

CIENCIA DE DATOS APLICADOS

# SEGMENTACIÓN Y ESTRATEGIA COMERCIAL

Integrantes:

- *Camila Constanza Aguilera Bustamante*
- *Diego Alejandro Herrera Villalobos*
- *Wladimir Richard Parada Rebolledo*
- *Néstor Patricio Rojas Ríos*
- *Ramiro Andrés Uribe Garrido*

Entrega: 24/10/2023

# Índice.

Índice.	1
Contexto.	2
Evaluación del conjunto de datos y Feature engineering.	2
Generación de los datasets de trabajo.	3
<i>Dataset 1: Dataset sin valores outliers.</i>	3
<i>Dataset 2: Dataset con variables clave pesadas.</i>	4
<i>Dataset 3: Dataset sin variables de poca varianza.</i>	4
<i>Dataset 4: Dataset de componentes principales.</i>	4
Segmentación.	5
<i>Selección del modelo de segmentación.</i>	5
<i>Selección de los parámetros del modelo K-means.</i>	5
<i>La segmentación escogida.</i>	7
<i>Análisis de componentes principales (PCA).</i>	8
Estrategia comercial.	10
<i>Descripción y alcance de la estrategia.</i>	10
<i>Recursos necesarios para la implementación.</i>	14
<i>Implementación.</i>	15
Palabras finales...	16
Anexos.	17

## Contexto.

En un entorno financiero altamente competitivo, la capacidad de un banco para retener y maximizar el valor de su cartera de clientes de alto valor se convierte en un factor crítico para su éxito y crecimiento sostenible. En este contexto, el presente informe se basa en un análisis exhaustivo de la cartera de clientes de un banco, donde el objetivo principal será agrupar o *clusterizar* a los clientes de forma de identificar potenciales segmentos para el desarrollo de un plan estratégico basado en sus características principales e intereses en los servicios financieros. Es así que, utilizando el lenguaje de programación **R**, realizamos una segmentación usando el modelo **K-means**. Esta técnica de análisis de datos ha permitido una segmentación precisa de los clientes en grupos homogéneos con características y necesidades financieras similares. La aplicación de **K-means** no solo ha revelado patrones significativos en la cartera de clientes, sino que también ha servido como base fundamental para la formulación de una estrategia comercial específica para los segmentos más relevantes.

A lo largo de este informe, se analizarán en detalle los componentes clave de esta campaña, desde los objetivos hasta la ejecución y el seguimiento, con el fin de proporcionar una comprensión integral de cómo el banco pretende llevar a cabo su estrategia comercial y mejorar la participación de sus clientes de alto valor. Además, se examinarán los resultados obtenidos y se destacarán las áreas para ajustes y mejoras continuas.

## Evaluación del conjunto de datos y *Feature engineering*.

Para desarrollar estrategias comerciales orientadas a los clientes de un banco es importante segmentarlos correctamente. Antes de cualquier proceso de segmentación es necesario conocer bien los datos con los cuales se trabajarán.

El conjunto de datos para realizar la segmentación (el *dataset*) tiene **45.206 registros** de clientes de banco, cada uno con información demográfica (***Fecha Nacimiento, Actividad laboral, Estado Civil y Nivel Educativo***) y de su interacción con productos y servicios del banco (***IdCliente, Saldo Medio Anual, Tiene Crédito Hipotecario, Tiene Crédito de Consumo, Tiene Inversiones, Tiene Mora, Contactos con su Ejecutivo y Medio de Contacto Preferente***). Afortunadamente para nosotros, no existen datos faltantes, por lo tanto, podríamos usar toda la información sin necesidad de eliminar registros incompletos o de imputar datos faltantes.

En cuanto al tipo de variables (*features*) con los que se trabajará, tenemos:

- 2 variables numéricas continuas: ***Saldo Medio Anual y Contactos con su Ejecutivo***.
- 4 variables categóricas binarias: ***Tiene Crédito Hipotecario, Tiene Crédito de Consumo, Tiene Inversiones y Tiene Mora***.
- 4 variables categóricas multiclase: ***Actividad laboral*** (7 clases), ***Estado Civil*** (3 clases), ***Nivel Educativo*** (4 clases) y ***Medio de Contacto Preferente*** (3 clases).
- 1 variable categórica de identificación: ***IdCliente***.

- 1 variable categórica de tiempo: **Fecha Nacimiento**.

Puesto que el método que escogimos para trabajar es la segmentación no supervisada mediante el modelo **K-means** (cuyas razones para su elección serán descritas más adelante), es importante hacer algunas modificaciones al *dataset*:

- El modelo funciona con variables cuantitativas, por lo que toda la información se debe transformar para obtener una escala numérica.
- Para evitar la distorsión originada por distintas magnitudes se deben escalar los resultados para que tengan rangos similares.

En primer lugar, las variables binarias se modificaron para que su valor fuera **1** cuando originalmente era **Sí**, y **0** cuando su valor inicial era **No**. Luego, a partir de las variables categóricas multiclase, se generaron variables categóricas binarias para cada una de las clases representadas (por ejemplo, la variable **Estado Civil** generó las variables binarias: **EC\_soltero**, **EC\_casado** y **EC\_divorciado**). Posteriormente, desde la variable **Fecha Nacimiento** se generó la variable **Edad**, calculando la edad al 1 de enero de 2010, fecha en que se habrían generado estos datos. Por último, se descartó la variable de identificación **IdCliente**, ya que es sólo un número correlativo.

Una vez que todas las variables contaron con valores de tipo numérico, se ajustaron sus escalas. Para eso se procedió a la **normalización** de estas, donde el valor más bajo se asignó como **0**, el valor más alto como **1**, y los demás valores se asignaron de forma proporcional. En el caso de las variables binarias derivadas de las categóricas multiclase, el valor se dividió por el número de clases de la variable original. Esto se hizo para mitigar la distorsión en el modelo que podría producirse si una de las variables originales estaba sobrerrepresentada al descomponerse en tantas variables como clases tenía. Con el mismo fin anterior, se eliminó una variable binaria derivada por cada variable multiclase original, ya que la información entregada por estas es redundante, pues corresponde a las otras variables binarias marcando **0**. La información que quedó por descarte de las otras fue: para **Actividad Laboral**, jubilado; para **Estado Civil**, divorciado; para **Nivel Educativo**, enseñanza básica; y para **Medio de Contacto Preferente**, teléfono particular.

## Generación de los *datasets* de trabajo.

Dataset 1: *Dataset sin valores outliers*.

Teniendo ya definidas las variables del *dataset*, se realizó una inspección de los datos numéricos continuos (**Saldo Medio Anual**, **Contactos con su Ejecutivo** y **Edad**) en busca de valores *outliers*. Al realizar gráficas de puntos bidimensionales entre estas tres variables, tanto en sus originales escalas lineales entre sí como en escala logarítmica entre sí, se pudieron definir algunos puntos de corte que, por contexto, descartaron algunos valores extremos:

- Para **Saldo Medio Anual** se descartaron aquellos registros con valores menores que 0,02 y mayores que 0,8.
- Para **Contactos con su Ejecutivo** se descartaron los registros con valores mayores a 0,18.
- Para **Edad** se descartaron los registros de valor 1 y aquellos cuyo valor de logaritmo fuese menor a -6.

