

# Grado en Estadística

**Título:** Estudio de la morosidad mediante Cuadros de Mando

**Autor:** Mariona García Díaz-Portales

**Director:** Héctor Rufino Alcalde

**Departamento:** Econometría, Estadística y Economía Aplicada

**Convocatoria:** Junio 2023



## RESUMEN I PALABRAS CLAVE

Uno de los motivos de pérdida de capital de las entidades financieras es la **morosidad**. Por ello, las entidades bancarias ponen hincapié en el **estudio, análisis y control de la mora**.

No sólo el comportamiento actual del usuario sino también el pasado, permite **predecir** este acto ahorrando así a la organización una buena parte de su capital.

Para el control, existen herramientas como **Qlik** capaces de plasmar estos datos de una manera dinámica y sencilla en un **Cuadro de Mando** permitiendo así, visualizar el porcentaje de impagos de la organización y en caso de ser elevado, poder aplicar medidas al respecto.

Con este estudio, se ha querido replicar un proceso veraz y real sobre el control de la morosidad en una entidad financiera.

One of the reasons for capital loss in financial institutions is **default**. Therefore, banking institutions place emphasis on the **study, analysis, and control of default in payment**.

Not only the current behavior of the user but also their past behavior allows **predicting** this occurrence, thus saving a significant portion of the organization's capital.

For control purposes, there are tools such as **Qlik** capable of dynamically and simply presenting this data in a **Dashboard**, enabling the visualization of the organization's default rate. If the rate is high, appropriate measures can be applied.

This study aims to replicate an accurate and realistic process regarding the control of default in payment in a financial institution.

# CLASSIFICACIÓN AMS

62-07 *Data analysis*

# ÍNDICE

<b>1. INTRODUCCIÓN</b> .....	2
<b>2. METODOLOGIA</b> .....	3
<b>3. CUADROS DE MANDO</b> .....	3
3.1. <i>Origen y definición</i> .....	3
3.2. <i>Selección de los indicadores (KPIs)</i> .....	6
3.2.1. Ejemplos indicadores financieros básicos.....	6
3.2.2. Ejemplos indicadores relevantes relacionados con los clientes .....	7
3.2.3. Ejemplos indicadores relevantes relacionados con los procesos internos .....	7
3.2.4. Ejemplos indicadores relevantes relacionados con la formación y crecimiento .....	8
3.3. <i>Clasificación de Cuadros de Mando y softwares</i> .....	8
3.3.1. CdM según su función .....	8
3.3.2. CdM según su contenido.....	10
3.4. <i>Cuadros de Mando y el mundo financiero</i> .....	11
3.5. <i>La importancia del control de la morosidad</i> .....	11
<b>4. CREACIÓN DEL CUADRO DE MANDO</b> .....	13
4.1. <i>Credit Bureau</i> .....	13
4.2. <i>Plataforma Qlik Sense</i> .....	14
4.3. <i>Nuestra Base de Datos</i> .....	14
4.3.1. Tablas y variables .....	16
4.4. <i>Misión y visión</i> .....	20
4.5. <i>Creación y transformación de variables</i> .....	21
4.6. <i>Asociación de las tablas</i> .....	23
4.7. <i>Perfil del cliente</i> .....	23
4.8. <i>Perfil del cliente en función de los contratos</i> .....	29
4.9. <i>La mora y el cliente</i> .....	34
4.9.1. Clientes conocidos .....	36
4.9.2. Clientes nuevos.....	36
<b>5. CONCLUSIONES</b> .....	39
<b>6. BIBLIOGRAFIA</b> .....	40
<b>7. ANEXO</b> .....	42
7.1. <i>Descripción de las variables</i> .....	42
7.2. <i>CdM contrato anterior</i> .....	49

# 1. INTRODUCCIÓN

El haber realizado prácticas en distintas entidades financieras hizo que todo con respecto a la morosidad llamara especialmente mi atención.

Es un tema tan sencillo (pues su definición es muy concreta) pero a la vez tan complejo por todo lo que abarca: pérdidas para las entidades, sobreendeudamiento en las familias, tipos de impagos y un largo etcétera que quise adentrar en el tema.

El objetivo por tanto, era acercarse al estudio y comportamiento del usuario y la mora de una manera sencilla y dinámica para que pudiese ser entendida por la mayoría.

Este estudio se expone y se sigue no solo mediante modelos estadísticos que permiten conocer el riesgo que supone un usuario, sino también mediante cuadros de mando, que es lo que se ha creado con este proyecto y como sucede en el mundo laboral.

Para dicha creación, se ha profundizado en la definición de cuadro de mando y su historia, de igual manera que se ha hecho con la morosidad.

En cuanto a la formación del Cuadro de Mando, era necesaria una base de datos muy cercana a la realidad y, gracias a unos datos que la plataforma Kaggle tenía subidos de una entidad australiana, se pudo llevar a cabo.

Estos datos fueron manipulados con diferentes herramientas: Excel y Qlik.

Qlik es una plataforma de análisis y visualización de datos que permite, sobre todo a las entidades, trabajar con grandes volúmenes de datos y transformarlos en información fácil de comprender.

Con ella, se ha creado un Cuadro de Mando en el que se conoce no sólo el perfil del usuario sino también una puntuación que ayude a conocer de antemano si será o no un buen pagador.

Por tanto, el proyecto se divide en tres grandes partes: introducción a los distintos conceptos necesarios para comprender el trabajo, explotación y transformación de los datos y creación del cuadro de mando.

## 2. METODOLOGIA

Lo más importante para este trabajo era un buen análisis de los datos.

Por ello, encontrar una base de datos de gran volumen que fuese lo más ajustada a la realidad era el primer reto. Por suerte, la plataforma Kaggle cuenta con un gran catálogo de datos en el que es difícil no encontrar lo que se busca.

Una vez encontramos la base de datos conformada por diez tablas con más de tres millones de filas en conjunto, había que tratarla. Por ello, primero se hizo una preselección de aquellas variables que se consideraron más adecuadas para lo que se quería analizar y una vez seleccionadas, fueron tratadas.

En un inicio, se transformaron algunas de ellas pero, conforme se avanzaba en el trabajo, se vio que era necesario no sólo tratar algunas de aquellas variables que no se tuvieron en cuenta en un inicio, sino que también se tuvieron que crear nuevas.

Estas nuevas variables surgen de otras que ya incluían los datos y se crearon, porque se consideró que su aportación era más enriquecedora a la de las variables que la conformaban de manera independiente.

Una vez se tuvieron los datos, se utilizó un software que utilizan muchas entidades para tratar y representar grandes volúmenes de datos: Qlik.

Este software permite asociar tablas con un gran volumen de datos, como ya se ha dicho, y obtener resultados de una manera sencilla y eficaz. Además, utiliza un lenguaje de programación bastante intuitivo y lógico, que tiene similitudes con el lenguaje SQL.

Al asociarse las tablas, este programa permite enlazar todas ellas y representarlas de manera visual en un cuadro de mando.

Las opciones son infinitas y el cuadro de mando realizado, es un ejemplo de muchos que se podrían haber construido ya que, consta de muchas opciones no solo de gráficos sino también de asociación de tablas y construcción de indicadores.

Cabe destacar que dicho software es de pago, pero gracias a cursar un grado universitario, puedes inscribirte y disfrutar de las ventajas que tiene prácticamente al completo. Los gráficos dinámicos, por ejemplo, no entran en esta versión para estudiantes.

## 3. CUADROS DE MANDO

### 3.1. Origen y definición

Para cualquier organización es importante avanzarse a la competencia y a los futuros acontecimientos que puedan afectar a su funcionamiento. Por ello, disponer de una visión global de la entidad es clave.

Una herramienta útil que permite monitorizar todos los parámetros de la empresa disponiendo así de una imagen real y completa de lo que sucede es el Cuadro de Mando (CdM), en inglés *Balanced Scorecard*. De hecho, un cuadro de mando es una herramienta de gestión empresarial que permite medir y visualizar el rendimiento de una entidad de una manera clara y sencilla.

Nace en los años 90, como necesidad de los directivos de disponer de información en tiempo real y de manera estructurada, para poder tomar decisiones estratégicas siguiendo su plan estratégico. Pero este no fue su verdadero inicio, anteriormente surgieron ideas similares.

Entre los años 50 y 80 existieron al menos cuatro proyectos parecidos. El primero de ellos y más antiguo, es el que se realizó en la compañía General Electric en el que se utilizó una única medida financiera junto con siete que no lo eran.

A este primer contacto con el estudio e inclusión de las métricas no financieras, le sigue el que realizó Herbert A. Simon junto con otros investigadores en el que detectaron varios propósitos para la información contable de las entidades. Este fue el primero en introducir el término *Scorecard* en la discusión de la gestión del rendimiento.

Pocos años después, Peter Drucker en su libro *The Practice of Management* argumentó que cada uno de los empleados debería tener unos objetivos de desarrollo personal y que, al mismo tiempo, estos estuviesen en línea con los de la organización.

Por último, Robert Anthony siguiendo el segundo proyecto anteriormente explicado, identificó tres tipos diferentes de sistemas de planificación y control: planificación estratégica, control de gestión y control operacional. Tras este estudio, concluyó que la planificación estratégica depende de una estimación que relaciona causa y efecto pero que esta, es un arte y no una ciencia.



Figura 1: Relevancia de la teoría de Peter Drucker

La idea más cercana de lo que hoy en día se conoce como un cuadro de mando surgió, como ya hemos mencionado, en los años 90 debido a la necesidad de buscar técnicas más efectivas e innovadoras para cumplir los objetivos establecidos por organizaciones de distintos sectores.

Los cuadros de mando iniciales se realizaban trimestralmente y estaban basados exclusivamente en datos financieros, lo que nos suponía una pérdida de visibilidad en el negocio.

Robert S.Kaplan y David P.Norton, dos economistas norteamericanos que trabajaron juntos en labores de consultoría, consideraban que para un mejor desarrollo de cualquier entidad, no eran suficiente los indicadores financieros tradicionales. Ellos creían que era necesario incluir también otro tipo de indicadores, como el tiempo de servicio al cliente o la calidad del producto en sí, ya que sin ellos se perdían de vista muchos activos intangibles.

Por ello, se designaron 4 perspectivas necesarias para tener una mayor amplitud del conocimiento del negocio que juntaban las medidas financieras, ya utilizadas hasta ese momento, con las medidas de los indicadores de rendimiento futuro. Este conjunto, es lo que crea el *Balanced Scorecard* (lo que hoy llamaríamos Cuadro de Mando) en el que, las medidas financieras te hablan del pasado mientras que, los nuevos indicadores te indican tu rendimiento futuro.

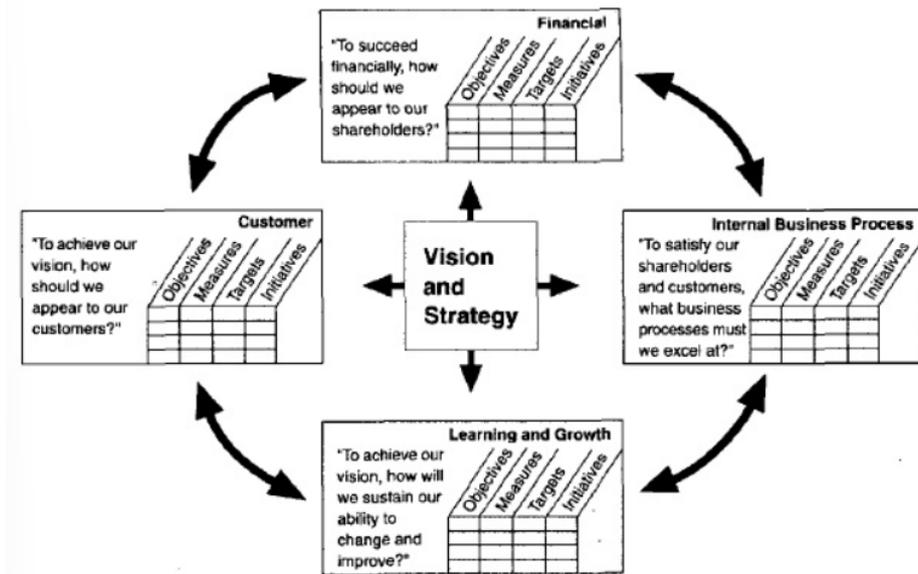


Figura 2: The Balanced Scorecard Provides a Framework to Translate a Strategy into Operational Terms

Hablemos de las anteriormente nombradas 4 perspectivas que establecieron:

- Perspectiva financiera. Directamente relacionada con los indicadores financieros, permiten medir las consecuencias económicas de las acciones que se toman con el objetivo de maximizar el valor y crecimiento del número de ventas, reducir costes y mejorar la productividad.
- Perspectiva del cliente. Entrarían medidas como la satisfacción del cliente, captación de estos o la rentabilidad. Refleja el posicionamiento de la empresa en los segmentos del mercado en los que se quiere competir.

El objetivo es la imagen, reputación, captación de nuevos clientes, así como de mantener a los que ya lo son.

- Perspectiva interna. Incluye los procesos operativos, desde que la empresa recibe el pedido hasta que este llega a él, maneras de hacer esta fase de una forma más eficaz, rápida, cumpliendo las necesidades y demandas de los clientes, pero al menor coste posible. Por este último motivo, esta perspectiva está muy relacionada con la parte financiera.

El objetivo es aplicar mejoras en la cadena de valor interna (innovaciones y servicios postventa).

- Perspectiva de aprendizaje y crecimiento. Tal y como su nombre indica, se enfoca en la vida futura de la organización. Para ello, se tiene en cuenta el capital humano (capacidades y competencias), el marketing (redes y capacidad de infraestructuras para informar) y la organización de la empresa para realizar todos estos procesos.

Tiene como objetivo mejorar la capacidad del personal, sistemas y otros procedimientos.

Tras la innovadora idea de los economistas, en los 4 años siguientes, varias empresas utilizaron su modelo y obtuvieron resultados al instante. De hecho, Kaplan y Norton descubrieron que no sólo utilizaban los CdM para analizar y preveer, sino que mediante ellos, comunicaban estrategias.

Desde entonces, el Cuadro de Mando ha ido ganando fuerza y ha sido implementado por gran parte de las organizaciones adaptando cada una el cuadro a sus necesidades, estrategias y previsiones ya que no tiene por qué ser un modelo fijo y se pueden añadir, modificar e incluso eliminar alguna de estas perspectivas explicadas. La idea es que, se utilicen o no las cuatro perspectivas, el conjunto de medidas que conforman el *Scorecard* ha de ser equilibrado.

Por tanto, se puede definir un cuadro de mando como un conjunto de medidas seleccionadas minuciosamente, derivadas de la estrategia empresarial establecida, que permite comunicar no solo a la entidad sino también a los competidores los resultados y objetivos cumplidos de esta.

A pesar de ser, "simplemente" un conjunto de medidas, si está bien construido debe incorporar relaciones causa-efecto e inductores de actuación:

- **Relaciones causa-efecto.** Las estrategias son hipótesis. Por ello, el CdM debe establecer de forma explícita las relaciones entre las diferentes perspectivas para que puedan ser gestionadas y validadas. Por ejemplo, podemos suponer que si se incrementa la fidelidad de los clientes, los ingresos también incrementarían.
- **Inductores de la actuación.** Un buen CdM debe tener tanto indicadores históricos, es decir, medidas de resultados, como también provisionales (inductores de actuación). Estos últimos son los indicadores que nos comunican la manera en la que conseguir resultados.

Las posibles razones por las que una entidad quiere integrar un CdM son:

- Fijar nuevos cumplimientos.
- Coordinar iniciativas de mejora.
- Clarificar la estrategia que se sigue en la actualidad.
- Implementar una nueva estrategia.
- Coordinar los objetivos de los empleados.
- Comunicar e informar.
- Existencia de crisis interna.

Los CdM se componen de una serie de indicadores clave de rendimiento (KPI, por sus siglas en inglés) que son los que nos permiten medir el desempeño de la organización en las diferentes perspectivas. Estos se representan en un panel o dashboard, que facilita la interpretación y análisis de la información por parte de los directivos.

### 3.2. Selección de los indicadores (KPIs)

Siguiendo la idea inicial, veremos los indicadores más relevantes que se utilizarían para realizar el seguimiento en cada uno de los cuatro ámbitos.

Un indicador de rendimiento es aquel que nos permite conocer si una tarea se está llevando a cabo de una manera eficiente o si, por el contrario, este está sufriendo una desviación. Estos, variarán su presencia en el cuadro ya que, los objetivos y necesidades de cada una de las distintas industrias cambian constantemente.

Esta selección es muy importante pues, si no se eligen correctamente, podríamos estar perdiendo aspectos importantes a estudiar para cumplir los objetivos de la entidad.

Se recomienda que el número de indicadores que compongan un CdM, no supere los veinte dado que, un sobre seguimiento de indicadores puede desviarnos de un correcto seguimiento de las medidas.

Una manera útil de seleccionarlos es utilizando el método del semáforo. Se basa en clasificar estos KPIs en tres colores: verde, amarillo y rojo, en función de su desviación. Estas métricas aparecerán de forma numérica y gráfica para facilitar su interpretación.

Valorar, tras esta pequeña explicación, diferentes indicadores que podemos incluir en las diferentes perspectivas.

#### 3.2.1. Ejemplos indicadores financieros básicos

Los podemos dividir en tres bloques:

- Indicadores de rentabilidad:
  - Margen de utilidad bruta: ganancias con las ventas
  - Margen operacional: relación de las ventas con los costes
  - Margen de utilidad neto: conocimiento de los beneficios netos
- Indicadores financieros de Liquidez:
  - Razón corriente: nos da una perspectiva de la capacidad de pago de las obligaciones de corto plazo del negocio
  - Capital de trabajo: nos permite conocer nuestras capacidades de llevar a cabo ciertas actividades dentro de una entidad con normalidad en un corto plazo.
  - Prueba ácida: nos permite conocer nuestra capacidad de pago (solvencia) en el corto plazo.
- Indicadores de Endeudamiento:
  - Endeudamiento: Ayuda a medir y comprender el riesgo y nos sirve para interpretar y analizar la proporción de bienes que están financiados por deuda.
  - Fondos de maniobra: nos permite saber el grado de solvencia del negocio y la capacidad de este para liquidar deudas.

### 3.2.2. Ejemplos indicadores relevantes relacionados con los clientes

- Satisfacción del cliente (CSAT): Es el indicador que nos permite conocer el grado de satisfacción del cliente de aquellos productos o servicios que ofrece la entidad. Normalmente, mediante una encuesta telefónica, se le permite al cliente valorar su experiencia en una escala del uno al cinco.
- Tiempo promedio de resolución: nos permite saber el tiempo promedio que un cliente necesita para solucionar completamente su problema o cuestión.
- Tasa de retención de clientes: Cuánto más tiempo lleve dado de alta un cliente en la entidad, menos probable será que confíe en otras organizaciones para adquirir productos que esta, ya le ofrezca.

