
Predicciones alimenticias

El objetivo de esto es ayudar al distribuidor a comprender las propiedades de los productos y los puntos de venta que desempeñan un papel crucial en la predicción de las ventas.

Sebastián Macías Gallego

Proyecto final 1 - Profundización Data Science Coding Dojo 2022

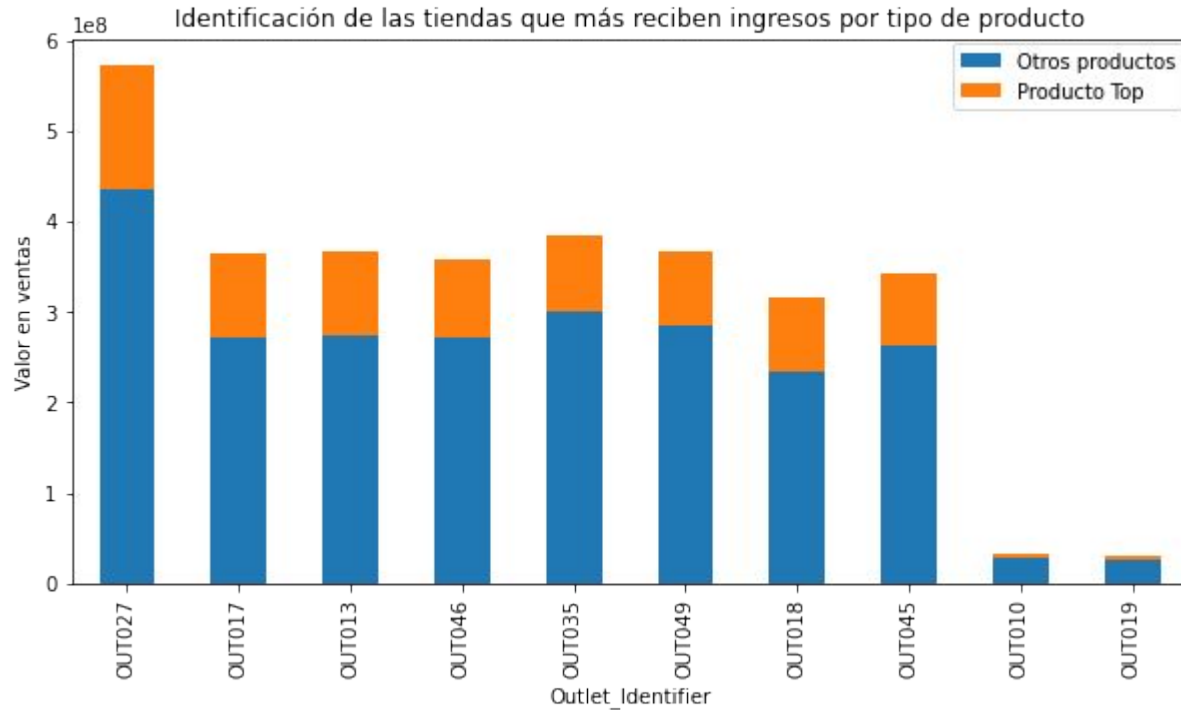
Aspectos clave a analizar

1. Cantidad de productos que generan el 80% de las ventas.
 2. Identificación de las características de las tiendas con mejor rendimiento en ventas y ganancias.
 3. Identificación de las características de los productos TOP, es decir, la menor cantidad de productos que soportan por sí solos el 20% de las ventas.
 4. Modelos predictivos.
-

GRÁFICA DE PARETO: Productos que generan el 80% de las ventas

- Del gráfico de pareto podemos identificar que del total de 1559 productos:
 - **149 productos generan el 20% de las ventas**, estos productos son foco para potenciar sus ventas, se espera que cada uno de estos productos tenga como mínimo 3800 ventas.
 - **772 productos generan el siguiente 60% de las ventas**, se espera que cada uno de estos productos tenga como mínimo 1680 ventas.
 - 638 productos que aportan el 20% de ingresos, podrían ser retirados.
-

Identificación de las características de las tiendas con mejor rendimiento en ventas y ganancia