De esta forma se descartaron 48 clientes, obteniendo un *dataset* bidimensional con **20 variables predictoras** y **45.148 registros de clientes**. Cuando en el texto se refiera al ***Dataset sin valores outliers*** se referirá a esta formación.

Dataset 2: *Dataset con variables clave pesadas*.

Con el equipo se discutió cuáles variables podrían ser más importantes para el negocio bancario. Se identificaron 6 características clave: la **Edad**, la **Actividad Laboral**, el **Saldo Medio Anual**, si **Tiene Crédito Hipotecario**, si **Tiene Crédito de Consumo** y, si **Tiene Mora**. Para resaltar la importancia que estos factores podrían tener en el ejercicio de segmentación es que se generó una variante del ***Dataset sin valores outliers***, en la cual todos los valores de las variables derivadas de estos aspectos fueron escalados por 2. De esta forma se generó el ***Dataset con variables clave pesadas***.

Dataset 3: *Dataset sin variables de poca varianza*.

Al observar los cuartiles del dataset llama la atención de que algunas variables tienen más del 75% de sus valores exactamente iguales, indicando que hay poca variación entre ellos y, probablemente, entreguen poca información a la generación de segmentos. Es por esto que se generó una variante del ***Dataset con variables clave pesadas*** en el cual se eliminan aquellas variables cuyos valores eran iguales en más de un 75% y que no pertenecieran al grupo de variables clave. Se descartaron ***Contactos con su Ejecutivo***; ***Tiene Inversiones***; para ***Nivel Educativo***, enseñanza media y técnico profesional; para ***Estado Civil***, todas las variables derivadas; para ***Medio de Contacto Preferente***, emailing. De esta forma se generó el ***Dataset con variables pesadas que descarta aquellas de poca importancia***, con **13 variables predictoras** y **45.148 registros de clientes**.

Dataset 4: *Dataset de componentes principales*.

Otro enfoque que se planteó para obtener el máximo de información posible del *dataset*, fue realizar un trabajo de disminución de dimensionalidad. Al ***Dataset sin valores outliers*** se le realizó un **Análisis de Componentes Principales (PCA**, por sus siglas en inglés) para obtener los componentes con mayor varianza y, en consecuencia, pudiesen aportar la mayor cantidad de información para entrenar el modelo de segmentación. Producto de ello, se obtuvieron los siguientes componentes:

Componente	Varianza (%)	Varianza acumulada (%)	Componente	Varianza (%)	Varianza acumulada (%)
Componente 1	38,972	38,97	Componente 11	0,748	98,61
Componente 2	21,096	60,07	Componente 12	0,508	99,12
Componente 3	10,829	70,90	Componente 13	0,271	99,39
Componente 4	7,223	78,12	Componente 14	0,204	99,60
Componente 5	6,866	84,99	Componente 15	0,136	99,73
Componente 6	6,124	91,11	Componente 16	0,098	99,83
Componente 7	2,917	94,03	Componente 17	0,089	99,92
Componente 8	1,618	95,65	Componente 18	0,065	99,98
Componente 9	1,427	97,07	Componente 19	0,009	99,99
Componente 10	0,793	97,87	Componente 20	0,007	100,00

El **PCA** se profundizará más adelante, pero en este momento es importante establecer que decidimos quedarnos con los 5 primeros componentes principales, pues juntos explican el **85% de la varianza** total del *dataset*. Con estas **5 columnas** y **45.148 registros de clientes** se definió el ***Dataset de componentes principales***.

## Segmentación.

### Selección del modelo de segmentación.

Una vez generados los *datasets* para trabajar, se procedió con la segmentación propiamente tal. Dentro del mundo del aprendizaje automático (*Machine Learning*), los modelos no supervisados, aquellos que no tienen un valor que predecir, nos ofrecen herramientas para hacer segmentación. El más conocido es el modelo ***K-means***, pero existen otros, tales como: el modelo de *Aglomeración jerárquica*, el modelo *K-prototipos*, el *Modelo de mezcla gaussiana* y el modelo de *Aglomeración espacial basado en densidad de aplicaciones con ruido* (*DBSCAN*, por sus siglas en inglés), entre otros. De entre todos estos modelos nos decantamos por usar ***K-means*** por las siguientes razones:

- ***Familiaridad***: es el modelo de aprendizaje no supervisado más popular para hacer segmentación, además de que el equipo lo conoce ampliamente por ya haber trabajado con él, por lo que podemos tomar con confianza los resultados obtenidos.
- ***Coste***: algoritmos basados en distancias, como el modelo de *Aglomeración jerárquica*, al necesitar generar una matriz de distancia de todos los datos, tiene un alto coste de memoria en su ejecución, lo cual limita el número de pruebas. ***K-means*** es bastante liviano de ejecutar y no toma mucho tiempo, por lo que realizar ajustes de parámetros buscando un resultado satisfactorio se hace muy fácil.
- ***Interpretabilidad***: una vez generados los segmentos, estos son fáciles de interpretar. Además, de las primeras pruebas realizadas con varios modelos, con ***K-means*** se obtuvieron los resultados que más sentido tenían para la lógica del negocio.

### Selección de los parámetros del modelo *K-means*.

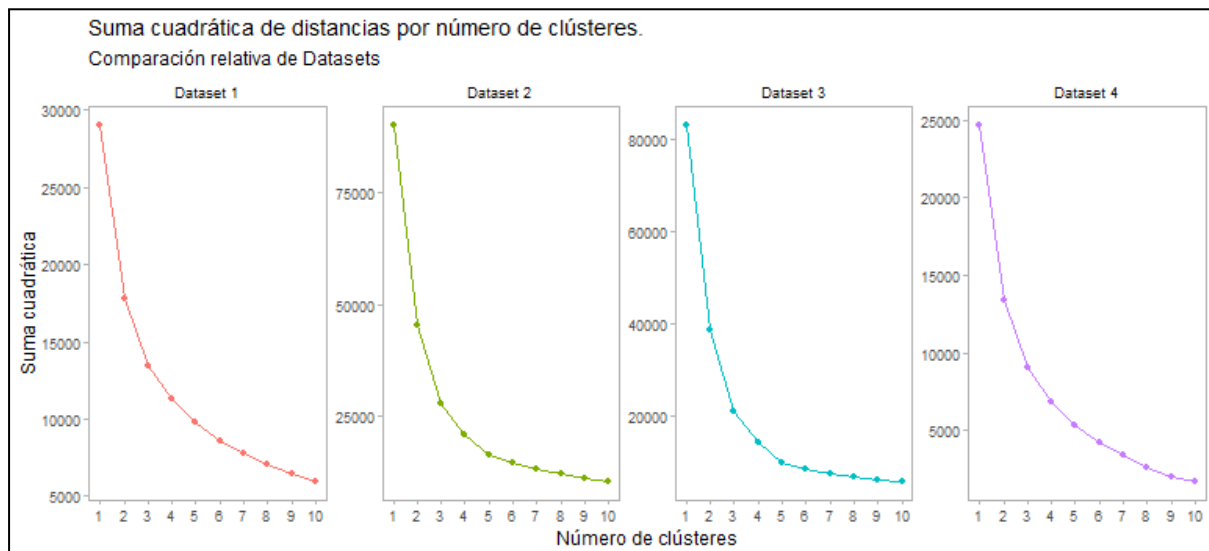
Con el modelo de segmentación ya seleccionado, pudimos enfocarnos en encontrar la mejor combinación de parámetros que nos diera un resultado satisfactorio. Los objetivos fijados para la segmentación fueron los siguientes:

- Los *clusters* generados deben tener buena ***cohesión***, es decir, la distancia entre sus registros constituyentes debe ser baja.
- Los *clusters* generados deben tener buena ***separación***, es decir, la distancia entre registros de distintos *clusters* debe ser alta.
- Los *clusters* generados deben dar ***información*** atinente para el negocio en el que estamos trabajando, en este caso, el sector bancario.

Buscando el equilibrio entre estos tres objetivos, se decidió probar con los **4 *datasets* ya descritos** y con **valores de *K* entre 2 y 10**. Dado que la asignación de *clusters* puede variar entre ejecuciones para aquellos registros que se encuentren en los bordes de sus segmentos, es esperable cierta variabilidad de los resultados con cada ejecución del algoritmo de ***K-means***. Es por esto que **cada una de las**

**combinaciones de valores de K y de datasets fue ejecutada 30 veces.** Las comparativas realizadas en los siguientes párrafos corresponden al valor promedio de esas 30 ejecuciones.

Una de las formas de encontrar el valor óptimo de **K** es mediante el cálculo de la suma cuadrática de las distancias de cada uno de los registros integrantes del *cluster* con el **centroide** del propio *cluster*. Recordemos que el **centroide** es un vector que tiene el valor promedio de cada una de las variables usadas en el *dataset* para cada uno de los *clusters* que se generaron mediante el modelo **K-means**, por lo que hay tantos **centroides** como *clusters* existan. Pues bien, para cada registro se mide la distancia de su valor en cada variable predictora con el correspondiente valor del **centroide**, se eleva al cuadrado y se suma. A mayor cantidad de *clusters* menor es la distancia que hay entre cada registro del *cluster* y su **centroide**. Sin embargo, al graficar la suma cuadrática total para cada *cluster*, se pueden identificar puntos en donde se produce un quiebre en la línea descendente. Ese punto se conoce como **el codo**. En el *gráfico 1* se comparan los valores de suma cuadrática de las distancias a escala relativa para cada uno de los *datasets* generados. Se puede observar que los *datasets* con variables claves pesadas (*Datasets 2 y 3*) presentan puntos candidatos a **codos** más notorios, pero en todos los casos se pueden identificar los **valores de K entre 3 y 5** como los mejores candidatos para el número óptimo de *clusters*.

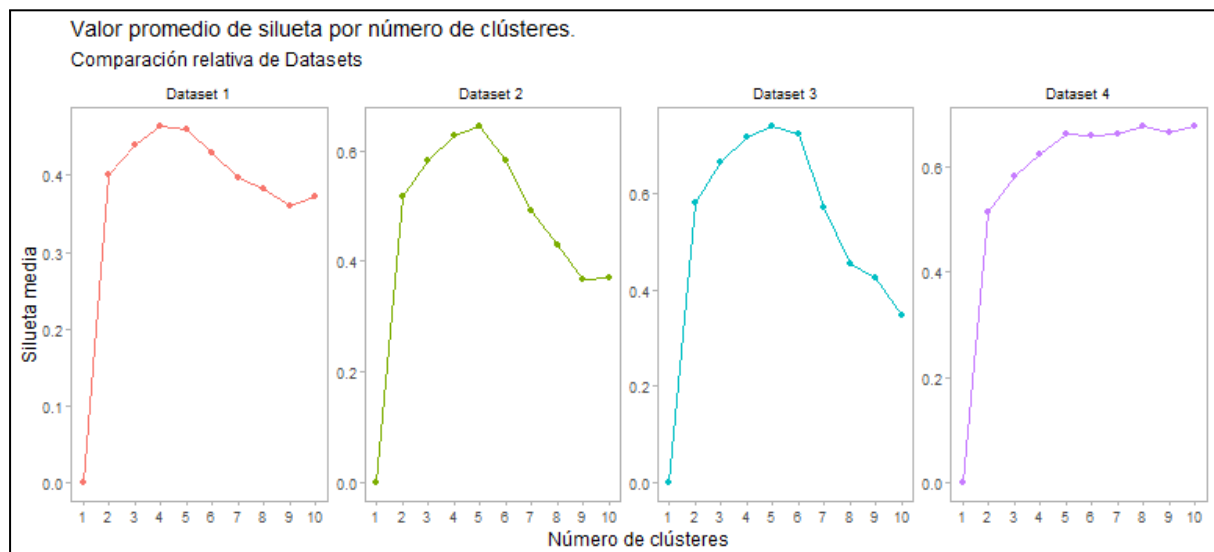


**Gráfico 1:** Comparación de la suma cuadrática promedio para cada valor de **K**, por *dataset*.

Por otro lado, podemos evaluar los *clusters* de acuerdo a su *cohesión* y su *separación* mediante el cálculo del **valor de silueta** para cada uno de los elementos del *cluster*. De esta forma, podemos calcular también el **valor de silueta** promedio para cada *cluster* y para el *dataset* completo, tomando valores entre 1 (la mejor *cohesión* y *separación*) y -1 (mejor pertenecer a otro *cluster*). En el *gráfico 2* se observa el **valor de silueta** para cada uno de los *datasets* y los valores de **K** propuestos, también con escala relativa. Se puede observar que, para todos los *datasets*, el **valor de K igual a 5** es uno de los más altos, siendo sólo superado por 4 en el *Dataset 1* y por unos máximos relativos en valores 8 y 10 para el *Dataset 4*.

