



Segmentación y Recomendador

Introducción

Solicitud

Realizar una segmentación de clientes que permita identificar desde los mejores clientes hasta aquellos que ya no lo son.

Implementar un sistema de recomendación de productos con el fin de incrementar las ventas y el monto de las mismas.

Base de datos

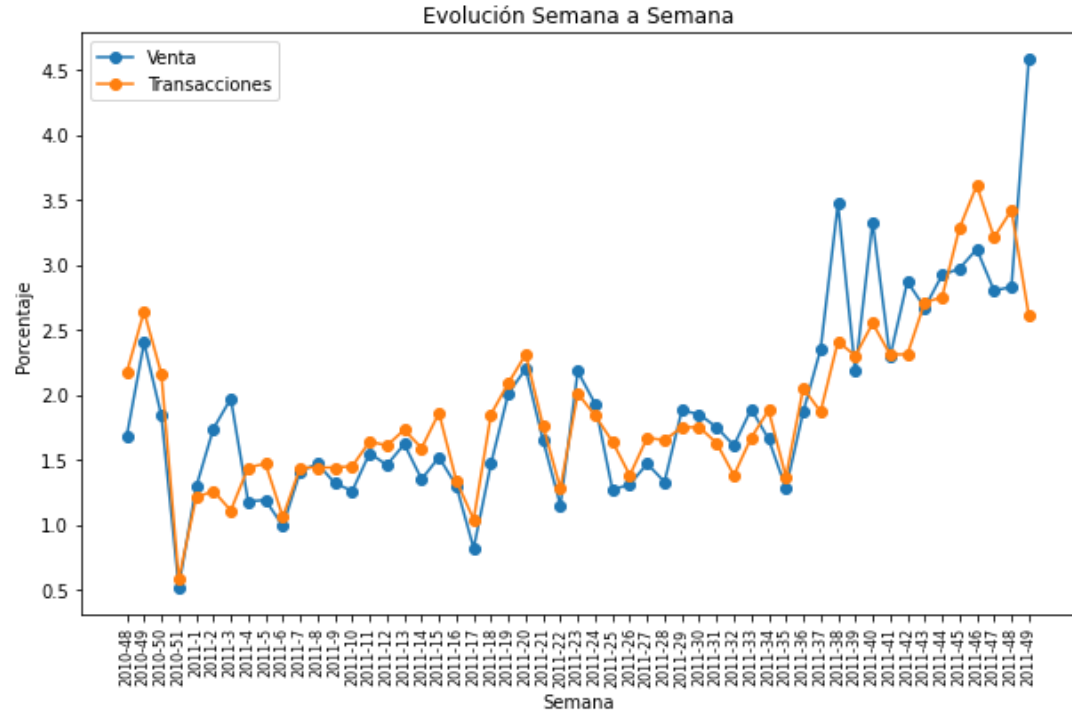
Se tienen datos de **4338** clientes, los cuales corresponden a todas las transacciones que se realizaron entre diciembre de 2010 y diciembre de 2011.

