

Proceso de Análisis de Datos – Cornershop

Introducción

En el presente documento, se explicará las principales conclusiones tras haber analizado las bases de datos (Order_Product, Orders, Shoppers, Storebranch).

Preguntas

- Calcular la cantidad de pedidos por día de la semana, diferenciando si los pedidos son on_demand.
- Calcular la cantidad promedio de productos distintos que tiene cada pedido agrupado por tienda (store).
- Calcular el found rate promedio de los pedidos agrupados por formato de producto y día de la semana.
- Calcular el error promedio y error cuadrático medio de nuestro modelo de estimación para cada hora del día.
- Calcular la cantidad de pedidos donde el picker_id y el driver_id son distintos.

Análisis

- La cantidad promedio de pedidos que se tiene al día es de 1042. Y se puede identificar que los pedidos que son “on demand” representan un 43% en promedio del total de todos los pedidos que se tienen en un día. Además, con la data de una semana que se tiene, se puede identificar que los pedidos disminuyen a lo largo de la semana. Es decir, suelen hacerse mayor cantidad de pedidos durante los primeros días y menos los fines de semana. Sin embargo, los pedidos “on-demand” aumentan los fines de semana.
- Se pudo identificar una tienda que cuenta con 26 productos diferentes en promedio por orden. Este es el máximo dentro de los resultados. Se puede inferir que esta tienda es un Supermercado ya que las personas suelen realizar sus compras semanales en este tipo de establecimientos.
- Un insight interesante que se pudo observar es que en todos los casos los usuarios solo escogen una tienda por orden. Y al revisar el aplicativo, se observó que al escoger diversas tiendas el precio se eleva en gran medida.
- Tras analizar la data en “Order Product”, pude identificar que había 14415 casos en el que el “quantity _found” en el formato de unidades era 0. Esto se debe a que los repartidores en ocasiones no pueden encontrar los productos en las tiendas (supermercados). Esto hace a que el found rate sea de 0.85. Pero si excluimos, estos casos del análisis, este llega a ser de 1 todos los días de la semana.
- El error promedio del modelo de estimación es de 42 minutos aproximadamente.
- Por último, existen 216 pedidos donde el picker_id y el driver_id son distintos.