

IDENTIFICACIÓN DE COMERCIOS PARA PLAN DE MANTENIMIENTO INTELIGENTE EN RED DE DATÁFONOS

• RESUMEN

En el presente proyecto se pretende encontrar perfiles de clientes (comercios) de una empresa de medios de pago colombiana que administra red de datáfonos, para inclusión dentro del plan de mantenimiento de la red. Las variables disponibles para el análisis proporcionan información valiosa sobre el comportamiento de los comercios, como el número de terminales vigentes, quejas y reclamos, transacciones realizadas, tasa de transacciones reversadas, tasas de transacciones negadas, entre otros. Mediante la aplicación de métodos de aprendizaje no supervisado como: Clustering DBSCAN y Jerárquico, se espera depurar, normalizar y procesar la data, en busca de obtener segmentos de comercios bien definidos. Esto ayudaría al negocio a enfocar mejor su plan de mantenimiento de la red, optimizar los recursos y garantizar la operación eficiente de los terminales de punto de venta (Datáfonos), mejorar la calidad de servicio y maximizar sus ingresos, beneficiando tanto al negocio como a sus clientes (comercios).

• INTRODUCCIÓN

En la gestión de una red de datáfonos, se presenta un desafío: la identificación de comercios para inclusión en un plan de mantenimiento inteligente. Este plan busca optimizar los recursos y garantizar la operación eficiente de los Datáfonos, beneficiando tanto al negocio como a sus clientes (comercios). Para abordar este problema, se recurre al aprendizaje no supervisado, mediante la aplicación de técnicas de Clustering, que permitirán identificar grupos de comercios con características similares en función de las variables proporcionadas. Al analizar estos grupos, se podrán identificar patrones problemáticos y definir criterios para la inclusión de comercios en el plan de mantenimiento.

El cliente potencial es la empresa encargada de la gestión de la red de datáfonos, cuyo objetivo es mejorar la calidad de servicio y maximizar los ingresos. El problema radica en identificar comercios con situaciones problemáticas para inclusión dentro de su plan de mantenimiento. Las variables disponibles para el análisis proporcionan información valiosa sobre el comportamiento de los comercios, como el número de terminales vigentes, quejas y reclamos, transacciones realizadas, tasa de transacciones reversadas, tasas de transacciones negadas, entre otros.

Bajo el contexto del negocio, es importante considerar lo siguiente:

Comportamiento de Quejas y Reclamos: Lo deseable para el negocio es que la "TASA_QR" de sus comercios tienda a cero.

Comportamiento Transaccional: Lo deseable es que DIFERENCIA_TRX sea positiva, lo que indica que el comercio realizó más trx el último mes respecto al anterior. 'TRX_ULT_MES' indicará el potencial actual transaccional del comercio.

Número y Tasa de Reversos: Un Reverso o trx reversada se refiere a una transacción previamente autorizada que se revierte, lo que significa que los fondos que se habían transferido se devuelven al titular de la cuenta o tarjeta. Esto sucede por varias razones como errores de procesamiento o de comunicación, entre otros. Aunque la tasa de reversos es muy

baja, puede generar molestias para comercios y/o tarjetahabientes ya que en algunos casos la devolución de los fondos no se refleja automáticamente. Lo deseable para el negocio es que la "TASA_REVERSOS_ULT_MES" de sus comercios tienda a cero.

Tasa de Negaciones: Una trx negada es una trx que es rechazada o denegada por la entidad o sistema Autorizador. El negocio busca minimizar su tasa de negaciones. Lo cual además de mejorar el servicio hacia sus clientes, favorece los ingresos del negocio quien recibe de los Bancos Emisores un pago mayor por las trx aprobadas respecto a negadas. Lo deseable para el negocio es que la tasa de negaciones de comercios tienda a cero.

Capacidad Operativa: Cada 'SECCIONAL_OPERACIONES' fijará una capacidad mensual del número de terminales a gestionar, dado el personal técnico disponible. Una vez seleccionados los comercios idóneos para inclusión en el Plan de Mantenimiento, cada Seccional validará aquellos que le corresponden para su gestión según capacidad.

Se espera que estas consideraciones contribuyan a lograr mayor interpretabilidad y constituyan criterios de decisión para la selección de los clústers de comercios a gestionar.

