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 10 interactive charts to represent key insights:
    1.1. **App Ratings Distribution:** Histogram of app ratings.
    1.2. **Sentiment Score Distribution:** Visualization of positive, neutral, and negative reviews.
    1.3. **Category-wise Average Rating:** Bar chart showing app performance by category.
    1.4. **App Installs vs. Ratings:** Scatter plot to observe relationships.
    1.5. **Most Reviewed Apps:** Bar chart ranking apps by review count.
    1.6. **Price vs. Rating:** Relationship between paid apps and their ratings.
    1.7. **Word Cloud of Most Common Review Terms.**
    1.8. **Trending App Categories:** Analysis of the most installed app categories.
    1.9. **Revenue Estimation:** Created a new revenue column for analysis.
    1.10. **Pie Chart of Sentiment Proportions.**

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.

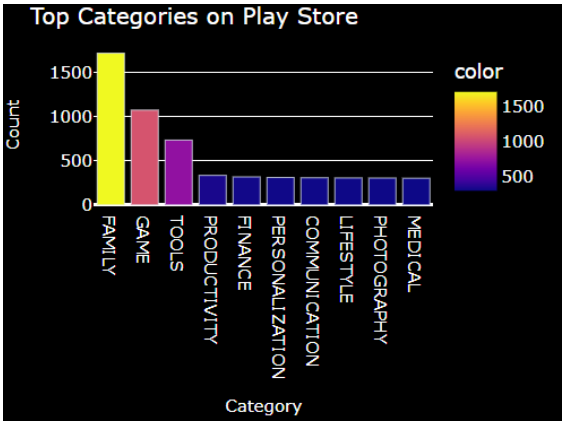
2. Accurate Sentiment Classification

- **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



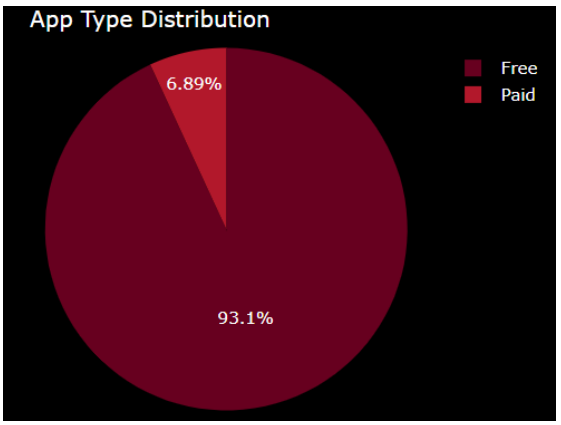


fig 1 **fig 2**

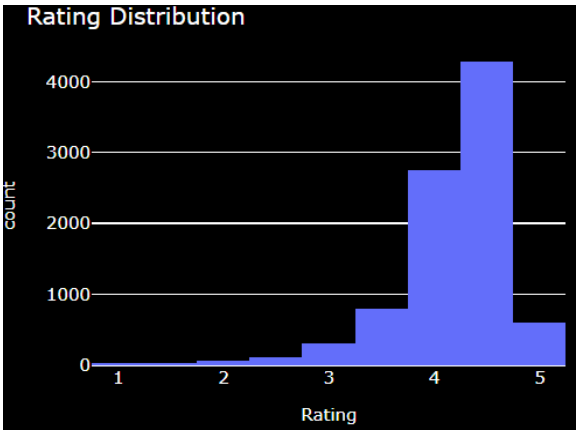
fig1 - Top Categories on Play Store:

- The "Family" category has the highest number of apps, followed by "Games" and "Tools".
- This indicates a strong market focus on entertainment and utility-based applications.

fig2 - App Type Distribution:

- A majority (93.1%) of apps on the Play Store are free, while only 6.89% are paid.

- This suggests that most developers rely on advertisements or in-app purchases for revenue.



