3 + salud[pyme_id] * 1.2)
    saldo = colchon
    for _, t in grupo.iterrows():
        saldo += t["importe"]
        fila = t.to_dict()
        fila["saldo_resultante"] = round(saldo, 2)
        nuevas_filas.append(fila)
        # Comisión por descubierto si el saldo es negativo.
        if saldo < 0:
            comision = min(abs(saldo) * 0.02, 50)
            saldo -= comision
            saldos_finales.append({
                "transaccion_id": f"{t['transaccion_id']}-COM",
                "pyme_id": pyme_id, "cuenta_id": t["cuenta_id"],
                "fecha": t["fecha"], "importe": -round(comision, 2),
                "concepto": "COMISION DESCUBIERTO", "categoria_real":
"Comisiones",
                "saldo_resultante": round(saldo, 2),
            })
    df = pd.DataFrame(nuevas_filas + saldos_finales)
    df = df.sort_values(["pyme_id", "cuenta_id", "fecha",
"transaccion_id"]).reset_index(drop=True)
    # Reordenar columnas
    cols = ["transaccion_id", "pyme_id", "cuenta_id", "fecha", "importe",
            "concepto", "saldo_resultante", "categoria_real"]
    return df[cols]

```

5. EJECUCIÓN

```

if __name__ == "__main__":
    pymes = generar_pymes(N_PYMES)
    transacciones = generar_transacciones(pymes)
    transacciones = calcular_saldos(transacciones, pymes)

    pymes.to_csv("pymes.csv", index=False)
    transacciones.to_csv("transacciones.csv", index=False)

    print(f"Pymes generadas: {len(pymes)}")
    print(f"Transacciones generadas: {len(transacciones)}")
    print(f"Periodo: {transacciones['fecha'].min().date()} a
{transacciones['fecha'].max().date()}")
    print("\nReparto por sector:")
    print(pymes['sector'].value_counts())
    print("\nReparto de movimientos por categoría:")
    print(transacciones['categoria_real'].value_counts())

```

```

print("\nEjemplo de transacciones de una pyme:")
ej = transacciones[transacciones['pyme_id'] == 'P0001'].head(8)
print(ej.to_string(index=False))

```

9.2. Categorización automática de movimientos

```
"""
```

Pipeline de categorización automática de movimientos bancarios.

Cada movimiento llega con un concepto en texto libre y un importe con signo. La función 'categorizar' asigna a cada movimiento una de las categorías de gasto o ingreso relevantes para el análisis posterior, aplicando un conjunto de reglas sobre el concepto y, como apoyo, sobre el signo del importe.

El acierto del pipeline se valida comparando su salida con la categoría real que el generador guardó como verdad oculta.

Salida:

- transacciones_categorizadas.csv : transacciones con la categoría asignada.

```
"""
```

```
import pandas as pd
```

```
# 1. REGLAS DE CATEGORIZACIÓN
```

```
# Cada categoría se asocia a una lista de palabras clave que, de aparecer en el
```

```
# concepto del movimiento, permiten clasificarlo. El orden importa: se evalúan
```

```
# primero las categorías más específicas.
```

```

REGLAS = [
    ("Nóminas", ["NOMINA"]),
    ("Alquiler", ["ALQUILER"]),
    ("Suministros", ["SUMINISTRO", "TELEFONIA"]),
    ("Proveedores", ["PROVEEDOR"]),
    ("Impuestos", ["IMPUESTO", "AEAT", "IVA", "IRPF"]),
    ("Deuda", ["PRESTAMO", "CUOTA", "HIPOTECA"]),
    ("Comisiones", ["COMISION", "DESCUBIERTO"]),
    ("Ventas", ["VENTA", "FACTURA CLIENTE", "TPV", "ABONO"]),
]

```

```
def categorizar(concepto, importe):
```

```
    """Asigna una categoría a un movimiento a partir de su concepto y su signo.
```

Si ninguna regla encaja, se aplica un criterio de respaldo basado en el signo del importe: un movimiento positivo no identificado se trata como un

ingreso por ventas, y uno negativo, como un gasto no clasificado.
"""