• REVISIÓN PRELIMINAR DE ANTECEDENTES EN LA LITERATURA

En la literatura investigada, se encuentran técnicas similares y enfocadas en el tipo de aprendizaje en cuestión que puede orientar y ver desde diferentes enfoques la posible solución al problema:

En el trabajo de suficiencia profesional titulado "Segmentación de clientes para mejorar la experiencia de compra de productos electrónicos en Falabella" (Aragón Gallegos, A. D. C., Cerquin Silva, S. I., Ecurra Yactayo, R. O., & Roncalla Viena, A. L., 2023), se busca a través de la aplicación de métodos de aprendizaje no supervisado como clustering jerárquico, K-medoids y K-Means; encontrar perfiles de consumidores y permitir potenciar las estrategias de marketing, con esto aumentar sus ventas y las preferencias.

Se puede identificar, que, a pesar del público objetivo diferente, también se aplican técnicas de aprendizaje no supervisado, específicamente en el contexto de clustering o agrupamiento de datos para identificar patrones sin necesidad de etiquetas predefinidas. Se tomará como guía en caso de requerir aplicar alguno de los tres algoritmos presentados en la tesis.

Por otro lado, en el artículo "Unsupervised Clustering Approach for Network Anomaly Detection" (Syarif, I., Prugel-Bennett, A., & Wills, G., 2012), se compara el enfoque de detección de intrusiones en redes de detección de anomalías con el enfoque de detección de uso indebido. Se analizan cinco algoritmos de agrupación (k-Means, improved k-Means, k-Medoids, EM clustering y distance-based outlier) en el contexto de la detección de anomalías. Los resultados indican que la detección de uso indebido falla en detectar intrusiones desconocidas, mientras que la detección de anomalías tiene un mejor rendimiento, especialmente el algoritmo de detección de valores atípicos basado en distancias. Este artículo proporcionará apoyo para la aplicación y validación del modelo seleccionado, evaluando otras opciones no estudiadas hasta el momento.

• DESCRIPCIÓN DETALLADA DE LOS DATOS

Fuente: Empresa de medios de pago colombiana. No se suministra el detalle de los datos, dada confidencialidad de los mismos.

```
# Obtener el tamaño del DataFrame
num_filas, num_columnas = df.shape
print(f'Número de filas: {num_filas}')
print(f'Número de columnas: {num_columnas}')
```

Número de filas: 135954
Número de columnas: 13

```
df.dtypes

ID                int64
SECCIONAL_OPERACIONES  object
TERMINALES_VIGENTES  int64
QR_ULT_MES         int64
TRX_MES_ANT        int64
TRX_ULT_MES        int64
REVERSOS_ULT_MES   int64
TASA_REVERSOS_ULT_MES float64
TASA_APROBACIONES_ULT_MES float64
TASA_NEGACIONES_ULT_MES float64
DIFERENCIA_TRX     int64
VARIACION_TRX      float64
TASA_QR            float64
dtype: object
```

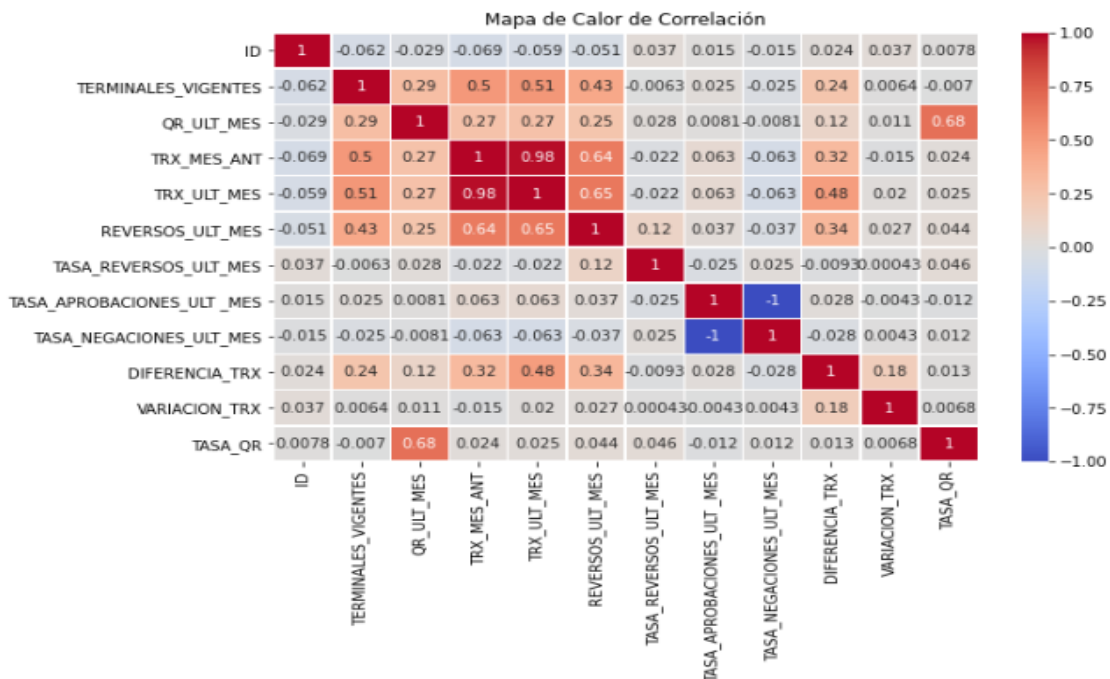
Descripción de las variables disponibles para análisis, según el contexto del negocio:

