FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

Grado en Business Analytics

Clusterización financiera de PYMEs: el caso DataBridge

Autor: Luis Ussía Arocena

Director: Carlos Bellón Nuñez-Mera

MADRID | Junio 2026

RESUMEN EJECUTIVO

Las PYMEs españolas generan información financiera relevante a través de sus movimientos bancarios, cobros, pagos y relaciones con clientes y proveedores. Sin embargo, estos datos suelen estar dispersos y poco explotados, lo que dificulta tanto el diagnóstico financiero de la empresa como la preparación de solicitudes de financiación comparables para bancos y otros financiadores.

Este Trabajo Fin de Grado analiza cómo la clusterización, como técnica de aprendizaje no supervisado, puede transformar movimientos financieros categorizados en perfiles accionables de PYMEs. El trabajo se enmarca en el proyecto DataBridge, una plataforma orientada a mejorar la preparación financiera de las empresas y facilitar su conexión preliminar con financiadores, a partir de datos proporcionados por Asfin.

La parte empírica se basa en una muestra de 89 empresas y 483,919 movimientos financieros categorizados. A partir de esta información se construyen variables agregadas por empresa, como ingresos medios, salidas, flujo neto, volatilidad, meses negativos, presión de caja, concentración comercial y peso de gastos financieros. Posteriormente, se aplica un modelo K-Means, que identifica 6 clusters finales con un coeficiente de silueta de 0.251.

Los resultados permiten clasificar las empresas en perfiles financieros diferenciados, como operativa estable, ingresos volátiles, concentración comercial elevada u operativa de alto volumen. El trabajo concluye que la clusterización no sustituye a un sistema de scoring ni a un rating crediticio, pero sí aporta valor como herramienta exploratoria para ordenar información financiera, generar un diagnóstico preliminar y apoyar un matching no vinculante entre PYMEs y financiadores potencialmente compatibles.

PALABRAS CLAVE

PYMEs, fintech, clusterización, K-Means, aprendizaje no supervisado, inteligencia financiera, análisis financiero, movimientos bancarios, financiación empresarial, DataBridge.

ABSTRACT

Spanish SMEs generate relevant financial information through bank transactions, collections, payments and relationships with customers and suppliers. However, this data is often fragmented and underused, making it difficult to assess the company's financial position and to prepare comparable financing applications for banks and other lenders.

This Bachelor's Thesis analyses how clustering, as an unsupervised learning technique, can transform categorised financial transactions into actionable SME profiles. The project is framed within DataBridge, a platform designed to improve the financial preparation of SMEs and support their preliminary connection with financing providers, using data provided by Asfin.

The empirical analysis is based on a sample of 89 companies and 483,919 categorised financial transactions. From this data, company-level variables are constructed, including average inflows, outflows, net cash flow, volatility, negative months, cash pressure, commercial concentration and the weight of financial expenses. A K-Means model is then applied, identifying 6 final clusters with a silhouette coefficient of 0.251.

The results make it possible to classify companies into differentiated financial profiles, such as stable operating businesses, volatile income profiles, companies with high commercial concentration and high-volume operators. The study concludes that clustering does not replace credit scoring or a formal rating system, but it provides value as an exploratory tool to organise financial information, support a preliminary diagnosis and facilitate non-binding matching between SMEs and potentially compatible financing providers.

KEY WORDS

SMEs, fintech, clustering, K-Means, unsupervised learning, financial intelligence, financial analysis, bank transactions, business financing, DataBridge.

ÍNDICE

1. INTRODUCCIÓN

1.1. Justificación del interés del tema

Las pequeñas y medianas empresas generan diariamente una gran cantidad de información financiera a través de sus movimientos bancarios, cobros, pagos, facturas, datos contables y relaciones con clientes y proveedores. Sin embargo, en muchas ocasiones esta información permanece dispersa y no se transforma en conocimiento útil para la toma de decisiones. En el contexto de la financiación empresarial, esta situación resulta especialmente relevante: una PYME puede tener datos suficientes para explicar su actividad, pero no siempre cuenta con las herramientas necesarias para convertirlos en un diagnóstico financiero claro, comparable y comprensible.

Esta limitación afecta tanto a las propias empresas como a los financiadores. Para la PYME, la falta de información estructurada dificulta entender su situación financiera real, anticipar necesidades de liquidez o preparar una solicitud de financiación sólida. Para bancos, fintech lenders u otros financiadores, recibir información incompleta o poco comparable incrementa el esfuerzo de análisis y reduce la calidad del proceso de originación. Por tanto, el problema no está únicamente en el acceso a financiación, sino también en la forma en la que los datos financieros de la empresa se preparan, interpretan y presentan.

En este contexto, el uso de técnicas de análisis de datos puede aportar valor. La clusterización, como técnica de aprendizaje no supervisado, permite agrupar empresas con comportamientos financieros similares sin necesidad de contar con una variable objetivo previa. Esto resulta útil cuando no se pretende predecir un impago ni emitir un rating crediticio, sino identificar tipologías de empresas a partir de sus patrones financieros. En el caso de DataBridge, la clusterización permite transformar movimientos categorizados en perfiles de PYMEs, como empresas de operativa estable, negocios con ingresos volátiles o compañías con concentración comercial relevante.

El interés del trabajo se sitúa, por tanto, en la intersección entre fintech, análisis de datos y financiación empresarial. DataBridge se plantea como una plataforma capaz de convertir información financiera dispersa en perfiles y dossiers útiles para una conexión preliminar entre PYMEs y financiadores. Este TFG no busca desarrollar una entidad

financiera ni un modelo de decisión crediticia, sino analizar cómo el tratamiento de datos puede hacer más operativa una propuesta de negocio basada en inteligencia financiera.

El trabajo cuenta, además, con el apoyo de datos proporcionados por Asfin. Asfin es una consultora especializada en financiación empresarial para PYMEs, que acompaña a empresas en el análisis, preparación y búsqueda de soluciones de financiación. En el marco de este TFG, Asfin actúa como socio sectorial y fuente de datos, aportando una base de empresas y movimientos financieros categorizados que permite desarrollar un caso práctico aplicado sobre información real del entorno PYME.

A partir de estos datos, el trabajo construye variables agregadas por compañía y aplica un modelo de clusterización para identificar perfiles financieros diferenciados. De esta forma, el análisis no se limita a una reflexión conceptual, sino que muestra cómo los datos financieros pueden convertirse en una herramienta concreta para apoyar el modelo de DataBridge. La aportación principal del trabajo reside precisamente en esa conexión: transformar una base transaccional en una tipología interpretable de empresas que pueda servir como punto de partida para diagnósticos preliminares, dossiers financieros y matching no vinculante con financiadores.

1.2. Objetivos del trabajo

El objetivo general de este trabajo consiste en aplicar técnicas de análisis de datos, especialmente clusterización, a una base de movimientos financieros categorizados de PYMEs con el fin de identificar tipologías de empresas y valorar su utilidad dentro del modelo DataBridge.

Este objetivo general se concreta en los siguientes objetivos específicos:

1. Transformar la base de movimientos financieros proporcionada por Asfin en una base agregada por empresa, mediante la limpieza de datos y la construcción de variables financieras comparables.
2. Aplicar un modelo de clusterización para identificar tipologías de PYMEs en función de su comportamiento financiero e interpretar los grupos resultantes desde una perspectiva empresarial.
3. Valorar la utilidad de estas tipologías dentro del modelo de negocio de DataBridge, identificando además las principales limitaciones del análisis.

1.3. Metodología

La metodología del trabajo combina revisión conceptual y análisis cuantitativo de datos. En primer lugar, se desarrolla un marco teórico sobre la clusterización como técnica de aprendizaje no supervisado, explicando su utilidad para identificar grupos de observaciones con características similares sin necesidad de contar con una variable objetivo previa. Esta aproximación resulta adecuada para el caso DataBridge, ya que el objetivo no es construir un modelo de scoring ni predecir impagos, sino identificar tipologías de PYMEs a partir de sus patrones financieros.

La parte empírica del trabajo se basa en dos bases de datos proporcionadas por Asfin: una base de empresas y una base de movimientos financieros categorizados. Asfin actúa como socio sectorial y fuente de datos del proyecto, aportando información real del entorno PYME que permite desarrollar un caso práctico aplicado. La unidad final de análisis es la empresa, no el movimiento individual, por lo que el primer paso consiste en transformar una base transaccional en una base agregada por compañía.

El tratamiento de datos se realiza en Python y sigue varias fases. En primer lugar, se limpian y preparan los datos, homogeneizando identificadores, fechas e importes. Posteriormente, los movimientos se clasifican como entradas o salidas de caja y se agregan por empresa. A partir de esta base agregada se construyen variables financieras comparables, como ingresos medios, salidas medias, flujo neto, volatilidad, presión de caja, concentración de contrapartes y peso de determinadas categorías de gasto.

Una vez construida la base analítica, se aplica un modelo de clusterización no supervisada mediante K-Means. Antes de ejecutar el modelo, las variables se escalan para evitar que las magnitudes absolutas condicionen artificialmente la distancia entre empresas. La selección del número de clusters se apoya en métricas habituales de validación, como la inercia y el coeficiente de silueta.

Finalmente, los clusters obtenidos se interpretan desde una perspectiva financiera y empresarial. La finalidad del análisis no es emitir un rating, recomendar financiación ni automatizar decisiones crediticias, sino generar perfiles orientativos de PYMEs que puedan apoyar el diagnóstico preliminar y el modelo de DataBridge. Los detalles técnicos del tratamiento de datos, las variables construidas, las métricas obtenidas y los resultados del modelo se desarrollan en el capítulo 4.

1.4. Estructura del trabajo

El trabajo se organiza en cinco capítulos, seguidos de la bibliografía. La estructura se resume en la siguiente tabla:

Ilustración 1. Estructura del trabajo por capítulos

Capítulo	Contenido principal	Finalidad dentro del trabajo
Capítulo 1. Introducción	Presenta la justificación del tema, los objetivos, la metodología y la estructura del trabajo.	Situar al lector y explicar el enfoque general del TFG.
Capítulo 2. Clusterización como herramienta analítica	Explica qué es la clusterización, su relación con el aprendizaje no supervisado, sus algoritmos principales y sus criterios de validación.	Construir el marco metodológico necesario para entender el análisis posterior.
Capítulo 3. Contexto: emprendimiento y fintech	Sitúa DataBridge dentro del ecosistema fintech, presenta referencias de uso intensivo de datos y resume el problema, la oportunidad y el modelo de negocio.	Conectar el análisis de datos con la oportunidad empresarial de DataBridge.
Capítulo 4. Análisis del cliente, tipología y clusterización	Describe las bases de datos de Asfin, el tratamiento de datos, la construcción de variables, la aplicación de K-Means y la interpretación de los clusters.	Desarrollar la parte empírica del trabajo y mostrar los resultados del análisis.
Capítulo 5. Conclusiones, limitaciones y futuras líneas de trabajo	Resume las principales conclusiones, identifica limitaciones y propone mejoras futuras.	Valorar la utilidad de la clusterización para DataBridge y delimitar el alcance del análisis.
Capítulo 6. Bibliografía	Recoge las fuentes académicas, técnicas e institucionales utilizadas.	Dar soporte documental al trabajo y permitir la trazabilidad de las fuentes.

Fuente: Elaboración propia.

La estructura del trabajo avanza desde el marco conceptual hasta la aplicación práctica. Primero se explica la clusterización como herramienta analítica; después se contextualiza su utilidad dentro del sector fintech; y finalmente se aplica sobre datos proporcionados por Asfin para construir perfiles de PYMEs. De esta forma, el trabajo conecta la teoría, el contexto empresarial y el análisis empírico en torno a una misma idea: demostrar cómo el tratamiento de datos puede hacer más operativa la propuesta de DataBridge.

