

Don't panic it's organic!

Trabajo final del curso de Data Science de CoderHouse
Camila Martegani - Comisión 29795





AGENDA

1. CONTEXTO
2. AUDIENCIA Y LIMITACIONES
3. PREGUNTAS Y OBJETIVOS DE LA INVESTIGACIÓN
4. ANÁLISIS EXPLORATORIO (EDA)
5. APLICACIÓN DE ALGORITMOS DE ML
6. CONCLUSIONES





CONTEXTO



CONTEXTO

La pandemia de **COVID-19** vino a romper los esquemas y nos llevó a cuestionar nuestra visión en una gran cantidad de aspectos, entre ellas, la *forma* en que **nos alimentamos**. De pronto, la idea de consumir **alimentos ultraprocesados** pareció el **apocalipsis** y la búsqueda de los **productos orgánicos**, nuestra **luz al final del túnel**.

Las grandes cadenas de supermercados han identificado esta nueva necesidad y han intentado actuar acorde a ello, pero, *¿han logrado "convencer" al público de que venden productos más sanos?*



CONTEXTO

En el presente proyecto se analizará una base de datos proveniente de una cadena de supermercados de Reino Unido que decide ofrecer beneficios al consumo de productos orgánicos para los clientes que forman parte del programa de fidelidad.



AUDIENCIA Y LIMITACIONES



AUDIENCIA Y LIMITACIONES

Este análisis intenta evaluar el impacto en el comportamiento de los consumidores de la inclusión de productos orgánicos en el surtido de cierta cadena de supermercados, por lo que, puede ser de gran utilidad para quién se dedique al área de Category Management y/o Analytics tanto de Retailers como de Manufacturers.

La información recopilada proviene de Reino Unido, por lo que es posible que los resultados obtenidos tengan un pequeño sesgo regional/cultural.



PREGUNTAS Y OBJETIVO DE LA INVESTIGACIÓN



PREGUNTAS Y OBJETIVOS DE LA INVESTIGACIÓN

PREGUNTAS CLAVE

- ¿En qué se diferencian los consumidores de productos orgánicos de los no orgánicos? ¿Hay alguna característica que los distinga?
- ¿Se puede segmentar la cartera de clientes de una forma más adecuada que por el programa de afiliación de sus integrantes?
- ¿Cómo identificar potenciales consumidores de productos orgánicos?
- Los programas de fidelidad de clientes, ¿Son efectivos? ¿Tienen algún impacto en los ingresos del retailer?
- ¿Cómo atraer a los consumidores de productos orgánicos a las tiendas de consumo masivo?



PREGUNTAS Y OBJETIVOS DE LA INVESTIGACIÓN

OBJETIVOS

- Comprobar si es posible hacer una caracterización del perfil de los compradores para segmentar la cartera de una forma más adecuada.
- Implementar acciones concretas que deriven en un mayor revenue para la empresa.
- Buscar un algoritmo que permita clasificar eficazmente nuevos clientes que se sumen a la cartera de la empresa



ANÁLISIS EXPLORATORIO (EDA)



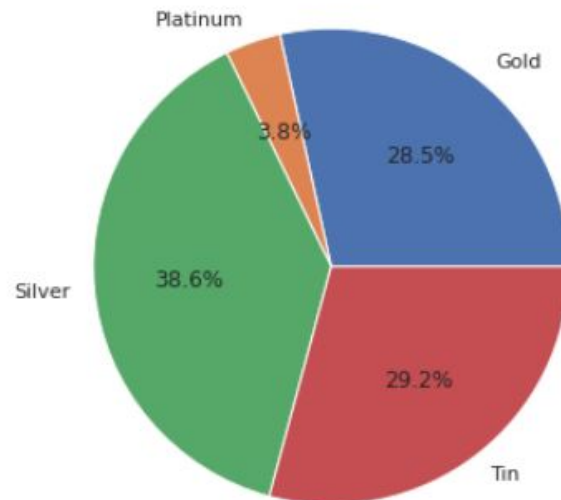
¿Cómo se **compone** la cartera de clientes?