### 3.2.3. Ejemplos indicadores relevantes relacionados con los procesos internos

- Innovaciones: innovar en servicios, en procesos, en organización y en comercialización, es clave para mantener a clientes que ya confían en la entidad, pero también para atraer a aquellos que buscan algo innovador que no todas las organizaciones de ese mismo sector ofrecen.
- Gestión del valor ganado (EVM): Empezar nuevos proyectos implica un coste. Conocer los gastos de dichos proyectos nos permite ver si estamos cumpliendo el presupuesto inicial establecido o si, por el contrario, nos estamos excediendo.

Para seguir este gasto, visualizamos gráficamente el tiempo y los costes incluyendo en dicho gráfico el BAC (presupuesto total establecido inicialmente), el PV (valor planificado, es decir, la cantidad pensada en gastar en el proyecto) y el EV (valor ganado).

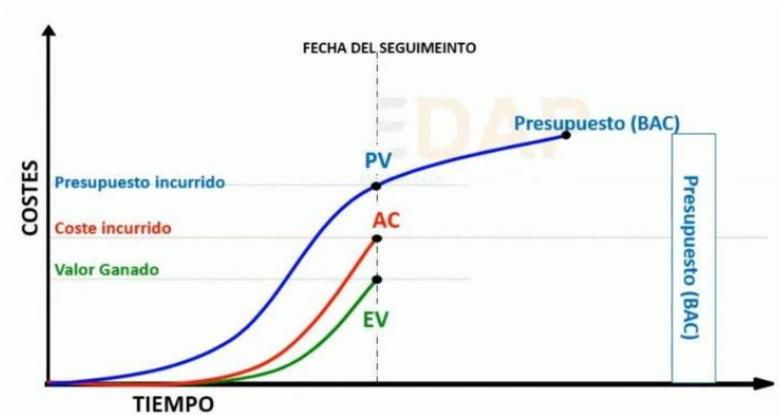


Figura 3: Gráfico Seguimiento Indicador EVM

### 3.2.4. Ejemplos indicadores relevantes relacionados con la formación y crecimiento

- Capacitación: Para poder tener procesos innovadores, hace falta que el personal este preparado para estas actualizaciones. Por ello, formar al personal no sólo para un mejor ambiente laboral sino también para que sea competente es clave en el crecimiento de una entidad. Posteriormente, ver si esta capacitación ha resultado eficiente, se pueden realizar encuestas a los empleados o ver el porcentaje de abandonos de los cursos ofrecidos.
- Inversión por empleado (ICE): saber hasta qué punto podemos invertir en el crecimiento de la plantilla es importante para no gastar más de lo que luego obtendremos de recompensa.

Se calcula como:  $ICE = \frac{\text{Cantidad de dinero invertido en capacitación}}{\text{Cantidad promedio de empleados}}$

- Competitividad: ver dónde se sitúa una empresa respecto a otras de su mismo sector es clave para saber cómo de relevantes somos en el mercado. Para el cálculo de la cuota de participación en el mercado, también conocida como Market Share, hacemos:

$$\text{Market Share} = \frac{\text{Venta total de la empresa en un periodo}}{\text{Venta total del mercado en un periodo}}$$

### 3.3. Clasificación de Cuadros de Mando y softwares

Existen dos maneras de poder clasificar los CdM: según su **función** y según su **contenido**.

#### 3.3.1. CdM según su función

Siguiendo esta clasificación encontramos 4 clasificaciones:

- Cuadro de Mando Integral
- Cuadro de Mando Ejecutivo
- Cuadro de Mando Analítico
- Cuadro de Mando Operativo

Veamos más detalladamente cada una de ellas.

- Cuadro de Mando Integral (CMI)

Tal y como se ha explicado en un inicio, el CMI es el concepto que creó Kaplan y Norton en el 1992.

A pesar de que a menudo se confunda el Cuadro de Mando Integral con el Cuadro de Mando o dashboard veamos en que consiste el primero mencionado y, en el siguiente apartado, encontraremos la definición del segundo.

Este cuadro se caracteriza por dar una visión total del rendimiento general de la organización, es decir, representa la actividad empresarial en su totalidad. Por ello, debe incluir los ya mencionados indicadores de rendimiento más relevantes para que cada departamento pueda analizarlos y cumplir así los objetivos empresariales establecidos. Esta información debe ser captada visualmente.

Un buen CMI debe incluir:

- Un mapa estratégico del plan estratégico.
- Una matriz que abarque todos los indicadores.
- Gráficos y otros visuales que permitan la comprensión de los datos sin analizarlos en profundidad.

Por tanto, el objetivo de forma resumida es evaluar el crecimiento, rendimiento y objetivos estratégicos de la entidad para alcanzar así los objetivos preestablecidos.

- Cuadro de Mando Ejecutivo

Con el fin de tomar decisiones de una forma rápida, clara y concisa, estos cuadros a diferencia de los anteriores, no contendrán toda la información de la entidad sino que nos focalizaremos en ciertos sectores, productos o departamentos.

Por ello, la misión en la que se basan es en representar la información de forma clara, concisa y en una sola *slide* por lo que la visualización de los indicadores y los comentarios que los acompañan son clave para que dicha misión se cumpla. De hecho, no debe suscitar la exploración profunda de los datos.

Concluimos por tanto que, un buen CdM Ejecutivo debe ser muy visual, claro, conciso y debería llevar a las mismas conclusiones independientemente del directivo que lo lea.

- Cuadro de Mando Analítico

A diferencia de los anteriores, el CdM Analítico está pensado no para dar una información clara y concisa, sino que, al ser creado para analistas, debe incluir la mayor cantidad de datos posibles de la totalidad de la organización.

De esta manera, los encargados pueden hacer predicciones, comparar variables, identificar nuevas tendencias, etc. Por ello, los evolutivos, la exploración en profundidad y la comparativa son elementos esenciales para estos cuadros.

Por estas razones, son más complejos e interactivos de manera que, desde el inicio, se pueda ir desglosando e indagando en los datos con el objetivo, de estudiar y analizar hacia qué dirección va el negocio.

En definitiva, su objetivo no es sólo dar información de una forma clara, sino que es más importante que incluya la mayor cantidad posible de datos para así acercar a la organización a sus objetivos estudiando posibles tendencias y comparando resultados con periodos anteriores.

- Cuadro de Mando Operativo

También conocido como cuadro operacional, permite controlar y hacer un seguimiento de todas las operaciones empresariales en curso. Para esto, es imprescindible la presencia de indicadores de rendimiento operacionales destacando así las métricas más relevantes relativas a dichos procesos en curso.

El objetivo de estos cuadros es poder tomar decisiones urgentes de estas operaciones en curso y, por ello, esta herramienta es importante para que se puedan detectar posibles anomalías, errores o áreas de mejora en el funcionamiento de estas operaciones.

Al igual que en los cuadros de mando analíticos, los datos históricos son imprescindibles con la diferencia de que las comparaciones se realizan en periodos de corto plazo comparando así datos desde hace un mes hasta un cuatrimestre o semestre.

En resumen, estos cuadros son una herramienta para seguir aquellos procedimientos que la entidad tiene activos y poder tomar decisiones al respecto de manera urgente.

### 3.3.2. CdM según su contenido

Siguiendo esta clasificación encontramos 4 clasificaciones:

- Business Activity Monitoring (BAM)
- Dashboarding
- Cuadro de Mando Comercial
- Cuadro de Mando Financiero

Veamos más detalladamente cada una de ellas.

- Business Activity Monitoring (BAM)

Tan importante es saber tu situación actual como poder predecir qué será de la entidad en un futuro.

Para ello, BAM nos permite conocer el comportamiento de un procedimiento y su evolución registrando toda la actividad empresarial mediante los KPIs, que pueden mantenerse, modificarse, incluirse nuevos o eliminarse, y analizándolos con la finalidad de maximizar su rentabilidad.

Gracias a este registro en tiempo real, podemos identificar posibles futuros escenarios que requieran de alguna modificación en las fases ya establecidas.

- Dashboarding

Suele confundirse con un CMI pero, sin embargo, un dashboard es utilizado para representar un conjunto de datos relacionadas con un aspecto concreto de la entidad y, normalmente, sirve para hacer un seguimiento de los objetivos específicos del proyecto o aspecto concreto. A diferencia del CMI que proporciona una visión global del plan estratégico y de acción.

Este tipo de cuadro de mando, podríamos definirlo como la idea de cuadro de mando inicial en el que solo se estudiaban métricas concretas hasta que Kaplan y Norton estudiaron la introducción de muchas y diferentes métricas para tener una visión más global de la organización.

- Cuadro de Mando Comercial

Es utilizado, tal y como indica su nombre, para medir y controlar el desempeño comercial de la organización, mediante la evaluación de indicadores como el número de clientes, ventas, margen de beneficio, etc.

- Cuadro de Mando Financiero

Como su nombre indica, se utiliza para controlar y medir el desempeño financiero, mediante la evaluación de KPIs económicos y financieros como ingresos, gastos, etc.

### 3.4. Cuadros de Mando y el mundo financiero

Ya hemos visto que los CdM son una herramienta muy útil para tener una visión actual y futura de la entidad, pero ¿cómo de relevante es para una entidad financiera este control?

Debido al hecho de que hablamos de una organización que trabaja con dinero para ganar más dinero, es mucho más complicado no focalizarse simplemente en indicadores financieros.

Para la banca, la definición más ajustada de un cuadro de mando sería que estos han de ser una representación gráfica del progreso a lo largo de los años no sólo de la totalidad de la entidad sino también de un empleado, un objetivo o alguna meta.

Veamos como la idea inicial de Kaplan y Norton toma forma dentro del sector bancario en función de las cuatro perspectivas que definieron.

- Perspectiva Financiera. Centrada en los ingresos y beneficios para compararlos con resultados anteriores, el presupuesto decidido, la competencia y también con las rentabilidades de otras sucursales de la misma organización.
- Perspectiva del Cliente. Focalizada en la relación de los clientes y el mercado. Se analiza el crecimiento de los clientes y la cuota de mercado y tiene como objetivo principal el aumento de la cuota de cartera del banco entre sus clientes.
- Perspectiva del Proceso Interno. Algunas medidas de inversión en empleados, nuevas tecnologías y edificios serían los activos o depósitos y hay que vigilar que estas no sean demasiado elevadas.
- Perspectiva de Aprendizaje y Crecimiento. Clasificado en tres componentes: capital humano, de la información y organizativo, es la perspectiva más difícil de seguir y la más subjetiva dado que todo depende de la empresa en la que nos encontremos y lo que se quiera potenciar de esta.

Debido al servicio que venden, tienen un riesgo superior al de otras organizaciones pertenecientes a otros sectores: el impago de sus clientes.

Incluir este indicador en el plan estratégico y derivar una parte del presupuesto a ello es esencial para no sufrir pérdidas con las que no se contaban y que, a causa de ello, se pierdan proyectos o incluso que se llegue a la banca rota. Pues para evitar estos impagos, a los clientes se les aplican varios filtros de riesgo y son clasificados en distintos grados de probabilidad de impago en función del salario o estado civil entre otras características. De esta manera, conseguimos que la probabilidad como empresa de sufrir impagos sea la más pequeña posible.

El impago de los clientes, que pueden ser tanto personas físicas como jurídicas, se recoge en los cuadros de mando como el indicador de la morosidad o mora.

### 3.5. La importancia del control de la morosidad

El control de la morosidad es fundamental para el buen funcionamiento de cualquier empresa que ofrezca productos o servicios a crédito.

Se conoce como morosidad el hecho de que un deudor, cliente, no realice los pagos en las fechas previamente establecidas. Cuando este impago sucede, el deudor pasa a formar parte de una lista o fichero de morosos. En dicha lista, no sólo se entra por un impago con una deuda bancaria sino también por un impago en cualquier otro recibo de cualquier otra compañía. Por ejemplo, un impago en tu factura del agua o gas.

Al tratarse de un tema delicado, no es sencillo entrar en dichas listas, para ello debe existir un documento vinculante en el que se muestren las condiciones y fechas de dicho pago de manera detallada.

Como ya se ha mencionado, antes de pedir un préstamo, contratar un seguro o cualquier otro servicio o producto que se ofrezca, es imprescindible para la entidad conocer el nivel de ingresos y gastos, pues no nos vale que una persona tenga un alto nivel de ingresos si esta vive por encima de sus posibilidades, y lo que determina la capacidad de endeudamiento.

Hay distintos tipos de morosos y se pueden clasificar en cinco grupos:

- Morosos fortuitos. A pesar de querer pagar, en el momento del pago estos carecen de liquidez.
- Morosos circunstanciales. También tienen intención de pagar, pero por un problema circunstancial no pueden hacerlo.
- Morosos intencionales. En este caso, las personas tienen capacidad para realizar el pago pero no intención de hacerlo. Por ello, retrasan al máximo el pago.
- Morosos desorganizados. Son aquellos que tienen intención de cumplir, pero al llevar una mala gestión financiera, desconocen las fechas de estos pagos y por ello no lo realizan a tiempo.
- Morosos negligentes. Nos encontramos en el mismo caso que los anteriores con la diferencia de que además de su desorden, tampoco tienen intención de pagar sus deudas.

Que un cliente sea considerado moroso, no implica la negación al pago sino que simplemente, por el momento, todavía no lo ha realizado. El problema surge si esta situación no se regula dado que es cuando se entra en los ficheros de morosidad.

En caso de que haya una incapacidad de pago de un préstamo o crédito bancario, se recurre a solicitar una refinanciación de la deuda del cliente por la entidad financiera siempre y cuando el cliente tenga capacidad de pago. En caso contrario, puede ejecutarse una garantía, como podría ser la vivienda, o ir a los juzgados.

Cuando una entidad financiera no controla y regula su tasa de mora, es decir, presenta elevados índices de esta, está bajo un proceso de insolvencia crediticia que puede ocasionar procesos irreversibles e incluso la quiebra del banco.

Cabe destacar que la morosidad de la banca en España se mantiene por debajo del 4%. Eso supone una bajada respecto el año pasado, en el que este porcentaje superaba el 4%.

Veamos, las principales razones por las que es crucial el control de este indicador.

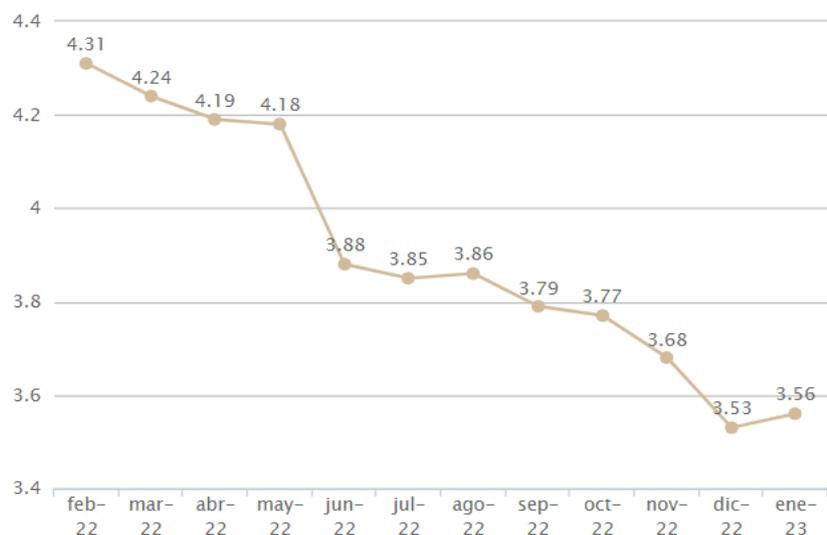


Figura 4: Morosidad española desde enero de 2022 hasta enero de este año. % de créditos dudosos sobre el total – Diario El Público, 24 de marzo de 2023

- 1. Mejora de la liquidez.** La mora afecta directamente a la capacidad de la empresa para generar efectivo y hacer frente a sus obligaciones financieras. Al controlarla, nos podemos asegurar que los ingresos lleguen en tiempo y forma, teniendo así una mejor gestión del flujo de caja.
- 2. Reduce los costos financieros.** El impago de los clientes genera una mayor carga financiera para la empresa, pues estos impagos suponen costes de personal dedicados a recobrar los impagos, costes judiciales, de cartas, etc.
- 3. Protección de la solvencia de la empresa.** La morosidad puede generar un desequilibrio en la estructura financiera de la empresa suponiendo un riesgo para su solvencia y estabilidad. Al estar controlada, podemos prevenir situaciones de impago que puedan afectar al patrimonio y capacidad de endeudamiento de la empresa.
- 4. Mejora de la reputación.** El impago, puede afectar negativamente a la imagen y reputación de la empresa ante los proveedores y mercado en general. Al controlarse, la imagen que se transmite es de solidez y confiabilidad en el cumplimiento de sus obligaciones financieras.

## 4. CREACIÓN DEL CUADRO DE MANDO

Para la parte práctica del trabajo se ha creado un cuadro de mando mediante que tiene como objetivo el estudio y evolución de la morosidad en una entidad financiera.

Para dicho estudio, se ha seleccionado una base de datos y un software concretos que se explicarán detalladamente en este apartado.

### 4.1. *Credit Bureau*

Antes de empezar con la introducción y explicación de la base de datos y origen de ellos, para entender parte de su procedencia, se explicará que es Credit Bureau.

Debido a que la idea del proyecto es estudiar la morosidad, conocer si el cliente impagará o no, es clave para el desarrollo del negocio pues puede suponer una gran pérdida económica para la entidad.

Por ello, saber si el cliente ha solicitado otros préstamos y si parte de ellos siguen en curso, entre otras casuísticas, ayuda a conocer la probabilidad de impago de la persona a la que se le ofrece el producto. Aquí es donde Credit Bureau hace acto de presencia en nuestros datos.

Los Credit Bureaus son instrumentos de apoyo clave en las decisiones de las concesiones o no de créditos pues contienen mucha información sobre el deudor permitiendo tener así el conocimiento sobre la probabilidad de impago. Usualmente, estos sistemas de medición de riesgo abastecen una predicción de 24 meses, no obstante, la estimación podría ser de la mitad.

Para facilitar el proceso de tratamiento de datos, las entidades cuentan con el *scoring de bureau* que convierte la gran cantidad de datos y variables de los deudores en perfiles de riesgo que calificarán la historia del deudor y se tendrá en cuenta para la decisión de concederle o no el servicio.

Algunas de las variables que se incluyen dentro del *scoring* podrían ser:

- Situación económica
- Tipos de consumos y hábitos
- Datos sobre el ciclo vital

Hay empresas que se dedican específicamente a la recopilación de estos datos, pero para poder hacer uso de los servicios que estas ofrecen, la entidad que así lo requiera, deberá compartir también los

datos de sus usuarios. Por ello, cuantos más asociados tenga la empresa, mayor importante será pues creará perfiles de riesgo mucho más ajustados a la realidad.