- **ID:** Código único o identificador. Id del punto físico o establecimiento comercial.
- **SECCIONAL_OPERACIONES:** Seccional a la que pertenece el comercio para su gestión operativa, según su ubicación.
- **TERMINALES_VIGENTES:** Terminales con las que cuenta el comercio (Datáfonos).
- **QR_ULT_MES:** Quejas o Reclamos radicados por el comercio en el último mes.
- **TRX_ULT_MES:** Transacciones realizadas por el comercio durante el último mes.
- **TRX_MES_ANT:** Transacciones realizadas por el comercio en el mes anterior.
- **REVERSOS_ULT_MES:** Transacciones del comercio reversadas en el último mes.
- **TASA_REVERSOS_ULT_MES:** % transacciones reversadas respecto al total de transacciones realizadas el último mes: $\text{REVERSOS_ULT_MES} / \text{TRX_ULT_MES}$.
- **TASA_APROBACIONES_ULT_MES:** % Transacciones aprobadas respecto al total de transacciones realizadas por el comercio durante el último mes.
- **TASA_NEGACIONES_ULT_MES:** % Transacciones negadas respecto al total de transacciones realizadas por el comercio durante el último mes.
- **DIFERENCIA_TRX:** Diferencia entre las trx realizadas por el comercio en los últimos 2 meses, esto es: $\text{TRX_ULT_MES} - \text{TRX_MES_ANT}$.
- **VARIACION_TRX:** Variación (%) del número de transacciones realizadas en los últimos 2 meses, esto es: $(\text{TRX_ULT_MES} / \text{TRX_MES_ANT}) - 1$
- **TASA_QR:** Tasa de quejas y reclamos por comercio según su número de terminales. Esto es: $\text{QR_ULT_MES} / \text{TERMINALES_VIGENTES}$.

Estadísticos descriptivos para los comercios en análisis:

	count	mean	std	min	25%	50%	75%	max
TERMINALES_VIGENTES	83937.0	2.207942	4.625107	1.00000	1.000000	1.000000	2.000000	446.000000
QR_ULT_MES	83937.0	0.185794	0.700365	0.00000	0.000000	0.000000	0.000000	26.000000
TRX_ULT_MES	83937.0	532.274384	1548.823112	1.00000	33.000000	119.000000	419.000000	75805.000000
REVERSOS_ULT_MES	83937.0	0.600295	2.119145	0.00000	0.000000	0.000000	0.000000	143.000000
TASA_REVERSOS_ULT_MES	83937.0	0.002099	0.015377	0.00000	0.000000	0.000000	0.000000	1.000000
TASA_NEGACIONES_ULT_MES	83937.0	0.077882	0.093913	0.00000	0.033333	0.057640	0.090909	1.000000
DIFERENCIA_TRX	83937.0	51.217270	295.095901	-9758.00000	-7.000000	5.000000	36.000000	17696.000000
VARIACION_TRX	83937.0	0.526204	9.782262	-0.99894	-0.092593	0.038674	0.222222	1616.333333
TASA_QR	83937.0	0.088930	0.326632	0.00000	0.000000	0.000000	0.000000	9.000000