Herramientas



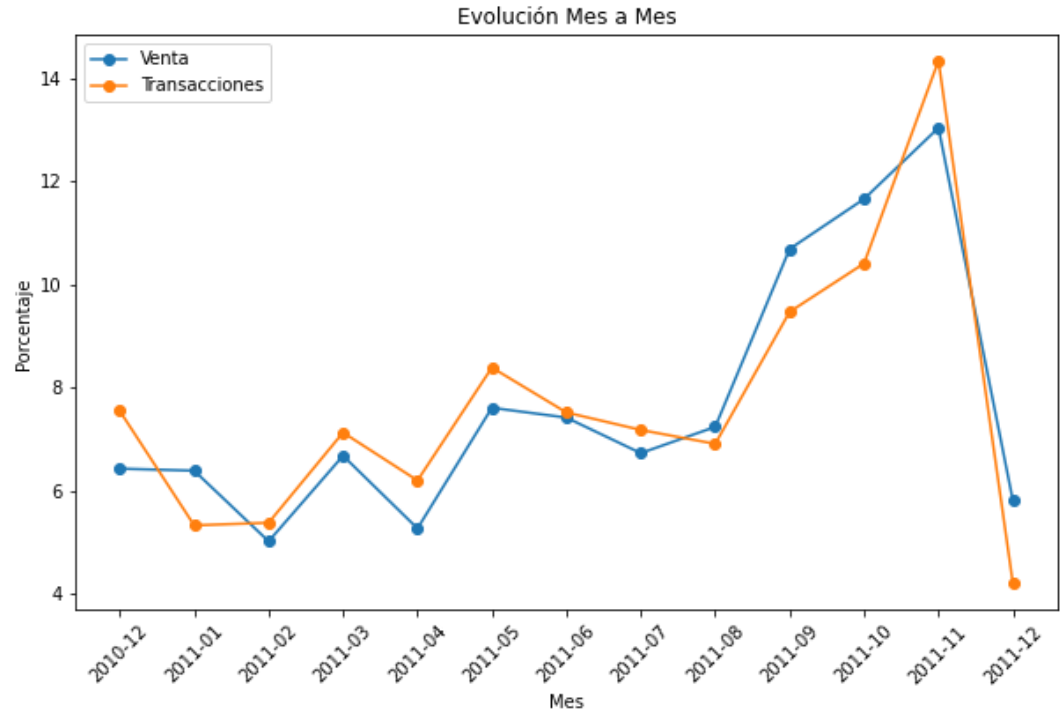
Comportamiento de Ventas y Transacciones

Se destacan las semanas 3, 38, 40 y 49, donde es notoria la separación entre la proporción de las ventas y la transacciones, indicando que en esos días se hicieron pocas transacciones pero de montos altos.



Comportamiento de Ventas y Transacciones

Al hacer una agregación mensual, no se aprecia un alejamiento de las transacciones y la venta en términos de proporción.



¿Quién es Nuestro Cliente Promedio?



Transacciones en el año



USD es su ticket promedio



USD es su compra anual



Días es su recompra



Días visita la tienda en el año



Días han pasado desde su ultima compra



Unidades en cada transacción



Productos diferentes en cada transacción



1.

Segmentación

RFM y K-means

Metodología

RFM

Calculo de las variables

Recencia

Frecuencia

Monto

Recompra

Selección del numero de Clústeres.
Evaluación de la agrupación.
Visualización.

K-Means

**Descripción
de Grupos**

¿Quiénes son los clientes de
cada clúster?

Recencia, Frecuencia y Monto

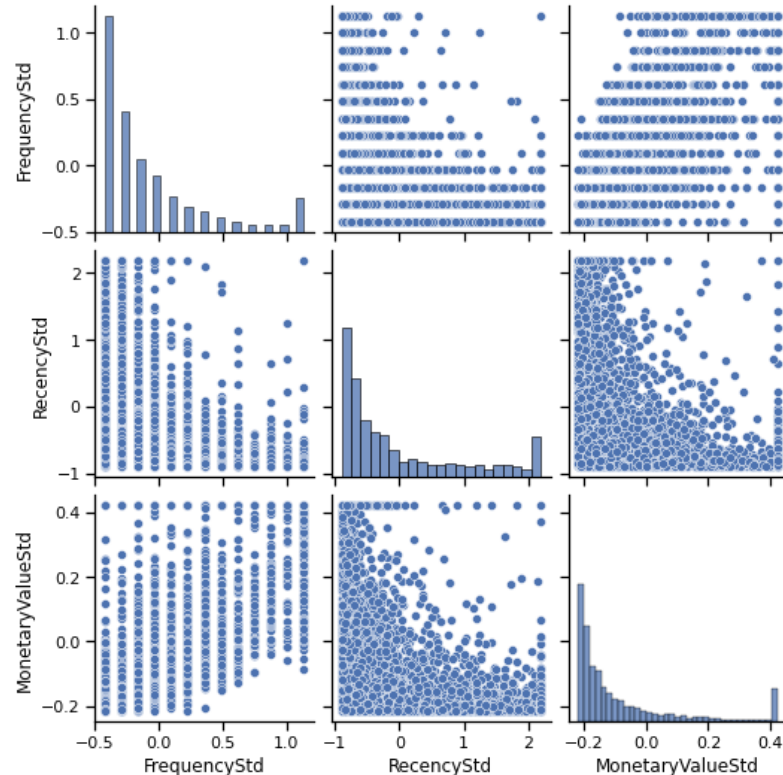
Recencia: Número de días entre la fecha máxima de los datos y la fecha de la última compra de cada cliente.