En España RAI o ASNEF-Equifax son dos de las más importantes pues la primera se centra solo en información referente a organizaciones mientras que la segunda suele ser el gestor más conocido ya que recoge los datos de todas las personas físicas; datos que otras empresas relacionadas con el sector financiero le han aportado.

#### 4.2. *Plataforma Qlik Sense*

Si bien era importante conocer la existencia y función de los Credit Bureau, es necesario también conocer la herramienta con la que se ha realizado el cuadro de mando.

Qlik Sense es una plataforma de análisis y visualización empresarial de datos que permite a los consumidores explorarlos y analizarlos. El objetivo es permitir a las entidades trabajar con grandes volúmenes de datos y convertirlos en información fácil de comprender.



*Ilustración 1: Logotipo Entidad Qlik*

Con su interfaz intuitiva y fácil de usar, se pueden crear visualizaciones interactivas, tableros de control y aplicaciones de análisis de datos permitiendo así la explotación de datos de una manera dinámica. También ofrece capacidades de BI (Business Intelligence) que permite incorporar datos desde diferentes fuentes (hojas de cálculo, archivos o servicios en la nube).

Debido a su capacidad de asociación de datos no existen restricciones predefinidas entre los datos de manera que, se pueden descubrir relaciones y conexiones ocultas entre ellos obteniendo así un análisis más intuitivo y flexible ya que es el propio consumido el que elija que camino de datos le interesa seguir en tiempo real.

Además, gracias a la inteligencia artificial y aprendizaje automático que incorpora, se proporcionan recomendaciones y sugerencias a los usuarios durante el análisis.

Además, Qlik Sense incorpora capacidades de colaboración lo que permite poder compartir visualizaciones, tableros y aplicaciones con otros usuarios facilitando así la toma de decisiones en equipo.

Junto a Qlik Sense tenemos Qlik View, una plataforma también de BI de la misma compañía pero que presenta distintas diferencias respecto a la primera.

Al igual que Qlik Sense, Qlik View presenta un enfoque de análisis asociativo similar al de Sense permitiendo también la presentación de los datos de forma interactiva.

Una característica distintiva es su modelo de datos de memoria pues, Qlik View permite cargar los datos en memoria teniendo así un fácil y rápido acceso a ellos para el análisis. Muy útil sobre todo cuando se trata de grandes volúmenes de datos.

Para este trabajo, se ha utilizado un programa para estudiantes, con duración de un año, llamado Academic Program (Qlik Cloud). Esta plataforma ofrece los servicios tanto de Qlik View como de Qlik Sense con la ventaja de que podemos trabajar los datos desde cualquier lugar o dispositivo ya que está en la nube y de forma gratuita por ser estudiantes.

#### 4.3. *Nuestra Base de Datos*

Para que el estudio fuera lo más cercano a un caso real que se puede encontrar una entidad financiera, la idea inicial era poder contar con una base de datos extensa, completa y relacionada con el mundo financiero.

Gracias a la plataforma Kaggle, una plataforma gratuita que pone a disposición de cualquier usuario una serie de problemas para solucionarse mediante ciencias de datos, análisis predictivo y *machine learning*, encontramos una base de datos que cumplía con dichas características.

La entidad no financiera Home Credit, enfocada en préstamos a plazos dirigidos a personas con un bajo o nulo historial crediticio, con el objetivo de ampliar la inclusión financiera de la población no bancarizada de una manera positiva y segura, utiliza una variedad de datos alternativos para predecir la capacidad de pago de sus clientes.

La empresa como tal, utiliza ya varios métodos estadísticos y de aprendizaje automático para hacer estas predicciones, pero subieron estos datos a internet con la finalidad de desafiar a los usuarios de Kaggle a encontrar todo el potencial de sus datos.

Para la elaboración del trabajo se decidió utilizar esta base de datos por tres grandes motivos:

- Al ser una base de datos de una entidad financiera, dispone de las características reales a nivel cliente y operación de los impagos que puede sufrir un banco.
- Es extensa, tanto en volumen de operaciones como en información para cada una de ellas.
- Permite tener una foto completa del riesgo de los clientes.

Dicha base se conforma de un total de 10 tablas que se describen a continuación:

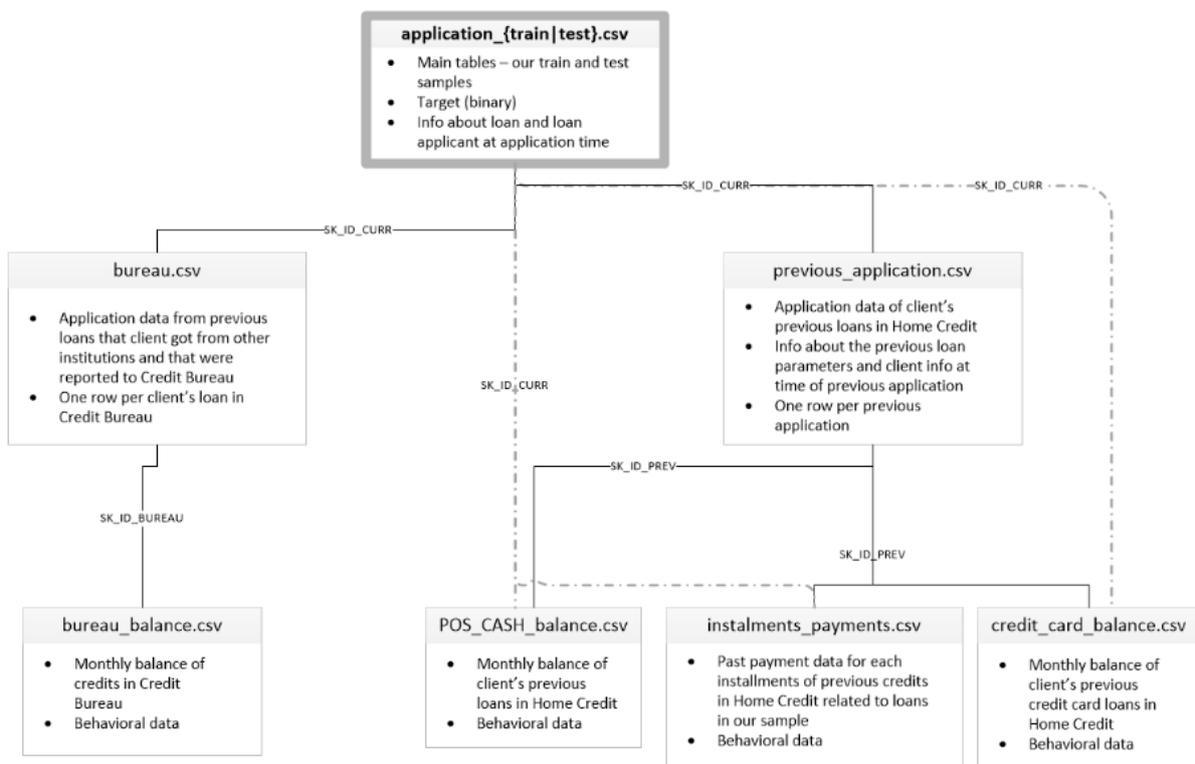


Figura 5: Diagrama entidad relación de las tablas de la base de datos - Kaggle

Tal y como muestra el diagrama anterior vemos que todo parte de *application\_{train|test}.csv* que incluye las tablas de entrenamiento y tests, el público objetivo y la información del préstamo y del prestatario en el momento de la solicitud de dicho préstamo.

De este documento, nacen dos vertientes: la primera relacionada con anteriores préstamos que ha hecho el cliente en otras entidades y el balance de créditos y la segunda que contiene datos de préstamos anteriores que el cliente ha solicitado en el propio banco incluido el tipo de cliente que era en el momento de dicha solicitud.

Veamos al detalle las tablas:

- *application {train/test}.csv*. Es la tabla principal que se divide en dos archivos uno para el entrenamiento que contiene el público objetivo y otro para el test que no contiene dicho público.

Cada línea de dichas tablas hace referencia a un préstamo diferente.

Como estos datos están pensados para la realización de un concurso en el que usuarios creen modelos para categorizar si un futuro cliente impagará o no, la base de datos de test no contiene toda la información necesaria como por ejemplo si el contrato finalmente se impaga. Por este motivo, únicamente se usará la tabla de *application train* que contiene más de 300 mil registros.

- *bureau.csv*. Contiene la información de otros préstamos pedidos por clientes que tienen con Home Credit un préstamo en otras entidades.

Por cada muestra hay tantas filas como número de créditos el cliente tenía en *Credit Bureau* antes de la fecha de solicitud.

- *bureau\_balance.csv*. Contiene balances mensuales de créditos previos de Credit Bureau.

La tabla contiene una fila por cada mes de historia de cada crédito previo que ya constaba en Credit Bureau. Es decir, la tabla tiene (nº de préstamos \* nº de préstamos previos \* nº de meses donde se tiene alguna operación previa del crédito) filas.

- *POS CASH balance.csv*. Contiene los balances mensuales de las operaciones e ingresos que el cliente tenía con Home Credit.

Esta tabla tiene una fila por cada mes recogido por cada crédito en Home Credit (créditos al consumo y préstamos), es decir, la tabla tiene (nº de préstamos \* nº de créditos previos \* nº de meses donde se tiene alguna operación previa del crédito) filas.

- *credit\_card\_balance.csv*. Contiene el balance mensual de las operaciones que se realizaron con la tarjeta.

Al igual que en las tablas anteriores, tenemos una fila por cada mes de historia de cada crédito previo que ya constaba, es decir, la tabla tiene (nº de préstamos \* nº de tarjetas de crédito previas \* nº de meses donde se tiene alguna operación previa al crédito) filas.

- *previous\_application.csv*. Todas las operaciones de los préstamos de clientes que tienen préstamos en la muestra.

Cada fila corresponde a una operación del crédito.

- *installments\_payments.csv*. Encontramos el historial de repago de los créditos desembolsados previamente relacionados con los créditos de la muestra.

A diferencia del resto de tablas, esta tiene una fila a por cada pago y una b por cada impago. Por tanto, cada fila equivale a un pago de una cuota o a una cuota correspondiente a un pago de un crédito de Home Credit anterior al relacionado con préstamos en la muestra.

#### 4.3.1. Tablas y variables

Una vez conocemos que información hay en cada tabla veamos las variables que recoge cada una de ellas y que nos permitirán realizar el CdM para el control de la morosidad.

La totalidad de las variables se encuentra en el anexo, en este apartado se explican aquellas que han dado forma a nuestro cuadro de mando.

Para más detalle sobre las variables restantes de la base de datos, consultar el anexo 7.1.

**1. application\_train.csv.** Estas son las variables que se han utilizado para crear el cuadro.

- AMT\_CREDIT: Monto del crédito del préstamo.
- AMT\_ANNUITY: Anualidad del préstamo.
- AMT\_GOODS\_PRICE: Es el precio de los bienes por los que se le dio al cliente el préstamo.
- CODE\_GENDER: El género del usuario.
- DAYS\_BIRTH: La edad del usuario cuando realizó el préstamo.
- DAYS\_EMPLOYED: Recuento de los días previos al contrato trabajados en el lugar de trabajo actual.
- NAME\_INCOME\_TYPE: Profesión del usuario.
- NAME\_EDUCATION\_TYPE: Último nivel de estudio alcanzado del usuario.
- NAME\_HOUSING\_TYPE: Situación de habitabilidad del cliente (si es propietario, vive de alquiler, con sus padres, ...).
- REGION\_RATING\_CLIENT: Indica la provincia en la que vive (dividido en 1, 2 y 3).
- REGION\_RATING\_CLIENT\_W\_CITY: Indica la provincia en la que vive incluyendo la ciudad (dividida en 1, 2 y 3).
- WEEKDAY\_APPR\_PROCESS\_START: El día de la semana en el que el cliente realizó el préstamo.
- SK\_ID\_CURR: ID del préstamo en nuestra muestra.
- TARGET: Variable objetivo en formato binario: 1 cuando el cliente presenta dificultades de pago, es decir, se atrasó más de X días en al menos una de las primeras Y cuotas del préstamo y 0 para el resto de casos.

Ya que eran dos tablas creadas para con una crear un modelo y entrenarlo y la otra para comprobar si este modelo funcionaba correctamente, se quiso fusionar ambas, es decir, añadir "debajo" de la de entrenamiento, la tabla de test.

Ambas tablas contenían las mismas variables a excepción de la variable *target* que sólo está presente en la de entrenamiento. Por este motivo, esta fusión no se realizó ya que como la parte del test era para comprobar si estos futuros clientes podrían llegar a impagar, no tenían la variable *target*, que es una de las que se implementó en el CdM.

**2. bureau.csv.** De dicha tabla se consideraron 13 variables como relevantes para conformar el CdM:

- SK\_ID\_CURR: ID del préstamo en nuestra muestra. Cada préstamo puede tener más de un crédito previo en Credit Bureau (CB).
- SK\_BUREAU\_ID: Recopilación de los préstamos previos de CB relacionados con nuestro préstamo (ID único para cada solicitud de préstamo).
- CREDIT\_ACTIVE: Estado de los créditos reportados por CB

- DAYS\_CREDIT: Número de días previos a la solicitud actual en los que el cliente solicitó un préstamo a CB.
- DAYS\_CREDIT\_ENDDATE: Duración de días restantes del crédito de CB en el momento de la solicitud en Home Credit.
- DAYS\_ENDDATE\_FACT: Días desde que se terminó el crédito en CB en el momento de la solicitud en Home Credit.
- AMT\_ANNUITY: Anuidad del crédito de Credit Bureau.
- AMT\_CREDIT\_SUM: Monto de crédito actual del crédito de Credit Bureau.
- AMT\_CREDIT\_SUM\_DEBT: Deuda actual en el crédito de CB.
- AMT\_CREDIT\_SUM\_LIMIT: Límite de crédito actual de la tarjeta reportado a CB.
- CNT\_CREDIT\_PROLONG: Número de veces que el crédito en Credit Bureau fue prolongado.
- CREDIT\_TPE: Tipo de crédito pedido en CB (tarjeta, efectivo, ...).
- CREDIT\_DAY\_OVERDUE: Número de días de mora en el crédito de CB en el momento de la solicitud del préstamo actual.

**3. POS\_CASH\_balance.csv** contiene ocho variables que son:

- SK\_ID\_PREV: ID de créditos previos en Home Credit relacionados con los préstamos de la muestra. Un préstamo en la muestra puede tener más de un préstamo en Home Credit.
- SK\_ID\_CURR: ID del préstamo en la muestra.
- CNT\_INSTALMENT: Plazo del crédito anterior (puede cambiar con el tiempo).
- CNT\_INSTALMENT\_FUTURE: Cuotas que quedan por pagar del crédito anterior.
- NAME\_CONTRACT\_STATUS: Estado del contrato durante el mes.
- SK\_DPD: Días de vencimiento durante el mes de los créditos previos.
- SK\_DPD\_DEF: Días de vencimiento durante el mes con tolerancia del crédito anterior (se ignoran las deudas con montos bajos).

**4. credit\_card\_balance.csv** contiene más de veinte variables pero se consideraron relevantes las que se explican a continuación:

- SK\_ID\_PREV: ID de créditos previos en Home Credit relacionados con los préstamos de la muestra. Un préstamo en la muestra puede tener más de un préstamo en Home Credit.
- SK\_ID\_CURR: ID del préstamo en la muestra.
- AMT\_DRAWINGS\_ATM\_CURRENT: Importe de veces que sacó dinero en ATM en el mes de los créditos previos.
- AMT\_DRAWINGS\_POS\_CURRENT: Cantidad de compras durante el mes de los créditos previos.

- AMT\_TOTAL\_RECEIVABLE: Total importe por cobrar en los créditos previos.
- CNT\_DRAWINGS\_ATM\_CURRENT: Número de operaciones en ATM durante el mes de los créditos previos.
- CNT\_DRAWINGS\_POS\_CURRENT: Número de operaciones de compras durante el mes en los créditos previos.
- CNT\_INSTALMENT\_MATURE\_CUM: Número de cuotas pagas de los créditos previos.
- NAME\_CONTRACT\_STATUS: Estado del contrato de los créditos previos (activo, firmado, ...)
- SK\_DPD: Días de vencimiento durante el mes de los créditos previos.
- SK\_DPD\_DEF: Días de vencimiento durante el mes con tolerancia del crédito anterior (se ignoran las deudas con montos bajos).

**5. previous\_application.csv** cuenta con prácticamente 40 variables pero se han seleccionado la mitad:

- SK\_ID\_PREV: ID de créditos previos en Home Credit relacionados con los préstamos de la muestra. Un préstamo en la muestra puede tener más de un préstamo en Home Credit.
- SK\_ID\_CURR: ID del préstamo en la muestra.
- AMT\_ANNUITY: Anualidad de los créditos anteriores.
- AMT\_APPLICATION: Importe que solicitó el cliente en los créditos previos.
- AMT\_CREDIT: Monto del crédito final en la solicitud anterior.
- AMT\_DOWN\_PAYMENT: Pago inicial de la solicitud anterior.
- AMT\_GOODS\_PRICE: Precio de los bienes que el cliente solicitó (si corresponde) en la solicitud anterior.
- CHANNEL\_TYPE: Canal por el que llegó el cliente en el crédito previo.
- CODE\_REJECT\_REASON: Porque se rechazó la solicitud anterior.
- NAME\_CLIENT\_TYPE: Indica si el cliente era ya conocido o no en el momento de la solicitud previa.
- NAME\_CONTRACT\_STATUS: Estado del contrato del crédito anterior (aprobado, cancelado, ...).
- NAME\_CONTRACT\_TYPE: Contrato del tipo de producto (préstamo, revolving, préstamo al consumo, ...) de los créditos anteriores.
- NAME\_GOODS\_CATEGORY: Qué bienes solicitó el usuario en el previo crédito.
- NAME\_PAYMENT\_TYPE: Método de pago que utilizó el usuario en su anterior crédito.
- NAME\_PORTFOLIO: Indica para que fue la solicitud anterior: coche, efectivo, ...

- NAME\_PRODUCT\_TYPE: Indica si el crédito anterior se solicitó sin cita previa (*walk-in*) o por adhesión a otro producto o servicio (*x-sell*).
- NAME\_YIELD\_GROUP: Interés (dividido en pequeño, medio y alto) de la solicitud anterior.
- NFLAG\_INSURED\_ON\_APPROVAL: Indica si el cliente solicitó seguro durante el ciclo de vida del crédito previo.
- NFLAG\_MICRO\_CASH: Indica si el préstamo es de un monto pequeño.
- PRODUCT\_COMBINATION: Agrupación detallada de la solicitud previa.
- RATE\_DOWN\_PAYMENT: Tasa de anticipo normalizada del crédito anterior.
- WEEKDAY\_APPR\_PROCESS\_START: Día de la semana en que el cliente solicitó la solicitud anterior.