- Los comercios tienen en promedio 2.2 **Terminales Vigentes** y máximo 446. El 25% de los comercios con mayor volumen, tienen más de 2 terminales.
- En promedio radicaron 0.18 **Quejas o Reclamos** el último mes, el que más QR registró tuvo 26. Al menos el 75% no radicó ninguna QR el último mes.
- Los comercios realizaron en promedio 532.3 **Trx el último mes**, con std de 1548.8 lo que indica alta dispersión de los datos. El 25% de los comercios con mayor número de trx realizadas, registraron más de 419 trx.
- El promedio del **Número de Reversos** por comercio en el último mes fue de 0.60. Al menos el 75% no tuvo ningún reverso.
- La **Tasa de Reversos** promedio del último mes fue de 0.2099% con std de 1.54%. Al menos el 75% tiene tasa del 0%.
- La **Tasa de Negaciones** promedio del último mes fue de 7.8% con std de 9.4%. El 25% con mayor tasa están por encima de 9.9%.
- La **Diferencia de trx** realizadas en el último mes respecto al anterior, es en promedio de 51.21 trx adicionales, reflejando aumento transaccional.
- La **Variación (%) de trx** realizadas en el último mes respecto al anterior, es en promedio de 52.6%, reflejando aumento transaccional.
- La **Tasa de Quejas y Reclamos** promedio es de 8.89%. Al menos el 75% tiene tasa del 0%. La tasa máxima fue de 9, corresponde a un comercio que registró un número de QR que es 9 veces mayor a su número de terminales vigentes.

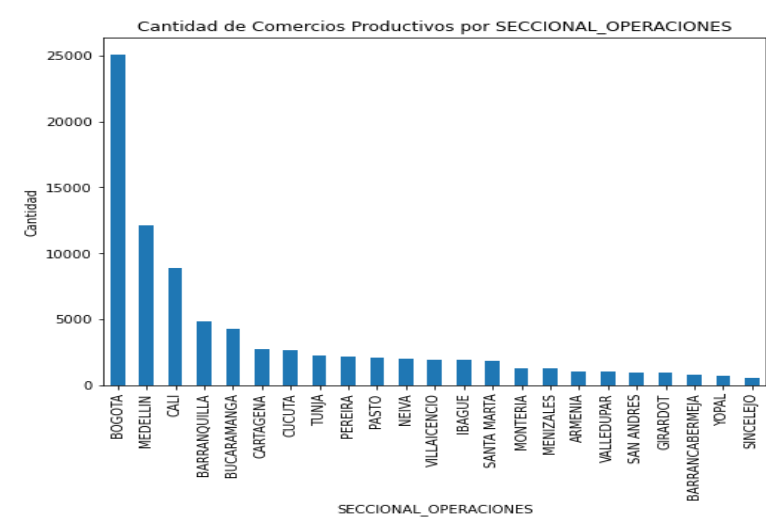


$TASA_NEGACIONES_ULT_MES = 1 - TASA_APROBACIONES_ULT_MES$. Por lo cual su correlación es = -100%. Dado lo anterior, se decide remover la variable "TASA_APROBACIONES_ULT_MES" y conservar "TASA_NEGACIONES_ULT_MES".

