Internship Report by Sai Sreeja

For “Build Real time Google Play store data analytics - python”

Introduction

To obtain meaningful research during my internship, app reviews and app information from the Google Play Store were examined. The aim was to analyse user opinion, rating and trends to support stakeholders in improving app efficiency and customer satisfaction. The project involved using Python, Pandas, natural language processing and interactive charts using Plotly to develop the research.

Background

The project required examining a dataset containing app attributes, user reviews and sentiment scores. The project's overall goal is to use data organisation, cleaning, and visualisation to show a broad picture of app popularity, sentiment patterns, and engagement levels. In the end, the project helps app developers, business analysts, and marketers better analyse user behaviour and improve their products.

Learning Objectives

- Perform data cleaning and transformation on raw datasets.
- Implement sentiment analysis using NLP techniques.
- Create interactive visualizations with Plotly.
- Identify trends in app ratings, reviews, and installations.
- Develop insights to improve app performance and user engagement.

Activities and Tasks

1. Data Cleaning and Transformation

- Removed missing values and duplicates from the dataset.
- Standardized numerical and categorical data fields.
- Converted data types and applied transformations for better analysis.

2. Sentiment Analysis on Reviews

- Applied Natural Language Processing (NLP) to analyze user sentiments.
- Classified reviews into positive, neutral, and negative categories using VADER sentiment analysis.

3. Data Visualization Using Plotly

1. Created 3 interactive charts to represent key insights:
    1.1. **Sentiment Distribution:** Stacked Bar Chart
    1.2. **App Rating Distribution :** Dual Axis Bar Chart
    1.3. **Review Trend Over Time:**Bubble Chart

Skills and Competencies

- **Data Wrangling:** Cleaning and transforming raw datasets using Pandas.
- **Sentiment Analysis:** Implemented NLP techniques for user sentiment evaluation.
- **Data Visualization:** Developed interactive dashboards using Plotly.
- **Statistical Analysis:** Explored relationships between app features and performance.

Feedback and Evidence

- The interactive dashboard provided stakeholders with meaningful insights into app performance.
- Developers could use sentiment trends to identify and address user concerns.
- Businesses leveraged installation and revenue trends for strategic planning.

Challenges and Solutions

1. Handling Large Datasets

- **Challenge:** Processing a large dataset with missing values and inconsistencies.
- **Solution:** Applied data preprocessing techniques such as imputation, filtering, and conversion.

2. Accurate Sentiment Classification

- **Challenge:** Some reviews contained ambiguous sentiments.
- **Solution:** Used NLP-based sentiment scoring for better classification.

3. Creating Meaningful Visualizations

- **Challenge:** Ensuring that insights are clearly represented.

- **Solution:** Used interactive visualizations and filtering options.

Results and Consequences

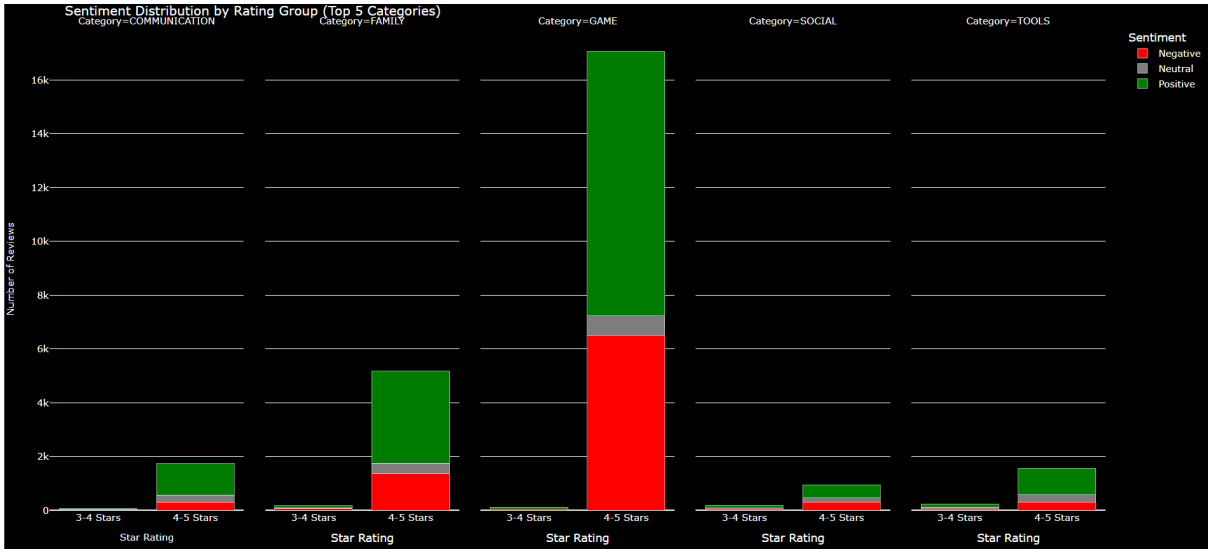

fig1

fig1 - Sentiment Distribution by Rating Group (Top 5 Categories):