Frecuencia: Número de transacciones de cada cliente.

Monto: Valor de compra para cada cliente

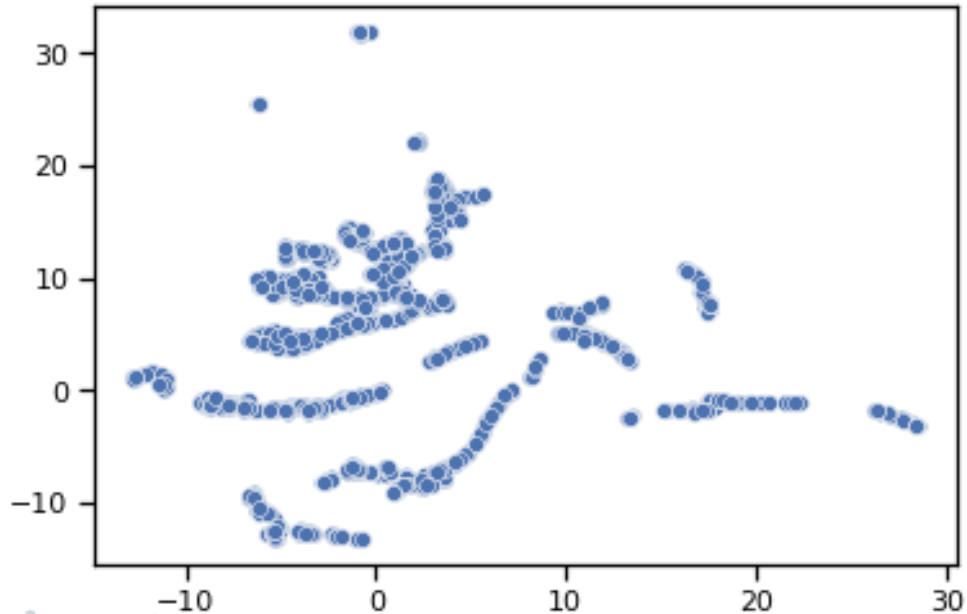
En la gráfica se observa la distribución da cada variable y la relación con las demás.

NOTA: Las variables están estandarizadas



Recencia, Frecuencia y Monto

Proyección de UMAP de los datos



Nótese la tendencia de agrupación natural de los datos.

K-Means

K-Means

El algoritmo *K-means* encuentra los K mejores clústeres, cuya varianza interna sea lo más pequeña posible.