**6. installments\_payments.csv** recopila ocho variables de las cuales se han seleccionado las siguientes:

- SK\_ID\_PREV: ID de créditos previos en Home Credit relacionados con los préstamos de la muestra. Un préstamo en la muestra puede tener más de un préstamo en Home Credit.
- SK\_ID\_CURR: ID del préstamo en la muestra.
- DAYS\_INSTALMENT: Fecha en la que se suponía que debía pagarse la cuota del crédito anterior (respecto a la fecha de solicitud del préstamo actual).
- DAYS\_ENTRY\_PAYMENT: Fecha en la que realmente se pagaron las cuotas del crédito anterior (en relación a la fecha de solicitud del préstamo actual).
- AMT\_INSTALMENT: Monto de la cuota prescrita del crédito anterior en dicha cuota.
- AMT\_PAYMENT: Lo que realmente pagó el cliente en el crédito previo en dicha cuota.

Se observa que todas las tablas presentan una variable en común, la variable *SK\_ID\_CURR*. Esta variable es la que nos permitirá relacionar todas las tablas entre sí.

Además, vemos que hay otras como *SK\_BUREAU\_ID* o *SK\_ID\_PREV* que nos permitirán relacionar con mayor fiabilidad algunas de ellas. Por ello, se puede considerar que hay más de una clave primaria en nuestros datos y que, dependiendo que la variable escogida para ello, nuestras asociaciones y por tanto, nuestro cuadro de mando, se conformarán de maneras diferentes.

#### 4.4. Misión y visión

Como entidad financiera, se debe intentar mitigar el riesgo de impago de los créditos o préstamos que se aprueban. Por ello, gracias a las variables presentes en nuestras tablas, podremos tener un seguimiento del comportamiento de los clientes. De esta manera, se pueden encontrar patrones que permitan conocer la capacidad de impago de un usuario previamente a la venta del producto.

Siguiendo esta idea, queremos estudiar el comportamiento del público objetivo (que encontramos en la tabla de *application\_train*) en función de todas las variables y casuísticas que se recogen.

Para ello, mediante Qlik y las claves primarias, se buscan asociaciones entre tablas que nos permitan ver de una manera fácil y dinámica el perfil del cliente general que presenta la organización y ver el comportamiento de cada uno de ellos frente a la morosidad.

Esta información, servirá posteriormente para clasificar a futuros clientes dentro de un *scoring* de más a menos probabilidad de impago, incrementando así la rentabilidad de la entidad impidiendo la entrada de clientes que no cumplirán con el pago de sus cuotas.

La idea de la creación de este cuadro de mando es que, a nivel práctico real, este cuadro estuviese asociado a una base de datos que se actualizase periódicamente y se pudiese seguir si el perfil, no sólo medio del cliente sino también de probabilidad de mora, cambia a lo largo del tiempo al igual que lo hacen las variables que nos pueden ayudar a predecir este acto.

Por tanto, la idea es que este cuadro estuviese en constante uso y que fuese modificado en función de si se añaden nuevos indicadores que ayuden a la predicción de la realización de impagos, no sólo en futuros clientes sino también en clientes ya conocidos.

#### 4.5. Creación y transformación de variables

Para la creación del CdM, se han debido transformar variables ya que, por ejemplo, la edad del usuario en la base de datos aparecía en días y en negativo, al igual que los días laborales trabajados, que también aparecen en negativo.

La variable edad se ha modificado no sólo con lo ya mencionado sino que también se ha representado en intervalos ya que a la hora de hacer cualquier tipo de análisis se ha considerado más práctico hacerlo por grupos de edad que no por edades individuales.

Los intervalos se han realizado con una amplitud de 7 unidades a excepción de los extremos. El primer intervalo, recoge las edades inferiores a 27 años (este último no incluido) y el último intervalo, recoge a aquellos usuarios con más de 63 años de edad (este incluido).

También se ha transformado la variable de días trabajados ya que para los pensionistas y parados, el valor que tomaba la variable era -365243. Ahora, el valor para ellos es 0, de manera que no influencia en el momento de hacer medias o calcular otros KPIs.

Además de estas transformaciones, basándonos en otras variables que ya se recogían en los datos, hemos creado nuevas variables, es decir, nuevos indicadores, que se han considerado útiles para el estudio del perfil del cliente y, posteriormente, para poder clasificarlo en si realizará o no algún impago.

Los indicadores creados son los que explican a continuación:

- **Ratio Endeudamiento.** El ratio de endeudamiento recoge el porcentaje de los ingresos que el titular debe destinar a pagar todos los préstamos que tiene con sus bancos. Para este caso, se utilizarán dos variables de la tabla de *application\_train*: *amt\_annuity* (anualidad del préstamo) y *amt\_income\_total* (renta del cliente).

De esta manera, la construcción de este indicador queda así:

$$\text{Ratio Endeudamiento} = \frac{\text{amt\_annuity}}{\text{amt\_income\_total}}$$

En este caso, cuanto mayor sea este ratio, el endeudamiento del cliente es mayor y por tanto, tiene una mayor probabilidad de tener problemas financieros y hacer frente a los pagos de sus cuotas.

- **Porcentaje amortizado.** El porcentaje amortizado hace referencia a la cantidad ya pagada a la entidad financiera y que por tanto, ya se ha descontado de la totalidad a devolver.

Para la creación de este indicador, se han utilizado dos variables de la tabla *bureau* : *amt\_credit\_sum\_debt* (hace referencia a la cantidad que aun se debe) y *amt\_credit\_sum* (importe concedido del préstamo).

Como la variable que tenemos no es el importe ya amortizado sino lo contrario, la fórmula del indicador queda de la siguiente manera:

$$\text{Porcentaje Amortizado} = 1 - \frac{\text{amt\_credit\_sum\_debt}}{\text{amt\_credit\_sum}}$$

Cuanto mayor capital amortizado, inferior será la deuda y por tanto, "mejor" cliente.

- **LTV (Loan to Value).** Es un indicador que comúnmente se usa en trámites hipotecarios para medir la cantidad del préstamo y el valor de la propiedad usada como garantía del préstamo.

Se ha calculado con dos variables de la tabla de *application* : *amt\_credit* (hace referencia a la anualidad del préstamo) y *amt\_goods\_price* (hace referencia al precio de los bienes por los que se le dio al usuario el préstamo).

De esta manera, quedaría el indicador de la siguiente manera:  $LTV = \frac{\text{amt\_credit}}{\text{amt\_goods\_price}}$

Esta variable es muy utilizada dado que da información de si la garantía, el inmueble, podrá cubrir la deuda del préstamo o no.

Si un cliente impaga durante un periodo prolongado, el banco, como último recurso, puede adjudicarse la vivienda a modo de garantía. Esta vivienda la podrá vender y, con ello, recuperar parte o la totalidad del préstamo concedido. En el transcurso de adjudicación y venta se pierde valor del inmueble dado que se tienen que pagar impuestos y costes.

Si el LTV es bajo, hay más posibilidades que la adjudicación del inmueble cubra la totalidad del préstamo y por lo tanto que el riesgo de pérdida sea menor.

- **Días de retraso en el pago de las cuotas.** Se vio que algunos clientes, a pesar de haber impagado, realizaban el pago días después. Por ello, con las variables *days\_entry\_payment* y *days\_instalment* de la tabla *installments\_payments* creamos este indicador.

Ambas variables están definidas por número de días de retraso y en negativo. Por tanto, para saber hace cuanto que se realizó el impago, se invierten las variables, es decir, pasan de ser días en negativo a positivo.

La variable *days\_entry\_payment* son los días atrás en los que se realizó el pago de la cuota mientras que *days\_instalment* indica los días atrás en los que se debería de haber realizado el pago de dicha cuota. De esta manera, se obtiene que:

$$\text{Días retraso en el pago de la cuota} = \text{days.entry.payment} - \text{days.instalment}$$

- **Impagos en otras entidades.** Para tener la capacidad de prever un impago, no solo se han de tener en cuenta los impagos realizados en la entidad sino también fuera de ella. Por este motivo, se ha decidido crear un indicador que nos indique que personas no han realizado ningún impago en nuestra entidad pero, por el contrario, si lo hicieron en otra.

Para construir este indicador se han utilizado también dos variables: *credit\_day\_overdue* de la tabla *bureau* y la variable *target* de la tabla *application\_train*. La primera, indica los días que el usuario se ha retrasado en el pago de la cuota y la segunda, indica si ha habido impagos en las primeras cuotas del préstamo en Home Credit.

Este indicador es binario y se asigna al 1 la existencia de impagos en otras entidades pero no en Home Credit y al 0 la no existencia de impagos tanto en otras entidades como en Home Credit. Es decir, será 1 si la variable *credit\_day\_overdue* es superior a 0 (ha impagado en otras ocasiones y fuera de Home Credit) y *target* = 0 (no ha impagado en alguna de las primeras cuotas del crédito actual). En caso contrario, la nueva variable adoptaría el valor 0.

#### 4.6. Asociación de las tablas

Antes de crear el cuadro, se han de crear asociaciones entre todas las tablas para poder así, cruzar los datos y obtener indicadores e información que a simple vista no veríamos tan claramente.

Por ello, se han utilizado las claves primarias para crear una intersección entre las tablas de manera que todas ellas queden enlazadas.

En la siguiente imagen se muestra de forma visual como quedaría la asociación utilizada.

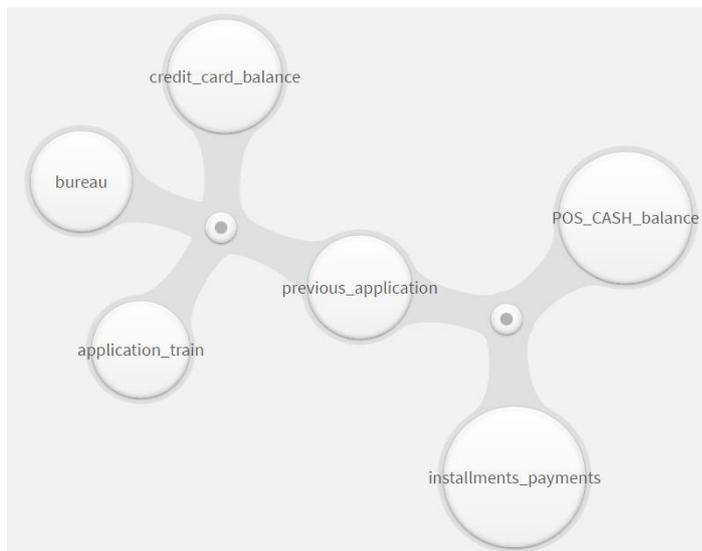


Ilustración 2: Asociación tablas Home Credit

Vemos como hay dos nexos que enlazan las tablas. Estos nexos representan la unión e indican mediante que variable se ha realizado dicha unión.

Además, vemos que la tabla *previous\_application* se encuentra en el centro, es decir, contiene ambas claves primarias.

En el caso de las tablas *bureau*, *application\_train* y *credit\_card\_balance*, la clave primaria que nos permite unirlas con *previous\_application* y, por tanto, con el resto de tablas es *SK\_ID\_CURR*.

Por otro lado, las tablas *POS\_CASH\_balance* y *installments\_payments* interseccionan

mediante la variable *SK\_ID\_PREV*.

Una vez creada la asociación, podemos ver el comportamiento e influencia de unas variables con otras que, sin dicha unión, no podríamos conocer.

#### 4.7. Perfil del cliente

Para iniciar con el estudio de la morosidad, primero se ven qué perfiles existen dentro del público objetivo de la base de datos.

Para ello, gracias al software implementado, se ha creado un perfil del cliente promedio de nuestra base de datos para partir de una idea base del tipo de cliente.

Para el análisis de este perfil general, se han seleccionado variables descriptivas y sobre todo socioeconómicas. Algunas de ellas han sido la edad, el género, la renta, situación laboral actual, región de residencia o antigüedad del cliente.

La idea es que, de una manera sencilla, clara y dinámica, podamos ver qué características cumplen los usuarios con un préstamo en Home Credit (HC).

En la siguiente imagen se muestran cinco variables de la tabla *application\_train* que nos permiten conocer mejor al usuario en este primer análisis. Además, dentro de esta hoja, podemos encontrar otra con cuatro variables más. Estas variables son:

- **Importe medio solicitado.** Este indicador es la media de la variable *amt\_goods\_price*.
- **Impagos en las primeras cuotas.** Es extraído de la variable *target*, al ser una variable binaria, queremos ver cuánto peso suponen los impagos (variable igual a 1) respecto al total de usuarios que han contratado un préstamo o crédito.
- **Edad del cliente.** La edad del cliente aparecía en la BBDD como ya se ha mencionado, en días y en negativo. Por ello, primero se realizó una transformación de la variable para expresarla en años y, posteriormente, se crearon los intervalos.
- **Región cliente.** Esta variable permite conocer la procedencia del cliente y ver que peso tiene cada una de ellas en la clientela.
- **Profesión del cliente.** Nos indica a que se dedica el usuario. Es importante conocer la cantidad de ingresos del usuario y ver si estos son fijos o no.
- **Habitabilidad del cliente.** Con este indicador se podrá conocer si es propietario de una vivienda, si vive con sus padres o cualquier otra casuística.
- **Días trabajados.** Por lo general, a mayor tiempo trabajado en un mismo lugar de trabajo más estable este se vuelve. Por ello, se ha añadido este indicador al CdM. Para el caso de los parados y pensionistas, esta variable toma el valor 0.
- **Renta del cliente.** Conocer los ingresos es importante para ver si el cliente será o no capaz de efectuar sus pagos.
- **Anualidad del préstamo.** Tan importante es conocer el nivel de ingresos como el nivel de gastos pues, una persona puede ganar menos pero vivir acorde con sus posibilidades y por el contrario, una persona a pesar de tener unos ingresos altos puede estar viviendo por encima de sus posibilidades aumentando así su ratio de endeudamiento. Más tarde se hablará de ello.

Cabe destacar que, la variable impagos en las primeras cuotas, se ha considerado relevante para poder ver de una manera rápida si aumenta o disminuye el porcentaje de clientes que impagan. Pues como ya se ha mencionado, la idea de este análisis es intentar disminuir la cantidad de usuarios en la cartera con mora.

Una vez introducidas las variables, en la siguiente imagen se muestran los primeros cinco indicadores que describen a los usuarios:

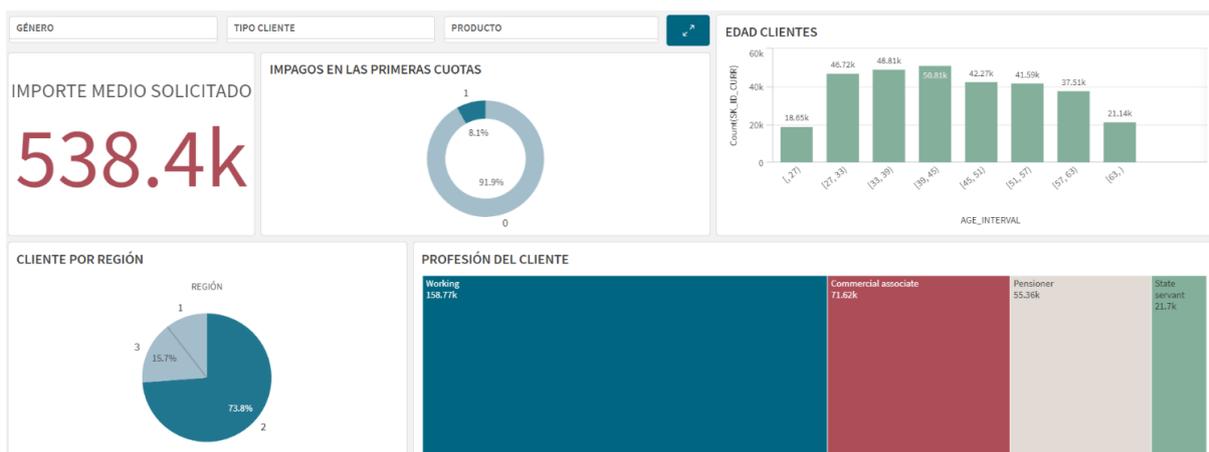


Ilustración 3: CdM Perfil del Cliente

Las tres pequeñas pestañas que encontramos en la parte superior, nos permitirán ir modificando los distintos gráficos y indicadores en función de las opciones que se seleccionen.

Vemos que, sin aplicar ningún filtro, es decir, teniendo en cuenta la totalidad de clientes que han solicitado un préstamo en Home Credit, la gran mayoría procede de la región dos mientras que la menos común es la uno.

Por otro lado, vemos que la mayoría de ellos se encuentran trabajando y tan solo una pequeña parte son funcionarios. Los pensionistas también suponen un gran peso en nuestra base de datos.

Moviéndonos a los gráficos de la parte superior, los rangos de edad en los que encontramos una mayor volumetría de usuarios son en aquellos que comprenden las edades de entre 27 y 45 años (esta última edad no está incluida).

Este rango de edad supone prácticamente el 48% del total de los clientes.

Por último, encontramos dos indicadores más relacionados con el importe y los impagos.

El primero, nos indica el importe medio solicitado por los usuarios. Esta cantidad supera el medio millón. En cambio, el segundo hace referencia a si el cliente ha realizado algún impago en sus primeras cuotas.

Este porcentaje de impagos se aproxima al 10%. A pesar de ser un porcentaje bajo, se debe intentar reducir y es por ello por lo que se ha realizado este análisis.

Como bien se ha mencionado, en el CdM se encuentran tres variables a modo de panel de filtrado que nos permitirán filtrar los datos y ver así los gráficos resultantes una vez aplicados.

Las tres variables elegidas para dicho filtraje son el género, el tipo de cliente (si ese era o no conocido previamente a la solicitud) y, por último, el producto solicitado en el préstamo con HC.

Ya que la idea inicial es evitar futuros impagos, se filtra por cliente y se crean dos cuadros uno haciendo referencia a los préstamos y otro que representa a los revolving.

Primero, se hace una selección de los clientes previamente no conocidos que han solicitado un revolving.