- The "Game" category has the highest number of reviews, with a significant portion being positive. However, it also has a notable share of negative reviews, indicating mixed user experiences.
- Other categories like "Communication" and "Tools" have fewer reviews but maintain a more balanced sentiment distribution, with positive reviews still being dominant.

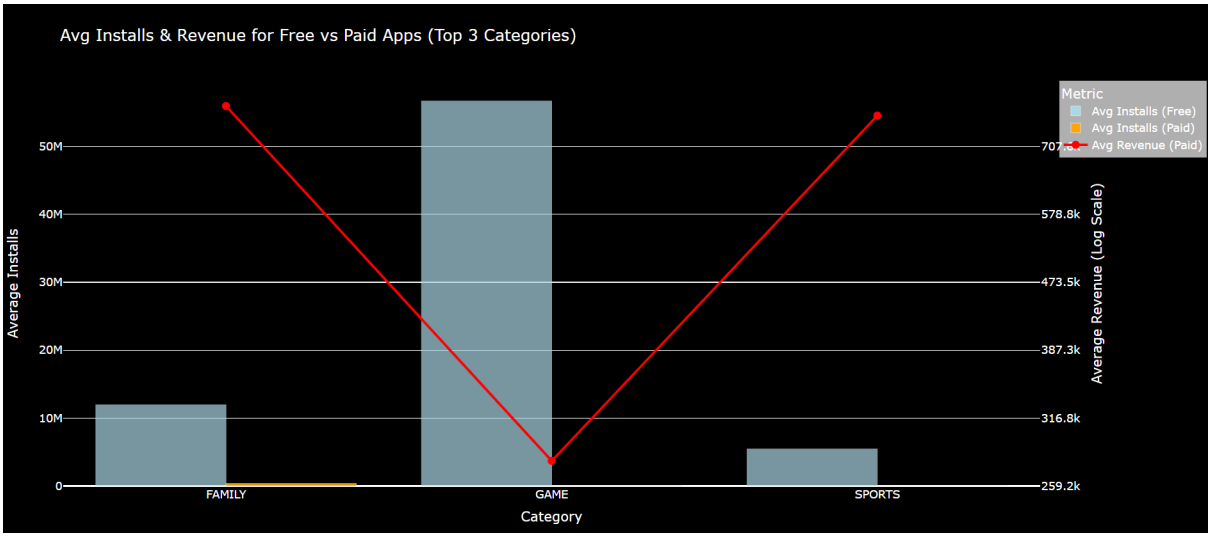

fig2.1

fig2.1 - Average Installs & Revenue for Free vs Paid Apps (Top 3 Categories)

- Free apps in the "Family" category receive the highest number of installs, significantly surpassing the other two categories. This suggests a high demand for family-oriented applications.
- Paid apps, however, generate more revenue despite having lower installs, indicating that monetization strategies for paid apps may be more effective in certain categories.