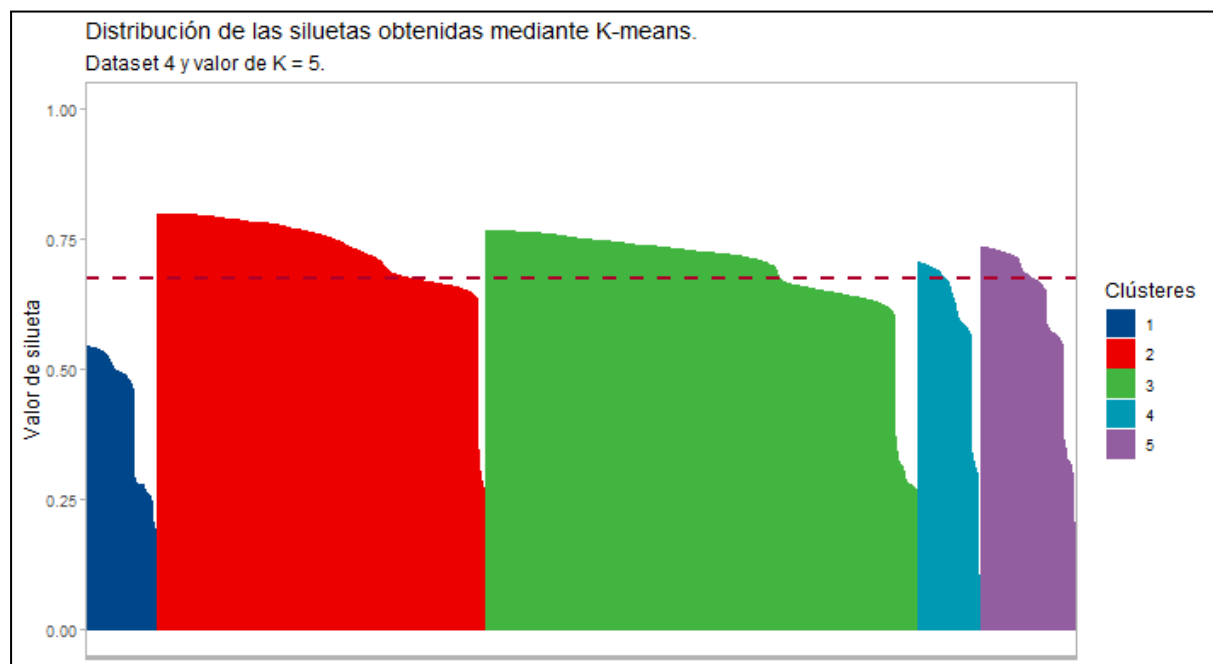
Por último, queda abordar la *interpretabilidad* de los resultados de la segmentación. Para este paso, con la información ya recopilada con los dos métodos previos, revisamos la **información de los clusters generados para los 4 datasets con valores de K iguales a 4 y 5** quedándonos, finalmente, con la combinación del **Dataset 4 y el valor de K de 5 clusters**, que obtuvo un valor promedio de

suma cuadrática de distancias de 5.315,75 y un valor promedio de silueta de 0,66. La *interpretabilidad* de esta segmentación, así como un resumen de la información entregada, serán evaluados más adelante en este informe.



**Gráfico 2:** Comparación del valor de *silueta* promedio para cada valor de **K**, por *dataset*.

La segmentación escogida.



**Gráfico 3:** Valores de silueta para cada registro del *Dataset 4*, agrupado por *cluster*.

Recordemos que el *Dataset 4* corresponde a los cinco primeros componentes principales generados mediante el modelo *PCA* al *Dataset sin valores outliers*, el cual explica el 85% de la varianza del *dataset* original. Al generar 5 segmentos mediante el modelo de **K-means** obtenemos lo siguiente:

- **Cluster 1:** 3.221 registros, el 7,13% del *dataset*, con un valor de silueta promedio de 0,43.
- **Cluster 2:** 15.005 registros, el 33,23% del *dataset*, con un valor de silueta promedio de 0,73.
- **Cluster 3:** 19.703 registros, el 43,63% del *dataset*, con un valor de silueta promedio de 0,69.



- **Cluster 4:** 2.873 registros, el 6,36% del *dataset*, con un valor de silueta promedio de 0,60.
- **Cluster 5:** 4.356 registros, el 9,65% del *dataset*, con un valor de silueta promedio de 0,63.

El *gráfico 3* muestra el **valor de silueta** para cada registro del *dataset* agrupado por *cluster*. En él se aprecia que el menos definido es el *cluster 1*, cuyo valor promedio es menor a 0,5; sin embargo, los otros *clusters* parecen bastante sólidos, especialmente los *clusters 2 y 3*. Cabe destacar que para esta segmentación ningún registro obtuvo un **valor de silueta** menor a 0, lo cual respalda la decisión de escoger este *dataset* con este número de *clusters*.

#### Análisis de componentes principales (PCA).

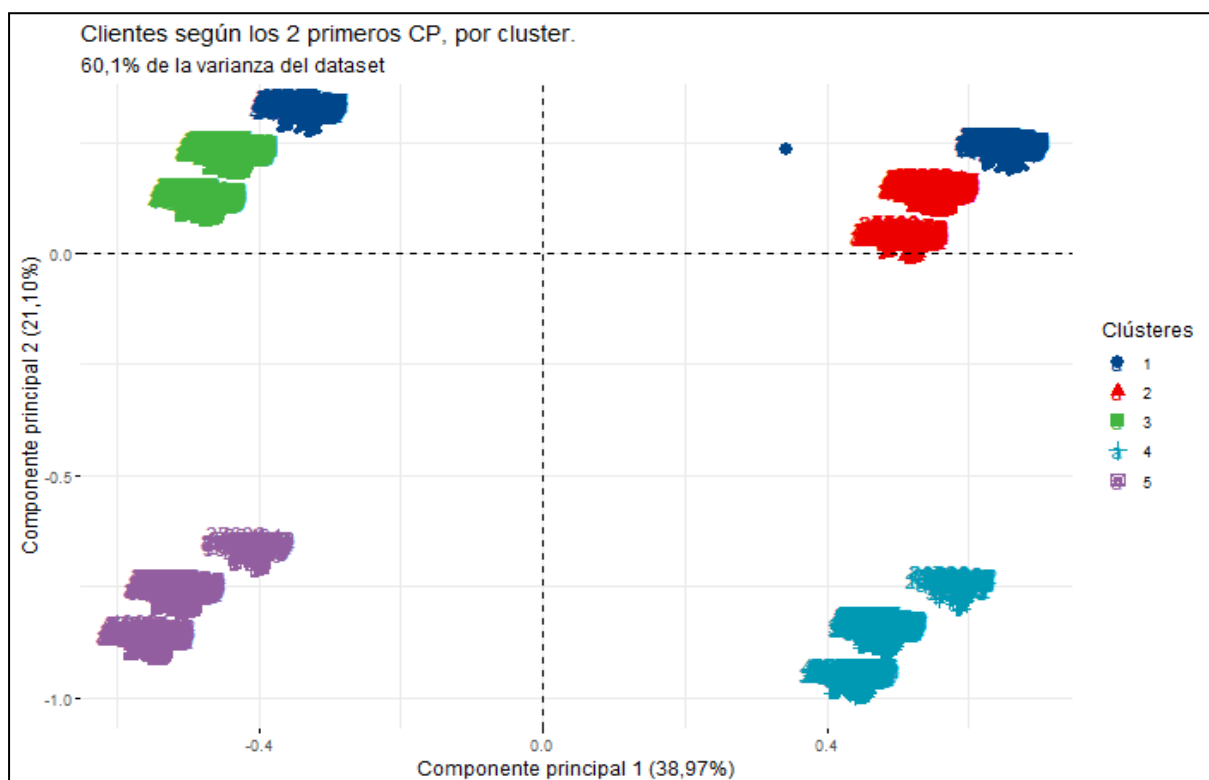
Como se mencionó anteriormente, se realizó un análisis de *PCA* para obtener la mayor información posible del *dataset* en base a la varianza que logra capturar cada componente principal. Ya se vieron los aportes de varianza de cada uno de los componentes, pero falta aún una interpretación de lo que representa cada uno de los componentes usados. Recordemos que para generar el *dataset* que se usó en el modelo *K-means*, se seleccionaron los 5 primeros componentes principales que explican el 85% de la varianza del *Dataset sin valores outliers*. Pues bien, al examinar los pesos que cada componente asigna a cada una de las variables, es posible identificar aquellas que más aportaron para cada uno. En la tabla siguiente se presentan los pesos de todas las variables para los primeros 5 componentes principales, resaltando en color verde aquellos cuyo valor absoluto es mayor de 0,1 y en negrita aquellos cuyo valor absoluto es mayor a 0,3:

Variables	Componente principal 1	Componente principal 2	Componente principal 3	Componente principal 4	Componente principal 5
Saldo Medio Anual	0,0043	0,0071	0,0122	0,0173	0,0059
Contactos con su Ejecutivo	-0,0004	0,0005	0,0033	-0,0012	0,0001
Edad	0,0303	0,0053	-0,0015	0,1988	-0,1188
Tiene Mora	-0,0382	-0,1039	-0,0930	<b>-0,6973</b>	<b>-0,5958</b>
Tiene Crédito Hipotecario	<b>-0,9879</b>	0,0857	0,1037	-0,0012	0,0041
Tiene Crédito de Consumo	-0,0730	<b>-0,9843</b>	0,1162	0,0660	0,0879
Tiene Inversiones	0,0992	0,0942	<b>0,9673</b>	-0,0089	-0,1289
Act. Lab. Empresario	-0,0005	-0,0017	-0,0007	0,0039	-0,0025
Act. Lab. Gerente	0,0062	0,0083	0,0230	0,0112	0,0122
Act. Lab. Obrero	-0,0079	-0,0065	-0,0112	-0,0347	-0,0306
Act. Lab. Técnico	0,0015	-0,0058	-0,0149	-0,0257	-0,0031
Act. Lab. Trab. Dependiente	-0,0058	0,0048	0,0005	0,0388	0,0336
Act. Lab. Trab. Independiente	0,0014	0,0003	-0,0013	-0,0001	-0,0017
Est. Civ. Casado	-0,0082	-0,0251	-0,0526	<b>0,4354</b>	<b>-0,5609</b>
Est. Civ. Soltero	0,0087	0,0298	0,0525	<b>-0,3919</b>	<b>0,5176</b>
Niv. Ed. Ens. Media	-0,0140	-0,0247	-0,0397	-0,0929	-0,0617
Niv. Ed. Tec. Profesional	0,0067	-0,0077	-0,0265	-0,0478	-0,0070
Niv. Ed. Universitario	0,0039	0,0349	0,0823	0,1418	0,1055
Med. Cont. Pref. Celular	0,0479	-0,0160	0,0899	-0,2341	-0,0525
Med. Cont. Pref. Emailing	-0,0608	0,0148	-0,0875	0,1922	0,0636

Basándonos en estos resultados, podemos inferir lo siguiente:

- El **Componente principal 1** muestra una fuerte correlación negativa con los clientes que tienen contratado un crédito hipotecario.
- El **Componente principal 2** muestra una fuerte correlación negativa con los clientes que tienen contratado un crédito de consumo y también, de forma menos robusta, con quienes tienen mora.
- El **Componente principal 3** muestra una fuerte correlación positiva con aquellos clientes que tienen inversiones y, en menor medida, con aquellos que tienen algún tipo de crédito.
- El **Componente principal 4** muestra una correlación negativa menos fuerte con los clientes que tienen mora y son solteros; esta última está en consonancia con la correlación positiva menos fuerte que tiene con los clientes casados.
- El **Componente principal 5** muestra una correlación negativa menos fuerte con los clientes que tienen mora y son casados, y, de manera inversa al Componente principal 4, una correlación positiva menos fuerte con los clientes solteros.

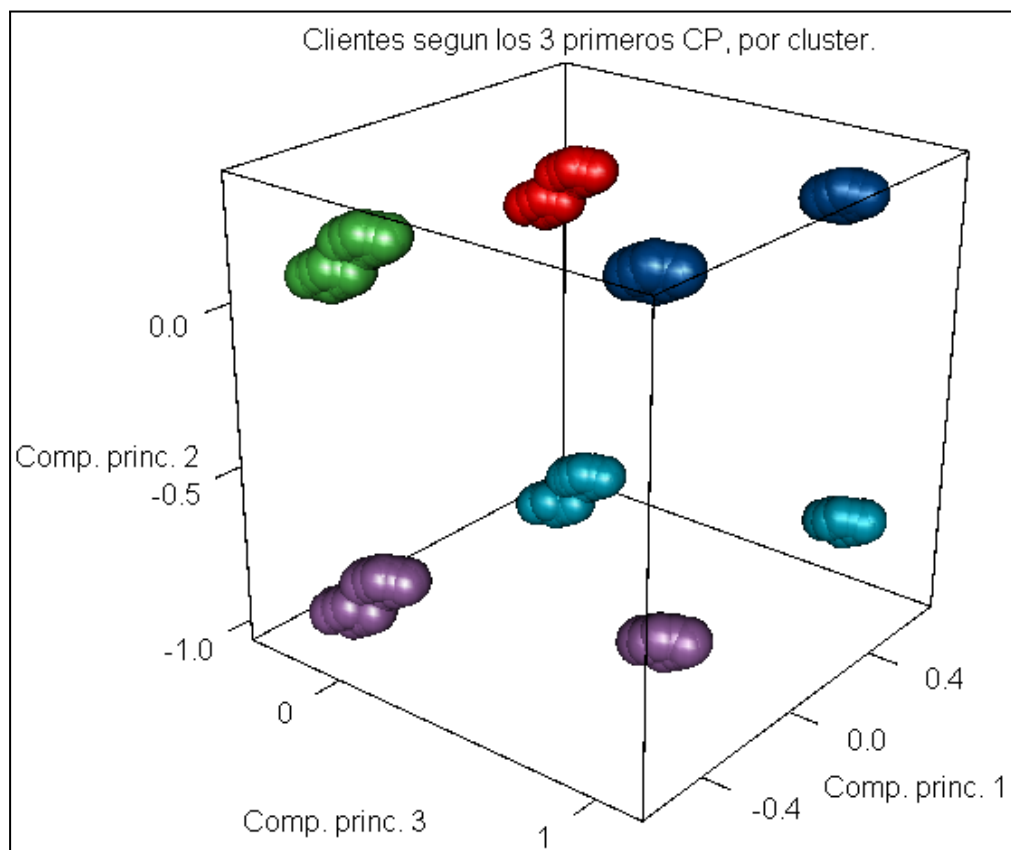
Teniendo esta información se pueden interpretar mejor los 2 gráficos siguientes:



**Gráfico 4:** Distribución de los clientes por *cluster*, según los 2 primeros componentes principales.

El *gráfico 4* muestra los registros de los clientes distribuidos en una gráfica de puntos usando como ejes los 2 primeros componentes principales: el *Componente principal 1*, que entrega el 38,97% de la varianza en el eje horizontal y el *Componente principal 2*, que entrega el 21,10% de la varianza, en el eje vertical. Se observa una clara distribución de los clientes en los 4 cuadrantes que se forman, con lo que podemos esperar que las variables ***Tiene Crédito Hipotecario*** y ***Tiene Crédito de Consumo*** serán relevantes en el resultado de la segmentación. Al asignar un color en función al segmento al cual pertenece, sorprende la concordancia entre estos y los cuadrantes: los clientes del ***cluster 2*** no tendrían contratado crédito hipotecario, y una proporción importante de estos, tampoco tendrían crédito de consumo; los clientes del ***clúster 3*** tendrían contratado crédito hipotecario, pero no de

consumo; los clientes del **clúster 4** tendrían contratado sólo crédito de consumo; los clientes del **clúster 5** tendrían contratados ambos tipos de crédito; finalmente, los clientes del **clúster 1**, si bien no tienen crédito de consumo, estarían repartidos en cuanto a la contratación del crédito hipotecario. En el *gráfico 5* se observa la distribución de los registros de clientes de manera tridimensional, donde el eje de profundidad corresponde al *Componente principal 3*, sumando un 10,83% de la varianza; específicamente, la información sobre la adquisición de inversiones. Ahora es claro que los clientes del **clúster 1** se diferencian por tener inversiones, así como una parte de los clientes de los **clusters 4 y 5**.



**Gráfico 5:** Distribución de los clientes por *cluster*, según los 3 primeros componentes principales (CP). Los colores de los *clusters* son los mismos del gráfico 4.

## Estrategia comercial.

### Descripción y alcance de la estrategia.

La segmentación de clientes bancarios realizada con el modelo ***K-means*** nos permite enfocar y direccionar esfuerzos para desarrollar e implementar razonablemente una estrategia comercial consistente en **aumentar el cruce de productos entre los clientes actuales del banco**, definiendo para este propósito, los productos y/o servicios a promocionar y los canales de contacto atingentes. La siguiente tabla muestra un resumen de cada segmento por el valor promedio de cada una de las variables:

Variables (valores promedio)	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Saldo.Medio.Anual (\$)	2.734.608	1.555.114	1.194.066	759.323	790.920
Contactos.con.su.Ejecutivo	1,26	0,40	0,61	0,33	0,64
Edad	42,25	42,43	40,57	41,81	40,38
Tiene_mora	0	0,02	0,05	0,10	0,11
Tiene_credito_hipotecario	0,32	0	1	0	1
Tiene_credito_de_consumo	0	0	0	1	1
Tiene_inversiones	1	0	0	0,04	0,03
AL_empresario	0,02	0,02	0,03	0,04	0,04
AL_gerente	0,27	0,14	0,11	0,09	0,08
AL_obrero	0	0,05	0,11	0,10	0,12
AL_tecnico	0,03	0,12	0,09	0,12	0,13
AL_trab_dep	0,58	0,59	0,63	0,59	0,60
AL_trab_ind	0,04	0,04	0,03	0,04	0,03
AL_jubilado	0,07	0,03	0	0,03	0,01
EC_divorciado	0,11	0,11	0,12	0,15	0,11
EC_casado	0,53	0,60	0,60	0,59	0,68
EC_soltero	0,36	0,29	0,29	0,27	0,20
NE_ens_basica	0,05	0,09	0,08	0,10	0,07
NE_ens_media	0,02	0,12	0,16	0,21	0,21
NE_tec_profe	0,07	0,16	0,11	0,14	0,14
NE_universit	0,87	0,63	0,65	0,55	0,58
MCP_fonopart	0,09	0,09	0,05	0,09	0,04
MCP_celular	0,82	0,73	0,62	0,76	0,66
MCP_emailing	0,09	0,18	0,33	0,15	0,29

En este sentido, a partir de los resultados obtenidos y del análisis realizado, es posible perfilar la siguiente segmentación de clientes:

- **Cientes Diamond:** Corresponde al **Cluster 1**, el cual contempla el 7,13% de la cartera, y agrupa a aquellos clientes de más alto valor para el banco. Poseen características como un patrimonio neto significativo, inversiones considerables, una proporción significativa tiene crédito hipotecario y sin historial de mora crediticia, lo que los hace financieramente estables. Mantienen un alto relacionamiento con su ejecutivo para tomar decisiones informadas sobre sus inversiones y/o productos. Suelen desempeñarse, principalmente, en roles ejecutivos en sus respectivas organizaciones y destacan por poseer un nivel educativo universitario.
- **Cientes Platinum:** Corresponde al **Cluster 2**, es decir, el 33,23% de la cartera, y comprende a aquellos clientes con un patrimonio neto sólido, con un muy bajo nivel de mora, que no invierten ni tienen créditos con la institución, por lo que pueden optar a esquemas de financiamiento con bajo riesgo crediticio para el banco. Además, poseen un nivel educativo mayoritariamente universitario.
- **Cientes Gold:** Corresponde al **Cluster 3**, contemplando el 43,63% de la cartera, y agrupa a los clientes que no alcanzan el estatus de **Diamond** o **Platinum**, pero que aún tienen un

patrimonio neto respetable. Utilizan los servicios bancarios para la contratación de créditos hipotecarios, pero no para la adquisición de productos de inversión. Presentan un muy bajo nivel de mora y poseen un nivel educativo mayoritariamente universitario.

- **Clientes Silver:** Corresponde al **Cluster 4**, compuesto por el 6,36% de la cartera, comprende a los clientes que tienen una situación financiera y un patrimonio neto bajo. Utilizan los servicios bancarios para la contratación de créditos de consumo y, escasamente, productos de inversión. Tienen un bajo nivel de mora y poseen un nivel educativo mayoritariamente universitario.
- **Clientes Bronze:** Corresponde al **Cluster 5**, conformado por el 9,65% de la cartera, y comprende a los clientes que tienen una situación financiera básica con un patrimonio neto bajo. Utilizan los servicios bancarios para la contratación de créditos hipotecarios y de consumo, así como escasamente productos de inversión. Tienen un bajo nivel de mora y poseen un nivel educativo mayoritariamente universitario.

De acuerdo con los resultados precedentes, el **76,86%** de la cartera del banco se concentra en los segmentos de clientes **Platinum** y **Gold**, por lo cual, la estrategia comercial a desarrollar e implementar se enfocará en estos clientes.

