

Challenge Técnico - Data Scientist - YOM

El desafío consta del desarrollo de un sistema de recomendación para la empresa Mayorista11 con el fin de optimizar las recomendaciones para comercios locales. Para esto contamos con 3 archivos de datos en formato CSV, los cuales se encuentran en la ruta '/content/data' y contienen:

1. Historial de transacciones de cada comercio
2. Información de los comercios
3. Información de productos

El trabajo desarrollado se abordó en 3 secciones, empezando por la exploración de la data, con el fin de entender su contenido, características estadísticas, calidad del dato y procesamiento de nulos, que para este caso no fue necesario.

En la segunda sección se aborda la división de los comercios por comuna, con el objetivo de cumplir con el enunciado, ya que las recomendaciones deben ser acorde a la comuna a la cual pertenece el comercio, y crear listados de los items mas vendidos a nivel de dinero y a nivel de unidades.

En la tercera sección se propone y desarrolla el sistema de recomendación. Si bien el enunciado habla de recomendación de productos de alta popularidad, la estrategia propuesta es recomendar productos que generen alto impacto en ventas. Asumimos que todos los items marginan porcentualmente igual, por lo que un item que tenga mayor volumen de ventas en dinero, impactaría positivamente tanto al comercio como a Mayorista11.

Sin embargo, es importante no descuidar los productos que tienen grandes cantidades de unidades vendidas, ya que puede que la estrategia del comercio no esté alineada con el primer criterio propuesto y puede que, por ejemplo, busquen una alta rotación del inventario.

Dicho lo anterior, se crean dos listas de recomendaciones. La principal, que se entrega los items que mas impactan en volumen de dinero, y luego una segunda lista que considera items que venden gran cantidad de unidades.

En ambos casos se exige que criterio de que sean items que el comercio no esté comprando/ vendiendo, ya que no tendría sentido la recomendación. Adicionalmente, se exige que la segunda lista no repita elementos de la primera lista, para que no haya duplicidad de recomendación.

Para probar el recomendador, basta con modificar la última celda, y darle el valor numérico correspondiente al id_commerce a la variable 'id_comercio'. De todas formas, la línea que se debe modificar tiene el comentario '#CAMBIAR ID POR EL QUE NECESITES CONSULTAR'.

Alejandro Espinoza
23 de septiembre, 2024