2. CLUSTERIZACIÓN COMO HERRAMIENTA ANALÍTICA

La clusterización, o análisis de conglomerados, es una técnica de aprendizaje no supervisado cuyo objetivo es agrupar observaciones en función de su similitud. A diferencia de los modelos supervisados, en los que existe una variable objetivo previamente definida, la clusterización trabaja con datos no etiquetados y busca identificar estructuras internas dentro del conjunto de datos. En términos prácticos, permite descubrir grupos de empresas, clientes u operaciones que comparten patrones comunes sin imponer una clasificación previa. Jain, Murty y Flynn (1999) definen el análisis de clusters como un proceso orientado a organizar un conjunto de objetos en grupos cuyos miembros presentan una mayor similitud entre sí que respecto a los objetos de otros grupos. Esta idea resulta especialmente útil en contextos exploratorios, donde el objetivo no es predecir una variable concreta, sino identificar patrones latentes en los datos.

En el contexto de este trabajo, la clusterización resulta especialmente útil porque el objetivo no es predecir si una empresa va a incumplir una obligación financiera ni emitir una calificación crediticia. El objetivo es identificar tipologías de PYMEs a partir de sus datos financieros y transaccionales, con el fin de entender mejor sus patrones de comportamiento, sus necesidades potenciales de financiación y su posible encaje con distintos financiadores. Por tanto, la clusterización se plantea como una herramienta exploratoria y descriptiva, no como un sistema automático de decisión.

2.1. Concepto de clusterización y aprendizaje no supervisado

La clusterización forma parte del aprendizaje no supervisado, una rama del aprendizaje automático en la que el algoritmo no recibe una respuesta correcta previamente definida. En lugar de aprender a predecir una etiqueta conocida, el modelo analiza las relaciones entre observaciones y agrupa aquellas que presentan características similares. En un análisis de empresas, estas características pueden ser variables como ingresos, gastos, estabilidad de flujos de caja, concentración de clientes, frecuencia de movimientos bancarios, peso de gastos financieros o intensidad de inversión.

La lógica central de la clusterización es que las observaciones dentro de un mismo grupo sean lo más parecidas posible entre sí y, al mismo tiempo, lo más diferentes posible respecto a las observaciones de otros grupos. En la práctica, esto exige definir primero

qué significa “parecido”. Para ello se utilizan medidas de distancia o similitud, como la distancia euclídea, la distancia Manhattan o medidas basadas en correlación, dependiendo de la naturaleza de las variables utilizadas.

En el caso de DataBridge, esta metodología permite pasar de una base de datos formada por movimientos bancarios categorizados a una lectura más útil desde el punto de vista empresarial. En lugar de analizar empresa por empresa de forma aislada, la clusterización permite identificar patrones comunes: empresas con ingresos estables pero tensión de circulante, empresas con elevada concentración de clientes, empresas con señales de inversión en inmovilizado, empresas con gastos financieros relevantes o empresas con baja calidad de información. Esta agrupación puede ayudar a construir una segmentación más accionable para la plataforma.

2.2. Utilidad de la clusterización en el análisis empresarial y financiero

La principal utilidad de la clusterización en un contexto empresarial es que permite segmentar una población heterogénea en grupos más interpretables. En el caso de las PYMEs, esta utilidad es especialmente relevante porque hablar de “PYMEs” como un único bloque puede resultar poco preciso. Una empresa industrial con necesidades de inversión en maquinaria, una compañía de servicios con ingresos recurrentes y una empresa comercial con fuerte necesidad de circulante pueden pertenecer todas al mismo universo PYME, pero presentar necesidades financieras muy diferentes.

Desde una perspectiva financiera, la clusterización permite identificar perfiles de comportamiento sin depender de una etiqueta previa de riesgo, impago o aprobación bancaria. Esto es importante para este trabajo porque la base de datos disponible no parte necesariamente de una variable objetivo supervisada, como “financiación concedida”, “financiación rechazada” o “default”. En ausencia de esa etiqueta, construir un modelo predictivo sería metodológicamente débil. La clusterización, en cambio, sí permite extraer valor de los datos disponibles mediante la identificación de patrones.

Aplicada al caso de DataBridge, la clusterización puede cumplir tres funciones. En primer lugar, ayuda a comprender qué tipos de PYMEs aparecen en la base de datos facilitada por Asfin. En segundo lugar, permite diseñar perfiles financieros que puedan relacionarse con distintas necesidades de financiación. En tercer lugar, puede servir como base para

un sistema de matching preliminar, en el que cada grupo de empresas se asocie a determinados productos o financiadores de forma no vinculante.

2.3. Proceso metodológico de un análisis de clusterización

Un análisis de clusterización requiere una serie de decisiones metodológicas previas que condicionan la calidad y utilidad de los resultados. No basta con ejecutar un algoritmo sobre una base de datos: antes es necesario definir qué observaciones se van a agrupar, qué variables se utilizarán para medir la similitud entre ellas y cómo se tratarán los datos antes de aplicar el modelo.

La primera fase consiste en seleccionar la unidad de análisis y las variables relevantes. En función del objetivo del estudio, las observaciones pueden ser clientes, empresas, operaciones o transacciones. Las variables elegidas deben reflejar los rasgos que se consideran importantes para comparar dichas observaciones. Una mala selección de variables puede generar clusters poco interpretables o dominados por dimensiones que no son relevantes para el problema analizado.

La segunda fase es la preparación de los datos. Esta etapa incluye la depuración de errores, el tratamiento de valores nulos, la eliminación de duplicados y la transformación de variables para trabajar con formatos homogéneos. La calidad de esta fase es especialmente importante porque los algoritmos de clusterización son sensibles a errores de entrada: si los datos están mal estructurados, los grupos obtenidos pueden responder más a problemas de calidad del dato que a patrones reales.

La tercera fase consiste en normalizar o escalar las variables. Muchos algoritmos de clusterización se basan en medidas de distancia, por lo que las variables con mayor escala numérica pueden tener un peso excesivo en la formación de los grupos. Por ejemplo, una variable expresada en euros puede dominar artificialmente a una variable expresada en porcentaje si ambas no se transforman previamente. Por ello, el escalado permite que las variables sean comparables y que ninguna magnitud condicione por sí sola el resultado.

La cuarta fase es la elección del algoritmo y del número de clusters. Esta decisión depende del tamaño de la muestra, la naturaleza de las variables, la presencia de valores extremos y el grado de interpretabilidad buscado. En análisis empresariales, suele ser preferible utilizar modelos que permitan explicar los grupos de forma clara, ya que el objetivo no

es solo obtener una partición estadística, sino construir perfiles que puedan tener sentido desde el punto de vista económico o de negocio.

Por último, la fase final consiste en validar e interpretar los clusters obtenidos. La validación puede apoyarse en métricas cuantitativas, como la inercia o el coeficiente de silueta, pero también exige una interpretación cualitativa. Un cluster solo resulta útil si puede explicarse de forma comprensible y si aporta información relevante para el objetivo del análisis. Por tanto, la calidad de una clusterización no depende únicamente de su resultado estadístico, sino también de su capacidad para generar categorías interpretables y accionables.

En este capítulo se presenta el proceso metodológico en términos generales. La aplicación concreta de estas fases al caso DataBridge, incluyendo las bases utilizadas, el tratamiento de datos, las variables construidas y los resultados obtenidos, se desarrolla posteriormente en el capítulo 4.

2.4. Principales algoritmos de clusterización

Existen distintos algoritmos de clusterización, cada uno con ventajas y limitaciones. La elección del método debe responder al objetivo del análisis y a las características de los datos. En este trabajo, los algoritmos más relevantes serían K-Means, clustering jerárquico y DBSCAN, aunque el análisis final puede centrarse en uno o varios de ellos según la calidad de los resultados obtenidos.

El algoritmo K-Means es uno de los métodos más utilizados por su sencillez e interpretabilidad. Su objetivo es dividir las observaciones en un número predefinido de grupos, minimizando la distancia entre cada observación y el centroide de su cluster. La documentación de scikit-learn explica que K-Means separa las muestras en grupos de varianza similar y minimiza la suma de cuadrados intra-cluster, también llamada inercia (scikit-learn, 2024).

De forma simplificada, K-Means busca minimizar la siguiente función:

$$\sum_{i=1}^{n} \min_{\mu_j \in C} ||x_i - \mu_j||^2$$

El clustering jerárquico permite construir una estructura de grupos en forma de árbol o dendrograma. En lugar de fijar desde el inicio una única partición, permite observar cómo se van agrupando las observaciones a distintos niveles de similitud. Esta técnica puede resultar útil en una fase exploratoria, ya que permite entender si existen grupos naturales dentro de la muestra antes de fijar una segmentación final.

El algoritmo DBSCAN, propuesto por Ester, Kriegel, Sander y Xu (1996), agrupa observaciones en función de la densidad de puntos y permite identificar observaciones atípicas o ruido. A diferencia de K-Means, no exige especificar previamente el número de clusters, ya que los grupos se forman a partir de regiones donde existe una densidad suficiente de observaciones. Para ello, el algoritmo utiliza dos parámetros principales: el radio de vecindad y el número mínimo de puntos necesarios para considerar que una zona es densa. Su principal ventaja es que puede detectar clusters con formas no necesariamente esféricas y separar observaciones aisladas como ruido, lo que lo convierte en una alternativa útil cuando los datos presentan grupos irregulares o valores atípicos (Ester et al., 1996).

Para el caso de DataBridge, K-Means puede ser una primera opción razonable por su claridad y facilidad de interpretación. No obstante, conviene contrastar sus resultados con técnicas alternativas, especialmente si los datos muestran outliers o grupos de tamaño muy desigual. En este trabajo, dado que la muestra presenta empresas con tamaños y volúmenes de movimiento muy distintos, el criterio final no se limita a la métrica estadística: se prioriza una solución que pueda explicarse con claridad y traducirse en perfiles de PYMEs accionables para DataBridge.

2.5. Validación e interpretación de los clusters

La validación de los clusters es una fase esencial, ya que cualquier algoritmo de clusterización puede generar grupos incluso cuando la estructura real de los datos es débil. Por ello, no basta con obtener una segmentación: es necesario evaluar si los grupos son consistentes, interpretables y útiles para el objetivo del trabajo.

Una primera herramienta de validación es el **método del codo**, que analiza cómo cambia la inercia al aumentar el número de clusters. La idea es identificar un punto a partir del cual añadir más grupos reduce poco la variabilidad interna. Ese punto puede sugerir un número razonable de clusters, aunque no debe interpretarse de forma mecánica.

Otra herramienta habitual es el **coeficiente de silueta**, propuesto por Rousseeuw. Este indicador mide hasta qué punto una observación está bien asignada a su cluster comparando su distancia media respecto a los miembros de su propio grupo con la distancia respecto al cluster más cercano. Su valor se mueve entre -1 y 1: valores próximos a 1 indican una asignación más clara, valores cercanos a 0 indican observaciones fronterizas y valores negativos sugieren posible mala asignación (Rousseeuw, 1987).

Además de estas métricas, en este trabajo será especialmente importante la validación interpretativa. Un cluster será útil si puede describirse con lógica financiera y si contribuye a transformar los resultados del modelo en perfiles empresariales comprensibles, útiles para el diagnóstico preliminar de DataBridge.. Por ejemplo, un grupo de empresas con ingresos estables pero fuerte presión de tesorería podría asociarse a necesidades de circulante; un grupo con inversión recurrente en inmovilizado podría tener mayor encaje con leasing o financiación de inversión; y un grupo con alta concentración de clientes podría resultar relevante para productos como factoring.

Por tanto, la validación debe combinar criterios cuantitativos y cualitativos. Los indicadores estadísticos ayudan a seleccionar una solución razonable, pero la utilidad final depende de que los grupos tengan sentido económico y puedan integrarse en la lógica de negocio de DataBridge

2.6. Aplicación al caso DataBridge y a la base de datos de Asfin

En este trabajo, la clusterización se utilizará para analizar la base de datos facilitada por Asfin, que contiene información de empresas y movimientos categorizados. El objetivo será transformar estos datos en variables financieras agregadas y, posteriormente, agrupar las empresas en perfiles homogéneos. La unidad de análisis será la empresa, no cada movimiento individual.

