

IDENTIFICACIÓN DE COMERCIOS PARA PLAN DE MANTENIMIENTO INTELIGENTE DE 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 una 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 y de sus datáfonos. Mediante la aplicación de métodos de aprendizaje no supervisado como: Clustering DBSCAN y Jerárquico, se espera obtener segmentos de comercios bien definidos. Esto ayudaría al negocio a enfocar mejor su plan de mantenimiento de, 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 al negocio y a sus clientes.

Los clusters identificados con Clustering Jerárquico lograron capturar comercios altas tasa de reversos o de quejas/reclamos, o con variación transaccional elevada. Con DBSCAN se identificaron clusters que capturan comercios con tasas de quejas o reclamos cercanas o mayores al 100% y outliers con problemáticas combinadas. Aunque ambos enfoques permiten identificar comercios con características propicias para inclusión en el plan de mantenimiento, se selecciona DBSCAN como algoritmo final dadas sus bondades.

INTRODUCCIÓN

En la gestión de una red de datáfonos, se presenta un desafío crucial: la identificación de comercios para su inclusión en un plan de mantenimiento inteligente. Este plan tiene como objetivo optimizar los recursos disponibles y garantizar la operación eficiente de los datáfonos, lo que beneficia tanto al negocio como a sus clientes. Para abordar esta problemática, se ha recurrido al aprendizaje no supervisado, mediante la aplicación de técnicas de Clustering.

El enfoque de Clustering permite identificar grupos de comercios que comparten características similares basadas en las variables proporcionadas. Al analizar estos grupos, se pueden detectar patrones problemáticos y definir criterios claros 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. Su objetivo principal es mejorar la calidad de su servicio y al mismo tiempo, maximizar sus ingresos. El desafío clave radica en la identificación eficaz de aquellos comercios que enfrentan situaciones problemáticas y que deben ser incluidos en el plan de mantenimiento. Para abordar este desafío, se dispone de un conjunto de variables que proporcionan información valiosa sobre el comportamiento de los comercios y sus datáfonos. Entre estas variables se incluyen el número de terminales vigentes (datafonos), quejas y reclamos registrados, transacciones realizadas y tasas de transacciones reversadas y negadas.

Dentro de la literatura revisada, se han identificado enfoques similares que se centran en el aprendizaje no supervisado y que pueden proporcionar valiosas perspectivas para la solución de este problema específico.

En un 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 aplican

métodos de aprendizaje no supervisado, como el clustering jerárquico, K-Medoids y K-Means, con el objetivo de identificar perfiles de consumidores. Este enfoque permite potenciar las estrategias de marketing, aumentar ventas y mejorar preferencias de clientes. Aunque el público objetivo es diferente, la aplicación de técnicas de aprendizaje no supervisado, específicamente de clustering, es una práctica relevante y puede guiar la elección de los algoritmos en el presente trabajo.

Por otro lado, en el artículo "Unsupervised Clustering Approach for Network Anomaly Detection" (Syarif, I., Prugel-Bennett, A., & Wills, G., 2012), se realiza una comparación entre el enfoque de detección de intrusiones en redes mediante la detección de anomalías y 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 revelan que la detección de anomalías tiene un mejor rendimiento, especialmente con el algoritmo de detección de valores atípicos basado en distancias.

La aplicación de técnicas de Clustering, respaldada por la literatura investigada, ofrece un enfoque prometedor para abordar este desafío que requiere la identificación precisa de comercios para inclusión en el plan de mantenimiento de la red de datafonos.

- Con Clustering Jerárquico se identificaron 5 clusters de los cuales 3 de ellos logran capturar 628 comercios con altas tasa de reversos o de quejas/reclamos, o con variación transaccional elevada.
- Con DBSCAN se identificaron 3/4 clusters que logran capturar 136 comercios con tasas de quejas o reclamos cercanas o mayores al 100% y 255 outliers con problemáticas combinadas de las variables consideradas.