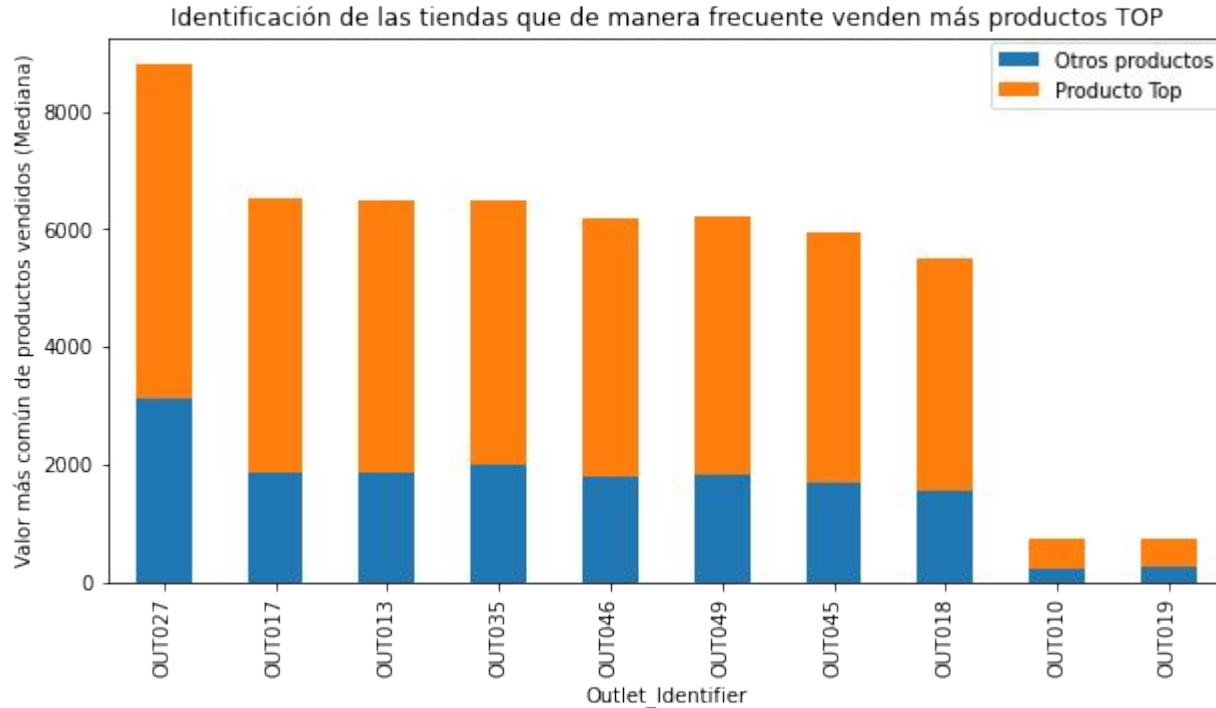


En orden descendente se presentan las tiendas que reciben más ingresos por la venta de productos TOP

En general sobresale la tiendas OUT027 por la venta superior de este tipo de productos.

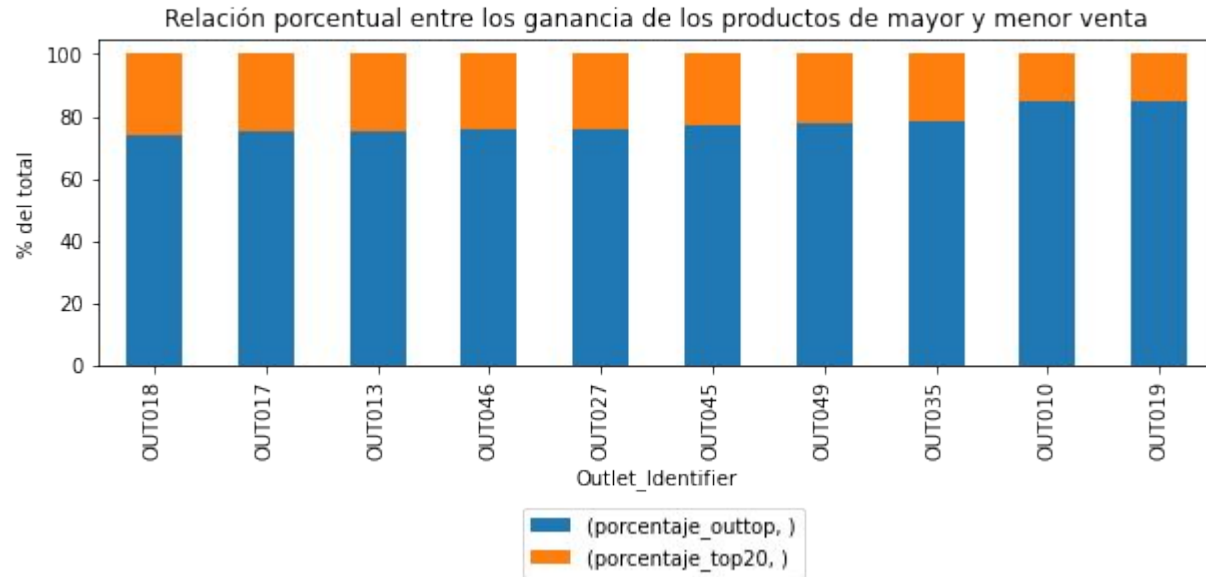
Con respecto al total se observa que los mayores ingresos totales están en las tiendas OUT: 027-035-049-013-017.

