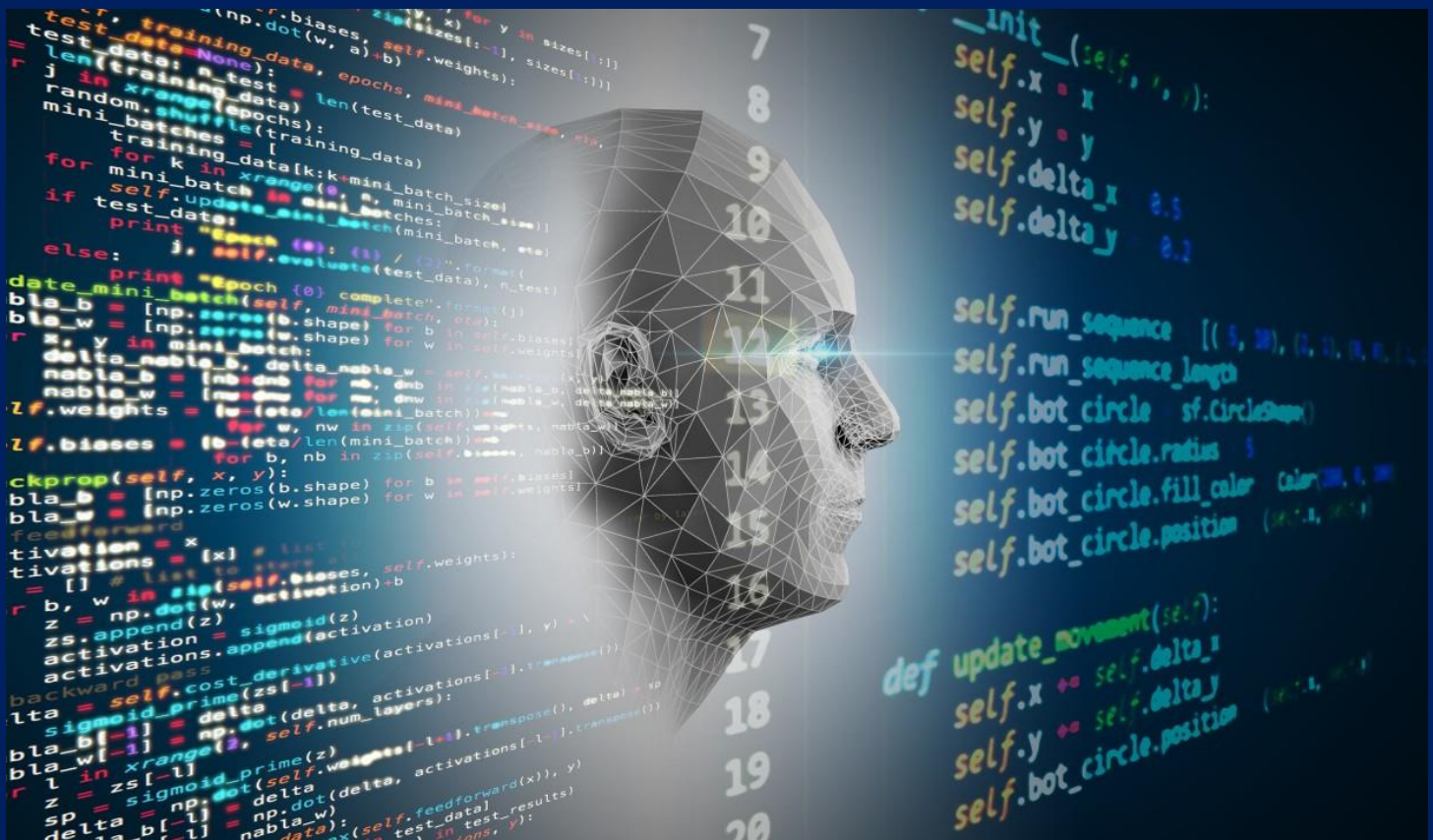




Segmentación de Clientes

by Ivan Alducin



Segmentación de Clientes

Estadística Descriptiva

Análisis Exploratorio

Durante esta semana nos vamos a enfocar en entender y trabajar un caso de uso para segmentación de clientes, pero antes de eso aquí una pequeña lista de algunas aplicaciones que se pueden trabajar con los datos recopilados de los clientes

- Estadística Descriptiva
- Segmentación de Clientes
- Predicción de Abandono
- Valor del Cliente a través del tiempo (CTLV)

Cabe mencionar que toda empresa (Con fines de lucro) tiene clientes, ¿No?.