Ambos enfoques permiten identificar comercios con características propicias para ser incluidos en el plan de mantenimiento de la red, sin embargo, una limitación del enfoque es que la interpretación de los clusters obtenidos por ambos métodos dependerá del conocimiento del negocio y de la habilidad del analista. Una vez identificados e interpretados los clusters a gestionar, se recomienda definir la estrategia de abordaje de comercios de acuerdo a los distintos escenarios u objetivos posibles tales como: Mantener el comportamiento actual del comercio (fidelización), Mitigar problemáticas transaccionales (altas tasas de reversos o de negaciones), Mitigar problemáticas sobre quejas y reclamos.

• MATERIALES Y MÉTODOS

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

Conjunto de datos inicial: 135.954 registros (comercios) y 13 variables.

Tipo y descripción de variables disponibles: Las primeras 12 numéricas, la última categórica:

- **ID** : Código único o identificador del establecimiento comercial.
- **TERMINALES_VIGENTES**: datáfonos instalados en el comercio.
- **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 el mes anterior.

- **REVERSOS_ULT_MES:** Transacciones del comercio reversadas en último mes.
- **DIFERENCIA_TRX:** $TRX_ULT_MES - TRX_MES_ANT$.
- **TASA_REVERSOS_ULT_MES:** % transacciones reversadas del total de trx realizadas el último mes: $REVERSOS_ULT_MES / TRX_ULT_MES$.
- **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.
- **VARIACION_TRX:** Variación (%) del número de transacciones realizadas en los últimos 2 meses, esto es: $(TRX_ULT_MES / TRX_MES_ANT) - 1$
- **TASA_QR:** Tasa de quejas y reclamos del comercio según su número de terminales. Esto es: $QR_ULT_MES / TERMINALES_VIGENTES$.
- **SECCIONAL_OPERACIONES:** Seccional a la que pertenece el comercio para su gestión operativa, según su ubicación geográfica.

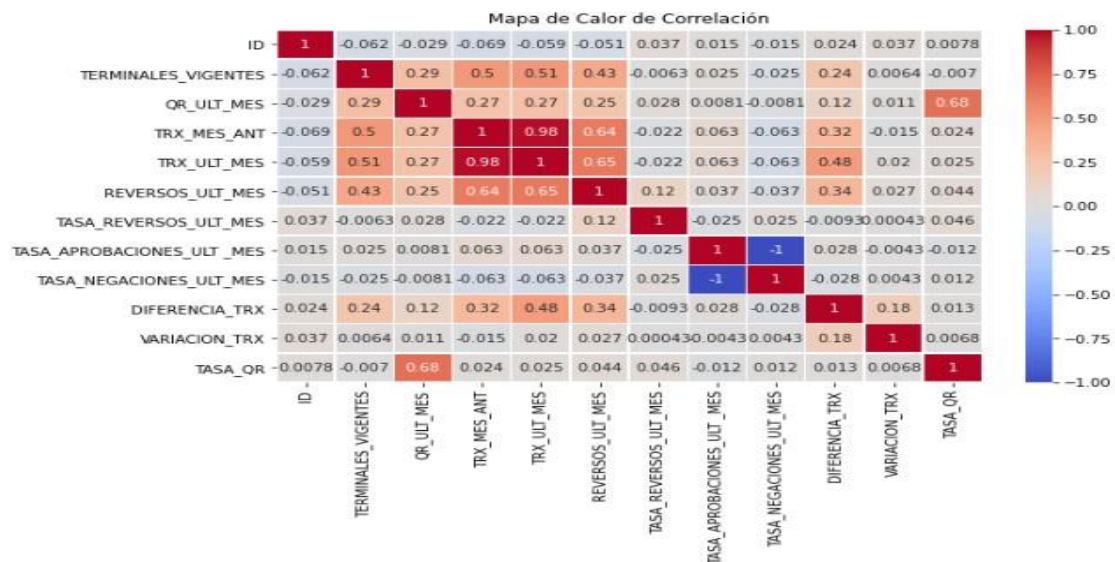
Un **Reverso o transacción reversada** se refiere a una transacción previamente autorizada que se revierte, lo que significa que los fondos que se habían transferido se devuelven al titular de la cuenta o tarjeta. Una **transacción negada** es una trx que es rechazada o denegada por la entidad o Autorizador. El negocio recibe un pago mayor por las trx aprobadas respecto a las negadas.