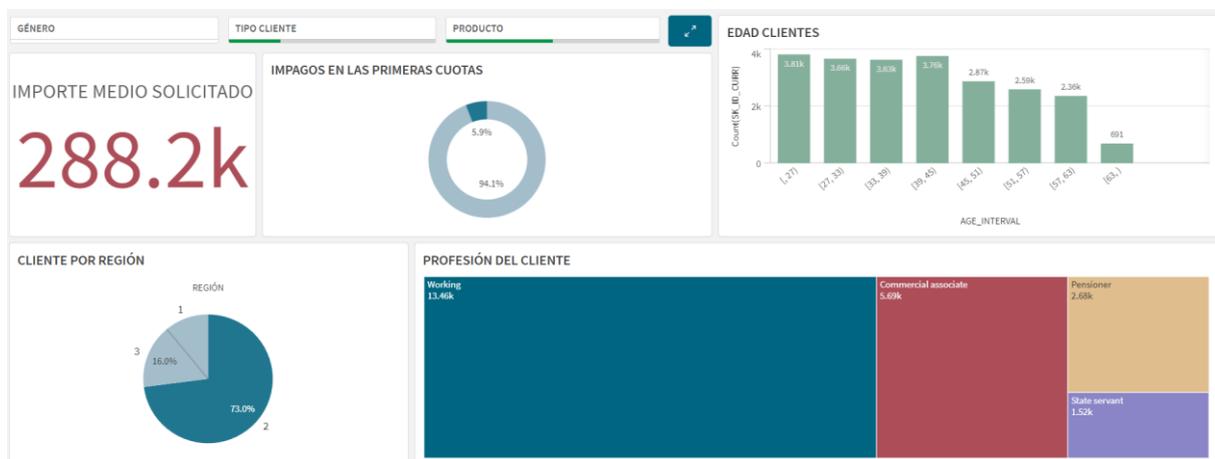


Ilustración 4: CdM filtrado por cliente y producto revolving

Fijándonos en la imagen, el porcentaje de impagos en los primeros recibos es inferior, apenas alcanzando el 6%.

En la variable región se aprecia una subida en la región 1, por lo que se puede considerar que en dicha región destaca la contratación del producto revolving.

Respecto al resto de variables, vemos que hay un gran cambio en las volúmetrías de edad: son los cuatro primeros grupos de edad los que contemplan el mayor volumen de solicitudes revolving. También destaca el bajo importe medio en comparación a la primera imagen.

Por otro lado, filtramos por nuevo cliente y por préstamo:

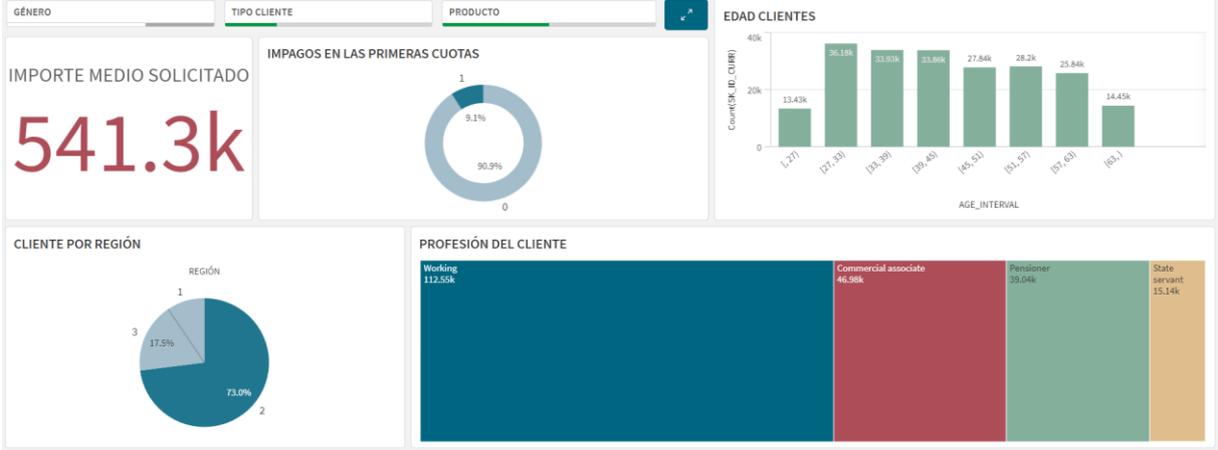


Ilustración 5: CdM filtrado por cliente y producto

En este caso, tenemos un porcentaje de impagos similar al de la ilustración 3.

Para el resto de variables, no hay grandes cambios pues las volúmetrías mayores de usuarios se encuentran en los mismos rangos que en un inicio y el importe medio solicitado aumenta levemente.

Se resume por tanto, que el nuevo cliente presenta una mayor tasa de impago con importes similares a la media total en el caso del préstamo y siguen siendo los trabajadores el mayor grupo de asalariados que solicita tanto un préstamo como un revolving.

Con el mismo CdM, se hace una comparación entre hombres y mujeres y se obtienen los siguientes cuadros:

En el caso de las mujeres, vemos cambios más sustanciales en comparación a los anteriores. El

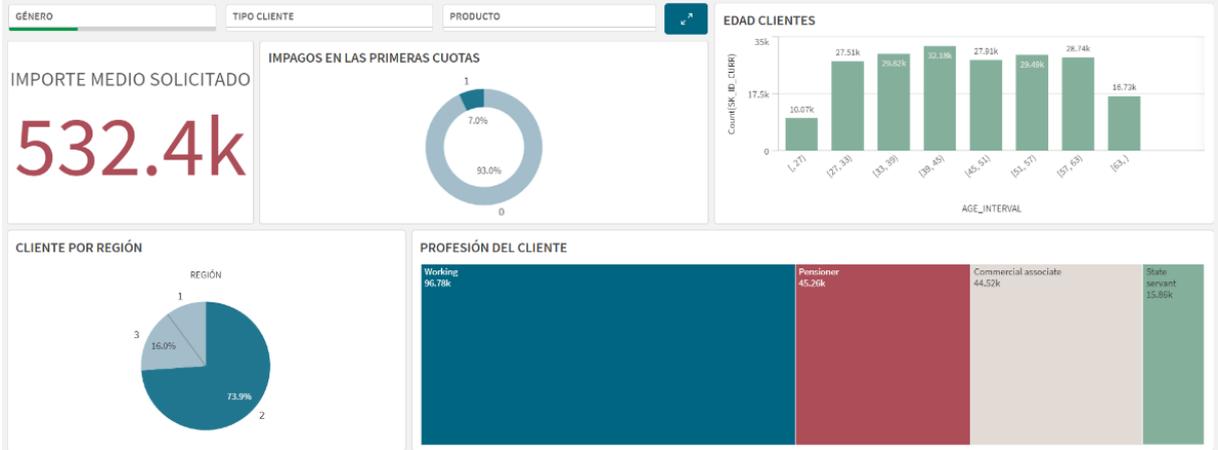


Ilustración 6: CdM filtrado por género femenino

porcentaje de impagos ha disminuido: se pasa de algo más de un 8% a un 7%. Por otro lado, los grupos de edades varían, las volúmetrías más altas ya no se encuentran entre los 27 y 45 años, sino que se encuentran entre los 33 y 45 años pero también entre los 51 y 57 años.

Por último, no se ven cambios relevantes en las regiones al igual que en el importe medio solicitado, pero sí en la profesión de las clientas: el peso del grupo de pensionistas es al que le corresponde la segunda volumetría mayor cuando, a nivel general, era la tercera volumetría mayor.

Fijándonos en los hombres, se mantiene la edad más común para solicitar los servicios de HC entre los 27 y 45 años y también, se mantiene la volumetría de las profesiones más comunes entre los usuarios y la región de la que proceden.

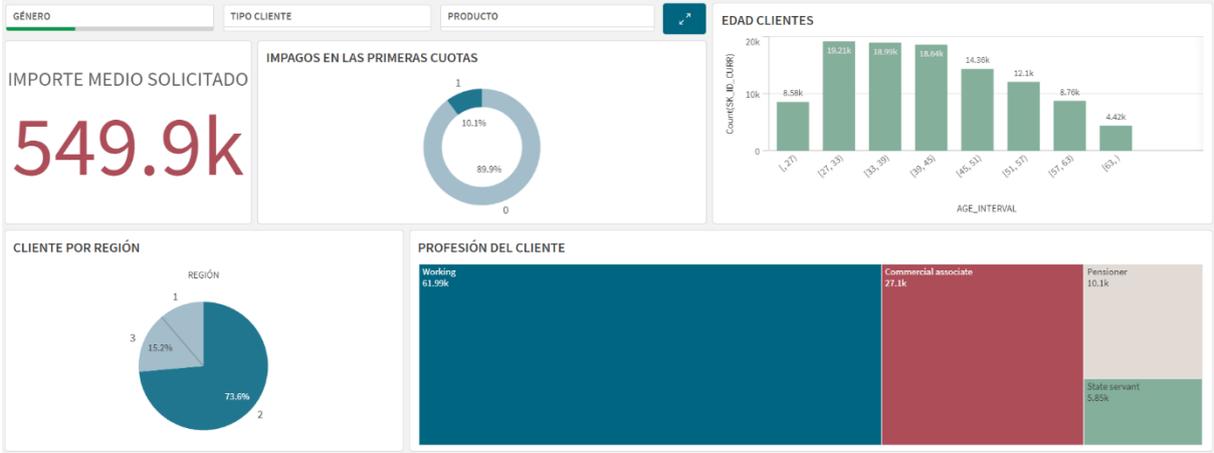


Ilustración 7: CdM filtrado por género femenino

Sin embargo, aumenta tanto el porcentaje de impagos que se incrementa un 2% respecto al primer CdM mostrado, como el importe medio solicitado.

Al ser cuadros dinámicos, se pueden hacer tantas combinaciones con los filtros como se desee. Por este motivo, como existen infinidad de posibilidades, sólo se explican algunas de ellas.

Como se ha mencionado en el inicio, este perfil del cliente tiene “dos hojas” en nuestro cuadro. En la siguiente ilustración se enseñan las variables anteriormente explicadas, pero aún no reflejadas. A diferencia de para el cuadro anterior se mostrará el cuadro sin aplicar ningún filtro y con la diferenciación por género.

Teniendo en cuenta todos los usuarios con un contrato activo en HC, se obtiene:

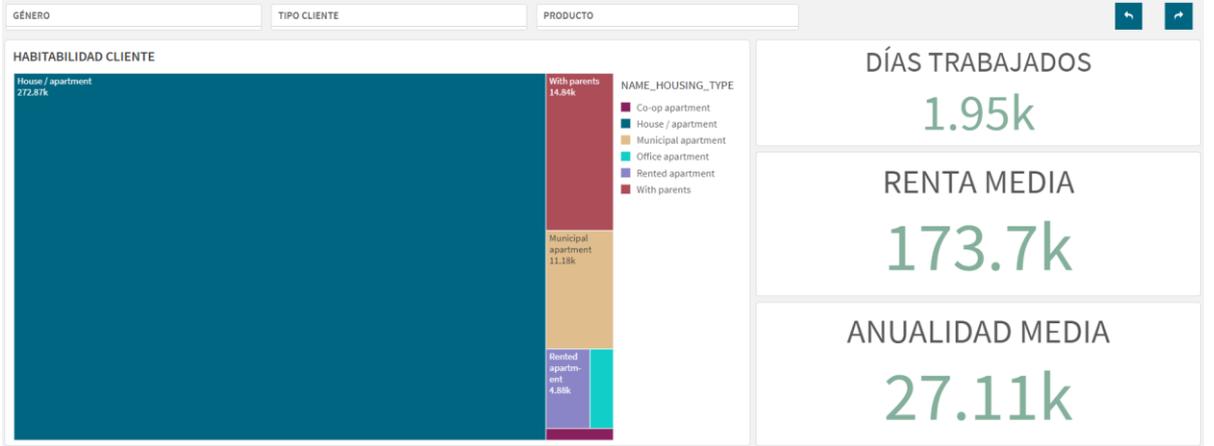


Ilustración 8: "hoja dos" del CdM

Se muestran seis tipos de habitabilidad siendo la de propietario la más común en la cartera de Home Credit seguida de viviendo con los padres.

Por otro lado, vemos que de media, el cliente lleva trabajados aproximadamente cinco años en su actual puesto de trabajo.

La anualidad media del crédito supera los veinticinco mil, como se ha explicado anteriormente, mirar este dato sin tener presente también la renta, no dice demasiado. Por este motivo, en esta hoja también se muestra la renta media del cliente que supera los ciento cincuenta mil. Por tanto, la cuota supone más del quince por ciento de la totalidad a devolver.

Se muestran ahora los Cdm en función del género del usuario:

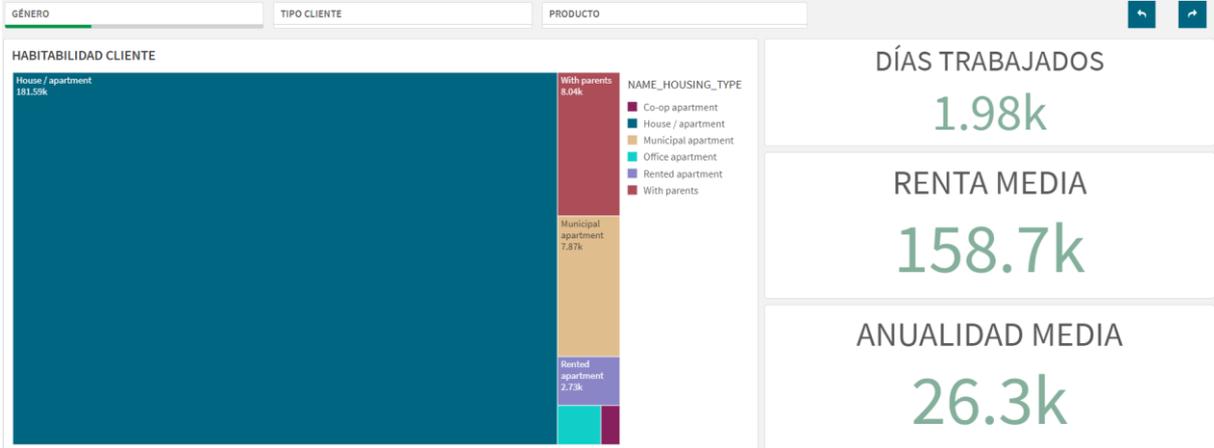


Ilustración 9: Cdm "hoja dos" filtrado por género femenino.

En el caso femenino, se ve un aumento en los días trabajados en el lugar actual de trabajo, pero con una renta media inferior a pesar de la anualidad que es prácticamente la que se obtiene de forma general.

En el caso masculino, ilustración 10, los días trabajados disminuyen, pero sube tanto la renta como la anualidad media. En este caso, la anualidad supone prácticamente un diecisiete por ciento de la totalidad a devolver.

En ambos casos, el usuario es propietario de una casa o piso.

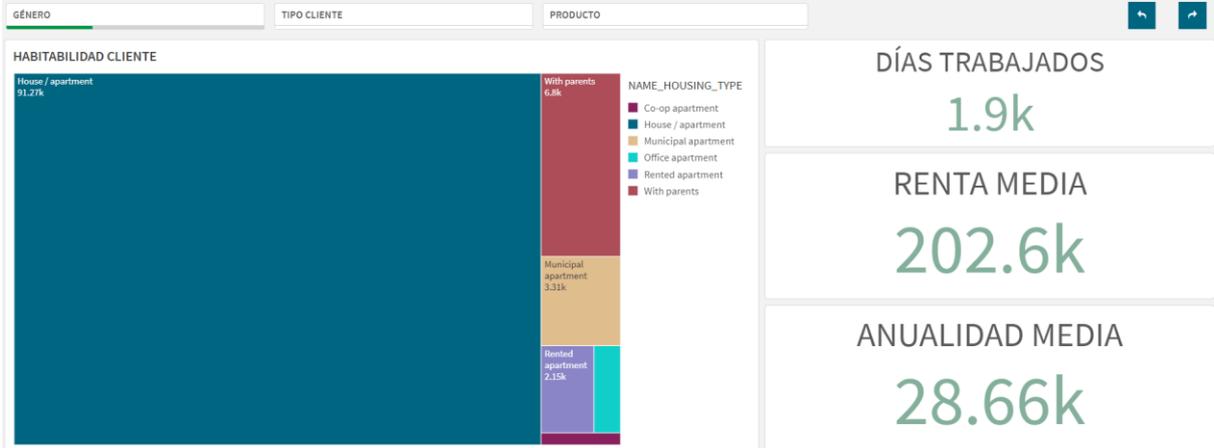


Ilustración 10: Cdm "hoja dos" filtrado por género masculino.

#### 4.8. Perfil del cliente en función de los contratos

Una vez conocemos de manera básica que tipos de usuarios conforman la entidad, se estudia la tipología de los contratos, es decir, que características presentan los contratos que previamente ha solicitado un cliente.

Conocer el comportamiento del cliente tanto en otras entidades como en HC ayuda a saber si es un buen pagador o no.

Para ver este comportamiento previo, se ha creado otro CdM pero con datos de contratos anteriores con la idea de conocer más al usuario.

En la ilustración once se muestran tres variables que dan información sobre los contratos que tiene aún activos el cliente:

- **Tipos de préstamos solicitados.** Conocer los motivos por los que el cliente solicita un importe de capital es útil para conocer mejor al usuario ya que, tal vez lo necesita de manera puntual o, por el contrario, utiliza los créditos o préstamos de manera recurrente.
- **Días de mora.** Que el cliente ya haya impagado en otras entidades, aumenta la probabilidad de riesgo en él. Aparece la media de este indicador, calculada a partir de la variable *credit\_day\_overdue*.
- **Días previos a la solicitud actual.** Indica el tiempo (en días) que hace que el cliente solicitó su anterior préstamo. También se muestra la media en el cuadro y se ha extraído de la variable *days\_credit*.
- **Días de finalización del contrato.** De la variable *days\_credit\_enddate* se ha calculado la media de días que faltan para que finalicen los contratos activos de los usuarios que han solicitado el crédito en Home Credit.

Además, al igual que en el CdM del apartado anterior, aparecen como filtro las variables género, tipo de producto y tipo de cliente.

Para seguir la misma línea que en el apartado anterior, en la ilustración siguiente no se ha aplicado ningún filtro.

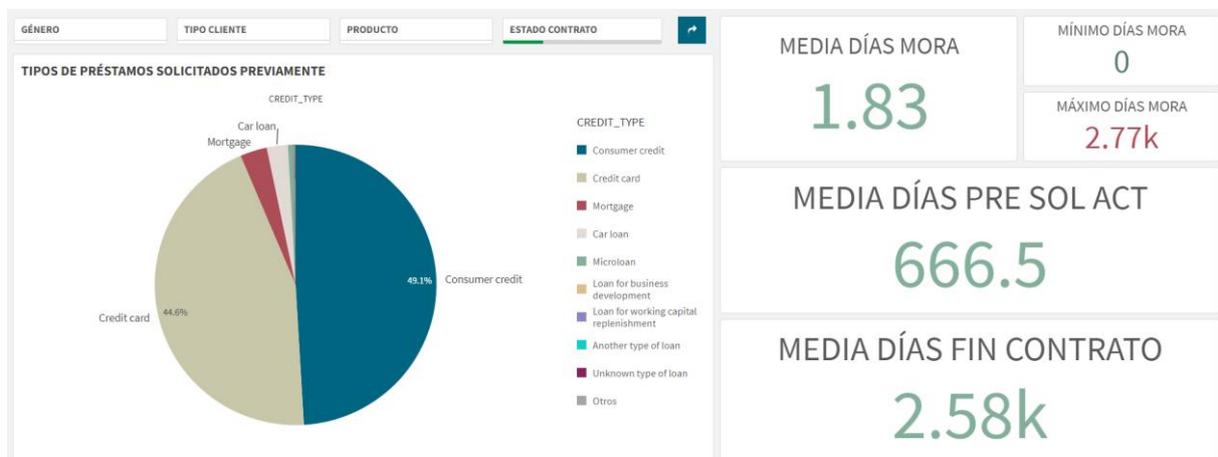


Ilustración 11: CdM datos contratos anteriores.