AFILIACIÓN

El programa de afiliación cuenta con 4 tipos de asociados: Tin, Silver, Gold, Platinum.

- **Silver** es el segmento con **mayor** porcentaje de clientes;
- **Tin y Gold** se encuentran levemente por debajo;
- **Platinum** es el segmento **más pequeño**.

Distribución de los clientes segun nivel de afiliación





¿Cómo se **compone** la cartera de clientes?

GÉNERO

- El segmento de **mujeres** representa el **55%** de la cartera;
- Los **hombres** representan el **25%**;
- Un **19%** de la cartera no ha detallado su género (U).



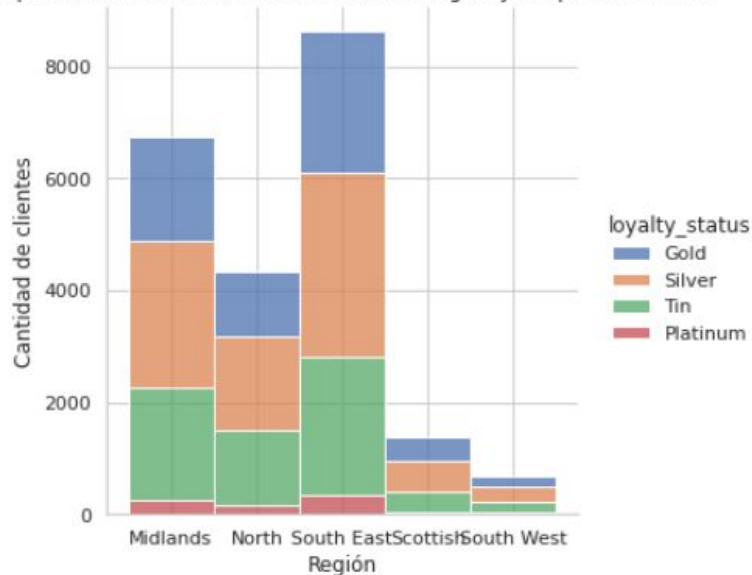


¿Cómo se **compone** la cartera de clientes?

DISTRIBUCIÓN REGIONAL

- **South East** (Inglaterra) es la región con **mayor concentración** de clientes, en tanto **South West** (Gales) se muestra con la **menor concentración**;
- El status **Silver** representa el **mayor porcentaje** de clientes a lo largo de **todas las regiones**;
- Las regiones con **menor porcentaje** de clientes **carecen** de miembros **Platinum** en su composición.

Composición de los clientes de acuerdo a la región y el tipo de miembro

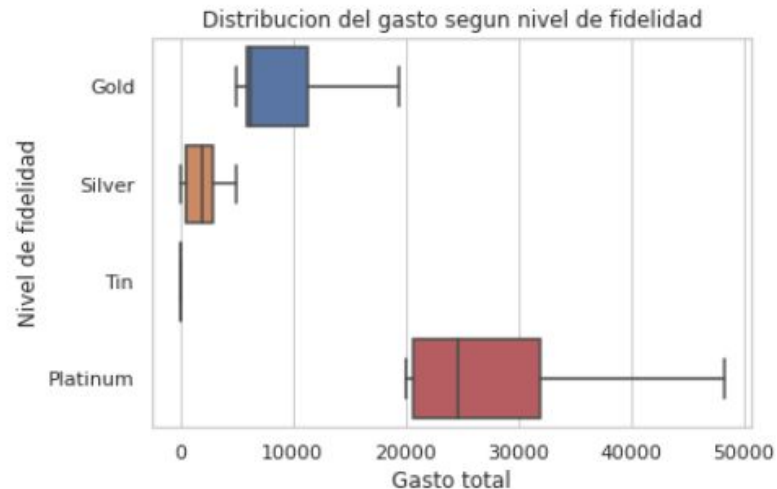




¿Cómo se **distribuye** el gasto de los clientes?

