Taller individual 1

Segmentación de clientes en un banco a partir del análisis de clúster

Leonardo Humberto Garcia Diaza,c,

Ana María Beltrán Cortés, MScb,c

aEstudiante de Maestría en Analítica para la Inteligencia de Negocios

bProfesor, Departamento de Ingeniería Industrial

cPontificia Universidad Javeriana, Bogotá, Colombia

1. **ENTENDIMIENTO DEL NEGOCIO**
   1. **Descripción de la situación:** El Banco ABC es un banco colombiano. La compañía brinda servicios financieros a personas y empresas, es una de las 10 entidades de su tipo más grandes en crédito de consumo en el país. La compañía ha identificado la necesidad de alcanzar un entendimiento apropiado de sus clientes aprovechando la información existente y está interesada en desarrollar promociones para grupos diferentes de clientes.
   2. **Objetivos del negocio:** Como se ha mencionado el banco quiere desarrollar promociones para grupos diferentes, y un método eficiente para conseguirlo es el descubrimiento de los posibles patrones de consumo en los tarjetahabientes mediante el análisis de clúster, lo cual sería de interés para mejorar los siguientes indicadores de rendimiento:

* *Indicador de satisfacción del cliente*, lograr un mejor entendimiento de su perfil y desarrollar estrategias que permitan lograr mayor lealtad. Mejorar en un 5% el indicador de retención.
* *Indicador de crecimiento en los ingresos*, aumentar las ganancias en los productos financieros Visa y MasterCard de los tarjetahabientes actuales. Alcanzar al menos un 10% de crecimiento en el consumo nacional de las franquicias en pesos.
* *Indicador de desempeño comercial*, idear campañas específicas para aquellos clientes con mejores probabilidades de consumo durante los fines de semana. Aumentar la rentabilidad de las campañas en un 10%.
  1. **Criterios de éxito del negocio:** Según la expectativa del banco se tiene éxito si estas acciones aportan a su éxito empresarial alcanzando los objetivos propuestos así:
* Conocer y comprender al cliente desde un nuevo enfoque, mejorando el porcentaje de satisfacción y la calidad de los servicios ofrecidos.
* Realizar proyecciones desde lo financiero y comercial con un alto porcentaje de asertividad.
  1. **Objetivos de minería de datos:**
* Desplegar un modelo de análisis para la segmentación de clientes a partir de sus ocasiones de consumo (promedio por transacción), por franquicia (Visa y MasterCard) y por días (fines de semana).
* Elaborar un análisis de los segmentos obtenidos, para tener una interpretación y que se pueda usar esta información para idear las campañas específicas.
  1. **Criterios de éxito de la minería de datos:** Según la expectativa del banco se tiene éxito cuando el modelo permita la segmentación de clientes con un alto porcentaje de fiabilidad a partir de los patrones de consumo en las tarjetas de crédito.

1. **ENTENDIMIENTO DE LOS DATOS**
   1. **Recolección de datos:** El Banco ABC ha puesto a nuestra disposición el archivo ***infoclientebanca.xlsx*** el cual contine contiene información de consumo (transacciones) de los tarjetahabientes en el último mes, cada cliente puede tener tarjetas de tres diferentes franquicias (Visa, MasterCard y una tercera que el banco maneja).
   2. **Descripción de datos:** Lo primero que hacemos es mirar al conjunto de datos como se muestra en la Figura 1.

Texto, Carta

Descripción generada automáticamente

*Figura 1.*

Tabla

Descripción generada automáticamente

*Figura 2.*

Encontramos que la base de datos contiene 47.871 con 26 variables, y que las variables son de *consumo* y *porcentaje*, salvo por la variable cliente que corresponde a un *valor único* para cada observación y la variable sitio consumo más frecuente que corresponde a una variable de tipo *caracter* de acuerdo con la Figura 2.

Además, para el conjunto de datos es conveniente mantener su tipo de variable, teniendo en cuenta el nombre que tienen; exceptuando cliente que lo convertimos a *caracter*, grupo de cliente y sitio consumo más frecuente que quizás sean variables *categóricas* de interés y número de transacciones que lo convertimos a *entero*.

* 1. **Exploración de los datos:** Para realizar una exploración de los datos es necesario separar las variables de consumo, porcentaje y categóricas. Para los objetivos de este análisis nos centraremos en las variables que se muestran en la Figura 3, ya que nos permitirá enfocar la segmentación de acuerdo con los criterios del banco.