Identificación de las características de las tiendas con mejor rendimiento en ventas y ganancia



La gráfica presenta en orden descendente las tiendas según su mediana en ventas, es decir, podemos indicar cuál tienda tendría la posibilidad de vender mayor cantidad de productos, estas son OUT: 027-017-013-035-046.

Identificación de las características de las tiendas con mejor rendimiento en ventas y ganancia

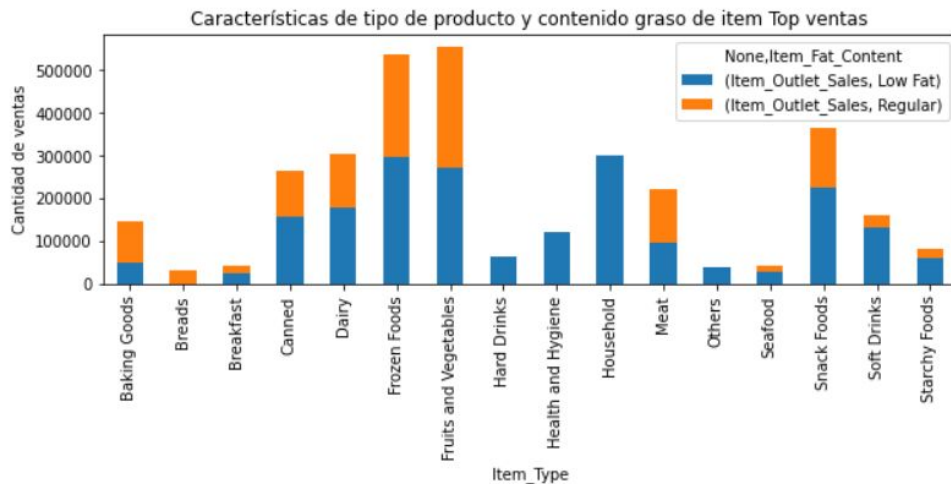


La gráfica nos muestra las tiendas que reciben mayor ganancia por la venta de productos TOP, dentro de las que más reciben están OUT: 018-017-013-046-027

Identificación de las características de las tiendas con mejor rendimiento en ventas y ganancia