El proceso partirá de la depuración de los movimientos, su agrupación por empresa y periodo, y la construcción de indicadores financieros. A partir de ahí, se aplicará la clusterización para identificar patrones de comportamiento entre empresas. Los clusters resultantes podrán interpretarse como tipologías de PYMEs con características financieras diferenciadas.

La utilidad de este análisis dentro del proyecto DataBridge es directa. En primer lugar, permite conocer mejor qué perfiles aparecen en la base de datos real o cedida. En segundo lugar, ayuda a traducir datos transaccionales en categorías comprensibles para el negocio. En tercer lugar, puede servir como base para generar un sistema de preclasificación y matching preliminar con financiadores, siempre manteniendo el carácter no vinculante del análisis.

La clusterización no debe entenderse como un modelo de riesgo crediticio ni como una herramienta de decisión automática. Su función en este TFG es exploratoria: identificar grupos, describir sus características y valorar cómo esa segmentación podría mejorar la propuesta de DataBridge. Esta delimitación resulta especialmente importante porque el proyecto DataBridge se ha desarrollado en el marco de Comillas Emprende, programa universitario orientado al impulso de iniciativas emprendedoras. Dentro de este proyecto, el estudio de la potencial startup se ha dividido en dos trabajos complementarios: un TFG de ADE, centrado en la oportunidad de negocio, la viabilidad empresarial y el plan estratégico de DataBridge; y el presente TFG de Business Analytics, centrado en el tratamiento de datos, la construcción de variables y la aplicación de clusterización sobre información financiera de PYMEs. De esta forma, este trabajo no pretende repetir el plan de negocio completo, sino profundizar en la capa analítica que puede hacer más operativa la propuesta de DataBridge.

3. CONTEXTO: EMPRENDIMIENTO Y FINTECH

Este capítulo sitúa a DataBridge dentro del ecosistema fintech y de la financiación empresarial. Se parte de una definición del concepto de fintech apoyada en fuentes institucionales, se analizan las tendencias tecnológicas que han hecho viable este tipo de soluciones y se revisan casos reales en los que el uso intensivo de datos y la segmentación de clientes han transformado servicios financieros. A continuación, se identifica el problema que aborda DataBridge —la dispersión de la información financiera de la PYME y las dificultades de acceso a financiación— y se cuantifica la oportunidad mediante un enfoque TAM/SAM/SOM, utilizado en metodologías de emprendimiento para distinguir entre mercado total, mercado accesible y mercado inicialmente capturable (Blank & Dorf, 2012). El capítulo se cierra con un modelo de negocio resumido en formato Business Model Canvas, herramienta propuesta por Osterwalder y Pigneur (2010) para representar de forma estructurada los principales bloques de un modelo de negocio, y se ilustra con el caso de Asfin, socio actual del proyecto y referencia metodológica en el ámbito de la consultoría financiera para PYMEs.

3.1. Qué es fintech

3.1.1. Concepto y delimitación

El término fintech procede de la unión de las palabras finance y technology y designa, en sentido amplio, a las soluciones tecnológicas aplicadas al diseño, prestación o mejora de servicios financieros. El Financial Stability Board (FSB, 2017) define fintech como la innovación financiera basada en tecnología capaz de generar nuevos modelos de negocio, aplicaciones, procesos o productos con efectos materiales sobre los mercados financieros, las instituciones o la prestación de servicios financieros. Esta definición es deliberadamente amplia: incluye tanto a entidades reguladas que digitalizan su operativa como a empresas tecnológicas no bancarias que intervienen en alguna fase de la cadena de valor financiera.

La delimitación es relevante para DataBridge porque el proyecto no aspira a convertirse en entidad de crédito ni a conceder financiación directamente. Su propuesta se sitúa en una fase previa: organizar información financiera de PYMEs, analizar patrones de comportamiento y facilitar una conexión preliminar con financiadores. Se encuadra, por tanto, dentro del ecosistema fintech como solución B2B de analítica financiera y apoyo a

la originación, próxima a las categorías de software financiero, infraestructura de open banking y soluciones de inteligencia de datos identificadas por el Banco de España (Sánchez y Quintanero, 2022).

3.1.2. Digitalización, open banking y dato financiero

El desarrollo del sector fintech responde a tres factores tecnológicos convergentes. El primero es la digitalización generalizada de la relación entre particulares, empresas y entidades financieras, que ha desplazado buena parte de la operativa hacia canales digitales. El segundo es la madurez de tecnologías como la computación en la nube, el big data, las APIs, los modelos analíticos y la inteligencia artificial, que permiten construir servicios financieros más ágiles, personalizados y escalables (Sánchez y Quintanero, 2022). El tercero es el desarrollo regulatorio del open banking en Europa a partir de la directiva PSD2 y, en una fase posterior, del paquete legislativo de acceso a datos financieros (open finance) propuesto por la Comisión Europea (Comisión Europea, 2023), que extiende el principio de portabilidad de datos más allá de las cuentas de pago.

En este marco, el dato financiero deja de ser un subproducto administrativo y se convierte en un activo. La información contable y transaccional de una empresa contiene señales sobre liquidez, estacionalidad, recurrencia de ingresos, concentración de clientes o capacidad de inversión. La oportunidad no reside únicamente en disponer de esos datos, sino en convertirlos en información accionable mediante procesos de categorización, modelización y segmentación. Esta tesis ha sido reiterada por organismos como la OCDE, que vincula la transformación digital del sistema financiero al uso intensivo de datos y a la capacidad de las pymes para participar en él en condiciones competitivas (OECD, 2020).

3.1.3. Fintech aplicada a la financiación de empresas

Durante la primera etapa de expansión del sector, la innovación fintech estuvo especialmente asociada a servicios orientados al consumidor final, como pagos digitales, banca móvil, inversión minorista y herramientas de finanzas personales. Esta evolución se explica por la digitalización de la relación entre clientes y entidades financieras y por la aparición de nuevos modelos de prestación de servicios financieros apoyados en tecnología (Arner, Barberis & Buckley, 2016; Financial Stability Board, 2017). Sin embargo, en los últimos años el foco se ha desplazado progresivamente hacia soluciones

empresariales y B2B, vinculadas a financiación alternativa, pagos entre empresas, contabilidad, facturación, gestión de tesorería, open banking y análisis financiero de empresas (Sánchez & Quintanero, 2022; OECD, 2020; Comisión Europea, 2023). Este desplazamiento responde, por un lado, a la madurez de muchas soluciones fintech de consumo y, por otro, a la persistencia de fricciones operativas y de acceso a financiación en el segmento PYME, donde la calidad, disponibilidad y estructuración de los datos financieros sigue siendo una barrera relevante (OECD, 2020; Aluffi et al., 2025). En este contexto, DataBridge se sitúa dentro de una lógica fintech B2B: no ofrece financiación directamente, sino que transforma información financiera empresarial en perfiles y dossieres útiles para mejorar el diagnóstico preliminar y la conexión con financiadores.

Las soluciones fintech aplicadas a la financiación empresarial adoptan formatos diversos. Entre ellas se encuentran las plataformas de crowdlending, que canalizan préstamos de inversores hacia empresas, y las de crowdfunding, orientadas a financiar proyectos mediante aportaciones colectivas. También existen marketplaces de financiación alternativa, que agregan distintas opciones de financiación y permiten comparar productos como préstamos, factoring, leasing o financiación de circulante. Otras soluciones se centran en herramientas de scoring y prevención de fraude, utilizando datos para estimar perfiles de riesgo, detectar anomalías o anticipar comportamientos sospechosos. A ello se suma el embedded finance, que consiste en integrar servicios financieros dentro de plataformas no financieras, como software de gestión, marketplaces o herramientas de facturación.

De forma creciente, también están surgiendo infraestructuras de inteligencia financiera que no conceden crédito directamente, sino que preparan, enriquecen y analizan la información empresarial para que sean los financiadores quienes tomen la decisión final. DataBridge se inscribe en esta última categoría: su valor diferencial no está en ofrecer un producto financiero propio, sino en transformar información dispersa de la PYME en perfiles segmentados, diagnósticos preliminares y dossieres comparables.

3.2. Tres referencias de uso intensivo de datos en servicios financieros

Para conectar el marco fintech con la propuesta de DataBridge, se analizan tres compañías que utilizan datos financieros o transaccionales como parte central de su modelo de negocio. Los casos seleccionados no pretenden ser una comparación directa con

DataBridge, sino ejemplos de cómo distintas fintech han convertido grandes volúmenes de datos en información útil para segmentar clientes, evaluar empresas, detectar patrones o mejorar procesos financieros.

La finalidad de estos casos no es afirmar que todas las compañías analizadas utilicen el mismo algoritmo de clusterización ni que apliquen exactamente la misma metodología que DataBridge. Su objetivo es mostrar que el sector fintech ya trabaja de forma habitual con una lógica común: partir de datos financieros o transaccionales, construir variables relevantes, identificar patrones de comportamiento y convertir esos patrones en información útil para el negocio. Esta lógica es precisamente la que se traslada a DataBridge, aunque aplicada al perfilado financiero de PYMEs y no a banca de consumo, concesión directa de crédito o prevención de fraude.

Por este motivo, se seleccionan tres referencias que permiten entender cómo el uso intensivo de datos puede generar valor en distintos ámbitos financieros. **Revolut** representa el uso de datos transaccionales en banca digital para personalizar servicios, mejorar la experiencia del usuario y reforzar mecanismos de seguridad. **Kabbage** resulta especialmente cercano a DataBridge, ya que utilizaba información bancaria, contable y operativa de pequeñas empresas para construir una lectura más dinámica de su situación financiera. **Feedzai**, por último, ilustra cómo el análisis de grandes volúmenes de transacciones permite detectar patrones anómalos y agrupar comportamientos de riesgo en el ámbito de la prevención de fraude.

En conjunto, estos casos permiten justificar que DataBridge se apoya en una tendencia ya presente en el ecosistema fintech: transformar datos financieros dispersos en perfiles interpretables y accionables. La diferencia está en el foco de aplicación. Mientras Revolut

se orienta al usuario bancario, Kabbage a la financiación digital y Feedzai al fraude, DataBridge aplica esa misma lógica analítica a la preparación financiera de PYMEs y a la generación de dossieres útiles para financiadores

Ilustración 2. Logo Revolut

3.2.1. Revolut

País: Reino Unido

Sector: Neobanco / Banca digital

Fuente: Página web

Descripción de la compañía:

Revolut es una fintech de banca digital que ofrece servicios financieros a particulares y empresas a través de una plataforma móvil. Su propuesta se basa en una experiencia bancaria digital, rápida y apoyada en funcionalidades como pagos, tarjetas, cambio de divisas, ahorro, inversión y herramientas de gestión del gasto. La propia compañía destaca su enfoque de "smart spending", servicios de seguridad y el uso de IA a través de AIR, su asistente personal dentro de la app

Uso de datos y relación con DataBridge:

El interés de Revolut para este trabajo no está en que replique exactamente la metodología de DataBridge, sino en que muestra cómo una fintech puede transformar información financiera cotidiana en conocimiento útil sobre sus usuarios. A partir de patrones de gasto, frecuencia de uso, categorías de consumo, localización de operaciones o comportamiento dentro de la aplicación, una entidad de este tipo puede segmentar perfiles, adaptar funcionalidades y detectar comportamientos anómalos.

La enseñanza para DataBridge es que los datos financieros no solo sirven para registrar operaciones, sino también para construir perfiles interpretables. Mientras Revolut aplica esta lógica al usuario bancario y a la personalización de servicios digitales, DataBridge la traslada al ámbito de las PYMEs, utilizando información financiera para identificar patrones de comportamiento empresarial y preparar dossieres más comparables para financiadores.

