Student: Isabella Mesquita

Notes on the "Metrics Analysis for Sales on Pinterest" Module 1 Project

1. Dataset Selection

The dataset came from a few sources. From kaggle, created by Oneli Wickramasinghe and last updated 9 months ago. I chose this data because: “This curated dataset encapsulates the essence of popularity and engagement on Pinterest, focusing on pins from the top five most followed accounts on the platform as of 2024. It provides an insightful glimpse into what captivates the audience on one of the world's leading social media platforms, highlighting the content that resonates most with its vast user base”; this helped me find some answers to my questions, like: What types of content (images, videos, infographics) generate the most engagement?.

Other resources were from Google searches, the resources that Pinterest provides and a survey in which I asked 8 questions to 20 people who use Pinterest to sell.

2. Data Cleaning Process

I used Python for data cleaning. First, I imported necessary libraries such as pandas, numpy and others. I then read the file and started a preliminary exploration. I noticed that there were spaces in the column names, so I cleaned this up by removing the spaces and sorting the data for better readability.

3. Data Exploration and Visualization

I performed some exploratory data analysis (EDA), using matplotlib and seaborn to create a bar plot to visualize the average engagement by content type, Calculate the correlation between title length and repin count and others.

4. Questions

These were the questions used for the research:

- What types of content (images, videos, infographics) generate the most engagement?
- Do keywords really have a big impact on sales?
- How do trends influence the performance of the type of content posted?

5. Conclusion

This project provided a comprehensive view of how EDA and Python can be combined to generate business insights effectively, using real-world data. With all the datasets I was able to answer the questions and put them in a slide presentation.