La metodología de Segmentación de Clientes es una herramienta poderosa que se usa en diferentes industrias, desde la banca, turismo y la industria de consumo masivo. Cada vez las empresas quieren conocer mejor a sus clientes por medio de las diferentes fuentes de datos que tienen disponibles con la información de estos clientes.

Nosotros desarrollaremos una metodología llamada **RFM**, que por sus siglas en inglés, hace referencia a los siguientes conceptos: *Recency*, la cual nos indica que tan reciente es una compra, *Frequency*, la cual hace referencia a que tan frecuente es una compra y por último *Monetary*, la cual hace referencia al gasto monetario que el cliente hace consumiendo los productos. Pero antes de todo, dejemos un poco de lado toda esta información y enfoquémonos en lo primero: La estadística descriptiva:

La estadística descriptiva nos da el soporte para poder contestar preguntas como

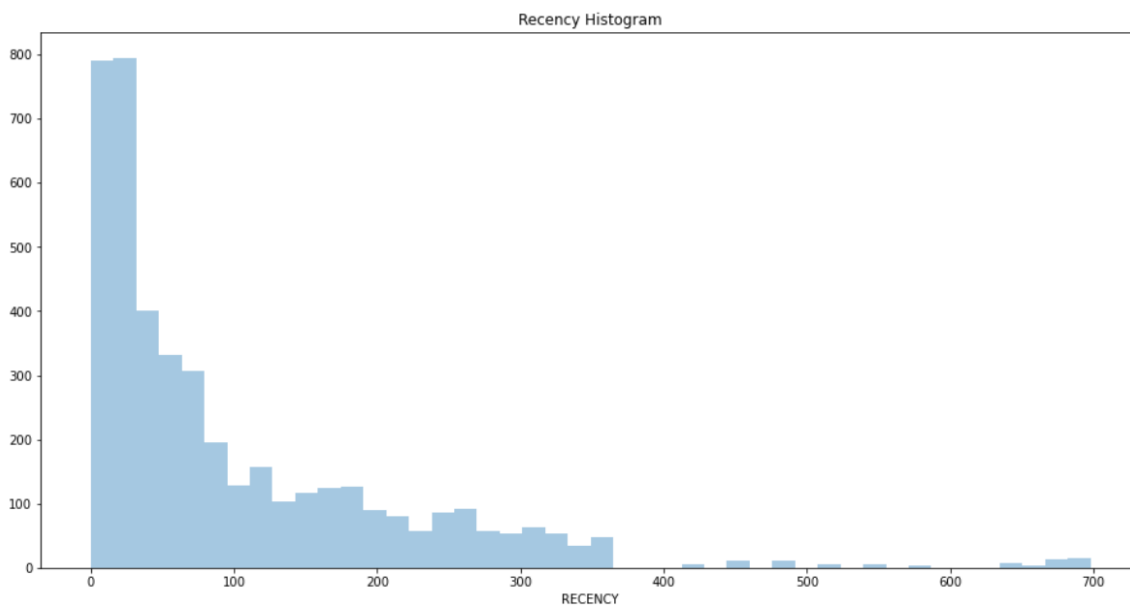
- ¿Qué posibilidades hay de que alguien compre un producto? ¿Es más probable que las personas lo compren si pueden usar un sistema de pago diferente?
- ¿Cuántos ocupantes se hospedarán hoy en mi hotel? ¿Se puede optimizar la ocupación?
- ¿Cuántas tallas de jeans se deben fabricar para que estas lleguen al 95% de la población? ¿Deberían producirse en la misma cantidad cada talla?
- Pruebas A/B: ¿Qué anuncio es más efectivo para que las personas compren mi producto?

Durante la primera etapa, te vas a encargar de encontrar respuestas a la mayor cantidad de preguntas que puedan dar valor y generar *insights* con la información que tienes, puedes tomar las preguntas anteriores como un ejemplo.