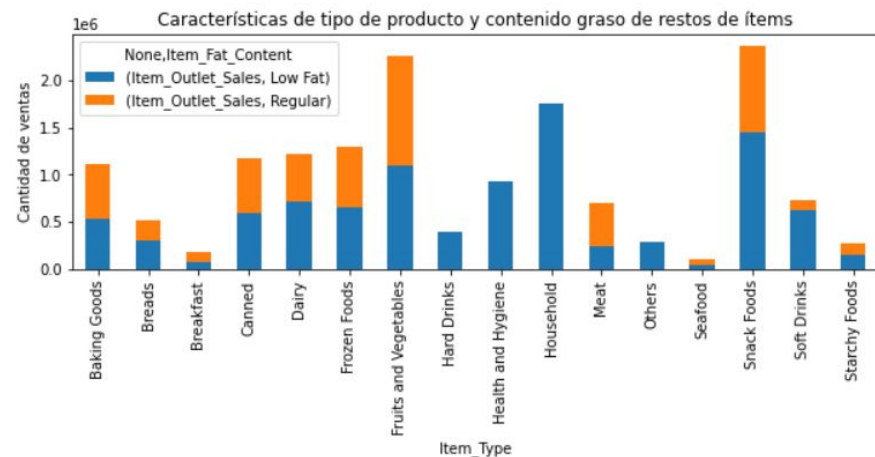
- De manera repetitiva en los análisis de ingresos y ventas se encuentran presentes en las siguientes tiendas 3 tiendas: OUT027 - OUT013 - OUT017 .
 - Las características de las tiendas donde más ventas e ingresos generan son:
 - El 67% son de tamaño **mediano** y un 33% de tamaño grande.
 - El 67% están en una locación **TIER 3** y un 33% se ubican en una locación TIER 2.
 - El 67% son tiendas **supermercado tipo 1** y el 33% restante es **supermercado tipo 3**.
 - La tienda OUT027 qué es la que mejor desempeño presenta en cantidad de ventas es de tamaño medio, en una locación tier 3 y es un supermercado tipo 3.
-

Identificación de las características de los productos TOP



Características de los 149 productos más vendidos:

- Son productos congelados , las frutas y los vegetales.
- Los snacks están más presentes en el resto de productos.
- La compra de estos productos no está sujeta al contenido graso.
- La comida casera está presente en menor cantidad.



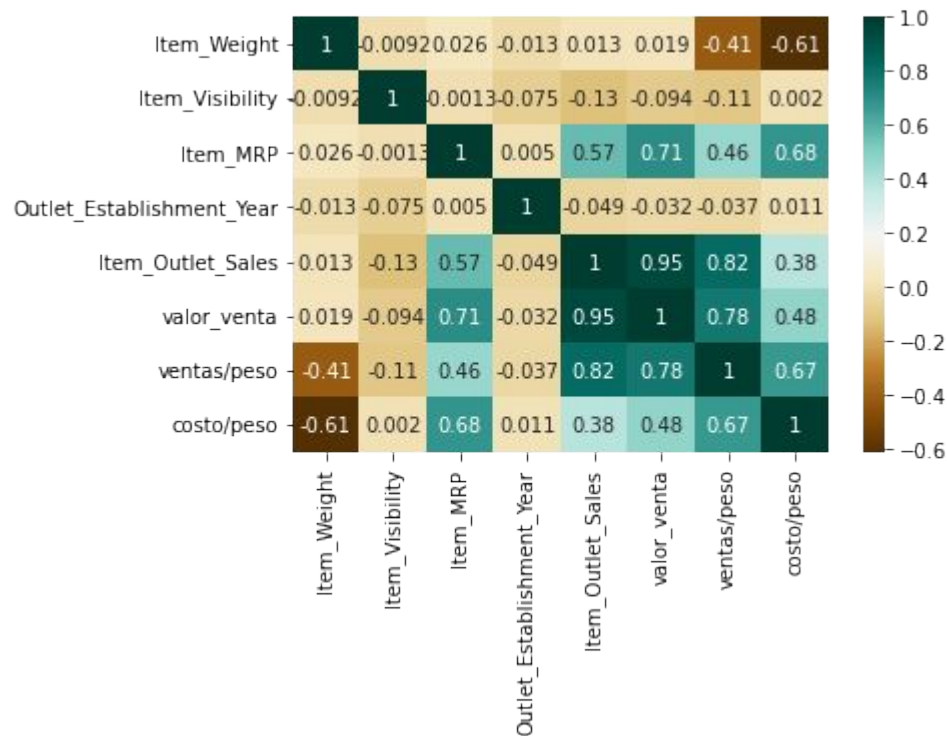
Características del resto de productos:

- Alta presencia de snacks.
- Las frutas y verduras aún son relevantes.
- La comida casera es mucho más relevante.
- Los desayunos están más presentes.