```
texto = str(concepto).upper()
for categoria, claves in REGLAS:
    if any(clave in texto for clave in claves):
        return categoria
# Criterio de respaldo
return "Ventas" if importe > 0 else "Otros gastos"
```

2. APLICACIÓN Y VALIDACIÓN

```
if __name__ == "__main__":
    tx = pd.read_csv("transacciones.csv")

    # Aplicar el categorizador
    tx["categoria"] = [
        categorizar(c, i) for c, i in zip(tx["concepto"], tx["importe"])
    ]

    # --- Validación contra la verdad oculta ---
    aciertos = (tx["categoria"] == tx["categoria_real"]).mean()
    print(f"Acierto global del categorizador: {aciertos:.1%}\n")

    print("Matriz de confusión (real en filas, asignada en columnas):")
    matriz = pd.crosstab(tx["categoria_real"], tx["categoria"])
    print(matriz, "\n")

    # Guardar el resultado sin la columna de verdad oculta, ya que en un
    # entorno real esa columna no existiría.
    salida = tx.drop(columns=["categoria_real"])
    salida.to_csv("transacciones_categorizadas.csv", index=False)
    print("Guardado: transacciones_categorizadas.csv")
    print(f"Movimientos categorizados: {len(salida)}")
```

9.3. Generación de indicadores de comportamiento

"""

De transacciones a indicadores de comportamiento financiero.

A partir de los movimientos ya categorizados, este script calcula para cada pyme un conjunto reducido de indicadores que resumen su comportamiento financiero. Estos indicadores son los que, en el capítulo siguiente, se combinan para construir el perfil de solvencia.

Salida:

- indicadores.csv : una fila por pyme, con sus indicadores.

"""

```
import numpy as np
import pandas as pd
```

```
CATEGORIAS_GASTO = ["Nóminas", "Alquiler", "Suministros", "Proveedores",
                    "Impuestos", "Deuda", "Comisiones", "Otros gastos"]
```

```
def calcular_indicadores(tx, pymes):
    tx = tx.copy()
    tx["fecha"] = pd.to_datetime(tx["fecha"])
    tx["mes"] = tx["fecha"].dt.to_period("M")

    filas = []
    for pyme_id, g in tx.groupby("pyme_id", sort=False):

        ventas = g[g["categoria"] == "Ventas"]
        gastos = g[g["categoria"].isin(CATEGORIAS_GASTO)]

        # Serie mensual de ventas
        ventas_mes = ventas.groupby("mes")["importe"].sum()
        n_meses = ventas_mes.shape[0]

        ingresos_totales = ventas["importe"].sum()
        gastos_totales = -gastos["importe"].sum()           # en positivo

        # 1. Ingresos medios mensuales (escala del negocio)
        ingresos_medios = ventas_mes.mean()

        # 2. Volatilidad de los ingresos (coef. de variación)
        volatilidad = ventas_mes.std() / ventas_mes.mean() if
ventas_mes.mean() else np.nan

        # 3. Índice de estacionalidad (mes pico frente a mes valle)
        estacionalidad = ventas_mes.max() / ventas_mes.min() if
ventas_mes.min() > 0 else np.nan

        # 4. Ratio gastos / ingresos (sostenibilidad)
        ratio_gastos = gastos_totales / ingresos_totales if ingresos_totales
else np.nan

        # 5. Flujo de caja neto medio mensual
        flujo_netos_medio = (ingresos_totales - gastos_totales) / n_meses

        # 6. Días de tesorería (liquidez)
        saldo_medio = g["saldo_resultante"].mean()
```

```

gasto_diario = (gastos_totales / n_meses) / 30.0
dias_tesoreria = saldo_medio / gasto_diario if gasto_diario else
np.nan

# 7. Frecuencia de descubiertos (tensión de liquidez)
meses_descubierto = g[g["saldo_resultante"] < 0]["mes"].nunique()
frec_descubiertos = meses_descubierto / n_meses if n_meses else
np.nan