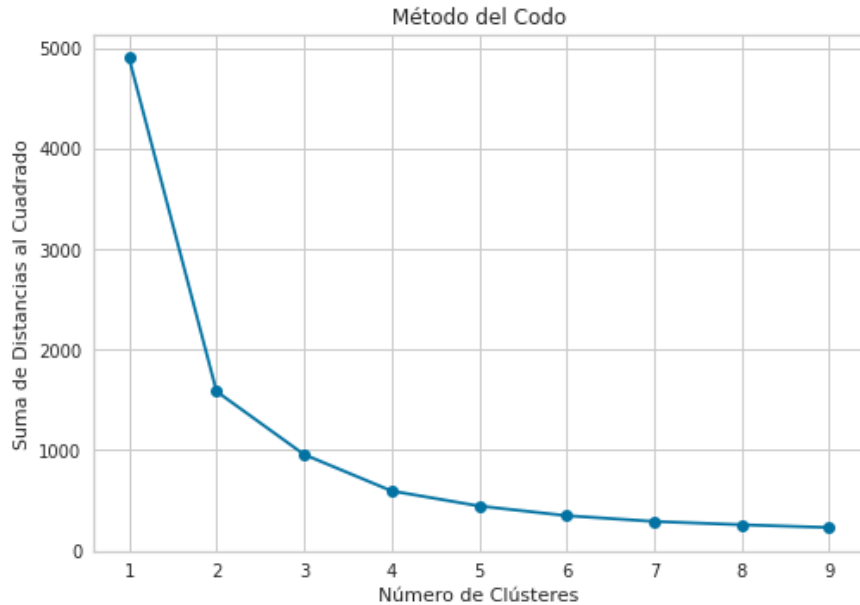
Método del Codo

El método del codo (Elbow) calcula la varianza total dentro clúster en función del número de clústeres y escoge el K óptimo aquel valor a partir del cual añadir más clústeres apenas consigue mejoría.

Coeficiente *Silhouette*

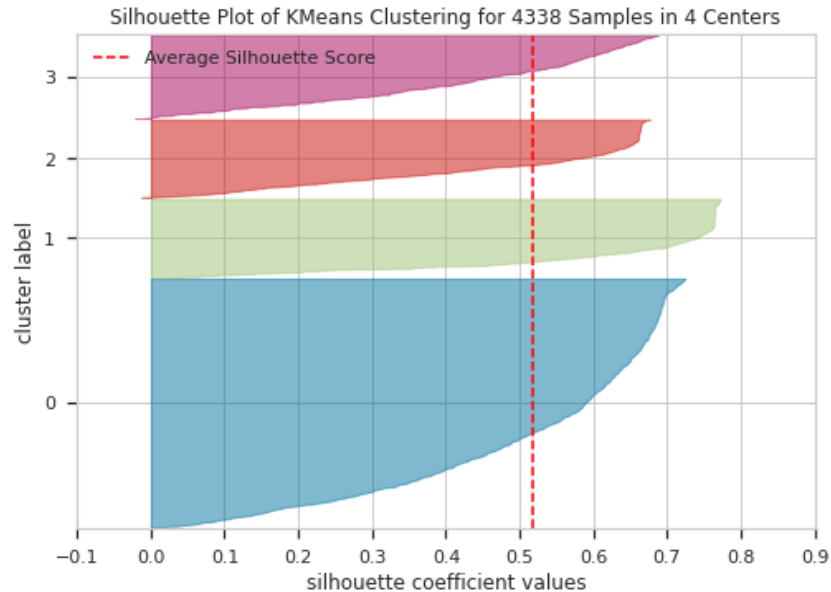
Cuantifica qué tan buena es la asignación de una observación a un clúster, comparando la distancia media a elementos del mismo clúster con la distancia media a elementos de los otros clústeres.

