

Primer entregable – Definición del Proyecto

Descripción clara del caso de negocio

En este proyecto me voy a poner en el rol de analista de Sephora, con el objetivo de ayudar en la toma de decisiones estratégicas relacionadas con el lanzamiento de un nuevo producto de maquillaje bajo su propia marca.

La idea es analizar en profundidad los productos de maquillaje que actualmente están disponibles en su tienda online, entendiendo qué tipos de productos hay, cómo son, cómo se valoran, que características tienen, etc. A partir de ahí, la finalidad es identificar oportunidades de mercado que Sephora podría aprovechar para lanzar un nuevo producto exitoso: uno que encaje bien con las tendencias actuales, que cubra algún hueco en su catálogo a poder ser y que tenga características que históricamente han estado asociadas a productos bien valorados.

El enfoque combina datos reales de una empresa real (Sephora), con un propósito claro que tiene sentido en el contexto de negocio: ayudar a lanzar un nuevo producto con mayor probabilidad de éxito.

Objetivos del proyecto y su impacto esperado

El objetivo principal sería definir, basándonos en datos reales, cuál podría ser el próximo producto de maquillaje que Sephora debería lanzar bajo su propia marca.

Objetivos más secundarios podrían ser:

- Crear un scraper que extraiga la información de los productos de maquillaje de la web de Sephora de manera automática.
- Montar una base de datos relacional funcional en PostgreSQL para almacenar esa información de forma estructurada.
- Automatizar el proceso de actualización de la base de datos, de forma que se mantenga al día con nuevos productos y cambios en los ya existentes.
- Mantener un histórico de datos, especialmente de precios y valoraciones, que permita enriquecer el análisis y detectar tendencias en el tiempo.
- Realizar un análisis exploratorio detallado con Python para entender el mercado y detectar patrones relevantes.
- Identificar características de productos que suelen estar asociadas a buenas valoraciones, así como las categorías y subcategorías de maquillaje donde Sephora aún no tenga tanta presencia, evaluar rangos de precios competitivos por tipo de producto, etc.
- Desarrollar un dashboard interactivo en Power BI para presentar los hallazgos de forma clara y visual.

Tecnologías y herramientas que se utilizarán

- Python: para hacer el scraping, procesar los datos, analizarlos y crear la base de datos.
- Selenium: para automatizar la extracción de productos desde la web de Sephora.
- Pandas / NumPy: para manejar y analizar los datos.
- PostgreSQL (usando DBeaver): para estructurar y guardar los datos de forma relacional.
- Power BI: para crear un dashboard visual con los resultados y la recomendación final.

Identificación y justificación de las fuentes de datos

En este caso no ha sido necesario buscar datasets externos porque la fuente de datos es clara: la propia página web de Sephora. Desde allí se puede acceder a la información de todos los productos de maquillaje, incluyendo datos básicos como nombre, marca o precio, y otros más específicos como el formato del producto, la cobertura, el tipo de piel para el que está recomendado, etc.

Gracias a que la web es accesible y estructurada, se puede scrapear sin problemas y se puede automatizar para mantener la base de datos actualizada cada semana. Esto además permite tener un histórico de cambios que resulta clave para detectar variaciones de precio, evolución de las valoraciones o aparición de nuevos productos, lo cual enriquece mucho más el análisis final.