Los tipos de préstamos que se solicitaron anteriormente se clasifican en diez grupos siendo el crédito al consumo y el crédito con tarjeta los que presentan más peso respecto al total.

El primer indicador que encontramos a la derecha es la media de días de mora que presentan los usuarios. Esta media obtiene un valor bajo a pesar de que existan usuarios cuya deuda sea de hace años (se muestra en el máximo de esta misma variable).

Que la media de días de impagos sea baja, es un indicador de que el control de riesgo que tiene la entidad es efectivo ya que la mayoría de los clientes de su cartera presenta tasa de morosidad cero.

Para añadir más datos del pasado, el cuadro presenta dos KPIs más. El primero, media "días pre sol act", hace referencia a la media de días en las que el usuario solicitó de nuevo un préstamo, es decir, los días que pasaron entre la última solicitud y la actual.

El segundo, hace referencia a los días que quedan para que finalicen aquellos otros contratos que el usuario tiene aún activos.

De ambos indicadores extraemos que, de media, el cliente tarda dos años en solicitar de nuevo un préstamo pero que no necesariamente este tiene que haber finalizado ya que, de media, el cliente tiene deuda pendiente (cerca de los siete años) por otros créditos que aún tiene activos.

En las siguientes ilustraciones se aplica el filtro por género.

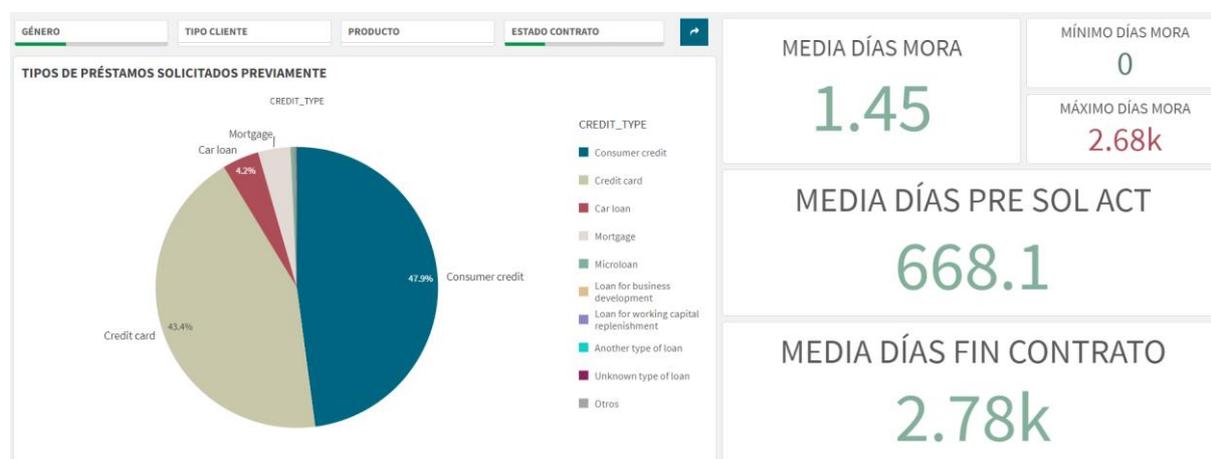


Ilustración 12: CdM contratos anteriores filtrado por género masculino.

De la imagen anterior, en la que se muestra la hoja aplicando el filtro masculino, destaca el incremento en la volumetría de los préstamos destinados al coche y la sutil disminución de la tarjeta.

La media de días entre solicitud de préstamos no se ve afectada por el género mientras que la media de días que quedan para finalizar los contratos ya activos incrementa.

A pesar de este incremento en los días que faltan por finalizar otros créditos, los días de mora disminuyen: no alcanza el día y medio.

Se muestra a continuación el caso femenino:

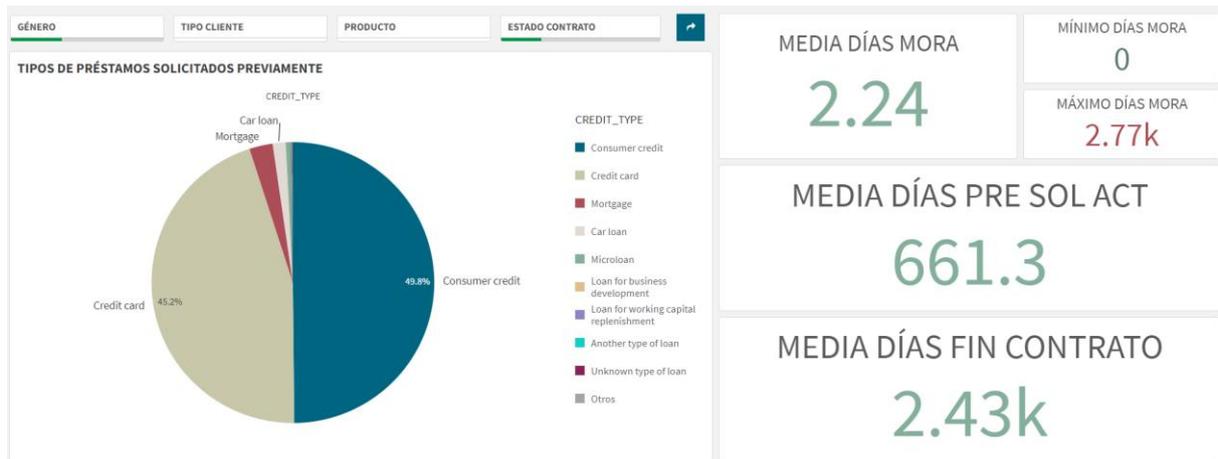


Ilustración 13: Cdm contratos anteriores filtrado por género femenino

El volumen de los tipos de préstamos más solicitados, conforman el noventa y cinco por ciento de la totalidad de los préstamos que solicitan las mujeres.

Se ve un incremento en los días de mora junto con una disminución de la media de días entre la solicitud actual y la anterior.

Respecto al resto de variables, no hay cambios sustanciales.

Antes de seguir con el comportamiento de los clientes en función de sus anteriores créditos, hemos querido saber en qué estado se encuentra el crédito anterior a la solicitud en Home Credit.

Por ello, se han extraído 3 indicadores: el estado del propio contrato, la media de cuotas que quedan aún por pagar de dicho contrato (junto con la mínima y la máxima) y la media de días de vencimiento durante el mes. Con esto, se ha obtenido lo siguiente:

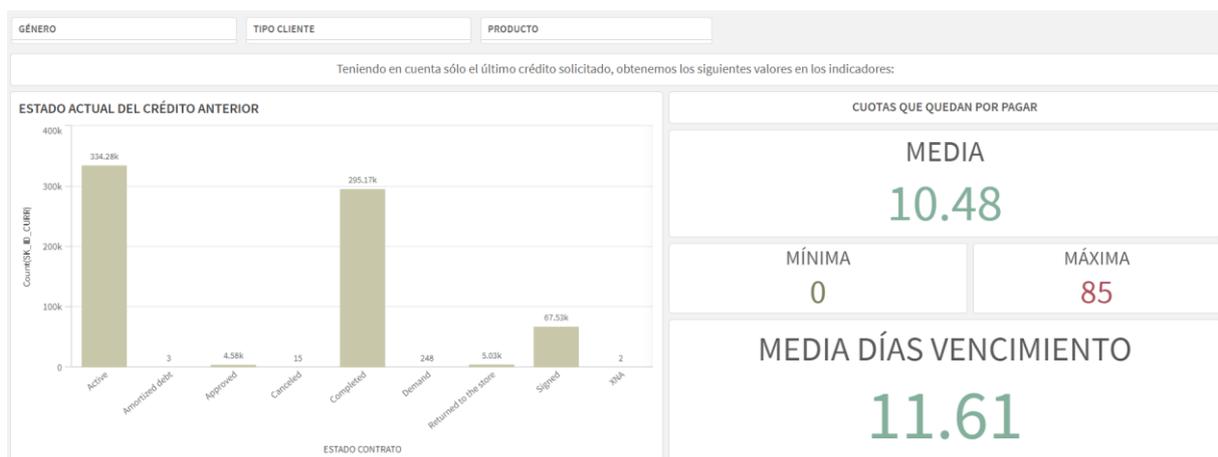


Ilustración 14: Cdm estado y características contrato anterior.

Se han mantenido los filtros que se han ido aplicando a lo largo de estos últimos apartados junto con las variables que se han indicado anteriormente.

Entrando en más detalle sobre los indicadores:

- **Estado actual del crédito anterior.** Esta variable consta de nueve categorías, entre ellas encontramos activo, completado, firmado y aprobado. Que los clientes tarden de media

aproximadamente tres años en solicitar otro préstamo, como se ve en una de las ilustraciones anteriores, no quiere decir que este préstamo previo haya finalizado.

- **Cuotas que quedan por pagar.** Conforme el usuario salde su deuda anterior, el ratio de endeudamiento de dicha persona, disminuirá lo que hace más improbable que impague por no tener liquidez.
- **Días vencimiento.** Los días de vencimiento hacen referencia al número de días desde el último pago atrasado que aún no ha sido abonado en su totalidad.

Una vez se conoce detalladamente lo que conforma el cuadro, se analizan los resultados.

Centrándonos primero en el estado del crédito, destacan tres categorías: activo, completado y firmado. Siendo los dos primeros mencionados los que más volumetría recogen.

Los tres bloques hacen referencia a que el contrato está vigente por lo que, la mayoría de los usuarios mantienen activa la deuda del préstamo solicitado previamente al de Home Credit.

Si se sigue analizando la imagen, la media de cuotas que quedan para que finalice el contrato no llega al año. Esto indica que, en caso de que el cliente no decida prolongar su contrato, sus gastos mensuales disminuirán.

El último indicador hace referencia a que, de media, estos contratos presentan un retraso en el pago de su última cuota de aproximadamente dos semanas.

Como en los anteriores casos, se filtra por género y se obtiene, en el caso de las mujeres:

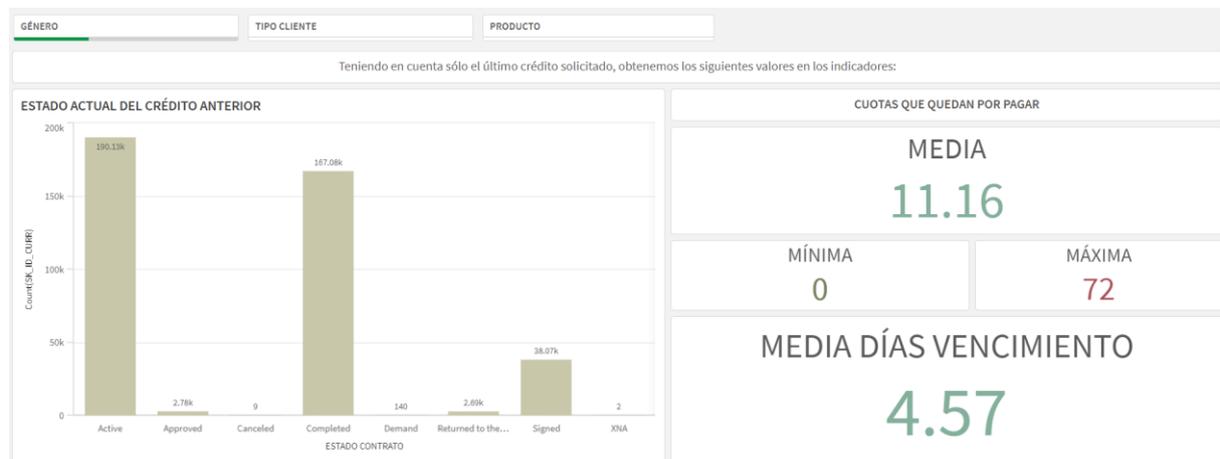


Ilustración 15: CdM estado y características contrato anterior filtrado por género femenino.

El primer cambio se observa en la disminución a ocho de los posibles estados de los contratos: desaparece la categoría de deuda amortizada.

Por otro lado, la media de cuotas restantes aumenta a pesar de que la máxima haya disminuido.

Finalmente, el retraso en el pago de la última cuota no llega a los cinco días respecto a prácticamente las dos semanas que se veían en el CdM sin aplicar ningún filtro.

Por otro lado, en el caso masculino, tal y como se observa en la ilustración quince, aparecen de nuevo las nueve categorías del estado del contrato.

La media de días de vencimiento se acerca también a los cinco días.

Por último, la media de cuotas que quedan por pagar no alcanza las diez.



Ilustración 16: CdM estado y características del contrato anterior filtrado por género masculino.

Para finalizar con el cuadro que hace referencia al crédito anterior, filtramos por cliente. En esta parte del documento solo se adjunta un indicador, pero se puede ver la totalidad de la hoja en el anexo 7.2.

El indicador que se muestra es el de días de vencimiento.

Para los nuevos clientes, este indicador toma el valor de aproximadamente nueve días mientras que, para aquellos clientes ya conocidos, este indicador es inferior a tres.

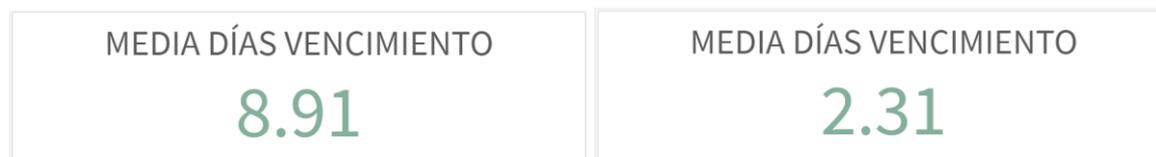


Ilustración 17: Días de vencimiento cliente nuevo vs. Cliente conocido.

Volviendo a la totalidad de los contratos previos, en otra hoja, se filtra por aquellos contratos que permanecen activos, al igual que en las ilustraciones once y doce, con la idea de analizar la anualidad media del cliente, el monto medio solicitado y la deuda pendiente.

Hay que tener en cuenta que un cliente puede tener más de un contrato abierto.

Como en el resto de los casos, se explican los indicadores que aparecen en este último análisis de los contratos previos:

- **Anualidad media.** Indica la mensualidad que el cliente paga por cada uno de los contratos que tiene actualmente activos. Por ello, cuánto más baja esta cantidad, menor endeudamiento para el cliente.
- **Monto medio solicitado.** Indica el importe medio que ha solicitado el cliente en otros contratos.
- **Deuda media pendiente.** Muestra la cantidad media que un usuario debe a la entidad.



*Ilustración 18: CdM contratos anteriores filtrado por contratos activos.*

Vemos que, solo filtrando por contratos previos activos, el cliente devuelve a la entidad de media más de sesenta y cinco mil mientras que el importe medio que se solicita por préstamo es de cuatro y medio (aproximadamente).

Por otra parte, encontramos la deuda a la entidad. Esta asciende hasta algo más de seiscientos mil.

#### 4.9. *La mora y el cliente*

Tal y como se comentó en uno de los apartados con los que se inicia la parte práctica, se han creado nuevas variables a partir de los datos que ya se tenían con la finalidad de ajustar más el perfil con la mora.

Por ello, en el siguiente CdM se han representado estas variables junto con algunas de las que ya se han mostrado para conocer un poco más a los usuarios.

De esta manera, veremos en un mismo cuadro el comportamiento de un cliente nuevo y uno ya conocido. Además, en el lado izquierdo encontramos indicadores que hacen referencia al contrato actual y, en el lado derecho, aquellos indicadores que hacen referencia a los contratos pasados aun activos.

Concretamente, esta hoja está formada por los siguientes indicadores o variables:

- En el caso de los indicadores relacionados con el cliente de Home Credit:
  - o Loan To Value (LTV).
  - o Productos contratados en Home Credit.
  - o Impagos en las primeras cuotas del crédito o préstamo solicitado en Home Credit.
  - o Indica si el cliente ha impagado en otras entidades pero no lo ha hecho en Home Credit.
  - o Renta media del cliente de la cartera de Home Credit.
  - o Anualidad del crédito o préstamo que se ha solicitado en la entidad.
  - o Capital ya amortizado de dicha solicitud.
- En el caso de los indicadores relacionados con otros contratos que tiene activos el cliente:
  - o Tipos de productos que solicitó el cliente previamente a la solicitud actual.

- Días de retraso en la última cuota pero solo del crédito anterior al actual.
- Media de días de mora del cliente existente en los datos de Credit Bureau.
- Media de días que quedan para que finalicen los contratos ya activos del usuario.

Por último, se ha añadido el ratio de endeudamiento teniendo en cuenta tanto el contrato último que ha abierto con HC como aquellos que ya tiene o en la misma entidad o en otras.

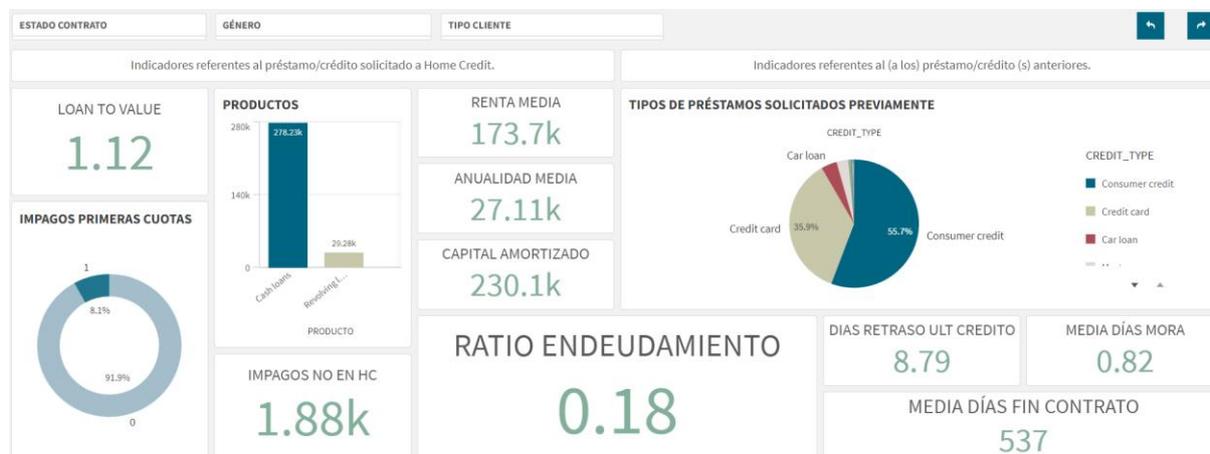


Ilustración 19: Cdm cliente y mora

Centrándonos primero en la parte izquierda, vemos que el LTV es cercano a uno, lo que nos indica que en caso de impago del cliente, hay más posibilidades que la adjudicación del inmueble cubra la totalidad del préstamo y por lo tanto que el riesgo de pérdida sea menor.