Ilustración 3. Logo Kabbage

3.2.2. Kabbage

Fuente: Página web

País: Estados Unidos

Sector: Financiación digital para pequeñas empresas

Descripción de la compañía:

Kabbage fue una fintech estadounidense especializada en financiación para pequeñas empresas. Su propuesta se basaba en agilizar el acceso a crédito mediante una plataforma automatizada que analizaba datos digitales de las empresas. Kabbage utilizaba información procedente de cuentas bancarias, servicios contables, plataformas de comercio electrónico, datos logísticos y otras fuentes conectadas con autorización del cliente para evaluar negocios de forma más rápida que la banca tradicional

Uso de datos y relación con DataBridge:

Kabbage es el caso más cercano a DataBridge porque parte de un problema similar: muchas pequeñas empresas tienen dificultades para demostrar su situación financiera mediante los canales tradicionales. Frente a un análisis basado únicamente en estados financieros históricos o documentación bancaria convencional, Kabbage utilizaba datos operativos y financieros actualizados para construir una lectura más dinámica del negocio.

La relación con DataBridge está en la lógica de transformar datos dispersos en perfiles útiles para financiación. Una PYME con ingresos recurrentes, otra con ventas estacionales, otra con alta volatilidad de caja o una empresa con fuerte crecimiento no presentan las mismas necesidades financieras. Kabbage ilustra que el análisis de datos puede ayudar a distinguir estos perfiles y a acelerar el diagnóstico. DataBridge toma esta lógica como referencia, aunque no concede crédito directamente: su papel es preparar y ordenar la información financiera para facilitar el análisis posterior por parte de los financiadores.

3.2.3. Feedzai

Ilustración 4. Logo Feedzai

Fuente: Página web

País: Portugal

Sector: Prevención de fraude y riesgo transaccional

Descripción de la compañía:

Feedzai es una fintech portuguesa especializada en inteligencia artificial, machine learning y big data analytics para identificar transacciones fraudulentas y reducir riesgos en servicios financieros, retail y comercio electrónico. La compañía desarrolla herramientas de machine learning en tiempo real para detectar pagos fraudulentos y minimizar riesgos operativos.

Uso de datos y relación con DataBridge:

Feedzai se incluye como caso porque muestra otra aplicación del análisis de patrones en datos financieros. En prevención de fraude, el valor no está solo en revisar una operación aislada, sino en compararla con muchas operaciones similares para identificar comportamientos normales, comportamientos atípicos o señales de riesgo. Variables como importe, frecuencia, horario, localización, dispositivo o historial transaccional pueden ayudar a distinguir patrones habituales de operaciones sospechosas.

La utilidad para DataBridge es conceptual: demuestra que el análisis de datos financieros permite pasar de operaciones individuales a patrones interpretables. Mientras Feedzai aplica esta lógica a la detección de fraude y riesgo transaccional, DataBridge la aplica al perfilado financiero de PYMEs. En ambos casos, el objetivo es convertir grandes volúmenes de información dispersa en señales útiles para interpretar mejor una situación financiera.

3.3. Problema y oportunidad de DataBridge

DataBridge se plantea como una plataforma fintech B2B orientada a mejorar la preparación financiera de las PYMEs y facilitar su conexión preliminar con financiadores. El modelo parte de una idea sencilla: la PYME aporta sus datos financieros, la plataforma los transforma en un diagnóstico preliminar y un dossier estructurado, y los financiadores reciben oportunidades comerciales mejor documentadas.

El problema que aborda DataBridge surge de una fricción muy concreta: muchas PYMEs generan información financiera relevante en su actividad diaria, pero esa información suele estar dispersa en extractos bancarios, documentos contables, modelos fiscales o registros internos. Movimientos bancarios, cobros, pagos y categorías de ingreso y gasto contienen señales útiles sobre la situación económica de una empresa, pero normalmente no se transforman en perfiles financieros comprensibles.

Desde el punto de vista de la PYME, esta falta de estructuración dificulta preparar una solicitud de financiación clara y defender adecuadamente sus necesidades ante terceros. Desde el punto de vista del financiador, obliga a analizar información heterogénea, incompleta o poco comparable. La oportunidad de DataBridge surge precisamente en ese punto intermedio: convertir datos financieros dispersos en información ordenada, interpretable y útil para ambas partes.

3.3.1. El problema: datos financieros dispersos en PYMEs

Las PYMEs suelen contar con información financiera en distintas fuentes: cuentas bancarias, programas contables, facturas, modelos fiscales o extractos de movimientos. Sin embargo, esta información no siempre está integrada ni preparada para ser analizada de forma conjunta. En muchas empresas, los datos existen, pero no se utilizan como una herramienta de diagnóstico financiero.

Esta situación genera dos problemas. En primer lugar, la propia empresa tiene dificultades para entender su posición financiera real más allá del saldo bancario o del resultado contable anual. En segundo lugar, cuando necesita financiación, no siempre puede presentar su información de forma ordenada y comparable ante bancos, fintech lenders u otros financiadores.

Por tanto, el problema no es únicamente la falta de financiación, sino la falta de una capa analítica que traduzca los datos de la empresa en información útil. Sin esa transformación, los movimientos financieros siguen siendo registros aislados, no señales accionables.

3.3.2. La oportunidad: convertir movimientos en perfiles accionables

La oportunidad de negocio consiste en transformar los movimientos financieros de las PYMEs en variables agregadas por empresa. A partir de datos como fechas, importes, categorías, entradas, salidas o contrapartes, es posible construir indicadores que reflejen mejor el comportamiento financiero de cada compañía.

Entre estos indicadores pueden incluirse ingresos medios mensuales, salidas medias, flujo neto, volatilidad de ingresos, porcentaje de meses negativos, presión de caja, concentración de clientes o proveedores, peso de gastos financieros y presencia de inversión en activos. Estas variables permiten pasar de una lectura desordenada de movimientos individuales a una visión estructurada de la empresa.

Ilustración 5. Datos disponibles, variables construidas y utilidad para DataBridge

Problema inicial	***Dato disponible***	***Variable construida***	***Utilidad para DataBridge***
Información financiera dispersa	*Movimientos bancarios*	*Ingresos, salidas y flujo neto*	*Diagnóstico preliminar de la empresa*
Dificultad para detectar estacionalidad	*Fechas e importes*	*Volatilidad de ingresos y meses negativos*	*Identificación de necesidades de circulante*
Dependencia de pocos clientes o proveedores	*Contrapartes*	*Concentración comercial*	*Detección de riesgo de dependencia*
Falta de comparación entre empresas	*Variables normalizadas*	*Indicadores comparables*	*Perfilado de PYMEs*
Solicitudes de financiación poco estructuradas	*Datos categorizados*	*Dossier financiero*	*Lead más cualificado para financiadores*

Fuente: Elaboración propia.

3.3.3. Papel de la clusterización en DataBridge

La clusterización permite agrupar empresas con comportamientos financieros similares sin necesidad de definir previamente una etiqueta de riesgo, aprobación o impago. Esto resulta especialmente adecuado para DataBridge, ya que el objetivo no es construir un scoring crediticio ni emitir una calificación financiera, sino identificar tipologías de PYMEs a partir de sus datos.

Aplicada al caso de DataBridge, la clusterización permite detectar perfiles como empresas de operativa estable, negocios con ingresos volátiles o estacionales, compañías con concentración comercial relevante o empresas con señales de presión de caja. Estos perfiles ayudan a entender mejor qué tipo de PYME está siendo analizada y qué información puede ser más relevante para su dossier financiero.

El valor de la clusterización está, por tanto, en convertir datos transaccionales en categorías comprensibles. Para la PYME, esto facilita una lectura más clara de su situación. Para el financiador, permite recibir oportunidades mejor estructuradas. Para DataBridge, supone una herramienta que conecta directamente el tratamiento de datos con la propuesta de valor del modelo de negocio.

Es importante precisar que la clusterización se utiliza como una técnica exploratoria y descriptiva. No decide si una empresa debe recibir financiación, no recomienda productos financieros regulados y no sustituye el análisis del financiador. Su función es mejorar el punto de partida: ordenar, clasificar e interpretar la información disponible.

3.3.4. Mercado objetivo y referencia TAM/SAM/SOM

El mercado potencial de DataBridge está formado por PYMEs con actividad económica recurrente, información financiera suficiente y posibles necesidades de financiación. No obstante, desde una perspectiva analítica, no todas las PYMEs tienen el mismo interés inicial para la plataforma. El foco debe situarse en aquellas empresas que generan datos suficientes para ser analizados y cuya situación financiera pueda beneficiarse de una preparación documental más ordenada.

Desde la lógica TAM/SAM/SOM, el mercado puede resumirse de la siguiente forma:

Ilustración 6. Mercado objetivo de DataBridge (TAM/SAM/SOM)

Nivel	*Definición*	*Enfoque para DataBridge*
TAM	*Total de PYMEs con actividad económica y potencial necesidad de financiación*	*Aproximadamente 1.6-1.7 millones de PYMEs con necesidades potenciales de financiación*
SAM	*PYMEs con datos suficientes, actividad recurrente y acceso viable a través de asesorías o financiadores*	*Entre 250,000 y 400,000 PYMEs con mayor encaje: actividad recurrente, datos financieros suficientes y cierta complejidad financiera*
SOM	*Primer segmento capturable en una fase inicial*	*Entre 1,000 y 2,500 PYMEs activas o dossieres generados en 3-5 años, mediante socios como Asfin, pilotos y financiadores especializados*

Fuente: Elaboración propia.

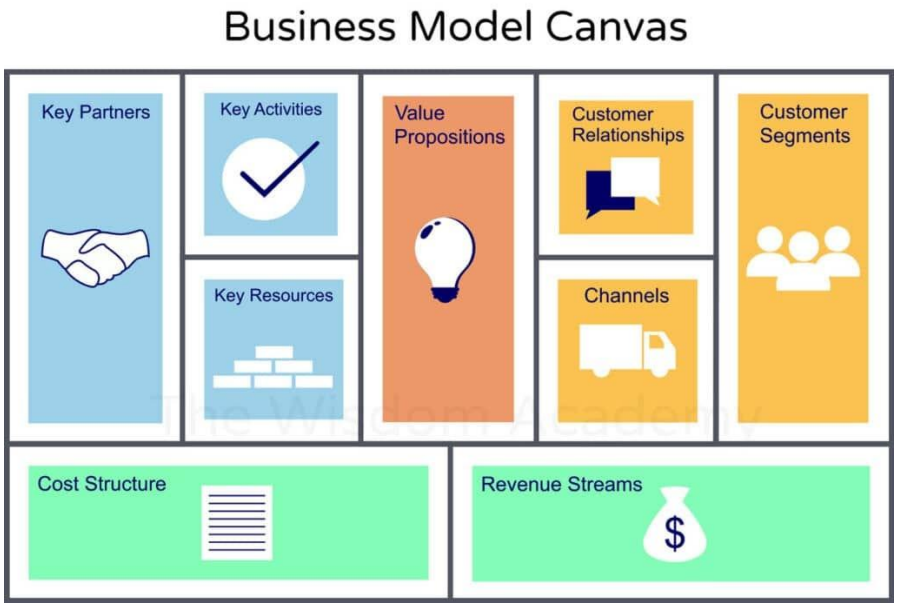
3.4. Modelo de negocio resumido

Una vez presentado el contexto fintech, la oportunidad de mercado y el papel de Asfin como socio del proyecto, este apartado resume el modelo de negocio de DataBridge desde la perspectiva del TFG de Business Analytics. No se pretende repetir el plan de negocio completo desarrollado en el TFG de ADE, sino mostrar cómo el componente analítico —especialmente la clusterización de PYMEs— encaja dentro de la propuesta empresarial

3.4.1. Business Model Canvas de DataBridge

El Business Model Canvas permite representar de forma ordenada los elementos principales del modelo de negocio: propuesta de valor, clientes, canales, relación con clientes, fuentes de ingresos, recursos clave, actividades clave, alianzas y estructura de costes (Osterwalder & Pigneur, 2010).