Un conjunto de letras blancas en un fondo blanco

Descripción generada automáticamente con confianza media

*Figura 3.*

Diagrama

Descripción generada automáticamente

*Figura 4.*

En la Figura 3, se observa que el mínimo para las variables de consumo (excepto número de transacciones y promedio por transacción) es cero. En los histogramas de la Figura 4 se observa asimetría positiva en las variables (excepto porcentaje MaterCard nacional) debido a las diferencias entre la media y la mediana, también cuentan con sesgo a la derecha a excepción de la variable de consumo MasterCard nacional.

Agregando a lo anterior, es importante mencionar que todas las variables cuentan con CV’s bastante altos, lo cual indica valores dispersos. Por lo tanto, para disminuir la presencia de valores atípicos podríamos aplicar un logaritmo natural a las variables y así evitar afectaciones en el algoritmo de segmentación.

1. **PREPARACIÓN DE LOS DATOS**

De acuerdo con los objetivos mencionados anteriormente, y las características del conjunto de datos, en la preparación de datos se siguieron los pasos explicados a continuación:

1. Se ha asignado un valor promedio a las variables que tienen porcentaje multiplicando dichas variables por el promedio por transacción.
2. Luego teniendo en cuenta que nuestro interés está en los fines de semana filtramos el Dataset obtenido y descartamos los valores iguales a 0 para viernes, sábado y domingo.
3. Descartamos la variable promedio por transacción ya que se encuentra distribuida en los porcentajes.
4. Aplicación de reducción de atípicos (logaritmo natural de (1+x) para tener en cuenta los ceros)
5. Estandarización de las variables para poder usar distancia euclídea

Tabla

Descripción generada automáticamente con confianza media

*Figura 5.*

En este apartado se expuso la preparación de los datos, ahora revisamos de nuevo nuestro conjunto de datos y observamos en la Figura 5 que la transformación fue en su mayor parte efectiva, sin embargo, vemos que en unas variables tuvo mejores resultados que en otras (promedio visa nacional mejor que en número de transacciones).

1. **MODELACIÓN**

Para nuestro modelo de segmentación utilizaremos el algoritmo k – medias, y debemos definir el valor de k, para ello usaremos el gráfico de codo, que consiste en las sumas de cuadrados dentro de los clústeres frente al número de clústeres. Al mirar un diagrama de codo, se desea ver una fuerte disminución de una k a otra seguida de una disminución más gradual en la pendiente. En el siguiente grafico vamos a usar entre 2 y 10 clústeres.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

1. **EVALUACIÓN**

**BIBLIOGRAFÍA**

Editorial La República S.A.S. (2021, 23 agosto). *Utilidades de bancos crecieron $1 billón en junio con un resultado de $4,9 billones.* Diario La República. <https://www.larepublica.co/finanzas/utilidades-de-los-bancos-crecieron-1-billon-en-junio-con-un-resultado-de-49-billones-3220226#:%7E:text=La%20Superintendencia%20Financiera%20de%20Colombia,dato%20fue%20%243%2C9%20billones>.

Editorial La República S.A.S. (2021b, octubre 15). *El último año aumentó 25% integración de nuevos usuarios a servicios financieros.* Diario La República. <https://www.larepublica.co/finanzas/en-el-ultimo-ano-aumento-25-la-integracion-de-nuevos-usuarios-a-servicios-financieros-3248350>

Grant, R. M. (2010). *Contemporary Strategy Analysis* (7.a ed.). John Wiley & Sons Inc.

*Guía de CRISP-DM de IBM SPSS Modeler.* (2019). IBM Corporation. <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=SS3RA7_sub/modeler_crispdm_ddita/modeler_crispdm_ddita-gentopic1.html>

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition.* Springer.

Kotler, P., & Lane, K. (2016). *Marketing Management* (15.a ed.). PEARSON.

Marr, B. (2021, 13 julio). *25 Crucial KPIs.* Bernard Marr. <https://bernardmarr.com/25-crucial-kpis/>

Marr, B. (2022, 3 febrero). *What is a KPI?* Bernard Marr. <https://bernardmarr.com/what-is-a-kpi/?paged1077=3>

Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2019). *Data Mining for Business Analytics: Concepts, Techniques and Applications in Python.* Wiley.

Tsiptsis, K. K., & Chorianopoulos, A. (2011). *Data Mining Techniques in CRM.* Wiley.

1. **DESPLIEGUE (RECOMENDACIONES DE NEGOCIO)**