# 8. Ratio de servicio de la deuda
deuda_total = -g[g["categoria"] == "Deuda"]["importe"].sum()
ratio_deuda = deuda_total / ingresos_totales if ingresos_totales else
0.0

# 9. Peso de los gastos fijos
fijos = -g[g["categoria"].isin(["Nóminas", "Alquiler",
"Suministros"])]["importe"].sum()
peso_fijos = fijos / ingresos_totales if ingresos_totales else np.nan

filas.append({
    "pyme_id": pyme_id,
    "ingresos_medios_mensuales": round(ingresos_medios, 2),
    "volatilidad_ingresos": round(volatilidad, 3),
    "indice_estacionalidad": round(estacionalidad, 3),
    "ratio_gastos_ingresos": round(ratio_gastos, 3),
    "flujo_netto_medio_mensual": round(flujo_netto_medio, 2),
    "dias_tesoreria": round(dias_tesoreria, 1),
    "frecuencia_descubiertos": round(frec_descubiertos, 3),
    "ratio_servicio_deuda": round(ratio_deuda, 3),
    "peso_gastos_fijos": round(peso_fijos, 3),
})

ind = pd.DataFrame(filas)
# Adjuntar atributos de la pyme y la salud latente (solo para validación
posterior)
ind = ind.merge(
    pymes[["pyme_id", "sector", "n_empleados", "antiguedad_anios",
"salud_latente"]],
    on="pyme_id", how="left")
return ind

if __name__ == "__main__":
    tx = pd.read_csv("transacciones_categorizadas.csv")
    pymes = pd.read_csv("pymes.csv")

    ind = calcular_indicadores(tx, pymes)
    ind.to_csv("indicadores.csv", index=False)

```

```

print(f"Indicadores calculados para {len(ind)} pymes.\n")
print("Estadísticos descriptivos de los indicadores:")
cols = ["volatilidad_ingresos", "indice_estacionalidad",
"ratio_gastos_ingresos",
        "dias_tesoreria", "frecuencia_descubiertos",
"ratio_servicio_deuda"]
print(ind[cols].describe().round(2).to_string())

print("\nCorrelación de cada indicador con la salud financiera real
(oculta):")
num = ind.select_dtypes(include=[np.number])
corr = num.corr()["salud_latente"].drop("salud_latente").sort_values()
print(corr.round(3).to_string())

```

9.4. La capacidad de discriminación del score en términos de negocio

"""

Análisis de lift: capacidad de discriminación del score en términos de negocio.

La pregunta es: si el banco ordena los leads por su puntuación de solvencia y atiende primero a los mejor puntuados, ¿captura más "buenas pymes" que si los contactara al azar? Para responderla se define como buena pyme a la mitad con mejor salud financiera real (la variable latente, que no interviene en el score) y se mide cuántas de ellas se concentran en los tramos altos de puntuación.

Salida:

- lift_por_decil.csv : tasa de buenas pymes y lift por decil de puntuación.

- fig_lift.png : curva de ganancia acumulada.

"""

```

import numpy as np
import pandas as pd
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
from matplotlib import font_manager

reg = "/usr/share/fonts/truetype/liberation/LiberationSerif-Regular.ttf"
bld = "/usr/share/fonts/truetype/liberation/LiberationSerif-Bold.ttf"
fp = font_manager.FontProperties(fname=reg)
fpb = font_manager.FontProperties(fname=bld)
for f in [reg, bld]:
    font_manager.fontManager.addfont(f)
plt.rcParams["font.family"] = "Liberation Serif"
NAVY, TEAL = "#1F3A5F", "#0F6E56"

```

```

# Datos: puntuación + definición de "buena pyme"

d = pd.read_csv("diagnosticos.csv")[["pyme_id", "puntuacion"]]
salud = pd.read_csv("indicadores.csv")[["pyme_id", "salud_latente"]]
d = d.merge(salud, on="pyme_id")

umbral = d["salud_latente"].median()
d["buena"] = (d["salud_latente"] >= umbral).astype(int)
base = d["buena"].mean()