Bajo el contexto de la red de datafonos, para el negocio es deseable: **Minimizar:** Tasa de quejas/reclamos, la tasa de reversos y la tasa de negaciones. **Maximizar:** Transacciones último mes, diferencia y variación trx últimos 2 meses.

Estadísticos descriptivos: Comercios con trx en el último mes (removiendo improductivos):

	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
- En promedio radicaron 0.18 **Quejas o Reclamos** el último mes, máximo 26.
- Realizaron en promedio 532.3 **Trx el último mes**, con std de 1548.8 con alta dispersión. El 25% de los comercios con mayor número de trx, registraron más de 419.
- El promedio del **Número de Reversos** por comercio en el último mes fue de 0.60.
- 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 (número de QR 9 veces mayor a terminales vigentes).



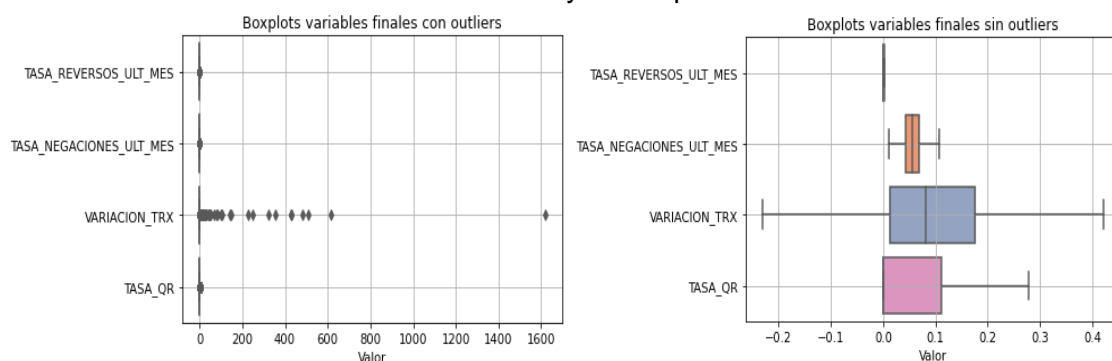
La correlación más fuerte se identifica entre las siguientes variables:

- TASA_NEGACIONES_ULT_MES, TASA_APROBACIONES_ULT_MES: -100%
- TRX_ULT_MES, TRX_MES_ANT: 98%.
- QR_ULT_MES, TASA_QR: 68%

Teniendo en cuenta los resultados del análisis de los datos, para la implementación de Clustering y con el objetivo de focalizar la segmentación en los comercios de mayor interés para el negocio, se define contemplar sólo aquellos que hacen parte del 10% con mayor número de transacciones en el último mes (percentil 90% = 1.276 trx) y la inclusión de 4 variables:

	TASA_REVERSOS_ULT_MES	TASA_NEGACIONES_ULT_MES	VARIACION_TRX	TASA_QR
60	0.000000	0.030921	0.021536	0.000000
61	0.002608	0.035994	0.099828	0.000000
125	0.001517	0.044259	0.021970	0.142857
194	0.001395	0.043236	0.056743	0.000000
202	0.000444	0.138605	0.202457	0.000000

Conjunto de datos final: 8.390 registros (comercios) y 4 variables. Las variables se encuentran en escalas de medida distintas y existe presencia de outliers:



Algoritmo seleccionado: DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Algoritmo de clustering ideal para realizar segmentación de clientes, descubriendo patrones y relaciones en los datos. Ventajas: Identificación de outliers, parametrización flexible, fácil interpretación de resultados y escalabilidad a grandes volúmenes de datos.