Además, como mencionamos arriba, vamos a crear también nuestros 3 indicadores claves que usaremos más adelante para construir el modelo de segmentación. Abajo algunas pistas:

- **Recency.** Esta métrica se puede construir tomando como base el día actual y a este restar los días hasta la última fecha de compra del cliente, puede ser días, semanas, años, esto depende del tipo de industria y la frecuencia promedio de compra (las personas no compran un auto nuevo cada 2 semanas, o tal vez sí, pero no las conocemos). Así que puedes usar la última fecha de compra del conjunto de datos como parámetro para estimar la diferencia de días del resto de compras.
- **Frequency.** Esta métrica se puede construir contando el número de compras que un cliente ha hecho, puede ser por compra, o por producto, depende del enfoque que uno le quiera dar.
- **Monetary.** Esta métrica es la más sencilla, representa simplemente el monto en dinero invertido por el cliente comprando un producto.

Puedes apoyar tu análisis exploratorio haciendo un estudio de las métricas anteriores. Los visuales son importantes y eso es justo lo que se tiene que mostrar al final del proceso. Trata de contar una historia con los datos que estas visualizando.



Machine Learning

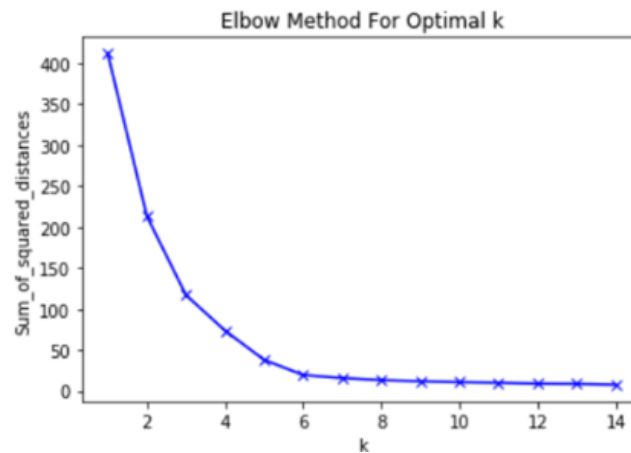
Clustering

Para esta siguiente fase, vamos a desarrollar un modelo de segmentación basado en las 3 variables de nuestra metodología RFM.

Para ello utilizaremos un algoritmo de aprendizaje no supervisado muy utilizado llamado **k-means**, el algoritmo k-means es un método de agrupamiento que divide un conjunto de datos en k grupos o clusters. Los datos se agrupan de tal manera que los puntos en el mismo clúster sean más similares entre sí que los puntos en otros clusters.

La idea es identificar a los clientes que representen alto valor de acuerdo con su patrón de compras, y clientes de bajo valor. Puedes agregar los grupos que tu consideres necesarios (Acabo de hacer dos grupos: alto valor y bajo valor). Pero lo ideal es tener una segmentación más específica que represente y diferencie cada grupo.

Una pregunta importante: ¿Cómo saber cuantos grupos debo crear? Para ellos se ocupará una técnica llamada **método del codo**. No daré muchos detalles ya que vendrán dentro del material auxiliar y además de que puedes hacer un poco de investigación o pedir a ChatGPT que te genere un código en Python para calcular este método ;)



También utilizaremos una función auxiliar en Python que nos va a permitir ordenar de menor a mayor el valor de nuestros clusters, donde el valor que corresponde al menor, como su nombre lo dice es el cluster que para nuestro efecto tenga el menor peso (bajo valor) y viceversa. Esto depende del número de clusters que crees, ejemplo si creas 4, el cluster 1 será asignado como el de más bajo valor mientras que de manera descendente el cluster 4 será asignado como el de más alto valor.



Python script

```
import pandas as pd

# Funcion para ordenar los clusters
def order_cluster(cluster_field_name, target_field_name, df, ascending):
    new_cluster_field_name = 'new_' + cluster_field_name
    df_new = df.groupby(cluster_field_name)[target_field_name].mean().reset_index()
    df_new =
df_new.sort_values(by=target_field_name,ascending=ascending).reset_index(drop=True)
df_new['index'] = df_new.index
df_final = pd.merge(df,df_new[[cluster_field_name,'index']], on=cluster_field_name)
df_final = df_final.drop([cluster_field_name],axis=1)
df_final = df_final.rename(columns={"index":cluster_field_name})
return df_final
```

Al final y como se mencionó en un principio la idea es tener la información detallada por cliente y el segmento al que este le corresponde de acuerdo a su patrón de compras (RFM).

Puedes representar tus resultados de manura tabular, pero lo ideal es aprovechar la ventaja y el poder de los visuales y plasmarlos dentro de una página web. Recuerda no todo son los visuales, sin una historia, esto son practicamente nada.

Gracias! :)