# Ordenar por puntuación descendente
d = d.sort_values("puntuacion", ascending=False).reset_index(drop=True)
n = len(d)
total_buenas = d["buena"].sum()

# Lift por decil

d["decil"] = pd.qcut(d["puntuacion"].rank(method="first", ascending=False),
                    10, labels=range(1, 11)).astype(int)
tabla = d.groupby("decil").agg(
    pymes=("buena", "size"),
    buenas=("buena", "sum"),
).reset_index()
tabla["tasa_buenas"] = tabla["buenas"] / tabla["pymes"]
tabla["lift"] = tabla["tasa_buenas"] / base
tabla.to_csv("lift_por_decil.csv", index=False)

# Ganancia acumulada

d["acum_buenas"] = d["buena"].cumsum()
x = np.arange(1, n + 1) / n * 100 # % pymes contactadas
y = d["acum_buenas"].values / total_buenas * 100 # % buenas capturadas

def capturado_en(pct):
    k = int(round(pct / 100 * n))
    return d["buena"].iloc[:k].sum() / total_buenas * 100

print(f"Tasa base de buenas pymes: {base:.0%}")
print(f"Lift del primer decil: {tabla.loc[0, 'lift']:.2f}")
print(f"Contactando el 20% mejor puntuado se captura el
{capturado_en(20):.0f}% de las buenas pymes.")
print(f"Contactando el 30% mejor puntuado se captura el
{capturado_en(30):.0f}% de las buenas pymes.")
print(f"Contactando el 50% mejor puntuado se captura el
{capturado_en(50):.0f}% de las buenas pymes.\n")
print("Lift por decil:")

```

```

print(tabla[["decil", "tasa_buenas",
"lift"]].round(2).to_string(index=False))

# Figura: curva de ganancia acumulada

fig, ax = plt.subplots(figsize=(8.2, 5.6), dpi=200)
ax.plot([0, 100], [0, 100], color="#999999", linestyle="--", linewidth=1.5,
        label="Contacto al azar")
ax.plot(np.concatenate([[0], x]), np.concatenate([[0], y]),
        color=NAVY, linewidth=2.4, label="Priorización por puntuación")

# Punto de referencia al 30%
cap30 = capturado_en(30)
ax.plot([30, 30], [0, cap30], color=TEAL, linestyle=":", linewidth=1.4)
ax.plot([0, 30], [cap30, cap30], color=TEAL, linestyle=":", linewidth=1.4)
ax.scatter([30], [cap30], color=TEAL, s=45, zorder=5)
ax.text(31, cap30 - 6, f"{cap30:.0f}% de las buenas pymes\nen el 30% mejor
puntuado",
        fontproperties=fp, fontsize=10.5, color=TEAL)

ax.set_xlabel("Pymes contactadas, ordenadas por puntuación (%)",
fontproperties=fp, fontsize=11.5)
ax.set_ylabel("Buenas pymes capturadas (%)", fontproperties=fp,
fontsize=11.5)
ax.set_title("Curva de ganancia acumulada del score",
fontproperties=fpb, fontsize=13, color=NAVY, pad=12)
ax.legend(prop=fp, fontsize=11, frameon=False, loc="lower right")
for lab in ax.get_xticklabels() + ax.get_yticklabels():
    lab.set_fontproperties(fp)
ax.set_xlim(0, 100); ax.set_ylim(0, 100)
ax.spines["top"].set_visible(False)
ax.spines["right"].set_visible(False)
ax.grid(color="#E5E8EC", linewidth=0.7)
plt.tight_layout()
plt.savefig("fig_lift.png", dpi=200, bbox_inches="tight", facecolor="white",
pad_inches=0.12)
print("\nGuardado fig_lift.png")

```

9.5. Cuantificación del valor para el banco frente a la recomendación sin análisis

"""

Cuantificación del valor para el banco frente a la recomendación sin análisis.

El modelo compara dos escenarios sobre una misma campaña comercial del banco:

- (A) Sin análisis: el banco trabaja leads no cualificados (orden aleatorio).
- (B) Con DataBridge: el banco trabaja leads cualificados y priorizados.

La diferencia clave es la TASA DE CONVERSIÓN de lead a operación cerrada, que es mayor en (B). Esa mejora no se inventa: se deriva de la concentración de buenas pymes medida en el análisis de lift del apartado anterior.

Parámetros económicos tomados del modelo de negocio (TFG de ADE) y supuestos explícitos, todos sometidos después a análisis de sensibilidad.

