

Desafío Técnico

Una cadena de supermercados opera un servicio de delivery ofreciendo sus productos a los clientes en distintas zonas de la ciudad. En el proceso de asignación del driver o flota se han identificado ineficiencias por falta de información respecto al peso y volumetría de los productos.

- Algunas ocasiones se sobreestima el tamaño y peso, asignando un driver grande para delivery pequeño
- En otras ocasiones se subestima, asignando un driver que no tiene la capacidad real para la entrega del pedido.

Lo anterior genera costos operativos elevados, rechazos de pedidos y mala experiencia en el servicio de entregas.

La ineficiencia en la correcta asignación del driver se debe a la inconsistencia en los datos, falta de información o de homogeneidad del catálogo.

Por tal razón, la cadena de supermercado te ha contratado a ti, un Data Scientist experimentado para que le ayudes a diseñar un proceso que le permita calcular para cada producto del catálogo el peso y la volumetría(alto, ancho, largo) del mismo.

Para dicho propósito se te ha proporcionado un dataset con 900 registros que contiene toda la información relacionada con el catálogo de productos. Tu función es diseñar un proceso reproducible que dado un input similar permita asignar las columnas de “peso” y “volumetría”(alto, ancho, largo) a cada ítem.

Entregables **mínimos** esperados:

1. Un repositorio con la estructura del proceso desarrollado e instrucciones para ejecutarlo.
2. Script/CLI que permita dockerizarse y por input reciba un .csv similar al compartido y de output esperado el mismo input expandido con las columnas de interés “peso” y “volumetría” como mínimo.
3. Un documento con la explicación del desarrollo del proceso, suposiciones/asunciones realizadas y explicación de los distintos métodos utilizados para la solución del problema. Opcional (Puedes entregar un EDA en un notebook).

Consejos

- Tienes libertad de, ante falta de información, hacer las suposiciones necesarias o usar las herramientas que creas convenientes para hacer la estimación (IA, Machine Learning, heurísticas, etc.)

Los datos para el test: **Dataset Prueba Técnica - market-products.csv**