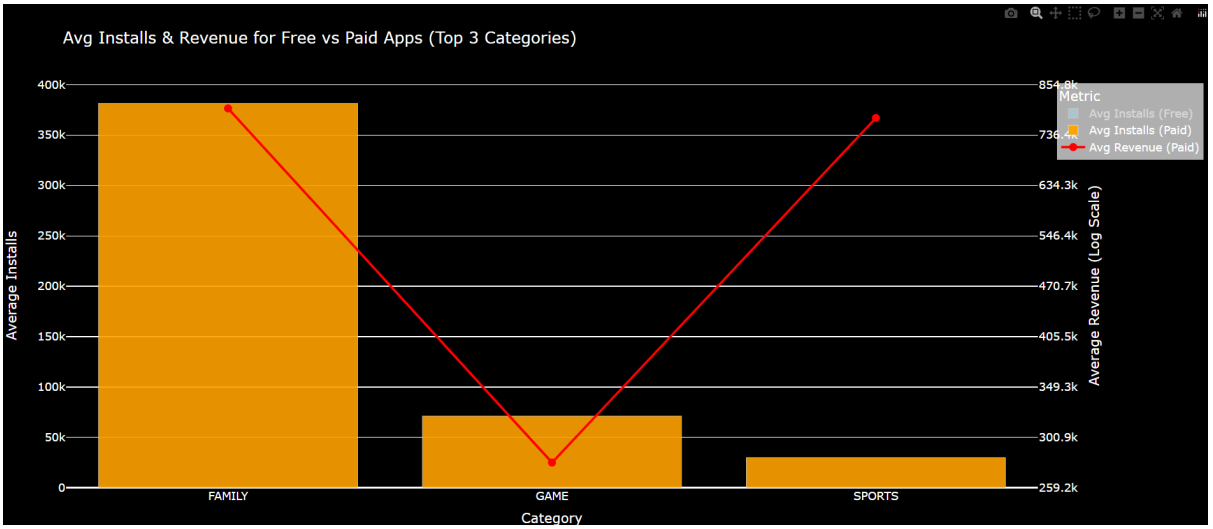

fig2.2

fig2.2 - Average Installs & Revenue for Free vs Paid Apps (Log Scale):

- When viewed on a log scale, the disparity between installs and revenue becomes more evident. "Game" apps have the highest number of installs, but their revenue is significantly lower compared to some paid applications.
- The "Sports" category shows the least engagement in terms of installs but generates a steady revenue stream, possibly due to niche user interest.

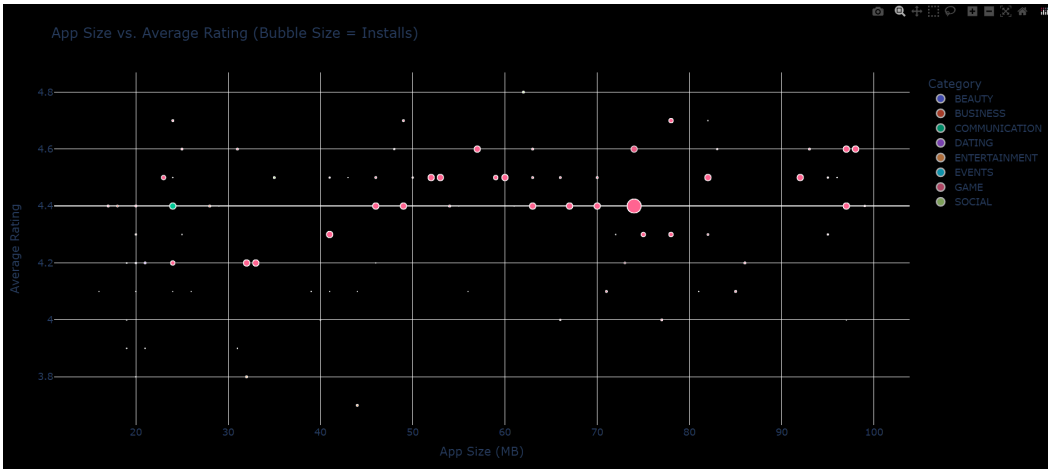

fig3

fig3 - App Size vs. Average Rating (Bubble Chart):

- The chart shows that smaller apps tend to have higher ratings, suggesting that users may prefer lightweight applications that do not take up too much storage.
- However, there are exceptions where larger apps also maintain high ratings, possibly due to better functionality and user satisfaction.

Consequences

1. Handling Large Datasets

- **Challenge:** Processing a large dataset with missing values and inconsistencies.
- **Solution:** Applied data preprocessing techniques such as imputation, filtering, and conversion.

2. Accurate Sentiment Classification

- **Challenge:** Some reviews contained ambiguous sentiments.
- **Solution:** Used NLP-based sentiment scoring for better classification.

3. Creating Meaningful Visualizations

- **Challenge:** Ensuring that insights are clearly represented.
- **Solution:** Used interactive visualizations and filtering options.

Outcomes and Impact

- **Improved App Performance Understanding:** Developers could analyze ratings and user feedback efficiently.

- **Business Insights for Stakeholders:** Helped businesses optimize app pricing, categories, and user experience.
- **Sentiment Analysis for Better Customer Engagement:** Provided an understanding of user concerns and satisfaction levels.
- **Data-Driven Decision Making:** Enabled informed decisions based on trends in app installs, reviews, and revenue.

Conclusion

This internship project provided hands-on experience in data analytics, NLP, and visualization techniques. The insights generated from Google Play Store data enabled stakeholders to improve app performance, customer satisfaction, and business strategies. By working with large datasets and implementing sentiment analysis, I strengthened my skills in data science and business intelligence, preparing for future analytical roles.

Submitted by-

Sai Sreeja Marupaka (sreejamarupaka444@gmail.com)