```
"""
```

```
import numpy as np
import pandas as pd
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
from matplotlib import font_manager

reg = "/usr/share/fonts/truetype/liberation/LiberationSerif-Regular.ttf"
bld = "/usr/share/fonts/truetype/liberation/LiberationSerif-Bold.ttf"
fp = font_manager.FontProperties(fname=reg)
fpb = font_manager.FontProperties(fname=bld)
for f in [reg, bld]:
    font_manager.fontManager.addfont(f)
plt.rcParams["font.family"] = "Liberation Serif"
NAVY, TEAL, OCRE = "#1F3A5F", "#0F6E56", "#B07A2E"

# PARÁMETROS

# Precios DataBridge (TFG de ADE)
TICKET = 80_000          # ticket medio de operación (€)
COM_EXITO_PCT = 0.003   # comisión a éxito (0,3 % del volumen)
COM_EXITO = TICKET * COM_EXITO_PCT  # = 240 € por operación cerrada

# Conversión: con leads cualificados, 25 % (TFG de ADE)
Q1 = 0.25              # conversión con DataBridge
# Sin análisis se deriva del lift: la proporción de buenas pymes baja del
# valor priorizado al aleatorio. Con buenas pymes convirtiendo a c_g y malas
# a c_b, y exigiendo que la mezcla priorizada (70 % buenas) dé el 25 %:
P1, P0 = 0.70, 0.50   # proporción de buenas pymes (priorizado vs
aleatorio)
C_B = 0.10            # conversión de una pyme no buena (supuesto)
C_G = (Q1 - (1 - P1) * C_B) / P1          # conversión de una buena pyme
(derivada)
Q0 = P0 * C_G + (1 - P0) * C_B          # conversión sin análisis
(derivada)

# Margen neto del banco por operación cerrada (supuesto): valor actual del
# margen financiero neto sobre una operación de 80.000 € a lo largo de su
vida.
```

```

MARGEN_OP = 4_000

LEADS = 1_000          # tamaño de la campaña

def escenario(q, con_databridge):
    ops = LEADS * q
    margen = ops * MARGEN_OP
    coste_comision = ops * COM_EXITO if con_databridge else 0.0
    return {"conversion": q, "operaciones": ops,
            "margen_bruto": margen, "comision_databridge": coste_comision,
            "margen_netto_de_comision": margen - coste_comision}

if __name__ == "__main__":
    print(f"Conversión derivada – sin análisis: {Q0:.1%} | con DataBridge:
    {Q1:.1%}")
    print(f"(buena pyme convierte {C_G:.1%}, no buena {C_B:.1%}; comisión a
    éxito {COM_EXITO:.0f} €)\n")

    a = escenario(Q0, con_databridge=False)
    b = escenario(Q1, con_databridge=True)

    print(f"{'':32}{'Sin análisis':>16}{'Con DataBridge':>18}")
    for k, etq in [("operaciones", "Operaciones cerradas"),
                  ("margen_bruto", "Margen bruto del banco (€)"),
                  ("comision_databridge", "Comisión a DataBridge (€)"),
                  ("margen_netto_de_comision", "Margen neto del banco (€)")]:
        print(f"{etq:32}{a[k]:>16,.0f}{b[k]:>18,.0f}")