En este contexto, para respaldar la estrategia comercial propuesta, **se fomentará la incursión de estos clientes en la contratación de productos financieros de inversión y créditos de consumo**, asegurando el cumplimiento responsable de las obligaciones asociadas a los productos previamente contratados con la institución bancaria, por ejemplo, créditos hipotecarios. Entre los productos y/o servicios a ofrecer para la implementación efectiva de la estrategia, se encuentran:

- Inversiones en acciones, bonos, y fondos mutuos, ofreciendo asesoría financiera personalizada y gratuita.
- Créditos de consumo a tasas preferenciales con la posibilidad de extender el beneficio si el cliente adquiere otros productos o servicios.
- Cuentas de ahorro y depósito a plazo a tasas de interés competitivas.
- Programa de membresía que permite acceso gratuito a las salas de espera *VIP* en aeropuertos nacionales e internacionales.
- Tarjetas de crédito con recompensas atractivas, como descuentos en comercios, millas aéreas u ofertas en restaurantes.
- Seminarios y talleres gratuitos de educación financiera.

**La estrategia comercial descrita busca maximizar la retención de clientes de alto valor para el banco** que se encuentran en los segmentos **Platinum** y **Gold**, que como se indicó, representan el **76,86%** de la cartera, **a través de su participación en la contratación de productos y servicios financieros que les resulten atractivos y con beneficios exclusivos**, al tiempo en que se fortalece la relación a largo plazo. Específicamente, esta estrategia comercial propende a:

- Fomentar la inversión en productos financieros de manera informada, lo cual, no solo aumenta los activos bajo gestión del banco, sino que también fortalece la relación con el cliente al brindarle un servicio personalizado.
- Promover el ahorro, lo que a su vez contribuye a la rentabilidad financiera del banco.
- Incentivar el endeudamiento responsable de estos clientes, lo que a su vez implica una adecuada gestión del riesgo crediticio para el banco.

- Mejorar la experiencia del cliente, buscando con ello fortalecer su lealtad hacia el banco.
- Demostrar el compromiso del banco con el bienestar financiero de sus clientes, al tiempo que les proporciona herramientas para tomar decisiones financieras informadas.

Para llevar a cabo la estrategia comercial descrita, se implementará una campaña que combina la comunicación por los medios digitales oficiales (correo electrónico, aplicación móvil, portal *online*), que es efectiva para informar y generar interés, con llamadas telefónicas personalizadas para cerrar acuerdos y proporcionar asesoramiento *ad hoc*. Los detalles de cada uno de estos canales de comunicación se describen a continuación:

- a. En relación con el canal de contacto **Correo Electrónico**, el 26% en promedio de los clientes perfilados en los segmentos **Platinum** y **Gold**, ha manifestado su preferencia a ser contactados por esta vía, por lo que:
  - i. Se enviará un correo de presentación a todos los clientes, explicando de manera práctica y visual, los beneficios de los productos y/o servicios.
  - ii. Se enviará una serie de correos electrónicos informativos y personalizados a lo largo de 6 semanas. Estos correos tratarán sobre temas como inversiones, ahorro, ventajas de créditos de consumo preferenciales y recompensas de tarjetas de crédito, e incluirá llamados a la acción que motiven a los clientes a agendar una cita con un asesor financiero del banco.
  - iii. Se complementará con mensajes informativos en ventanas emergentes (o *pop-up*) en los distintos medios digitales que posee el banco, como son la aplicación móvil y el portal *online* de cada cliente, generando una comunicación masiva evitando así el efecto *spam* de los correos electrónicos.
- b. En relación con el canal de contacto **Llamadas Telefónicas**, el 67% en promedio de los clientes perfilados en los segmentos **Platinum** y **Gold**, ha manifestado su preferencia a ser contactados por esta vía, por lo que:
  - i. Después de que los clientes hayan recibido los correos anteriormente descritos o notificaciones masivas, se realizarán llamadas telefónicas a aquellos que han mostrado interés o han interactuado con los mensajes informativos. Durante estas llamadas, los ejecutivos ofrecerán asesoramiento personalizado y responderán las preguntas derivadas por los clientes.
  - ii. Durante las llamadas, se abordarán las ofertas diseñadas para estos clientes, en relación con la consulta financiera gratuita, las tasas preferenciales en cuentas de ahorro o créditos de consumo, o los beneficios de unirse al programa de membresía con acceso a salas *VIP* en aeropuertos.

En ambos casos, se realizará un seguimiento de la efectividad de la campaña, mediante **tasas de respuesta a los correos electrónicos, tasas de interacción en los medios digitales (aplicación móvil y portal online), tasas de conversión de llamadas, y el número de productos o servicios contratados como resultado de la campaña**. Cabe destacar que valores **mayores al 50% en estos indicadores permitirá corroborar la asertividad de la campaña**. Asimismo, se recopilarán comentarios de los clientes a través de encuestas o entrevistas telefónicas cortas, permitiendo mejorar los resultados e implementar la retroalimentación obtenida, generando acciones para ajustar la estrategia comercial y las ofertas de productos y/o servicios para futuras campañas.

Ahora bien, para evaluar el éxito de la estrategia se establecieron ciertos objetivos cuantificables en relación a los productos ofrecidos para los clientes perfilados en los segmentos **Platinum** y **Gold**, los que son:

1. ***Aumento del 5% en clientes del segmento Gold con crédito hipotecario.***
2. ***Aumento del 20% en clientes ingresados al sistema de membresías del banco, tanto para clientes Platinum como clientes Gold.***
3. ***Aumento del 10% en inversiones realizadas en productos financieros del segmento Gold.***
4. ***Aumento del 15% en cuentas de ahorro o fondos mutuos, dirigidas a clientes Platinum.***

Estos valores fueron establecidos en relación al tiempo de campaña anteriormente mencionado (6 semanas). En caso de no lograr los objetivos anteriormente definidos, existe la posibilidad de una extensión del periodo por 2 semanas más.

Recursos necesarios para la implementación.

En todo proceso comercial se debe trabajar bajo un presupuesto previamente aceptado por la gerencia. En este caso, no existe un presupuesto establecido, sin embargo, no se incurrirá en gastos excesivos o injustificados, con el objetivo principal de **maximizar la retención de clientes de alto valor para el banco**. Estos gastos se pueden dividir en 2 categorías: **materiales y profesionales**, ambos descritos a continuación:

- ***Gastos materiales:*** la campaña implica enviar material informativo con el fin de despertar el interés de los clientes. Para lograr este objetivo, es clave el desarrollar un diseño específico que cumpla con generar cierto impacto en los receptores. Tanto los correos electrónicos, notificaciones en la aplicación móvil y mensajes *pop-up* en el portal *online* son necesarios en todo este proceso.
- ***Gastos profesionales:*** para implementar de manera correcta esta campaña será necesario trabajar con equipo multidisciplinario compuesto por: **Diseñadores UX, Analistas BI, Programadores, Product Owners y Ejecutivos de cuentas**, entre otros. Todos cumplirán un papel relevante al momento de fomentar la contratación de productos y servicios financieros, junto con beneficios exclusivos, y en la generación de la máxima fidelización y retención de clientes. En el caso de no estar contratados en la empresa, será necesario gestionar su reclutamiento y proceso de selección, para asegurarse de contar con las personas con perfiles adecuados.