K-Means

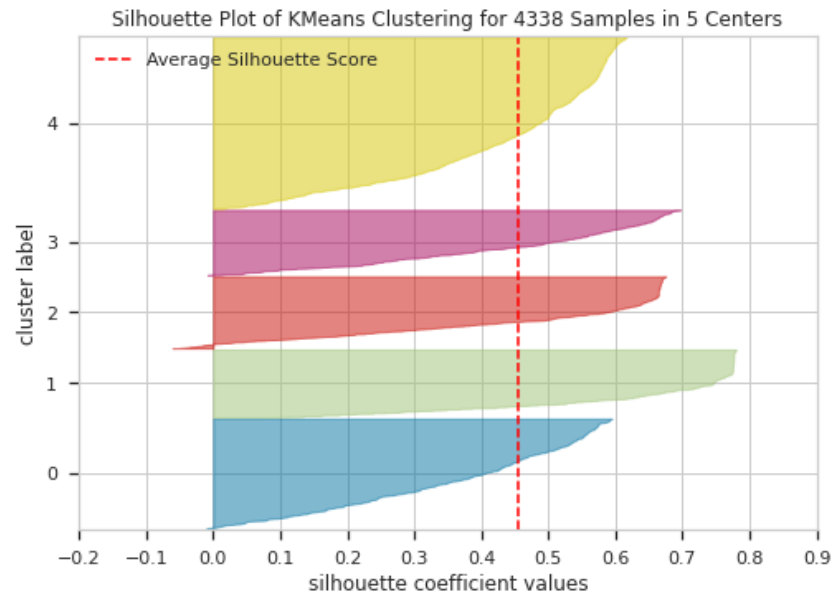


El método del Codo sugiere que el K óptimo está entre 4 y 5.

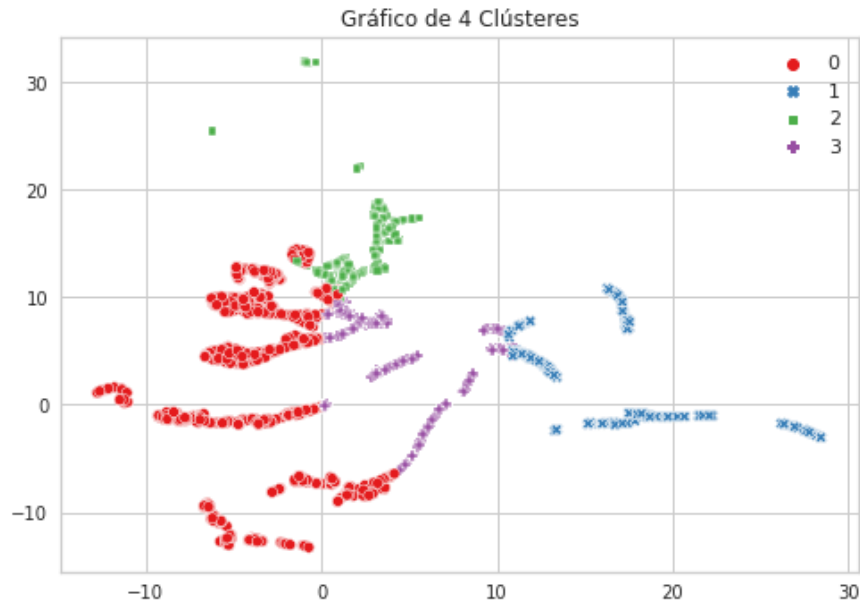
K-Means



Un numero aceptable para el índice Silhouett es > 0.5 , por tanto el K óptimo es igual 4.



K-Means



Representación en dos dimensiones de los datos con en los diferentes clústeres.

Descripción de Grupos

Premium



R: 16 días



F: 14 Transa.



R: 8,008 USD



692 **clientes**

El **16%** de los
clientes hacen el
62% de la venta

Fieles



R: 36 días



F: 3 Transa.



R: 1,069 USD



2206 **clientes**

El **51%** de los
clientes hacen el
26% de la venta

Potenciales



R: 146 días



F: 2 Transa.



R: 800 USD



741 **clientes**

El **17%** de los
clientes hacen el
7% de la venta

Ocasionales



R: 285 días



F: 1 Transa.



R: 597 USD



699 **clientes**

El **5%** de los
clientes hacen el
16% de la venta

Conclusiones

- ❑ Los datos proporcionaron suficiente información para responder a la solicitud.
- ❑ Es importante tener cuidado con los clientes “Premium”, dado que el abandono por parte de uno de ellos, puede representar pérdidas para la compañía.
- ❑ Implementar estrategias comerciales de bajos costo con el grupo de clientes “Potenciales” puede significar un aumento importante de las ganancias.