    valor_total = b["margen_netto_de_comision"] - a["margen_netto_de_comision"]
    valor_lead = valor_total / LEADS
    op_extra = b["operaciones"] - a["operaciones"]
    valor_bruto = op_extra * MARGEN_OP
    print(f"\nOperaciones adicionales: {op_extra:.0f}")
    print(f"Valor bruto generado (margen de las operaciones extra):
    {valor_bruto:,.0f} €")
    print(f"Comisión que cobra DataBridge: {b['comision_databridge']:,.0f}
    €")
    print(f"VALOR NETO creado para el banco: {valor_total:,.0f} € por {LEADS}
    leads "
          f"({valor_lead:,.0f} € por lead)")
    print(f"La comisión de DataBridge equivale al "
          f"{b['comision_databridge']/valor_bruto:.0%} del valor bruto que
    genera.\n")

    # -----
    #
    # SENSIBILIDAD (tornado): valor neto por lead frente a 3 supuestos clave

```

```

# -----
#
def valor_por_lead(margen=MARGEN_OP, q0=Q0, q1=Q1, com=COM_EXITO):
    return (q1 - q0) * margen - q1 * com

base = valor_por_lead()
variables = [
    ("Margen neto por operación (€)", "margen",
     valor_por_lead(margen=2500), valor_por_lead(margen=6000), "2.500 /
6.000"),
    ("Conversión sin análisis", "q0",
     valor_por_lead(q0=0.23), valor_por_lead(q0=0.18), "23% / 18%"),
    ("Ticket medio (→ comisión)", "com",
     valor_por_lead(com=60_000*COM_EXITO_PCT),
valor_por_lead(com=100_000*COM_EXITO_PCT),
     "60k / 100k €"),
]

fig, ax = plt.subplots(figsize=(9.0, 4.4), dpi=200)
y = np.arange(len(variables))
for i, (nombre, _, lo, hi, etq) in enumerate(variables):
    izq, der = min(lo, hi), max(lo, hi)
    ax.barh(i, der - izq, left=izq, height=0.55, color=TEAL, alpha=0.85,
zorder=3)
    ax.text(izq - 4, i, f"{izq:.0f}", va="center", ha="right",
           fontproperties=fp, fontsize=10, color="#333333")
    ax.text(der + 4, i, f"{der:.0f}", va="center", ha="left",
           fontproperties=fp, fontsize=10, color="#333333")
ax.axvline(base, color=NAVY, linewidth=1.8, linestyle="--", zorder=4)
ax.set_yticks(y)
ax.set_yticklabels([v[0] for v in variables], fontproperties=fp,
fontsize=11)
ax.set_xlabel("Valor neto creado para el banco (€ por lead)",
fontproperties=fp, fontsize=11.5)
ax.set_title("Sensibilidad del valor a los supuestos clave",
             fontproperties=fpb, fontsize=13, color=NAVY, pad=12)
for lab in ax.get_xticklabels():
    lab.set_fontproperties(fp)
ax.spines["top"].set_visible(False)
ax.spines["right"].set_visible(False)
ax.grid(axis="x", color="#E5E8EC", linewidth=0.7, zorder=0)
ax.set_xlim(0, 245)
ax.set_ylim(-0.75, len(variables) - 0.25)
ax.text(base, -0.62, f"caso base: {base:.0f} \u20ac / lead", ha="center",
va="center",
        fontproperties=fpb, fontsize=10, color=NAVY,
        bbox=dict(boxstyle="round,pad=0.3", facecolor="white",
edgecolor=NAVY, linewidth=1))
plt.tight_layout()

```

```
plt.savefig("fig_sensibilidad.png", dpi=200, bbox_inches="tight",  
           facecolor="white", pad_inches=0.12)  
print("guardado fig_sensibilidad.png")
```

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, Gonzalo Leopoldo Puig de Marcos, estudiante del Grado en Administración de Empresas y Grado en Análisis de Negocio (E2-Analytics) de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "DESARROLLO DEL MODELO ANALÍTICO DE LA STARTUP: DATABRIDGE ", 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. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

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: 17/06/2026

Firma: ____ G.L.P. _____