Ilustración 7. Estructura del Business Model Canvas.



Fuente: elaboración propia a partir de Osterwalder y Pigneur (2010).

El elemento diferencial de DataBridge no es únicamente conectar PYMEs y financiadores, sino hacerlo a partir de información financiera estructurada. Por ello, el análisis de datos no es una función secundaria del modelo, sino uno de sus recursos y actividades clave.

Ilustración 8. Business Model Canvas de DataBridge

Bloque	*Aplicación a DataBridge*
Propuesta de valor	*Convertir datos financieros dispersos de PYMEs en diagnósticos, perfiles y dossiers útiles para procesos de financiación.*
Segmentos de clientes	*PYMEs con necesidades de financiación, financiadores especializados y asesorías como canal de captación.*
Canales	*Asesorías, gestorías, acuerdos con financiadores, captación directa y socios sectoriales como Asfin.*
Relación con clientes	*Modelo digital asistido, con onboarding guiado, diagnóstico preliminar y soporte en la preparación documental.*
Fuentes de ingresos	*Suscripciones o fees pagados por financiadores por recibir leads cualificados; posibles planes premium para asesorías.*
Recursos clave	*Base de datos, algoritmo de clusterización, plataforma tecnológica, conocimiento financiero y red de financiadores.*
Actividades clave	*Tratamiento de datos, construcción de variables, generación de perfiles, elaboración de dossiers y matching preliminar.*
Alianzas clave	*Asfin, asesorías, financiadores, proveedores tecnológicos y posibles integraciones contables o bancarias.*
Estructura de costes	*Desarrollo tecnológico, almacenamiento de datos, mantenimiento, cumplimiento normativo, captación comercial y soporte.*

Fuente: Elaboración propia a partir de Osterwalder y Pigneur (2010).

3.4.2. Asfin: socio del proyecto y referencia metodológica

Ilustración 9. Logo Asfin

Fuente: Página web

Asfin ocupa un papel relevante dentro del proyecto como socio sectorial y fuente de datos para el desarrollo del análisis. En el marco de este TFG, Asfin proporciona una base de información vinculada a empresas y movimientos financieros categorizados, lo que permite construir un caso práctico de aplicación de clusterización sobre datos empresariales.

La importancia de Asfin no reside solo en aportar datos, sino también en su conocimiento del entorno PYME y de las necesidades de financiación empresarial. Este tipo de conocimiento resulta especialmente útil para interpretar los resultados del modelo, ya que una clusterización puramente estadística necesita ser traducida a perfiles financieros comprensibles y útiles desde el punto de vista de negocio.

Ilustración 10. Rol de Asfin en el proyecto DataBridge

Elemento	*Descripción*
Rol en el proyecto	*Socio sectorial y fuente de datos para el análisis académico.*
Conocimiento aportado	*Experiencia en necesidades financieras de empresas y procesos de financiación PYME.*
Datos proporcionados	*Información de empresas y movimientos financieros categorizados.*
Utilidad analítica	*Permite construir variables agregadas por empresa y aplicar clusterización.*
Utilidad de negocio	*Ayuda a comprobar cómo DataBridge podría transformar datos reales en perfiles financieros.*

Fuente: Elaboración propia.

Desde el punto de vista del TFG, Asfin funciona como el puente entre el planteamiento conceptual de DataBridge y su aplicación práctica. Sin una base de datos empresarial, la propuesta quedaría en un plano teórico. Con los datos proporcionados, es posible demostrar cómo la plataforma podría transformar movimientos financieros en tipologías de PYMEs.

3.4.3. Encaje entre el modelo de negocio y el análisis de clusters

El encaje entre el modelo de negocio y el análisis de clusters es el punto central del TFG de Analytics. DataBridge necesita identificar qué tipo de empresa tiene delante para generar un diagnóstico útil y un lead más cualificado. La clusterización permite precisamente eso: transformar datos de movimientos en perfiles empresariales.

El proceso puede resumirse en cinco fases:

Ilustración 11. Fases del proceso analítico de DataBridge

Fase	*Descripción*	*Resultado*
1. Captura de datos	*La PYME aporta movimientos financieros o información categorizada.*	*Base de datos transaccional.*
2. Tratamiento	*Se limpian, ordenan y agregan los movimientos por empresa.*	*Variables financieras comparables.*
3. Clusterización	*Se agrupan empresas con comportamientos similares.*	*Tipologías de PYMEs.*
4. Diagnóstico	*Se interpreta el perfil de cada empresa.*	*Dossier financiero preliminar.*
5. Matching	*Se comparte la información con financiadores potencialmente compatibles.*	*Lead cualificado y documentado.*

Fuente: Elaboración propia.

Este proceso demuestra que la clusterización no es un añadido técnico aislado, sino una pieza necesaria para que el modelo de negocio tenga sentido operativo. La plataforma no puede generar leads de calidad si antes no entiende mínimamente el perfil financiero de cada empresa.

En este sentido, el TFG de Analytics complementa el TFG de ADE. Mientras el trabajo de ADE justifica la oportunidad de negocio y el modelo empresarial, este trabajo muestra cómo el tratamiento de datos puede hacer esa oportunidad más concreta. DataBridge no solo propone conectar PYMEs con financiadores; propone hacerlo mediante una capa

analítica capaz de ordenar información, detectar patrones y generar perfiles financieros no vinculantes.

4. ANÁLISIS DEL CLIENTE, TIPOLOGÍA Y CLUSTERIZACIÓN

En este capítulo se presenta el análisis realizado sobre la base de datos proporcionada por Asfin, con el objetivo de construir una tipología de PYMEs a partir de sus movimientos financieros categorizados. El análisis se ha desarrollado en Python y se apoya en técnicas de tratamiento de datos, construcción de variables financieras y clusterización no supervisada.

El objetivo no es predecir impagos ni emitir una calificación crediticia. La finalidad del ejercicio es distinta: transformar una base de datos formada por movimientos individuales en una segmentación interpretable de empresas. De esta forma, DataBridge puede identificar perfiles financieros de PYMEs y generar información útil para un diagnóstico preliminar y un dossier financiero más ordenado.

La lógica seguida en el capítulo es la siguiente: primero se describen las bases de datos utilizadas; después se explica el proceso de limpieza y transformación; posteriormente se detallan las variables construidas; a continuación, se presenta la metodología de clusterización; y, finalmente, se interpretan los clusters obtenidos y su utilidad dentro del modelo de negocio de DataBridge.

4.1. Descripción de las bases de datos utilizadas

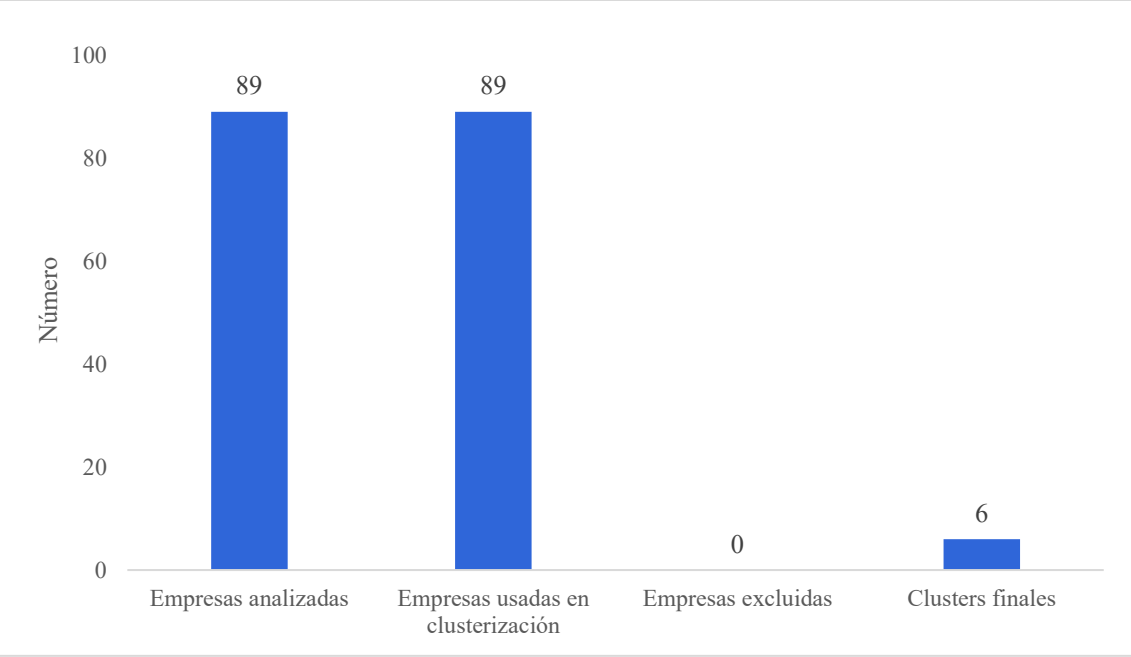
El análisis parte de dos bases de datos proporcionadas por Asfin. La primera contiene información identificativa de las empresas, mientras que la segunda recoge movimientos financieros categorizados. La unidad final de análisis no es cada movimiento individual, sino cada empresa. Por tanto, el primer paso del análisis consiste en transformar una base transaccional en una base agregada por compañía.

La base de movimientos incluye información como el identificador de la empresa, la fecha de operación, el importe, la categoría del movimiento, la subcategoría, la entidad bancaria, la cuenta asociada y, en algunos casos, el cliente o proveedor relacionado. Esta estructura permite analizar no solo el volumen total de ingresos y salidas, sino también la recurrencia, volatilidad, concentración y composición de los flujos financieros de cada empresa.

Tras el tratamiento inicial, la muestra utilizada en el modelo quedó formada por 89 empresas y 483,919 movimientos financieros. Aunque el código incorporaba filtros

mínimos de calidad, número de movimientos y meses activos, para evitar observaciones poco representativas, en la base final todas las empresas cumplían dichos criterios, por lo que no fue necesario excluir ninguna compañía. Como se observa en el gráfico de composición de la muestra, las 89 empresas disponibles fueron finalmente utilizadas en la clusterización, dando lugar a una solución final de 6 clusters

Ilustración 12. Resumen de la muestra analizada



Fuente: elaboración propia.

Este punto es importante porque condiciona el alcance del análisis: el modelo permite identificar patrones dentro de la muestra disponible, pero no debe interpretarse como una representación completa de todas las PYMEs españolas. Es una segmentación exploratoria basada en datos reales de empresas, pero limitada al universo de información proporcionado

4.2. Tratamiento de datos y construcción de variables

Una vez cargadas las bases de datos, el primer paso consistió en depurar y preparar la información. Para ello, se transformaron los formatos de fecha, se convirtieron los importes a formato numérico, se identificaron entradas y salidas de caja y se agruparon los movimientos por empresa y mes.