Tras validar los resultados con distintos valores para eps (0.25, 0.5, 0.75, 1) y min_samples (4, 8, 16), se escogen: eps=0.5 y minsamples=8, obteniendo 4 clusters y 255 outliers.

```
eps=0.5, min_samples=8, Número de clústeres: 4
Cluster_DBSCAN_Final
-1      255
0      7999
1      112
2       10
3       14
dtype: int64
```

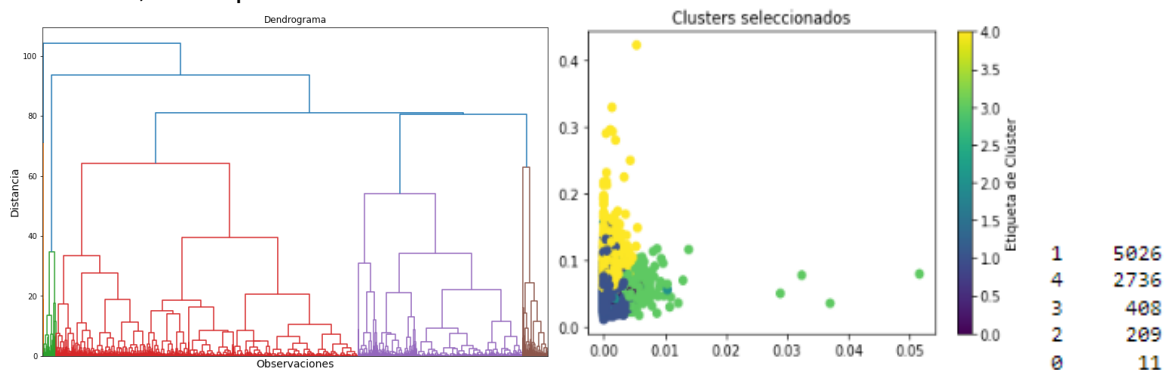
• RESULTADOS Y DISCUSIÓN

Cluster_DBSCAN_Final		-1	1	2	3	Cluster_DBSCAN_Final		-1	1	2	3
TASA_REVERSOS_ULT_MES	count	255.000000	112.000000	10.000000	14.000000	VARIACION_TRX	count	255.000000	112.000000	10.000000	14.000000
	mean	0.003509	0.000687	0.000546	0.000430		mean	27.384924	0.262758	0.093845	0.395626
	std	0.005260	0.000654	0.000240	0.000213		std	129.711134	0.787985	0.235379	1.188696
	min	0.000000	0.000000	0.000000	0.000000		min	-0.535839	-0.337922	-0.315117	-0.094788
	25%	0.000707	0.000000	0.000419	0.000294		25%	0.021572	0.010770	-0.013314	0.040759
	50%	0.002128	0.000528	0.000604	0.000508		50%	0.119527	0.091598	0.130197	0.054883
	75%	0.004632	0.001264	0.000726	0.000556		75%	0.379218	0.185048	0.144037	0.122835
	max	0.051569	0.002387	0.000762	0.000772		max	1616.333333	5.393443	0.609113	4.505226
TASA_NEGACIONES_ULT_MES	count	255.000000	112.000000	10.000000	14.000000	TASA_QR	count	255.000000	112.000000	10.000000	14.000000
	mean	0.087237	0.051848	0.040529	0.048009		mean	0.778767	1.002778	1.500000	1.303231
	std	0.057761	0.015959	0.006672	0.007071		std	0.846804	0.021536	0.000000	0.063714
	min	0.016742	0.017574	0.028780	0.036352		min	0.000000	1.000000	1.500000	1.200000
	25%	0.051487	0.040985	0.036359	0.043346		25%	0.080128	1.000000	1.500000	1.250000
	50%	0.071131	0.050295	0.040799	0.047262		50%	0.500000	1.000000	1.500000	1.333333
	75%	0.102641	0.061959	0.045199	0.054129		75%	1.333333	1.000000	1.500000	1.333333
	max	0.422182	0.092318	0.049505	0.059007		max	4.666667	1.200000	1.500000	1.428571