Cabe recalcar que no se posee información sobre los cargos preexistentes por lo que, se plantea una propuesta de personal a contratar:

- **1 Product Owner.**
- **1 Diseñador UX.**
- **2 Programadores.**
- **2 Analistas BI.**
- **1 Supervisor o Ejecutivo Senior.**
- **Ejecutivos del banco (Call Center).**

Finalmente, de ser necesario, se realizarán capacitaciones a los ejecutivos que formarán el *Call Center* para que posean el conocimiento necesario y la habilidades de comunicación indispensables al momento de buscar generar interés en un cliente.

## Implementación.

Una vez establecidos los recursos necesarios, se describirá el paso a paso para llevar a cabo una correcta implementación de la estrategia. A continuación se muestra un resumen de **Carta Gantt**, que ayuda a comprender el desarrollo del plan (la carta en extenso está disponible en los anexos):

ETAPA	S E M O 1	S E M O 2	S E M O 3	S E M O 4	S E M O 5	S E M O 6	S E M O 7	S E M O 8	S E M O 9	S E M O 10	S E M O 11	S E M O 12	S E M O 13	S E M O 14	S E M O 15	S E M O 16	S E M O 17	S E M O 18	S E M O 19	S E M O 20	S E M O 21	S E M O 22	S E M O 23	S E M O 24	
ETAPA 1: Planificación de campaña, estimación de costos y tiempo																									
ETAPA 2: Equipo de trabajo, material, coordinación de áreas e implementación																									
ETAPA 3: Análisis de resultados y valorización del impacto																									

También, se describirán las etapa del plan:

**Etap 1, Planificación de campaña, estimación de costos y tiempo:** en esta etapa se construirá la estrategia, se establecerán los objetivos y las acciones a llevar a cabo. Se realizará una estimación de los aportes por áreas y, finalmente, se presentará el proyecto a la gerencia con el fin de evaluar el impacto positivo que generará la propuesta al banco, junto con el costo que implica desarrollarla. Una vez recibido el visto bueno de la gerencia, se revisará nuevamente la estimación de costos, inversiones y gastos que se deberán llevar a cabo para implementar la campaña y los tiempos asociados a cada etapa del plan.

**Etap 2, Equipo de trabajo, material, coordinación de áreas e implementación:** una vez conseguido el consentimiento de los gerentes y líderes de las áreas relacionadas, se debe proceder a la publicación, selección del personal, entrevista y capacitación de los empleados nuevos que formarán parte del equipo (este paso se realizará solo en el escenario donde el banco no posee los profesionales necesarios). En el caso de que los profesionales ya sean parte de la institución, se procederá al diseño y creación del material relacionado con la campaña y a la implementación de los *pop-up* y/o notificaciones a incorporar en la aplicación móvil y portal online de los clientes, lo cual será realizado mediante *Agile Methodology (Sprint)* por los programadores y diseñadores, quienes serán liderados por el *Product Owner/Product Manager*.

A su vez, se generarán los equipos de *Call Center* quienes serán guiados por un *Ejecutivo Senior* en lo concerniente al relacionamiento con los clientes y de la concreción de las solicitudes de productos. Los ejecutivos de cuenta de cada cliente estarán en comunicación



directa para evaluar la situación comercial y crediticia de cada uno, además de hacer el envío de la documentación necesaria para obtener el producto en cuestión.

**Etapas 3, Análisis de resultados y valorización del impacto:** una vez implementada la campaña, es relevante ir midiendo los resultados. Para esto, un equipo de *analistas BI* se dedicará semanalmente a evaluar el impacto con base al seguimiento de las métricas previamente mencionadas (*tasas de respuesta a los correos electrónicos, tasas de interacción en los medios digitales, tasas de conversión de llamadas y el número de productos o servicios contratados como resultado de la campaña*). Estos resultados serán analizados y representados en alguna herramienta de tablero de reportes (*dashboards* en *Power BI*, por ejemplo), lo que permitirá entregar de forma oportuna informes de resultados a la gerencia. Por consiguiente, la campaña se implementará por un plazo de 6 semanas, teniendo presente que, de no lograr los objetivos establecidos, se extenderá por 2 semanas más, obteniendo un plazo final de estudio de 8 semanas. Una vez concluido este período, se procederá a evaluar su impacto y generar la comparativa con los costos incurridos, lo que entregará el verdadero valor de la estrategia comercial para el banco.

## Palabras finales...

La estrategia comercial delineada en este informe representa una respuesta orientada a los datos para maximizar el valor de la cartera de clientes del banco. A través de una segmentación precisa y un enfoque cuidadosamente elaborado en los segmentos **Platinum** y **Gold**, se ha logrado una estrategia que no solo fortalece la retención de los clientes de alto valor, sino que también incrementa su participación en una variedad de productos y servicios financieros. La combinación de canales de contacto, la inversión en productos financieros, el ahorro, los créditos de consumo, y una experiencia enriquecida, respaldados por el análisis de datos, conforman una estrategia integral y centrada en el cliente. Este enfoque no solo fortalece la posición competitiva del banco, sino que también cimienta relaciones sólidas y a largo plazo con sus clientes de mayor valor, allanando el camino hacia un crecimiento sostenible y una rentabilidad duradera.

# Anexos.

## Carta Gantt

ETAPA	S E M O 1	S E M O 2	S E M O 3	S E M O 4	S E M O 5	S E M O 6	S E M O 7	S E M O 8	S E M O 9	S E M O 10	S E M O 11	S E M O 12	S E M O 13	S E M O 14	S E M O 15	S E M O 16	S E M O 17	S E M O 18	S E M O 19	S E M O 20	S E M O 21	S E M O 22	S E M O 23	
ETAPA 1: Planificación de campaña, estimación de costos y tiempo																								
Tarea 1. Construcción de la estrategia																								
Establecer el objetivo																								
Definir las acciones necesarias																								
Redactar el plan y primera presentación																								
Establecer métricas de seguimiento																								
Definición de objetivos cuantitativos																								
Tarea 2. Planificación de las etapas																								
Evaluar costos																								
Medir aporte por área																								
Planificación de personal																								
Evaluar proceso de selección																								
Tarea 3. Generación de presentación final																								
Corrección de errores																								
Estimación del impacto a lograr																								
Creación de presentación																								
Creación de Carta Gantt																								
Creación de presupuesto final																								
Tarea 4.Aceptación y revisión																								
ETAPA 2: Equipo de trabajo, material, coordinación de áreas e implementación																								
Tarea1. Organización de los equipos de trabajo (considerando que todos los roles ya pertenecen al banco)																								
Formar equipo de Call Center																								
Asignación de listado de clientes por ejecutivo Call Center																								
Formar equipo de Tecnología/Producto																								
Reasignación de tareas de programadores a otros profesionales																								
Tarea 2. Definición de Sprint																								
Diseño de material																								
Presentación de diseños																								
Corrección de errores																								