En cuanto a producto vemos que el más solicitado es el préstamo, de hecho, el revolving solo ha sido solicitado por aproximadamente treinta mil personas comparado con los más de doscientos setenta mil clientes que han solicitado un préstamo.

De los usuarios actuales en la cartera de Home Credit, prácticamente dos mil clientes han impagado en otras entidades, pero no en esta.

De media, ya se ha devuelto más de doscientos mil a la entidad de la totalidad de contratos en Home Credit.

Si nos fijamos en los tipos de préstamos solicitados previamente, más del noventa por ciento lo forman los créditos ya sea con tarjeta o al consumo.

La media de días que restan para que se finalicen los contratos previos, es de menos de dos años.

Por último, si se mira la media de días de mora de la totalidad de los créditos previos del cliente, esta no llega a dos mientras que, si sólo nos fijamos en los días que el usuario se retrasó en pagar su cuota del contrato previo al contratado actualmente, esta cifra supera los ocho días.

El ratio de endeudamiento de los usuarios no llega al veinte por ciento, muy por debajo de lo que aconseja el Banco de España.

Una vez conocemos el cuadro sin aplicar ningún filtro, se analiza en función del tipo de cliente.

#### 4.9.1. Clientes conocidos

En este apartado se muestra como quedaría el cuadro de mando filtrando por clientes ya conocidos con contratos previos activos.

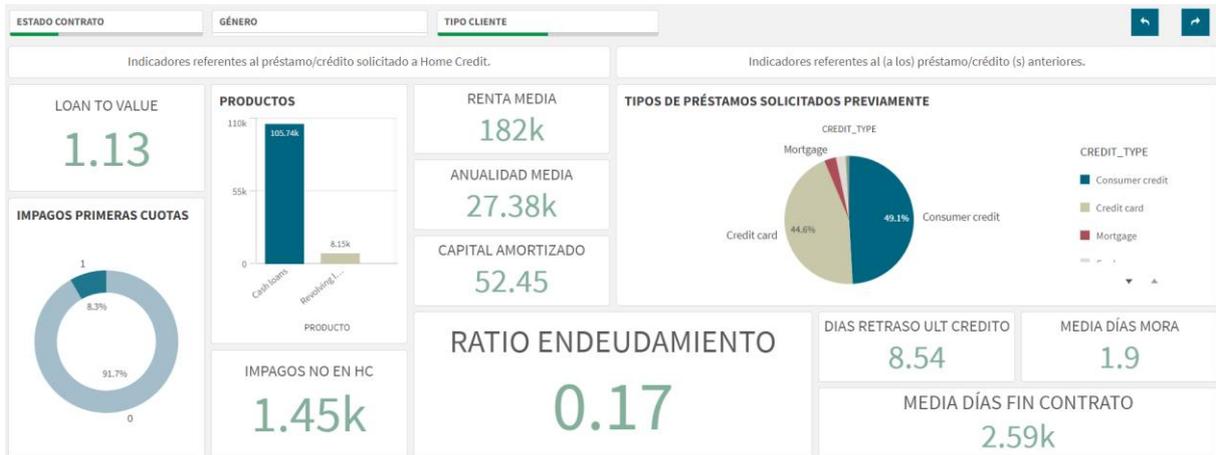


Ilustración 20: CdM cliente conocido y mora

Vemos que una gran parte de los clientes de la cartera de la entidad ya eran conocidos.

Incrementa levemente el *Loan To Value*, la renta media y la anualidad pero disminuye el capital amortizado. Disminuye también el número de impagos y se mantiene el porcentaje de impagos en las primeras cuotas.

En lo que respecta a los indicadores relacionados con los contratos pasados, aumenta la media de días de mora junto con los días que restan para finalizar los contratos previos.

Se observa una bajada en los días de retraso en el pago de la cuota del crédito previo.

No se perciben cambios en los tipos de préstamos contratados con anterioridad: los porcentajes se aproximan bastante a los de la ilustración dieciocho y tampoco con el ratio de endeudamiento.

#### 4.9.2. Clientes nuevos

Una vez se conoce el análisis de los clientes de los que previamente se tenía constancia, se filtra por aquellos que se desconoce.

En la siguiente imagen, se ve como el LTV coincide con el de clientes ya conocidos.



Ilustración 21: CdM cliente nuevo y mora

La volumetría de clientes es inferior como también lo es el número de impagos en otras entidades y los impagos en las primeras cuotas.

La renta media disminuye en comparación al caso anterior pero aumenta el capital amortizado en comparación a los clientes ya conocidos.

Aumenta levemente el porcentaje de contratos de créditos al consumo y disminuye el de tarjeta.

Por otro lado, a pesar de que los días de mora disminuyan, alcanza los doce los días en los que el usuario se retrasó en su última cuota del préstamo anterior.

Aumenta también, la media de días que restan para finalizar el contrato.

El ratio de endeudamiento coincide con el primer caso: es del dieciocho por ciento.

Una vez se conoce el tipo de cliente en función de si es o no conocido, se aplica un score para saber si un nuevo cliente es un buen pagador, independientemente de su tipología.

Se ha hecho una selección de variables, en función de si estas reflejaban de una manera correcta el nivel de morosidad del cliente, y se le ha dado una puntuación a cada una creando así una nota media para el cliente.

Las variables seleccionadas para formar este score son:

- **Ratio de endeudamiento.** Es la variable más importante ya que es indica directamente su capacidad de pago.
- **Impagos.** Que el cliente siempre haya sido un buen pagador, no pone en antecedente a dicho cliente creando así una seguridad mayor a la entidad respecto a una persona que es recurrente en dicha acción.
- **Impagos actuales.** No sólo hay que conocer el historial de impagos del usuario, sino también si ya ha realizado algún impago en el préstamo que actualmente tiene con Home Credit.
- **Capital Amortizado.** Cuanto mayor sea la cantidad ya devuelta, menor número de cuotas quedan por pagar por lo que, la probabilidad de impago se reduce.
- **Renta alta.** Si el cliente presenta una renta superior a la media, por lo general, debería tener un ratio de endeudamiento menor y, por consiguiente, una buena capacidad de pago.
- **Tiempo entre préstamos.** A mayor tiempo entre la contratación de préstamos, menor necesidad de recurrir a ellos por lo que, se considera que la persona tiene un menor ratio de endeudamiento.

Una vez conocidas las variables que conforman el score, cada usuario obtiene su puntuación si cumple los siguientes requisitos:

- El cliente obtendrá dos puntos si su ratio de endeudamiento es inferior al 35% (es el marcado por el Banco de España).
- Si no se recogen impagos del cliente en cualquier entidad se obtienen dos puntos.
- La no constancia de impagos actuales, hace que el usuario sume tres puntos más.
- En caso de que el cliente tenga amortizado más del 60% de su préstamo actual, obtiene un punto más.
- Cuando la renta del usuario sea superior a la media, se añade un punto.

- Que el usuario tarde más que la media en haber solicitado un préstamo, le añadirá un punto.

Tras aplicar estas condiciones a cada usuario, se obtiene que el score medio de la entidad es de siete sobre diez.

Esta puntuación a pesar de ser elevada se puede mejorar. Es por ello por lo que se deberán aplicar medidas de riesgo u otras restricciones en el momento de la contratación para evitar los impagos y aumentar así este score.

## 5. CONCLUSIONES

Estudiar, analizar y controlar la morosidad ha sido el objetivo de este proyecto.

A pesar de conocer el concepto de morosidad, se ha explicado de una manera más extensa y detallada todo lo que la engloba.

Entender que es un cuadro de mando, donde se originó, los distintos tipos que se pueden encontrar y sobre todo, la gran importancia que tienen en todas las entidades ha sido esencial para llegar al final del trabajo destacando su antigüedad y ventajas que ha supuesto para estas.

Por otro lado, gracias a haber comprendido los dos grandes conceptos del trabajo y con el uso de Qlik, se ha podido crear un cuadro de mando que ha permitido transformar un gran conjunto de datos en respuestas visuales y dinámicas, comprensibles para todos.

Con datos actuales de Home Credit y del pasado de los usuarios, proporcionados por Credit Bureau, se ha podido ver el comportamiento de los distintos perfiles que existen en la cartera de la entidad. Esto ha servido para conocer al cliente y ser capaces así de prever si será o no un buen pagador mediante un score.

Además del objetivo mencionado inicialmente, se quería plasmar lo que sucede realmente en una entidad, es decir, verdaderamente la morosidad es un indicador de muchas organizaciones y utilizan estos cuadros de mando, conectados a tablas que se actualizan periódicamente, para poder seguir dicho indicador.

Por ello, se considera que, a pesar de que el CdM creado sea estático, con una foto, podría funcionar bien en un banco en el que, dicho cuadro de mando, se fuera actualizando de forma recurrente con la BBDD.

De esta manera, la entidad sería capaz de seguir este indicador, actualizando sus variables, agregando otras o eliminando aquellas que vayan quedando obsoletas para poder prever de una mejor manera la morosidad de sus clientes y ahorrarse así capital que puede ser destinado a otros recursos.

En definitiva, tras conocer lo que se quería estudiar, tratar los datos y transformarlos en un cuadro de mando, se es capaz de seguir un indicador de una manera rápida y efectiva ya que, conocer cierta información que aportan distintos indicadores ayudará a la entidad a adelantarse a acontecimientos, distribuir mejor su capital y proponerse nuevos objetivos.

## 6. BIBLIOGRAFIA

- Montaño, N. (s. f.-b). *¿Qué es y para qué sirve un cuadro de mando?* <https://www.ambitbst.com/blog/qu%C3%A9-es-y-para-qu%C3%A9-sirve-un-cuadro-de-mando>
- Teoría de cuadros de mando. Tarjetas de puntuación y Dashboard.* (s. f.-b). <https://www.dataprix.com/es/book/export/html/1265>
- Roncancio, G. (2022, 5 octubre). *Historia del Cuadro de Mando Integral - CMI (contada por Kaplan).* <https://gestion.pensemos.com/historia-cuadro-de-mando-integral-por-robert-kaplan>
- Lazarus management. (s. f.). *Principales indicadores Financieros y de gestión que debes conocer. – Lazarus Management.* <https://www.lazarusmanagement.es/principales-indicadores-financieros-y-de-gestion-que-debes-conocer/>
- Narvaez, M. (2023, 23 febrero). *Indicadores de servicio al cliente: Qué son y ejemplos.* QuestionPro. <https://www.questionpro.com/blog/es/indicadores-de-servicio-al-cliente/>
- Ortega, C. (2021, 15 diciembre). *KPIs para procesos internos: Qué son, ejemplos, importancia y ventajas.* <https://tudashboard.com/kpis-para-procesos-internos/>
- Moya, D. P. (2021, 5 mayo). *Indicadores de innovación básicos para pymes.* Gestionar Fácil. <https://www.gestionar-facil.com/indicadores-de-gestion-todo-lo-que-debes-saber/indicadores-de-innovacion-para-pymes/>
- Mancuzo, G. (2022, 22 febrero). *EVM: ¿Qué es la Gestión del Valor Ganado?* Blog - ComparaSoftware. <https://blog.comparasoftware.com/evm-que-es-la-gestion-del-valor-ganado/>
- Da Silva, D. (2022, 27 junio). *4 indicadores de calidad en una empresa triunfante.* Zendesk MX. <https://www.zendesk.com.mx/blog/indicadores-calidad-empresa/>
- Andrés, Á. (2022, 28 marzo). *Cómo crear un buen plan de capacitación laboral.* Blog de Recursos Humanos de Bizneo HR: práctico y actual. <https://www.bizneo.com/blog/capacitacion-laboral/>
- Rodrigues, N. (2023, 20 enero). *Los 5 indicadores de procesos más relevantes para las empresas.* <https://blog.hubspot.es/sales/que-son-indicadores-de-procesos>
- Emilio, N. (2022, 18 enero). *Tipos de cuadros de mando.* <https://blog.bismart.com/tipos-de-cuadros-de-mando>
- Cuadros de Mando (dashboards y scorecards).* (2022, 13 diciembre). ABAST. <https://www.abast.es/business-intelligence-analytics/cuadros-de-mando-dashboards-scorecards/>
- Team, P. F. (2021, 27 agosto). *Business Activity Monitoring: Tomando decisiones en tiempo real.* Pandora FMS - The Monitoring Blog. <https://pandorafms.com/blog/es/bam-business-activity-monitoring/>

- Emilio, N. (2021, 29 junio). Diferencias entre un cuadro de mando integral y un dashboard. Bismart. <https://blog.bismart.com/diferencias-entre-un-cuadro-de-mando-integral-y-un-dashboard>
- Communications. (2023, 18 enero). Qué es la morosidad y qué consecuencias tiene. BBVA NOTICIAS. <https://www.bbva.com/es/salud-financiera/que-es-la-morosidad-y-que-consecuencias-tiene/>
- Home Credit Default Risk | Kaggle. (s. f.). <https://www.kaggle.com/competitions/home-credit-default-risk/overview>
- Europa Press. (2023, 24 marzo). La morosidad de la banca española se mantiene prácticamente estable en enero en el 3,5%. Público. <https://www.publico.es/economia/morosidad-banca-espanola-mantiene-practicamente-estable-enero-3-5.html>
- Inductores de Actuación - Vinculación de los Indicadores del Cuadro de Mando Integral con la. (s. f.). <https://1library.co/article/inductores-actuaci%C3%B3n-vinculaci%C3%B3n-indicadores-cuadro-mando-integral.yn9xr9kq>
- Admin. (2016, 24 mayo). *Credit bureau, mucho más que una calificadora de riesgo* | AIS Group | Aplicaciones de inteligencia artificial. AIS Group | Aplicaciones de inteligencia artificial. <https://ais-int.com/credit-bureau-mucho-mas-que-una-calificadora-de-riesgo/>
- ¿Cuáles son los bureaus de crédito en España? (2023, 12 mayo). MiSolvencia. <https://misolvencia.es/bureaus-de-credito-en-espana#que-es-un-bureau-de-credito>

## 7. ANEXO

### 7.1. Descripción de las variables

1. **application\_{train|test}.csv** es la tabla que presenta más variables y son las siguientes:
  - SK\_ID\_CURR: ID del préstamo en nuestra muestra.
  - TARGET: Variable objetivo en formato binario: 1 cuando el cliente presenta dificultades de pago, es decir, se atrasó más de X días en al menos una de las primeras Y cuotas del préstamo y 0 para el resto de casos.
  - NAME\_CONTRACT\_TYPE: Indica si el préstamo es revolving o cash.
  - CODE\_GENDER: El género del usuario.
  - FLAG\_OWN\_CAR: Indica si el cliente es propietario de un coche.
  - FLAG\_OWN\_REALTY: Indica si el cliente es propietario de una casa o apartamento.
  - CNT\_CHILDREN: Número de hijos del cliente.
  - AMT\_INCOME\_TOTAL: Renta del cliente.
  - AMT\_CREDIT: Monto del crédito del préstamo.
  - AMT\_ANNUITY: Anualidad del préstamo.
  - AMT\_GOODS\_PRICE: Es el precio de los bienes por los que se le dio al cliente el préstamo.
  - NAME\_TYPE\_SUITE: Nombre del acompañante del usuario en el momento de la realización del préstamo.
  - NAME\_INCOME\_TYPE: Profesión del usuario.
  - NAME\_EDUCATION\_TYPE: Último nivel de estudio alcanzado del usuario.
  - NAME\_FAMILY\_STATUS: Estatus familiar del cliente.
  - NAME\_HOUSING\_TYPE: Situación de habitabilidad del cliente (si es propietario, vive de alquiler, con sus padres, ...).
  - REGION\_POPULATION\_RELATIVE: Variable normalizada de población donde vive el usuario (el número más alto indica que el cliente vive en una región más poblada).
  - DAYS\_BIRTH: La edad del usuario cuando realizó el préstamo.
  - DAYS\_EMPLOYED: Recuento de los días previos al contrato trabajados en el lugar de trabajo actual.
  - DAYS\_REGISTRATION: Recuento de los días previos al contrato que el cliente cambió su registro.
  - DAYS\_ID\_PUBLISH: Recuento de los días previos al contrato en los que el cliente cambió su documento de identidad con el que firmó el contrato.
  - OWN\_CAR\_AGE: Años del coche del cliente.