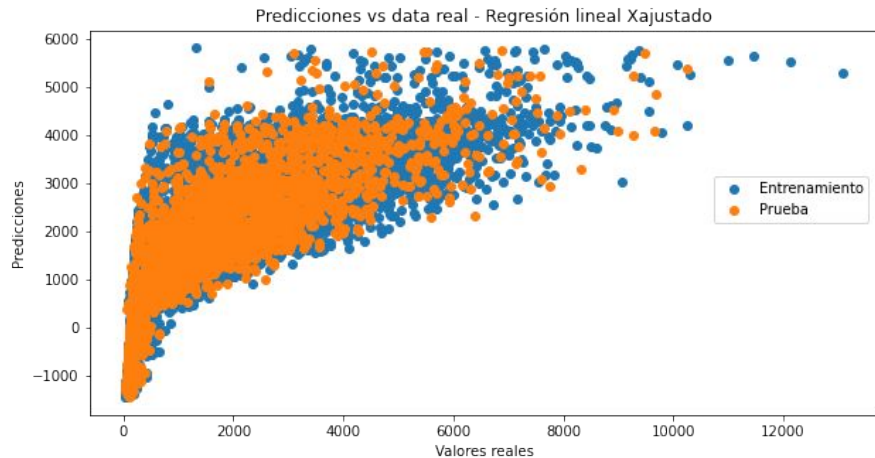
Identificación de las características de los productos TOP

A partir de los datos presentados se observa que la venta de productos por tienda no presenta una correlación fuerte con variables tales como el peso, la visibilidad, la relación más alta la tiene con el valor del producto; por tanto, al menos bajo la perspectiva de la correlación no es posible identificar otras variables de peso para la predicción de las ventas.

Dado lo anterior se recurre a los modelos predictivos.

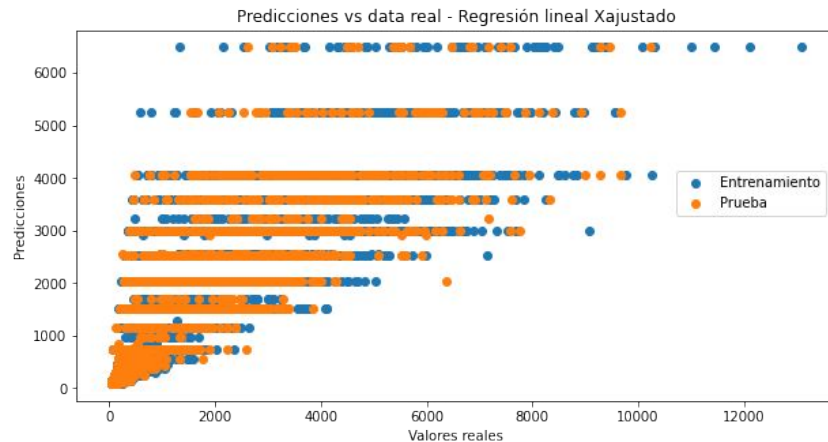


Modelos predictivos



Regresión lineal:

- Presenta inconvenientes con variables que generan ruido al modelo como los nombres de los productos.
- **R2:**
 - Entrenamiento: 0.56234 // Prueba: 0.56695
- **RMSE:**
 - Entrenamiento: ≈ 1138 unidades // Prueba: ≈ 1093 unidades



Un árbol de decisiones optimizado:

- Es mucho más estable ante la presencia de variables que generan ruido como los nombres de los productos.
- **R2:**
 - Entrenamiento: 0.60374 // Prueba: 0.59432
- **RMSE:**
 - Entrenamiento: ≈ 1083 unidades // Prueba: ≈ 1058 unidades.

CONCLUSIONES:

- Se recomienda centrar la atención en los 921 productos que generan el 80% de ingresos.
 - Es aconsejable que las tiendas nuevas sean del tipo medio, localización tier 3 y podrían estar en supermercados tipo 1 o 3.
 - El tipo de producto que más se vende son las frutas, verduras, congelados y snacks, aún así dentro del top de productos no hay gran diferencia en la venta de productos con alto o bajo contenido graso.
 - Los modelos de predicción ambos presentan un mal comportamiento en el conjunto de pruebas y entrenamiento, ello quiere decir que es probable que los datos no sean los adecuados o las suposiciones realizadas sobre los mismos se pueden mejorar mucho más, a pesar de ello los valores imputados del peso y el tamaño de la tienda al parecer no tienen una gran incidencia en la proyección de ventas.
-