El tratamiento de datos siguió cuatro fases principales. En primer lugar, se realizó una limpieza básica de la información, eliminando errores de formato y homogeneizando los identificadores de empresa. En segundo lugar, se clasificaron los movimientos como entradas o salidas, utilizando tanto el signo del importe como las columnas de categorización disponibles. En tercer lugar, se agregaron los movimientos a nivel mensual para capturar la evolución temporal de ingresos, pagos y flujo neto. Por último, se construyeron variables agregadas por empresa, que son las que posteriormente alimentan el modelo de clusterización

La transformación más relevante consiste en pasar de una base de movimientos a una base de empresas. Cada compañía queda representada por un conjunto de indicadores que resumen su comportamiento financiero: volumen de operaciones, ingresos medios, salidas medias, volatilidad, presión de caja, concentración comercial y peso de determinadas categorías de gasto

Entre las variables más relevantes se encuentran las siguientes:

- **Ingresos medios mensuales**, que permiten aproximar el tamaño operativo de la empresa.
- **Salidas medias mensuales**, que reflejan su estructura de pagos.
- **Flujo neto mensual medio**, calculado como diferencia entre entradas y salidas.
- **Volatilidad de ingresos**, que mide la estabilidad o irregularidad de los cobros.
- **Porcentaje de meses negativos**, que identifica empresas con presión recurrente de caja.
- **Ratio de presión de caja**, calculado como salidas totales sobre entradas totales.
- **Concentración de contrapartes**, que mide la dependencia respecto a clientes o proveedores principales.
- **Peso de gastos financieros**, útil para detectar empresas con carga financiera visible.
- **Peso de inversión o inmovilizado**, que puede indicar necesidades de financiación de activos.

- **Número de categorías y subcategorías**, que aproxima la diversidad operativa de la empresa.

Este paso es el núcleo del análisis. Sin esta construcción de variables, la clusterización solo agruparía movimientos aislados sin significado empresarial. Al convertir los movimientos en indicadores financieros, el modelo puede comparar empresas entre sí y detectar patrones comunes.

4.3. Metodología de clusterización

La técnica utilizada para segmentar las empresas ha sido **K-Means**, un algoritmo de aprendizaje no supervisado que agrupa observaciones en función de su similitud. En este caso, cada observación es una empresa y cada variable representa una dimensión de su comportamiento financiero.

K-Means busca dividir las empresas en grupos de forma que las compañías dentro de un mismo cluster sean lo más parecidas posible entre sí y, al mismo tiempo, diferentes de las empresas pertenecientes a otros clusters. El algoritmo asigna cada empresa al centroide más cercano y recalcula los centroides hasta minimizar la variabilidad interna de cada grupo.

De forma simplificada, el algoritmo trata de minimizar la siguiente función:

$$\sum_{i=1}^{n} \min_{\mu_j \in C} ||x_i - \mu_j||^2$$

donde $x_i$representa cada empresa, $\mu_j$representa el centroide del cluster y Cel conjunto de clusters definidos

Antes de aplicar el algoritmo, las variables fueron escaladas mediante RobustScaler, una técnica de preprocesamiento que transforma cada variable restando su mediana y dividiendo el resultado entre su rango intercuartílico. Es decir, para cada observación se calcula una versión escalada de la variable como:

x escalado = (x − mediana) / IQR,

donde IQR es el rango intercuartílico, calculado como la diferencia entre el tercer cuartil y el primer cuartil (Q3 − Q1). A diferencia de una estandarización basada en media y desviación típica, esta técnica es más robusta ante valores extremos, porque la mediana y

los cuartiles se ven menos afectados por observaciones atípicas. Esta decisión se justifica porque las variables financieras suelen presentar importes muy desiguales entre empresas. Sin este tratamiento, el modelo podría quedar dominado por el tamaño absoluto de la empresa en lugar de capturar patrones de comportamiento financiero.

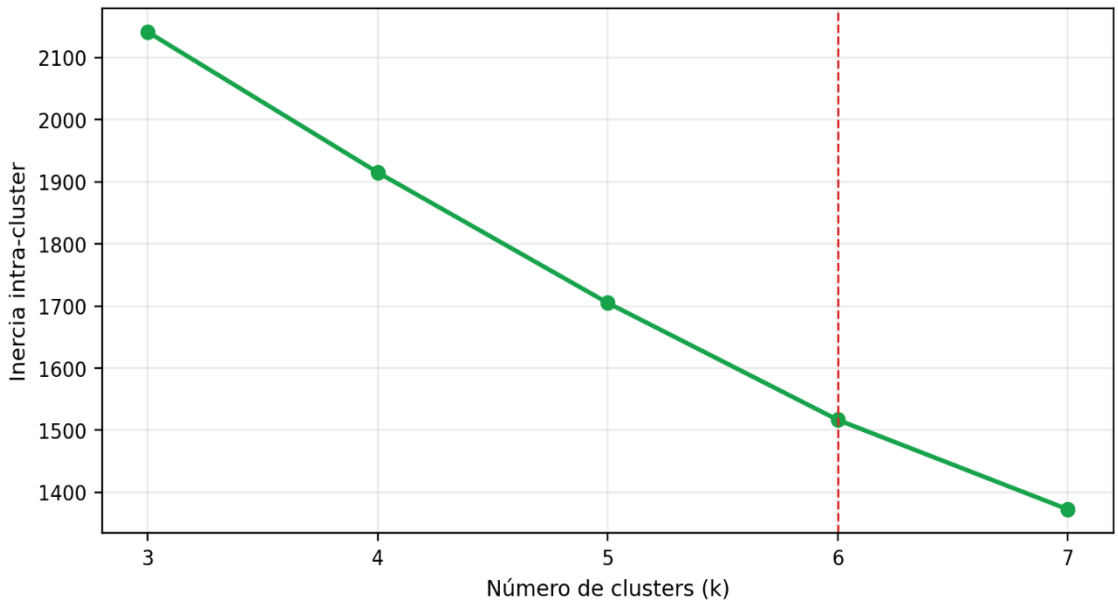
También se limitaron los valores extremos mediante winsorización en los percentiles 1 y 99. En concreto, los valores situados por debajo del percentil 1 se sustituyeron por el valor de dicho percentil, y los valores por encima del percentil 99 se sustituyeron por el valor del percentil 99. De esta forma, se reduce el impacto de observaciones extremas sin eliminar empresas de la muestra ni modificar de forma excesiva la distribución de las variables. Este punto es importante porque la muestra contiene empresas con tamaños y volúmenes de movimiento muy distintos.

4.4. Selección del número de clusters y validación

Para seleccionar el número de clusters se probaron distintas soluciones, desde 3 hasta 7 grupos. La evaluación se realizó mediante dos métricas: la inercia y el coeficiente de silueta.

La inercia mide la suma de distancias internas dentro de los clusters. Cuanto menor es la inercia, más compactos son los grupos. Sin embargo, la inercia siempre mejora al aumentar el número de clusters, por lo que no puede utilizarse de forma aislada. Por ello, se complementa con el método del codo, que permite observar a partir de qué punto añadir más grupos aporta una mejora marginal

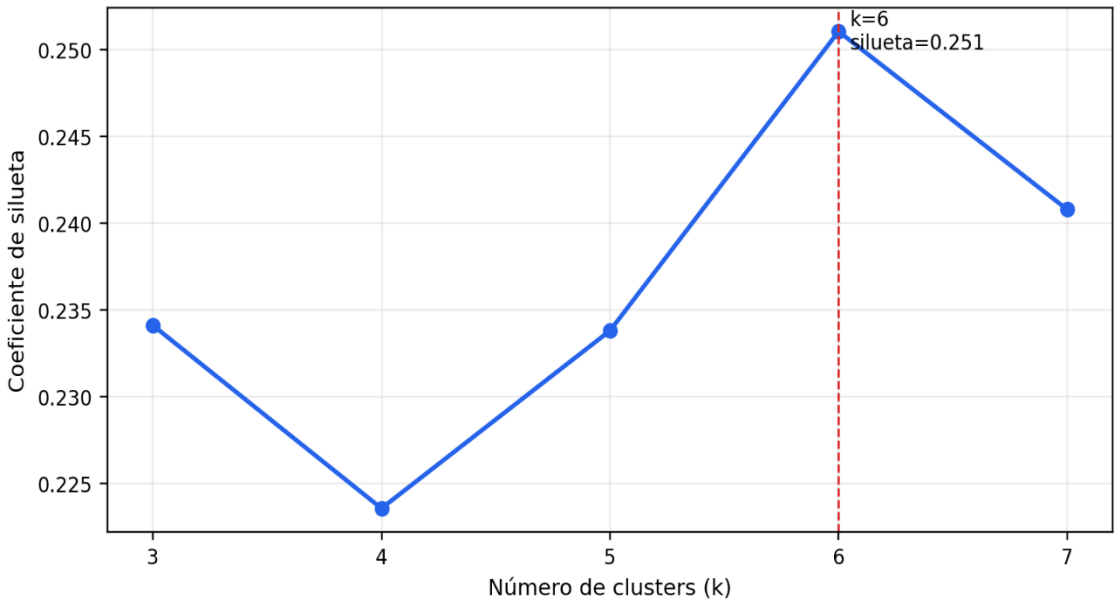
Ilustración 13. Método del codo para la selección del número de clusters.



Fuente: elaboración propia.

El segundo criterio utilizado fue el coeficiente de silueta, que mide hasta qué punto cada empresa está bien asignada a su cluster. Este indicador se mueve entre -1 y 1. Valores más altos indican una mayor separación entre grupos y una asignación más clara de las observaciones.

Ilustración 14. Selección del número de clusters: Coeficiente de silueta



Fuente: elaboración propia.

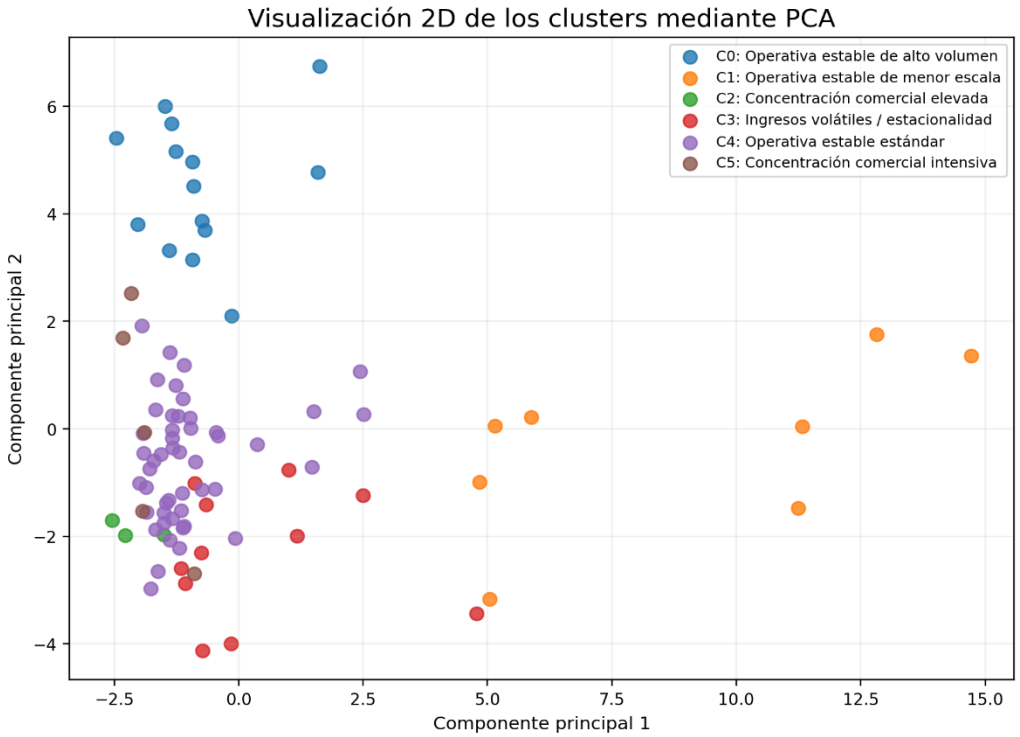
A partir de estas métricas, se seleccionó una solución de 6 clusters, con un coeficiente de silueta de 0.251. Este resultado debe interpretarse con prudencia. No indica una separación perfecta entre grupos, pero sí permite obtener una segmentación razonable para un análisis exploratorio. En una base de datos empresarial real, con empresas heterogéneas y variables financieras ruidosas, no es habitual encontrar separaciones completamente limpias.