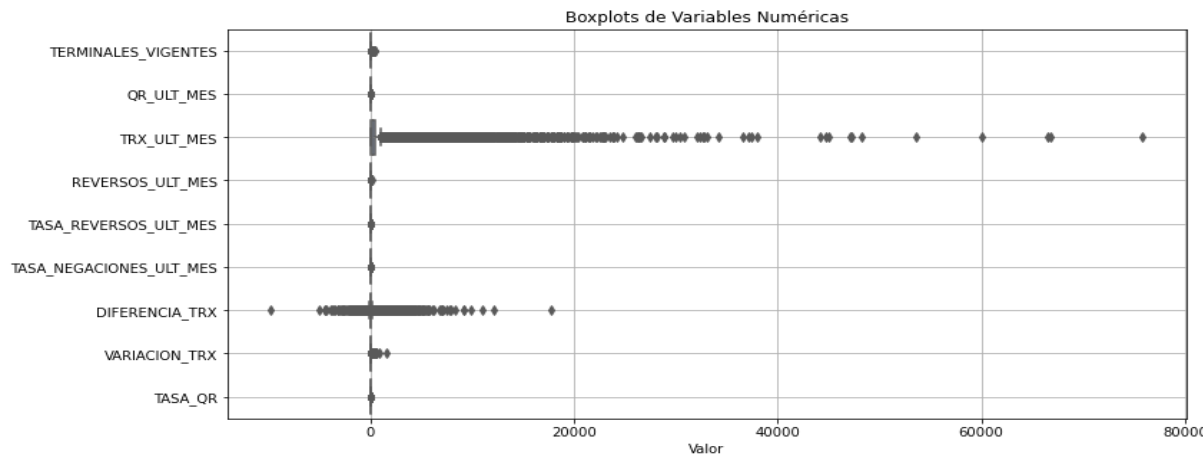
El número de Trx del mes anterior tiene alta correlación con el número de Trx del último mes: 98%. Se decide remover las variables "TRX_MES_ANT" y conservar TRX_ULT_MES".

Por otro lado: existen 23 seccionales a nivel nacional, para la gestión operativa de los comercios: La seccional con mayor volumen de comercios es **Bogotá** con: 25.092 comercios

(29.9%). Las 3 seccionales principales agrupan el 55% de los comercios: **Bogotá, Medellín y Cali**:



En la gráfica a continuación es posible identificar: Las variables se encuentran en escalas de medida distintas y que existe presencia de outliers:



```
: # Obtener el tamaño del DataFrame Final, luego de validaciones y depuración:
num_filas, num_columnas = df.shape
print(f'Número de filas: {num_filas}')
print(f'Número de columnas: {num_columnas}')
```

Número de filas: 83937
Número de columnas: 6

	TRX_ULT_MES	TASA_REVERSOS_ULT_MES	TASA_NEGACIONES_ULT_MES	DIFERENCIA_TRX	VARIACION_TRX	TASA_QR
1	1076	0.004647	0.040892	86	0.086869	0.0
2	6	0.166667	0.166667	6	0.000000	1.0
3	153	0.006536	0.026144	24	0.186047	0.0
4	75	0.000000	0.080000	-17	-0.184783	0.0
5	53	0.000000	0.037736	-6	-0.101695	1.0

- **PROPUESTA METODOLÓGICA**

Se incorpora aprendizaje no supervisado en el proyecto al implementar un algoritmo que logre identificar patrones en comercios críticos a nivel de su comportamiento transaccional o de la funcionalidad de terminales (datáfonos).

De forma inicial se planea implementar el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise) ya que es un algoritmo de clustering ideal para realizar segmentación de clientes, descubriendo patrones y relaciones en los datos; además, tiene ventajas como: Identificación de outliers, parametrización flexible, fácil interpretación de resultados y escalabilidad a grandes volúmenes de datos.

Por su parte, el algoritmo de Clustering Jerárquico es candidato opcional para utilizarse debido a la identificación de estructuras jerárquicas en los datos, la capacidad de visualizar claramente los resultados a través de dendrogramas, la flexibilidad en la interpretación de agrupaciones y su capacidad para manejar conjuntos de datos de diferentes tamaños y escalas. Además, el clustering jerárquico es particularmente útil cuando se busca identificar subgrupos dentro de grupos más grandes, lo que puede proporcionar información valiosa sobre la estructura de los datos de clientes.

- **BIBLIOGRAFÍA**

Aragón Gallegos, A. D. C., Cerquin Silva, S. I., Ecurra Yactayo, R. O., & Roncalla Viena, A. L. (2023). *Segmentación de clientes para mejorar la experiencia de compra de productos electrónicos en Falabella* [Trabajo de Suficiencia Profesional]. Tesis presentada en satisfacción parcial de los requerimientos para obtener el título profesional de Ingeniero(a) Industrial y Comercial.

Syarif, I., Prugel-Bennett, A., & Wills, G. (2012). *Unsupervised Clustering Approach for Network Anomaly Detection*. En P. Revesz y M. L. Gavrilova (Eds.), *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA 2012)* (Vol. 7335, pp. 76-90). Springer. DOI: 10.1007/978-3-642-30507-8_7.