

MEMORIA DEL ANÁLISIS EXPLORATORIO (THE BRIDGE)

ASPECTOS MÁS COMPLICADOS DEL EDA:

- Realizar algunas tareas de Data Cleaning fue un tanto complejo. Sin embargo, el primer dataset que limpié fue el de Apple, y todo salió bastante rodado, por lo que pude aprovechar muchas de las funcionalidades que utilicé para luego limpiar el de Google, que aunque fue un poco distinto, lo saqué sin demasiados problemas.
- Paradójicamente, la mayor parte del tiempo la he dedicado a hacer el Análisis propiamente dicho y no al Data Cleaning, ya que quise analizar en profundidad los datos, y tuve que tener en cuenta más aspectos de los originalmente creía. Estuve varios días con un Notebook paralelo 'en sucio', haciendo cálculos y gráficos hasta que todo quedase perfecto —o parecido a lo que yo quería— .
- Otra cosa que me costó un poco fue adaptarme a Plotly, ya que es una librería de visualización que nunca había utilizado anteriormente, y tuve que estar permanentemente mirando la documentación para saber cómo configurar los distintos gráficos en función de mis necesidades. Aunque he de decir que el resultado final me ha dejado bastante satisfecho, y seguiré utilizando Plotly en el futuro.
- Por último, destacar que nunca había realizado una presentación por diapositivas, por lo que espero que entendáis que no será una presentación muy profesional. Cuando vimos Streamlit empecé a desarrollar una presentación, pero me di cuenta de que (quizás por desconocimiento por mi parte) no se ajustaba tanto a una presentación propiamente dicha, sino que está más enfocado a realizar proyectos interactivos, por lo que decidí hacerla con Google Presentaciones, y también me llevó unas cuantas horas.

TÉCNICAS QUE HE UTILIZADO:

Como habéis podido comprobar, he utilizado multitud de funcionalidades de todo tipo (Pandas, NumPy, estadística, matemáticas, y Python puro).

Me gusta profundizar mucho en los datos, limpiarlos muy a fondo, y he intentado ser lo más objetivo y riguroso posible.

- DATA CLEANING:

Mucho filtrado con Pandas sobre todo, mapas de calor y de correlación, manipulación de datos de todo tipo.

- VISUALIZACIÓN Y ANÁLISIS

Tuve que abstraerme bastante para poder mostrar resultados que realmente me aportasen valor, ya que no me servía de nada analizar directamente los datos que tenía. Tuve que realizar múltiples agrupaciones, sacar muchos porcentajes e intentar relacionarlos con otros valores para obtener conclusiones serias.

Respecto a los gráficos, que al final es lo que representa los resultados (por lo que es muy importante), quería enfocarme en la sencillez para el espectador. No quería gráficos complejos, por muy bonitos que sean, porque muchas veces no son los que mejor se entienden. Quería crear gráficos muy sencillos de entender para cualquier persona, tenga conocimientos técnicos o no, y por eso he tratado de hacer un buen uso de los colores y no hacerlos pesados de visualizar. Quería que en un vistazo de cinco segundos cualquier persona pudiese asimilar la información que quiero transmitir. Por eso he utilizado mayoritariamente gráficos de barras (creo que son muy sencillos de comprender) y he querido representar Google con el color verde de Android, y Apple con el color azul de la App Store.

FUENTES QUE HE UTILIZADO:

A lo largo del Proyecto voy nombrando todas las fuentes que utilizo, incluso las imágenes de los Notebooks, pero las voy a resumir aquí a grandes rasgos:

- Datasets de statcounter
- Datasets de Kaggle
- Documentación de todas las librerías
- Apuntes y proyectos propios
- Notebooks del bootcamp
- stackoverflow

CONCLUSIONES DEL EDA:

He invertido muchas más horas de las que esperaba en el EDA, pero me ha encantado todo el proceso de búsqueda de datos, limpieza y acondicionamiento, y el posterior análisis.

Una de las cosas que más me satisfecho me ha dejado ha sido el poder comparar los valores reales con los valores esperados en los ratings de las Apps, es algo que saqué estrujándome el cerebro y queriendo hacer un análisis riguroso de los datos, y cuando hice los cálculos por separado para comprobar que estaba en lo cierto, quedé muy contento. Además, los gráficos que he realizado también me han gustado mucho, y he aprendido mucho Plotly, que es algo que quería hacer.