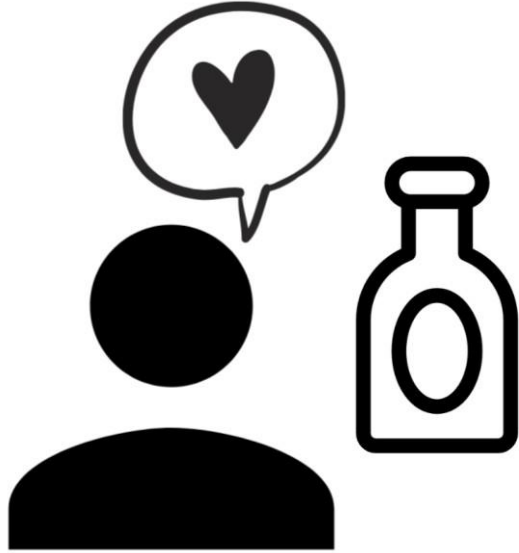


2.

Recomendador

Productos Como el Mío

Enfoque



Buscar el ítem más
próximo alojado en D



Matriz de similitudes D

Recomendador

Similaridad de Coseno

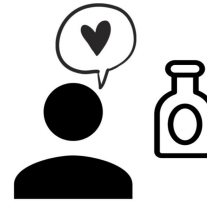
Está definida como:

$$K(x, y) = \frac{xy^T}{\|x\| \|y\|}$$

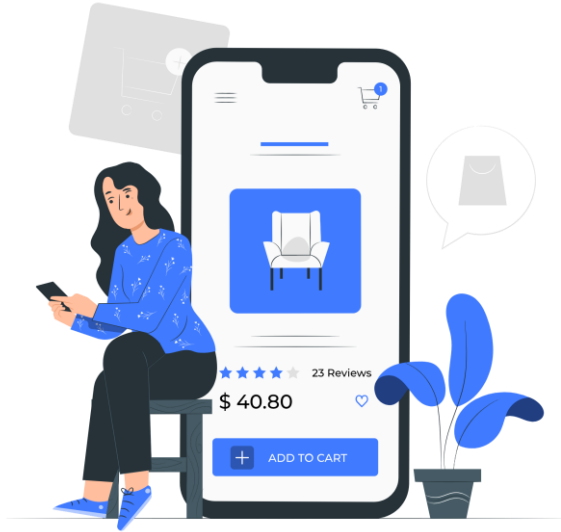
Donde x y y son vectores fila. La similitud de coseno es $CD = 1 - k$ donde $0 \leq CD \leq 1$.

Producto Preferido

El producto preferido está definido como aquél que el cliente más lleva en termino de unidades.



Resultado



Conclusiones

❑ El sistema de recomendación desarrollado es de un enfoque sencillo, tanto en los técnico como para la puesta en producción, aun así responde a la solicitud y sin duda puede hacer aumentar transacciones y montos.

❑ Es un recomendador que no permite hacer validaciones antes de la puesta en producción, por tanto es un buen modelo base pero se podrían implementar algunos más sofisticados.

Referencias

- © Ashishkumar Singh, Grace Rumantir, Annie South and Blair Bethwaite. *Clustering Experiments on Big Transaction Data for Market Segmentation*. The evolution of direct, data and digital marketing, Richard Webber, Journal of Direct, Data and Digital Marketing Practice (2013).
- © Iboukadel Kassambara. *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*, volume 1. STHDA, 2017.
- © Jonathan Galindo Estrada , Reinaldo Uribe. *Basket-similarity-based Model For Product Recommendation In Retail*. XII Coloquio de Estadística Medellín, Colombia 19 al 22 de Noviembre 2019.

!Gracias!



Stiven Pérez Pulgarín
CIENTÍFICO DE DATOS