AFILIACIÓN

- Miembros **Platinum** evidencian un **mayor** nivel de **gasto** aunque también una **mayor variabilidad** del mismo;
- Miembros **Gold** y **Silver** muestran un **gasto** más **moderado y parejo**;
- Miembros **Tin** gastan **montos muy pequeños** en comparación a los demás segmentos;



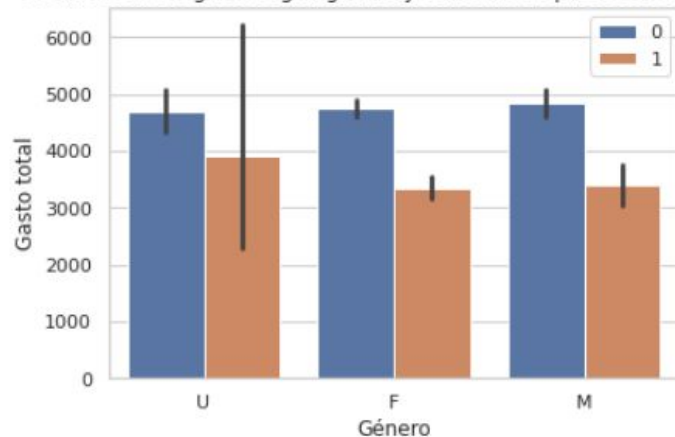


¿Cómo se **distribuye** el gasto de los clientes?

GÉNERO Y CONSUMO

- Quienes **consumen** productos orgánicos (1) **gastan menos** dinero que quienes no lo hacen (0);
- **No** se evidencian grandes **diferencias** de comportamiento en los distintos **géneros** analizados.

Distribucion del gasto segun género y consumo de productos orgánicos



Nota: 1 indica que el cliente compra productos orgánicos ; 0 indica que no los compra



APLICACIÓN DE ALGORITMOS DE ML



Primera etapa: K-Means

- ✓ Se corre inicialmente un modelo no supervisado, es decir, sin una variable objetivo / a predecir, con el fin de encontrar patrones en la información que permitan clasificar de algún modo a los clientes existentes en la base de datos.
- ✓ Se avanza con el K-Means y se divide la cartera en 3 grandes grupos: A, B y C

A	Nadie consume orgánicos	<ul style="list-style-type: none">• 46% de la cartera compuesta por mujeres• Miembros Tin y Silver
B	Todos consumen orgánicos	<ul style="list-style-type: none">• Más del 75% de la cartera compuesta por mujeres• Miembros Tin y Silver principalmente, seguidos por Gold
C	Consumo mixto	<ul style="list-style-type: none">• 48% de la cartera compuesta por mujeres• Miembros Gold mayormente, también Silver y Platinum



Segunda etapa: Random Forest

- ✓ Utilizando la clasificación obtenida con el K-Means como variable objetivo, se construye un modelo supervisado que permite clasificar nuevos clientes que se sumen a la cartera.
- ✓ Las variables más relevantes para clasificar nuevos clientes son:





CONCLUSIONES



Conclusiones

- ✓ Gracias al K-Means, es posible encontrar patrones que distinguen a los clientes de la cartera, los cuales permiten segmentarla de una manera eficiente y con una gran exactitud.
- ✓ Además, con el Random Forest es posible predecir a qué grupo pertenecerán los nuevos clientes que se sumen con un gran nivel de asertividad y precisión.
- ✓ Las variables utilizadas para definir cada grupo son fáciles de obtener y no representan mayores costos para el detallista.
- ✓ Con los resultados obtenidos podrían ejecutarse acciones de marketing enfocadas en las características de cada uno de los segmentos, que luego podrían traducirse en mayores ingresos para el negocio y mayor visibilidad de la empresa.



GRACIAS POR SU ATENCIÓN