Por tanto, el resultado se considera adecuado para el objetivo del trabajo: construir una tipología inicial de PYMEs que ayude a interpretar patrones financieros y a conectar el análisis de datos con el modelo de negocio de DataBridge.

4.5. Visualización de clusters

Para representar visualmente los resultados se utilizó un análisis de componentes principales (PCA), reduciendo el conjunto de variables a dos dimensiones. Esta técnica no se utiliza para construir los clusters, sino para visualizarlos en un plano bidimensional y facilitar su interpretación.

Ilustración 15. Visualización bidimensional de los clusters mediante PCA



Fuente: Elaboración propia.

La figura muestra que los clusters no se distribuyen como bloques completamente separados, sino como grupos parcialmente solapados. Este resultado es coherente con la naturaleza del problema: las PYMEs no pertenecen a categorías financieras totalmente cerradas, sino que suelen compartir rasgos comunes. Por ejemplo, una empresa puede tener operativa estable y, al mismo tiempo, cierta concentración comercial; o puede tener ingresos relativamente estables, pero episodios puntuales de presión de caja.

La utilidad del PCA no está en demostrar una separación perfecta, sino en comprobar que el algoritmo identifica zonas diferenciadas dentro del espacio de variables. En este caso, la visualización permite observar grupos de empresas con características relativamente próximas, que posteriormente se interpretan desde una perspectiva de negocio.

4.6. Resultados obtenidos: tipologías de PYMEs

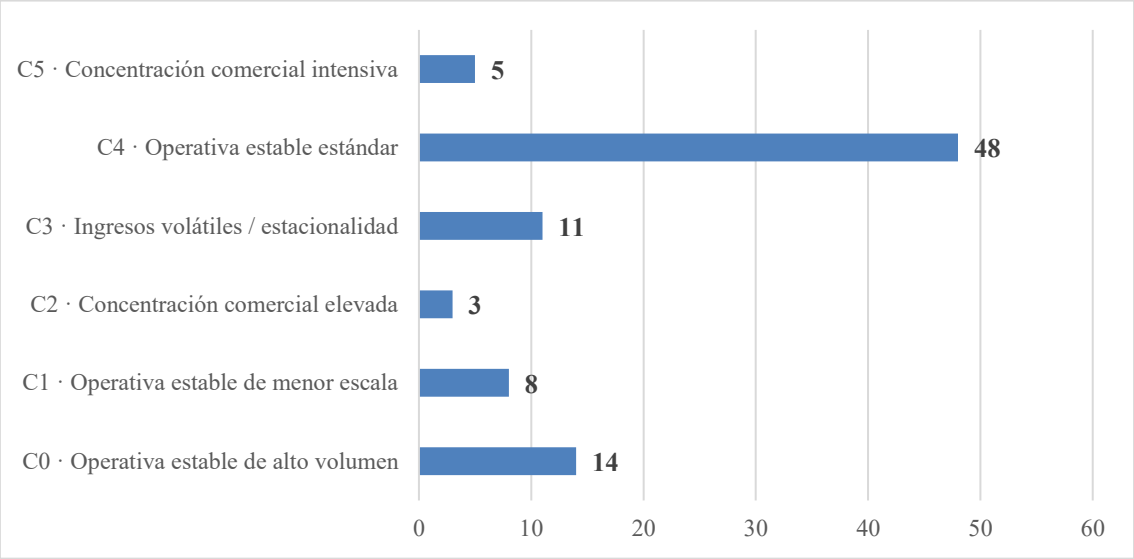
El modelo final clasifica las empresas en **6 clusters**. A cada cluster se le ha asignado una denominación empresarial a partir del análisis de sus variables medianas. Esta interpretación es necesaria porque el algoritmo únicamente devuelve grupos numéricos; el valor para DataBridge aparece cuando esos grupos se traducen en perfiles financieros comprensibles.

Ilustración 16. Tipologías de PYMEs identificadas

Cluster	Nombre propuesto	Nº empresas	Interpretación
0	Operativa estable de alto volumen	14	Empresas con actividad financiera relevante y comportamiento relativamente equilibrado.
1	Operativa estable de menor escala	8	Empresas con menor volumen transaccional, pero sin señales claras de tensión financiera.
2	Concentración comercial elevada	3	Empresas con dependencia relevante respecto a determinadas contrapartes.
3	Ingresos volátiles / estacionalidad	11	Empresas con mayor irregularidad en ingresos y posible necesidad de financiación flexible.
4	Operativa estable estándar	48	Grupo mayoritario, con comportamiento financiero normalizado y sin señales extremas.
5	Concentración comercial intensiva	5	Empresas con concentración significativa en clientes o proveedores concretos.

Fuente: Elaboración propia.

Ilustración 17. Perfil e interpretación de los clusters obtenidos.



Fuente: elaboración propia.

El cluster más numeroso es el Cluster 4: Operativa estable estándar, con 48 empresas. Esto indica que una parte relevante de la muestra presenta un comportamiento financiero relativamente normalizado, sin señales extremas de volatilidad, concentración o tensión de caja. Este grupo puede representar el perfil base de PYME que DataBridge podría analizar mediante un dossier estándar.

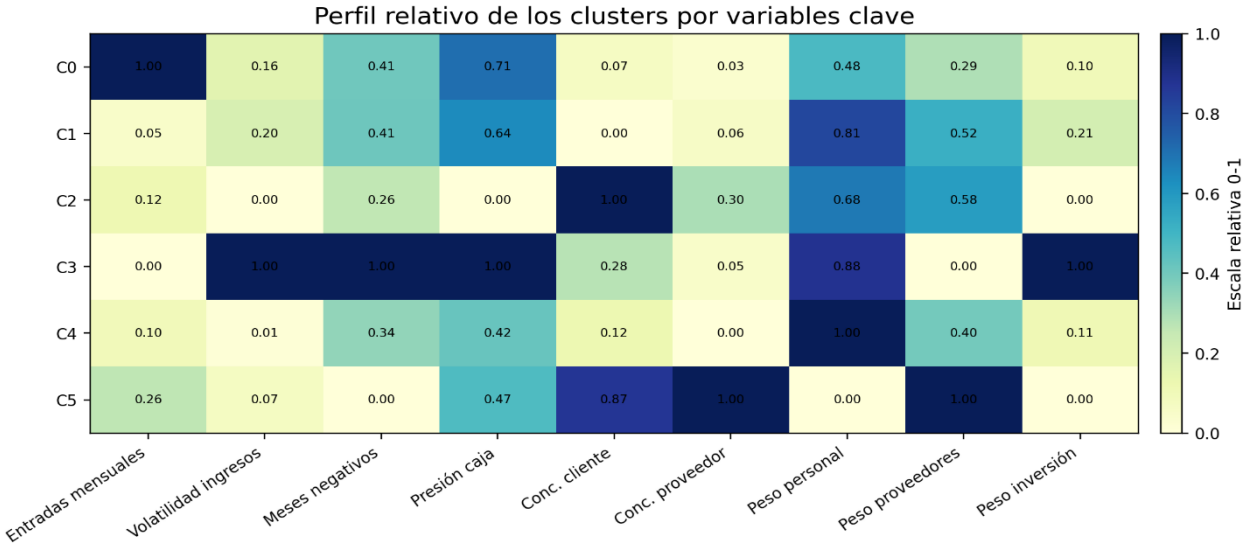
Los clusters 0 y 1 también pertenecen a la familia de operativa estable, aunque con diferencias de escala. El Cluster 0 agrupa empresas con mayor volumen operativo, mientras que el Cluster 1 recoge empresas de menor dimensión transaccional. Esta distinción es útil porque permite no tratar de la misma forma a empresas estables pero de tamaños distintos.

El Cluster 3, denominado Ingresos volátiles / estacionalidad, es especialmente relevante para DataBridge. Agrupa empresas cuyos ingresos muestran mayor irregularidad, lo que puede estar asociado a negocios estacionales, ciclos de cobro más variables o dependencia de proyectos concretos. Desde el punto de vista financiero, este perfil puede requerir soluciones flexibles de circulante o líneas adaptadas a picos y valles de actividad.

Los clusters 2 y 5 se han interpretado como perfiles de concentración comercial. Aunque tienen menor número de empresas, son relevantes desde el punto de vista de riesgo y financiación. Una empresa con alta dependencia de pocos clientes puede tener ingresos

elevados, pero también una exposición significativa si alguno de esos clientes retrasa pagos o reduce pedidos. En este tipo de perfil, productos como factoring, seguro de crédito o análisis de concentración pueden ser especialmente relevantes.

Ilustración 18. Perfil relativo de los clusters por variables clave



Fuente: elaboración propia.

La Ilustración 18 resume el perfil relativo de cada cluster en las principales variables utilizadas para la interpretación financiera. La escala está normalizada entre 0 y 1, por lo que los valores más altos no deben interpretarse como importes absolutos, sino como una mayor intensidad relativa de esa variable dentro del conjunto de clusters. Esta visualización permite identificar de forma rápida qué rasgos diferencian a cada grupo de empresas.

En términos generales, se observa que el cluster C3 concentra los valores más elevados en volatilidad de ingresos, meses negativos, presión de caja y peso de la inversión, lo que apunta a empresas con mayor inestabilidad operativa y posibles necesidades de financiación más complejas. El cluster C0 destaca por un mayor nivel relativo de entradas mensuales y presión de caja, mientras que los clusters C2 y C5 presentan una mayor concentración comercial: C2 principalmente por concentración de clientes y C5 por concentración de proveedores. Por su parte, los clusters C1 y C4 muestran perfiles más equilibrados, aunque con cierto peso relativo de personal y proveedores.

Por tanto, la figura refuerza la interpretación de que la solución de clusterización genera grupos con rasgos financieros diferenciados. Esta lectura facilita la posterior descripción de cada cluster y su posible utilidad dentro del modelo DataBridge.

En conjunto, aunque el algoritmo genera seis clusters, desde el punto de vista empresarial pueden agruparse en tres grandes familias: PYMEs de operativa estable, PYMEs con ingresos volátiles o estacionales y PYMEs con concentración comercial relevante. Esta simplificación es importante para DataBridge, ya que el objetivo no es crear una taxonomía excesivamente compleja, sino generar perfiles que puedan ser entendidos por una PYME, una asesoría o un financiador.

4.7. Interpretación financiera de los clusters

La interpretación de los clusters permite conectar el análisis técnico con la utilidad empresarial. En este sentido, cada grupo puede asociarse a una posible necesidad financiera, siempre de forma orientativa y no vinculante

Ilustración 19. Relación entre clusters y posibles necesidades financieras

Cluster	***Perfil identificado***	***Posible necesidad financiera***	***Encaje preliminar no vinculante***
0	*Operativa estable de alto volumen*	*Financiación según necesidad declarada: inversión, crecimiento o refinanciación*	*Dossier financiero estándar y comparación de alternativas*
1	*Operativa estable de menor escala*	*Necesidades puntuales de liquidez o financiación de menor importe*	*Dossier simplificado y análisis básico de capacidad operativa*
2	*Concentración comercial elevada*	*Dependencia relevante de clientes o proveedores concretos*	*Factoring, seguro de crédito o análisis de concentración*
3	*Ingresos volátiles / estacionalidad*	*Desfases temporales entre cobros y pagos*	*Línea de circulante, financiación estacional o productos flexibles*
4	*Operativa estable estándar*	*Necesidad financiera no asociada a señales extremas*	*Dossier estándar y matching general con financiadores*
5	*Concentración comercial intensiva*	*Alta dependencia de una contraparte principal*	*Factoring, confirming, seguro de crédito o análisis específico de riesgo comercial*

Fuente: Elaboración propia.

Las empresas de operativa estable pueden beneficiarse de un dossier financiero estándar que ordene su información y facilite su presentación ante financiadores. En estos casos, la necesidad de financiación no surge necesariamente de una señal de tensión, sino de una

necesidad concreta declarada por la empresa: inversión, crecimiento, refinanciación o mejora de condiciones.