- FLAG\_MOBIL: Indica si el cliente tiene móvil (1 = Sí, 0 = No).
- FLAG\_EMP\_PHONE: Indica si el cliente tiene móvil de empresa (1 = Sí, 0 = No).
- FLAG\_WORK\_PHONE: Indica si el cliente tiene teléfono fijo (1 = Sí, 0 = No).
- FLAG\_CONT\_MOBILE: Indica si el móvil es desechable (1 = Sí, 0 = No).
- FLAG\_PHONE: Indica si el cliente tiene teléfono fijo (1 = Sí, 0 = No).
- FLAG\_EMAIL: Indica si el cliente tiene e-mail (1 = Sí, 0 = No).
- OCCUPATION\_TYPE: Indica que hobbies tiene el cliente.
- CNT\_FAM\_MEMBERS: Indica el número de miembros en la familia.
- REGION\_RATING\_CLIENT: Indica la provincia en la que vive (dividido en 1, 2 y 3).
- REGION\_RATING\_CLIENT\_W\_CITY: Indica la provincia en la que vive incluyendo la ciudad (dividida en 1, 2 y 3).
- WEEKDAY\_APPR\_PROCESS\_START: El día de la semana en el que el cliente realizó el préstamo.
- HOUR\_APPR\_PROCESS\_START: La hora aproximada en la que se realizó el préstamo.
- REG\_REGION\_NOT\_LIVE\_REGION: Indica si la dirección no coincide con la dirección de contacto (1 = diferente, 0 = igual, a nivel provincial).
- REG\_REGION\_NOT\_WORK\_REGION: Indica si la dirección no coincide con la dirección del trabajo (1 = diferente, 0 = igual, a nivel provincial).
- LIVE\_REGION\_Not\_WORK\_REGION: Indica si la dirección de contacto no coincide con la del trabajo (1 = diferente, 0 = igual, a nivel provincial).
- REG\_CITY\_NOT\_LUVE\_CITY: Indica si la dirección no coincide con la dirección de contacto (1 = diferente, 0 = igual, a nivel de ciudad).
- REF\_CITY\_NOT\_WORK\_CITY: Indica si la dirección no coincide con la dirección del trabajo (1 = diferente, 0 = igual, a nivel de ciudad).
- LIVE\_CITY\_NOT\_WORK\_CITY: Indica si la dirección de contacto no coincide con la del trabajo (1 = diferente, 0 = igual, a nivel de ciudad).
- ORGANIZATION\_TYPE: Tipo de organización en la que trabaja el cliente.
- EXT\_SOURCE\_1: Puntuación normalizada de otra fuente de datos.
- EXT\_SOURCE\_2: Puntuación normalizada de otra fuente de datos.
- EXT\_SOURCE\_3: Puntuación normalizada de otra fuente de datos.
- APARTMENTS\_AVG, BASEMENTAREA\_AVG, YEARS\_BEGINEXPLUATATION\_AVG, YEARS\_BUILD\_AVG, COMMONAREA\_AVG, ELEVATORS\_AVG, ENTRANCES\_AVG, FLOORSMAX\_AVG, FLOORSMIN\_AVG, LANDAREA\_AVG, LIVINGAPARTMENTS\_AVG, LIVINGAREA\_AVG, NONLIVINGAPARTMENTS\_AVG, NONLIVINGAREA\_AVG, APARTMENTS\_MODE, BASEMENT\_MODE, YEARS\_BEGINEXPLUATATION\_MODE, YEARS\_BUILD\_MODE, COMMONAREA\_MODE, ELEVATORS\_MODE,

ENTRANCES\_MODE, FLOORSMAX\_MODE, FLOORSMIN\_MODE, LANDAREA\_MODE, LIVINGAPARTMENTS\_MODE, LIVINGAREA\_MODE, NONLIVINGAPARTMENTS\_MODE, NONLIVINGAREA\_MODE, APARTMENTS\_MEDI, BASEMENT\_MEDI, YEARS\_BEGINEXPLUATATION\_MEDI, YEARS\_BUILD\_MEDI, COMMONAREA\_MEDI, ELEVATORS\_MEDI, ENTRANCES\_MEDI, FLOORSMAX\_MEDI, FLOORSMIN\_MEDI, LANDAREA\_MEDI, LIVINGAPARTMENTS\_MEDI, LIVINGAREA\_MEDI, NONLIVINGAPARTMENTS\_MEDI, NONLIVINGAREA\_MEDI, FONDKAPREMONT\_MODE, HOUSETYPE\_MODE, TOTALAREA\_MODE, WALLSMATERIAL\_MODE y EMERGENCYSTATE\_MODE : Información normalizada sobre el edificio en el que reside el usuario: cuál es la media, moda y mediana del apartamento, área común, sala de estar, edad del edificio, número de ascensores, número de entradas, estado del edificio, número de la planta.

- OBS\_30\_CNT\_SOCIAL\_CIRCLE: Número de observaciones del entorno social del usuario con mora observable después de 30 días vencidos.
- DEF\_30\_CNT\_SOCIAL\_CIRCLE: Número de observaciones del entorno social del usuario incumplidas después de 30 días vencidos.
- OBS\_60\_CNT\_SOCIAL\_CIRCLE: Número de observaciones del entorno social del usuario con mora observable después de 60 días vencidos.
- DEF\_60\_CNT\_SOCIAL\_CIRCLE: Número de observaciones del entorno social del usuario incumplidas después de 60 días vencidos.
- DAYS\_LAST\_PHONE\_CHANGE: Recuento de los días previos al contrato desde que el cliente se cambió el teléfono.
- FLAG\_DOCUMENT (desde el 2 hasta el 21): Indica si el cliente entregó la documentación desde la 2 hasta la 21, respectivamente.
- AMT\_REQ\_CREDIT\_BUREAU\_HOUR: Número de consultas que se hicieron a Bureau Credit sobre el usuario una hora antes de la solicitud.
- AMT\_REQ\_CREDIT\_BUREAU\_DAY: Número de consultas que se hicieron a Bureau Credit sobre el usuario un día antes de la solicitud.
- AMT\_REQ\_CREDIT\_BUREAU\_WEEK: Número de consultas que se hicieron a Bureau Credit sobre el usuario una semana antes de la solicitud (excluyendo la hora de antes de la solicitud).
- AMT\_REQ\_CREDIT\_BUREAU\_MON: Número de consultas que se hicieron a Bureau Credit sobre el usuario un mes antes de la solicitud (excluyendo el día de antes de la solicitud).
- AMT\_REQ\_CREDIT\_BUREAU\_QRT: Número de consultas que se hicieron a Bureau Credit sobre el usuario 3 meses antes de la solicitud (excluyendo el mes de antes de la solicitud).
- AMT\_REQ\_CREDIT\_BUREAU\_YEAR: Número de consultas que se hicieron a Bureau Credit sobre el usuario un año antes de la solicitud (excluyendo los últimos 3 meses de antes de la solicitud).

**2. bureau.csv** contiene diecisiete variables. Son las siguientes:

- SK\_ID\_CURR: ID del préstamo en nuestra muestra. Cada préstamo puede tener más de un crédito previo en Credit Bureau (CB).

- SK\_BUREAU\_ID: Recopilación de los préstamos previos de CB relacionados con nuestro préstamo (ID único para cada solicitud de préstamo).
- CREDIT\_ACTIVE: Estado de los créditos reportados por CB
- CREDIT\_CURRENCY: Recopilación de la moneda del crédito en CB.
- DAYS\_CREDIT: Número de días previos a la solicitud actual en los que el cliente solicitó un préstamo a CB.
- CREDIT\_DAY\_OVERDUE: Número de días de mora en el crédito de CB en el momento de la solicitud del préstamo actual.
- DAYS\_CREDIT\_ENDDATE: Duración de días restantes del crédito de CB en el momento de la solicitud en Home Credit.
- DAYS\_ENDDATE\_FACT: Días desde que se terminó el crédito en CB en el momento de la solicitud en Home Credit.
- AMT\_CREDIT\_MAX\_OVERDUE: Monto máximo vencido en el CB hasta el momento, es decir, a fecha de la nueva solicitud de préstamo.
- CNT\_CREDIT\_PROLONG: Número de veces que el crédito en Credit Bureau fue prolongado.
- AMT\_CREDIT\_SUM: Monto de crédito actual del crédito de Credit Bureau.
- AMT\_CREDIT\_SUM\_DEBT: Deuda actual en el crédito de CB.
- AMT\_CREDIT\_SUM\_LIMIT: Límite de crédito actual de la tarjeta reportado a CB.
- AMT\_CREDIT\_SUM\_OVERDUE: Monto vencido actual en el crédito de CB.
- CREDIT\_TPE: Tipo de crédito pedido en CB (tarjeta, efectivo, ...).
- DAYS\_CREDIT\_UPDATE: Cuántos días antes de la solicitud llegó la última información del crédito de Credit Bureau.
- AMT\_ANNUITY: Anulidad del crédito de Credit Bureau.

**3. bureau\_balance.csv** contiene las siguientes tres variables:

- SK\_BUREAU\_ID: ID de los créditos de CB (código único para cada solicitud).
- MONTHS\_BALANCE: Mes de saldo relativo a la fecha de la solicitud. El -1 indica la fecha de saldo más reciente.
- STATUS: Estado del préstamo del Credit Bureau durante el mes. Se clasifica como: activo, cerrado, DPD0-30, ... donde C significa cerrado, X significa estado desconocido, 0 que no hay DPD (vencimiento), 1 que se hizo el máximo entre el día 1 y 30, el 2 significa DPD 31-60, ..., 5 significa DPD 120+ o vendido/cancelado.

**4. POS\_CASH\_balance.csv** contiene ocho variables que son:

- SK\_ID\_PREV: ID de créditos previos en Home Credit relacionados con los préstamos de la muestra. Un préstamo en la muestra puede tener más de un préstamo en Home Credit.

- SK\_ID\_CURR: ID del préstamo en la muestra.
- MONTHS\_BALANCE: Mes de saldo en relación a la fecha de la solicitud. El -1 indica la fecha de saldo más reciente, 0 la información en la solicitud que a menudo será igual a -1 ya que los bancos no actualizan con regularidad la información en Credit Bureau.
- CNT\_INSTALMENT: Plazo del crédito anterior (puede cambiar con el tiempo).
- CNT\_INSTALMENT\_FUTURE: Cuotas que quedan por pagar del crédito anterior.
- NAME\_CONTRACT\_STATUS: Estado del contrato durante el mes.
- SK\_DPD: Días de vencimiento durante el mes de los créditos previos.
- SK\_DPD\_DEF: Días de vencimiento durante el mes con tolerancia del crédito anterior (se ignoran las deudas con montos bajos).

**5. credit\_card\_balance.csv** contiene más de veinte variables que se explican a continuación:

- SK\_ID\_PREV: ID de créditos previos en Home Credit relacionados con los préstamos de la muestra. Un préstamo en la muestra puede tener más de un préstamo en Home Credit.
- SK\_ID\_CURR: ID del préstamo en la muestra.
- MONTHS\_BALANCE: Mes de saldo relativo a la fecha de la solicitud. El -1 indica la fecha de saldo más reciente.
- AMT\_BALANCE: Balance durante el mes de los créditos previos.
- AMT\_CREDIT\_LIMIT\_ACTUAL: Límite de la tarjeta de crédito durante el mes de los créditos anteriores.
- AMT\_DRAWINGS\_ATM\_CURRENT: Importe de veces que sacó dinero en ATM en el mes de los créditos previos.
- AMT\_DRAWINGS\_CURRENT: Cantidad sacada durante el mes en los créditos previos.
- AMT\_DRAWINGS\_OTHER\_CURRENT: Importe de otras extracciones durante el mes en los créditos previos.
- AMT\_DRAWINGS\_POS\_CURRENT: Cantidad de compras durante el mes de los créditos previos.
- AMT\_INST\_MIN\_REGULARITY: Cuota mínima del mes de los créditos anteriores.
- AMT\_PAYMENT\_CURRENT: Importe pagada por el cliente durante el mes en los créditos previos.
- AMT\_PAYMENT\_TOTAL\_CURRENT: Cantidad total pagada por el cliente durante el mes en los créditos previos.
- AMT\_RECEIVABLE\_PRINCIPAL: Importe a cobrar al titular en los créditos anteriores.
- AMT\_RECIVABLE: Importe a cobrar en los créditos previos.
- AMT\_TOTAL\_RECEIVABLE: Total importe por cobrar en los créditos previos.

- CNT\_DRAWINGS\_ATM\_CURRENT: Número de operaciones en ATM durante el mes de los créditos previos.
- CNT\_DRAWINGS\_CURRENT: Número de operaciones durante el mes de los créditos anteriores.
- CNT\_DRAWINGS\_OTHER\_CURRENT: Número de otras operaciones durante el mes de los créditos anteriores.
- CNT\_DRAWINGS\_POS\_CURRENT: Número de operaciones de compras durante el mes en los créditos previos.
- CNT\_INSTALMENT\_MATURE\_CUM: Número de cuotas pagas de los créditos previos.
- NAME\_CONTRACT\_STATUS: Estado del contrato de los créditos previos (activo, firmado, ...)
- SK\_DPD: Días de vencimiento durante el mes de los créditos previos.
- SK\_DPD\_DEF: Días de vencimiento durante el mes con tolerancia del crédito anterior (se ignoran las deudas con montos bajos)

**6.previous\_application.csv** cuenta con prácticamente 40 variables que hacen referencia a lo siguiente:

- SK\_ID\_PREV: ID de créditos previos en Home Credit relacionados con los préstamos de la muestra. Un préstamo en la muestra puede tener más de un préstamo en Home Credit.
- SK\_ID\_CURR: ID del préstamo en la muestra.
- NAME\_CONTRACT\_TYPE: Contrato del tipo de producto (préstamo, revolving, préstamo al consumo, ...) de los créditos anteriores.
- AMT\_ANNUITY: Anualidad de los créditos anteriores.
- AMT\_APPLICATION: Importe que solicitó el cliente en los créditos previos.
- AMT\_CREDIT: Monto del crédito final en la solicitud anterior.
- AMT\_DOWN\_PAYMENT: Pago inicial de la solicitud anterior.
- AMT\_GOODS\_PRICE: Precio de los bienes que el cliente solicitó (si corresponde) en la solicitud anterior.
- WEEKDAY\_APPR\_PROCESS\_START: Día de la semana en que el cliente solicitó la solicitud anterior.
- HOUR\_APPR\_PROCESS\_START: Hora aproximada del día en el que se realizó la solicitud anterior.
- FLAG\_LAST\_APPL\_PER\_CONTRACT: Indica si fue la última solicitud del contrato anterior. A veces, por error del cliente o del empleado, pueden aparecer más de una solicitud para un único contrato.
- NFLAG\_LAST\_APPL\_IN\_DAY: Indica si la solicitud fue la última del día del cliente. A veces, el cliente realiza más de una solicitud en un mismo día. También, en raras

ocasiones, podría ser un error en el sistema en el que una aplicación aparezca duplicada.

- NFLAG\_MICRO\_CASH: Indica si el préstamo es de un monto pequeño.
- RATE\_DOWN\_PAYMENT: Tasa de anticipo normalizada del crédito anterior.
- RATE\_INTEREST\_PRIMARY y RATE\_INTEREST\_PRIVILEGED: Ratio de interés normalizado en el crédito previo.
- NAME\_CASH\_LOAN\_PURPOSE: Objetivo del préstamo en efectivo.
- NAME\_CONTRACT\_STATUS: Estado del contrato del crédito anterior (aprobado, cancelado, ...).
- DAYS\_DECISION: Relacionado con la solicitud actual, indica cuándo se tomó la decisión sobre la solicitud anterior.
- NAME\_PAYMENT\_TYPE: Método de pago que utilizó el usuario en su anterior crédito.
- CODE\_REJECT\_REASON: Porque se rechazó la solicitud anterior.
- NAME\_TYPE\_SUITE: Nombre del acompañante del cliente en el momento de la solicitud del crédito anterior.
- NAME\_CLIENT\_TYPE: Indica si el cliente era ya conocido o no en el momento de la solicitud previa.
- NAME\_GOODS\_CATEGORY: Qué bienes solicitó el usuario en el previo crédito.
- NAME\_PORTFOLIO: Indica para que fue la solicitud anterior: coche, efectivo, ...
- NAME\_PRODUCT\_TYPE: Indica si el crédito anterior se solicitó sin cita previa (*walk-in*) o por adhesión a otro producto o servicio (*x-sell*).
- CHANNEL\_TYPE: Canal por el que llegó el cliente en el crédito previo.
- SELLERPLACE\_AREA: Lugar de la venta del crédito anterior.
- NAME\_SELLER\_INDUSTRY: Sector del vendedor.
- CNT\_PAYMENT: Términos del crédito anterior.
- NAME\_YIELD\_GROUP: Interés (dividido en pequeño, medio y alto) de la solicitud anterior.
- PRODUCT\_COMBINATION: Agrupación detallada de la solicitud previa.
- DAYS\_FIRST\_DRAWING: Relacionado con la fecha de la solicitud de la solicitud actual, cuando fue el primer desembolso del crédito anterior.
- DAYS\_FIRST\_DUE: En relación con la fecha de solicitud de la solicitud actual, cuando se suponía que era la primera fecha de vencimiento del anterior crédito.
- DAYS\_LAST\_DUE\_1ST\_VERSION: En relación con la fecha de solicitud de la solicitud actual, cuando fue la primera fecha de vencimiento del anterior crédito.

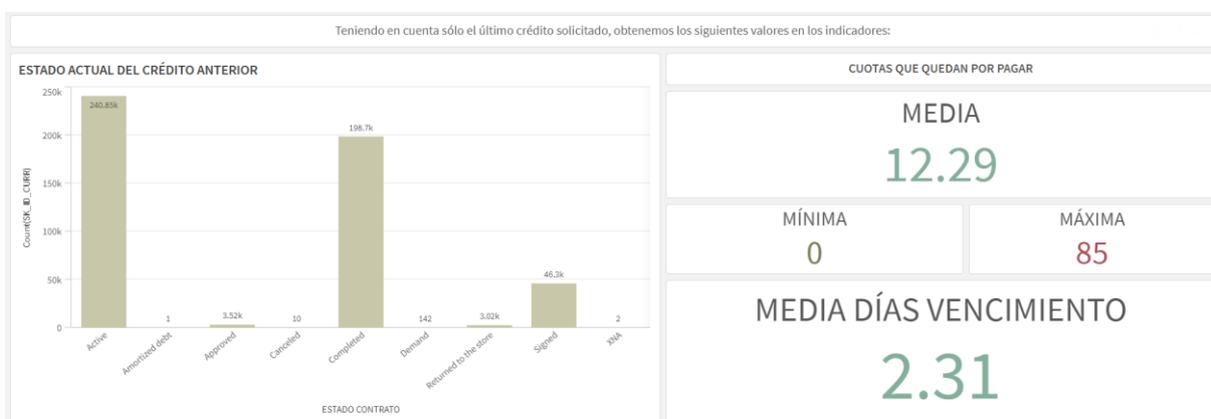
- DAYS\_LAST\_DUE: En relación con la fecha de solicitud de la solicitud actual, cuando fue la última fecha de vencimiento del anterior crédito.
- DAYS\_TERMINATION: En relación con la fecha de solicitud de la solicitud actual, indica cuando se esperaba finalizar el anterior crédito.
- NFLAG\_INSURED\_ON\_APPROVAL: Indica si el cliente solicitó seguro durante el ciclo de vida del crédito previo.

**7. installments\_payments.csv** recopila ocho variables que se explican a continuación:

- SK\_ID\_PREV: ID de créditos previos en Home Credit relacionados con los préstamos de la muestra. Un préstamo en la muestra puede tener más de un préstamo en Home Credit.
- SK\_ID\_CURR: ID del préstamo en la muestra.
- NUM\_INSTALLMENT\_VERSION: Versión del calendario del crédito anterior (0 indica tarjeta de crédito). El cambio de versión de la cuota mensual quiere decir que algún parámetro del calendario de pago ha cambiado.
- NUM\_INSTALLMENT\_NUMBER: Cuota en la que se observa el pago.
- DAYS\_INSTALLMENT: Fecha en la que se suponía que debía pagarse la cuota del crédito anterior (respecto a la fecha de solicitud del préstamo actual).
- DAYS\_ENTRY\_PAYMENT: Fecha en la que realmente se pagaron las cuotas del crédito anterior (en relación a la fecha de solicitud del préstamo actual).
- AMT\_INSTALLMENT: Monto de la cuota prescrita del crédito anterior en dicha cuota.
- AMT\_PAYMENT: Lo que realmente pagó el cliente en el crédito previo en dicha cuota.

**7.2. Cdm contrato anterior**

Hoja filtrada por cliente repetido y renovado.



Hoja filtrada por cliente nuevo