Los clusters de interés (-1, 1, 2, 3) contienen 391 comercios, correspondiente al 4.66% de los comercios en la bd final, para inclusión dentro del plan de mantenimiento:

- Los **cluster 1, 2, 3** logran capturar comercios con tasa de quejas o reclamos cercana o mayor al 100%, esto es: el número de quejas o reclamos radicados el último mes iguala o supera el número de datafonos instalados en el comercio.
- El **Cluster -1** que identifica los outliers, logra capturar comercios con situaciones de interés combinadas: tasa de reversos y negaciones por encima de la tasa general, alta tasa de quejas y reclamos (promedio de 77.9%) y/o alta variación transaccional.

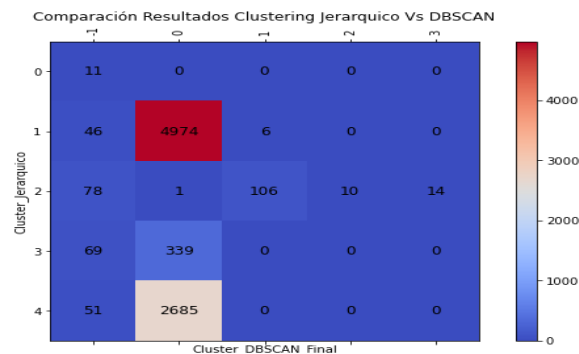
Algoritmo alternativo: Clustering Jerárquico: Los 3/5 clusters de interés (0, 2, 3) contienen 628 comercios, correspondiente al 7.49% de los comercios en la BD final.



- **Cluster 0** Captura comercios con alta variación transaccional Vs. el mes anterior.
- **Cluster 2:** Comercios con alta tasa de quejas y reclamos en el último mes.
- **Cluster 3:** Comercios con alta tasa de reversos alta en el último mes.

Ambos enfoques permiten identificar comercios con características propicias para inclusión en el plan de mantenimiento de la red, sin embargo: Clustering Jerárquico separa

aquellos comercios con tasas extremas respecto a alguna de las variables, pero de forma individual, agrupando en 2 grandes clústers el resto de los comercios. Mientras que DBSCAN además logra capturar comercios con varias problemáticas combinadas que, aunque no son propiamente valores extremos, presentan varios valores de interés simultáneamente.



La interpretación de los clusters dependerá del conocimiento del negocio y habilidad del analista, las acciones a implementar en los clusters de interés requieren definir estrategias de abordaje de comercios de acuerdo a la oportunidad que capture cada cluster. En análisis posteriores se pueden incluir nuevas variables o se sugiere implementar un plan de mantenimiento piloto con los resultados obtenidos, para afinar el entendimiento de los clusters identificados, vital para la toma de decisiones sobre su correcta intervención.

• CONCLUSIONES

- En este proyecto se abordó el desafío de identificar perfiles de clientes (comercios) en una empresa de medios de pago colombiana de datáfonos, para inclusión en el plan de mantenimiento de la red, para mejorar el servicio y optimizar recursos.
- Mediante la aplicación de técnicas de aprendizaje no supervisado, como Clustering DBSCAN y Jerárquico, se obtuvieron segmentos de comercios bien definidos, usando una BD final que incluye los 8.390 comercios de mayor transaccionalidad y 4 variables.
- Aunque ambos enfoques permiten identificar comercios con características propicias para inclusión en el plan de mantenimiento, se seleccionó DBSCAN el cual logra capturar además de comercios con tasas extremas, aquellos que combinan tasas indeseadas las cuales no son necesariamente valores atípicos.
- La interpretación de los clusters dependerá del conocimiento del negocio y de la habilidad del analista. Se recomienda definir estrategias de abordaje por cluster, afinar la comprensión de los mismos y definir los cursos de acción correspondientes.

• 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.