Las empresas con ingresos volátiles o estacionales presentan una lectura distinta. En este caso, la utilidad del análisis está en identificar que la empresa puede necesitar financiación flexible, adaptada a ciclos de cobro o a picos de actividad. Este perfil puede tener encaje con líneas de circulante, financiación estacional o instrumentos que permitan cubrir desfases temporales entre ingresos y pagos.

Las empresas con concentración comercial requieren una lectura más específica. La dependencia de pocos clientes o proveedores puede afectar a la estabilidad de caja, incluso si la empresa tiene una actividad aparentemente normal. En estos casos, DataBridge podría destacar en el dossier la concentración detectada y orientar el matching preliminar hacia financiadores o productos más adecuados, como factoring, confirming, seguro de crédito o análisis de riesgo comercial.

Es importante insistir en que esta asociación no equivale a una recomendación financiera. DataBridge no decide qué producto debe contratar la empresa ni sustituye el análisis del financiador. La clusterización actúa como una herramienta de perfilado preliminar que ayuda a ordenar la información y a mejorar el punto de partida del proceso.

4.8. Aplicación de los resultados al modelo DataBridge

Los resultados obtenidos tienen una aplicación directa dentro del modelo de negocio de DataBridge. La plataforma necesita transformar datos financieros dispersos en información útil para dos perfiles: la PYME y el financiador.

Para la PYME, la clusterización permite obtener una lectura más clara de su situación financiera. En lugar de recibir simplemente una tabla de movimientos, la empresa puede entender si su perfil se parece más al de una compañía estable, estacional, concentrada o con señales de presión operativa. Esto facilita la comprensión de su propia situación y puede ayudarle a preparar mejor una solicitud de financiación.

Para el financiador, el valor está en recibir oportunidades más ordenadas. Un lead no sería únicamente el nombre de una empresa interesada en financiación, sino una empresa acompañada de un dossier, variables financieras agregadas y una tipología preliminar.

Esto puede reducir la fricción inicial del análisis y mejorar la calidad del pipeline comercial.

5. CONCLUSIONES, LIMITACIONES Y FUTURAS LÍNEAS DE TRABAJO

5.1. Conclusiones del análisis

El trabajo se planteó tres objetivos específicos. A continuación se sintetizan las conclusiones obtenidas en relación con cada uno de ellos.

5.1.1. Transformación de la base de datos

El primer objetivo consistía en transformar la base de movimientos financieros proporcionada por Asfin en una base agregada por empresa. Partiendo de 483.919 movimientos pertenecientes a 89 empresas, se realizó un proceso de limpieza, categorización y agregación que permitió pasar de un registro transaccional a una visión por compañía.

Sobre esta base agregada se construyeron variables financieras como ingresos medios, salidas medias, flujo neto, volatilidad de ingresos, porcentaje de meses negativos, presión de caja, concentración de contrapartes y peso de distintas categorías de gasto. El resultado es una base analítica comparable entre empresas, lista para aplicar técnicas de clusterización y suficientemente robusta para sustentar el resto del análisis.

5.1.2. Tipologías de PYMEs identificadas

El segundo objetivo consistía en aplicar un modelo de clusterización para identificar tipologías de PYMEs e interpretarlas desde una perspectiva empresarial. El modelo final identifica seis clusters, que pueden agruparse en tres grandes familias: empresas de operativa estable, empresas con ingresos volátiles o estacionales y empresas con concentración comercial relevante.

Esta clasificación permite reconocer perfiles diferenciables dentro del universo PYME y asociar cada grupo a posibles necesidades financieras, siempre con carácter orientativo y no vinculante. La interpretación no se limita a etiquetar grupos numéricos, sino que

traduce los patrones detectados por el algoritmo en categorías comprensibles para una asesoría, un financiador o la propia empresa.

5.1.3. Utilidad para el modelo DataBridge

El tercer objetivo consistía en valorar la utilidad de estas tipologías dentro del modelo de negocio de DataBridge. Las tipologías obtenidas tienen una aplicación directa: permiten elaborar diagnósticos preliminares de la situación financiera de cada PYME, preparar dossiers más completos y generar leads más cualificados para los financiadores.

En este sentido, la clusterización funciona como una capa analítica previa al proceso de financiación, no como una herramienta de decisión crediticia. DataBridge no sustituye al financiador en su análisis de riesgo, sino que reduce la asimetría de información en las primeras fases del proceso, aportando contexto financiero estructurado sobre cada empresa.

5.2. Limitaciones del trabajo

El análisis presenta varias limitaciones que deben tenerse en cuenta. La primera es el tamaño de la muestra. Aunque se han analizado 89 empresas y casi medio millón de movimientos, el número de compañías sigue siendo reducido para extraer conclusiones generalizables al conjunto del mercado PYME español.

La segunda limitación está relacionada con la calidad y profundidad de los datos. El análisis depende de la correcta categorización de los movimientos y de la disponibilidad de información sobre clientes, proveedores, fechas e importes. Si los datos de entrada contienen errores, categorías incompletas o movimientos mal clasificados, los clusters resultantes pueden verse afectados.

La tercera limitación es metodológica. El coeficiente de silueta obtenido, **0.251**, indica que la segmentación es útil desde un punto de vista exploratorio, pero no muestra una separación fuerte entre grupos. Esto no invalida el análisis, pero obliga a interpretarlo con prudencia. Los clusters deben entenderse como perfiles orientativos, no como categorías cerradas.

La cuarta limitación es de negocio. La clusterización permite detectar patrones, pero no demuestra por sí sola que una PYME quiera compartir sus datos ni que un financiador esté dispuesto a pagar por recibir estos perfiles. Esa validación comercial debería

realizarse en una fase posterior mediante pilotos, entrevistas o pruebas reales con usuarios.

Por último, debe señalarse una limitación regulatoria. El análisis no debe utilizarse como scoring crediticio ni como rating. DataBridge debe mantener su posicionamiento como herramienta de análisis preliminar y preparación documental, dejando siempre la decisión final de financiación en manos del financiador.

5.3. Futuras líneas de trabajo

Como futura línea de trabajo, sería recomendable ampliar la muestra con más empresas, más sectores y un mayor periodo temporal. Esto permitiría comprobar si los clusters obtenidos se mantienen en una base de datos más amplia.

También sería útil incorporar nuevas fuentes de información, como datos contables, fiscales o de endeudamiento, para construir perfiles financieros más completos.

Otra línea de mejora sería comparar K-Means con otros algoritmos de clusterización, como clustering jerárquico o DBSCAN, con el fin de analizar si los resultados son consistentes.

Finalmente, sería importante validar los clusters con expertos de Asfin, asesores financieros o financiadores. Esta validación permitiría comprobar si las tipologías identificadas tienen sentido desde el punto de vista práctico y si podrían integrarse en un prototipo real de DataBridge.

6. BIBLIOGRAFÍA

Aluffi, P. A., Brandi, J., Bazzi, M., Kennedy, K., Arderne, M., Rodrigues, D., & Lotz, M. (2025). *Categorising SME bank transactions with machine learning and synthetic data generation*. arXiv. https://arxiv.org/abs/2508.05425

Arner, D. W., Barberis, J., & Buckley, R. P. (2016). *The evolution of FinTech: A new post-crisis paradigm?* Georgetown Journal of International Law, 47, 1271-1319.

Asfin. (2025). *Base de datos de empresas y movimientos financieros categorizados* [Conjunto de datos no publicado proporcionado para el Trabajo de Fin de Grado].

Asfin. (s. f.). *Página web corporativa*. Recuperado el 30 de mayo de 2026, de https://asfin.ai/

Blank, S. G., & Dorf, B. (2012). *The startup owner's manual: The step-by-step guide for building a great company*. K&S Ranch.

Comisión Europea. (2023). *Financial data access and payments package*. Directorate-General for Financial Stability, Financial Services and Capital Markets Union. https://finance.ec.europa.eu/publications/financial-data-access-and-payments-package_en

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. En *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226-231). AAAI Press. https://cdn.aaai.org/KDD/1996/KDD96-037.pdf

Feedzai. (s. f.). *AI-native fraud and financial crime prevention platform*. Recuperado el 30 de mayo de 2026, de https://www.feedzai.com/es/

Financial Stability Board. (2017). *Financial stability implications from FinTech: Supervisory and regulatory issues that merit authorities' attention*. Financial Stability Board. https://www.fsb.org/uploads/R270617.pdf

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys, 31*(3), 264-323. https://doi.org/10.1145/331499.331504

Kabbage. (s. f.). *Kabbage from American Express: Small business loans, lines of credit, and checking*. Recuperado el 30 de mayo de 2026, de https://www.kabbage.co/

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470316801

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. En L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281-297). University of California Press.

Organisation for Economic Co-operation and Development. (2020). *The digital transformation of SMEs*. OECD Publishing. https://doi.org/10.1787/bdb9256a-en

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. John Wiley & Sons.

pandas development team. (s. f.). *pandas documentation*. Recuperado el 30 de mayo de 2026, de https://pandas.pydata.org/docs/

Revolut. (s. f.). *Banking & beyond*. Recuperado el 30 de mayo de 2026, de https://www.revolut.com/es-ES/

Reynolds, E. (2016, 23 de junio). Kabbage can get you a small business loan in seven minutes. *WIRED*. https://www.wired.com/story/kathryn-petralia-kabbage-wired-money-2016/

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65. https://doi.org/10.1016/0377-0427(87)90125-7

Sánchez, C., & Quintanero, J. (2022). *Las empresas fintech: panorama, retos e iniciativas* (Documentos Ocasionales, n.º 2214). Banco de España. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosOcasionales/22/Fich/do2214.pdf

scikit-learn developers. (2024a). *Clustering*. scikit-learn documentation. Recuperado el 30 de mayo de 2026, de https://scikit-learn.org/stable/modules/clustering.html

scikit-learn developers. (2024b). *Principal component analysis*. scikit-learn documentation. Recuperado el 30 de mayo de 2026, de https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

scikit-learn developers. (2024c). *RobustScaler*. scikit-learn documentation. Recuperado el 30 de mayo de 2026, de https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html

The Matplotlib development team. (s. f.). *Matplotlib documentation*. Recuperado el 30 de mayo de 2026, de https://matplotlib.org/cheatsheets/

7. DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA

Por la presente, yo, Luis Ussía Arocena, estudiante del Grado en Administración y Dirección de Empresas + Business Analytics de la Universidad Pontificia Comillas, al presentar mi Trabajo Fin de Grado titulado "Clusterización financiera de PYMEs: el caso DataBridge", declaro que he utilizado herramientas de Inteligencia Artificial Generativa, principalmente ChatGPT y Claude, únicamente como apoyo en las actividades descritas a continuación:

1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Crítico: Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
5. Interpretador de código: Para realizar análisis de datos preliminares.
6. Estudios multidisciplinares: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
7. Constructor de plantillas: Para diseñar formatos específicos para secciones del trabajo.
8. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
9. Generador previo de diagramas de flujo y contenido: Para esbozar diagramas iniciales.
10. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
11. Generador de datos sintéticos de prueba: Para la creación de conjuntos de datos ficticios.
12. Generador de problemas de ejemplo: Para ilustrar conceptos y técnicas.
13. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

14. Generador de encuestas: Para diseñar cuestionarios preliminares.
15. Traductor: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación, análisis y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes. Las herramientas de inteligencia artificial generativa han sido utilizadas únicamente como apoyo auxiliar en tareas de estructuración, revisión, redacción preliminar y mejora formal del documento, sin sustituir el análisis propio ni la revisión de las fuentes utilizadas.

Asimismo, declaro que las decisiones sobre el enfoque del trabajo, la selección de contenidos, la interpretación de los resultados, las conclusiones y la versión final del documento han sido responsabilidad exclusiva del autor. Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación de esta declaración.

Fecha: 02/06/2026

Firma:

Luis Ussia Arocena