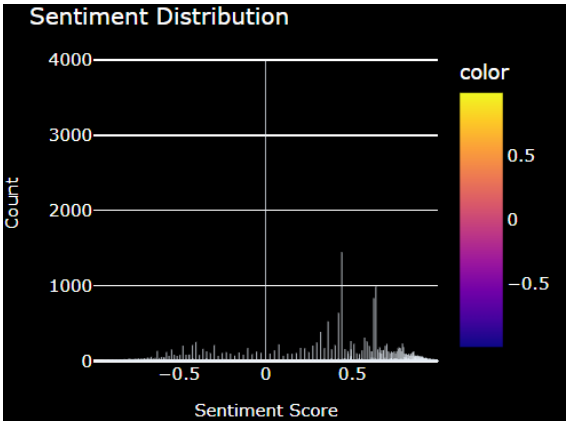


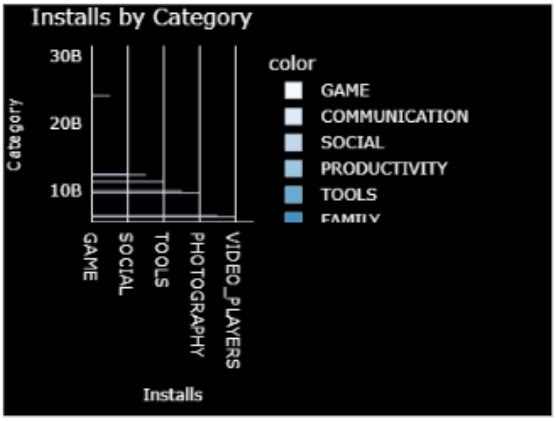
fig3 **fig4**

fig3 - Rating Distribution:

- Most apps receive ratings between 4 and 5, indicating general user satisfaction.
- Very few apps fall below a rating of 3, suggesting a quality threshold for app sustainability.

fig4 - Sentiment Distribution:

- The majority of reviews have positive sentiment scores, with a skew towards positive experiences.
- Some negative spikes indicate areas where improvements may be needed.



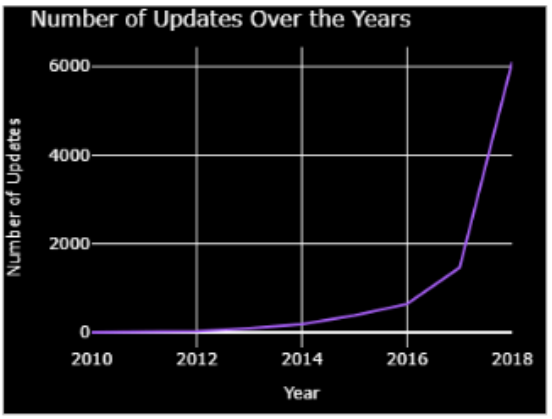


fig5 **fig6**

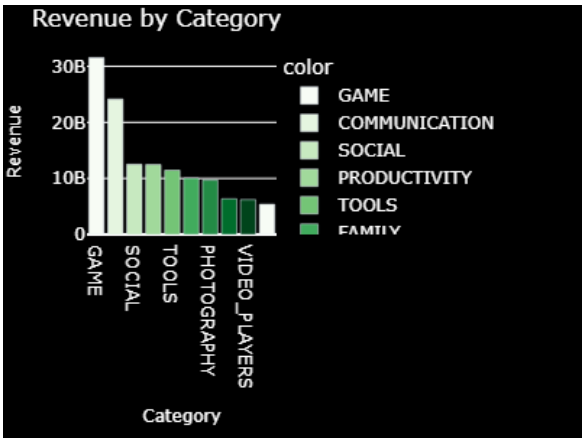
fig5 - Installs by Category:

- "Games" lead in installations, followed by "Social" and "Tools", highlighting user preference.

- Categories like "Photography" and "Video Players" also have significant user engagement.

fig6 - Number of Updates Over the Years:

- The number of app updates has increased significantly after 2016.
- This could indicate an industry trend of frequent app maintenance and feature enhancements.



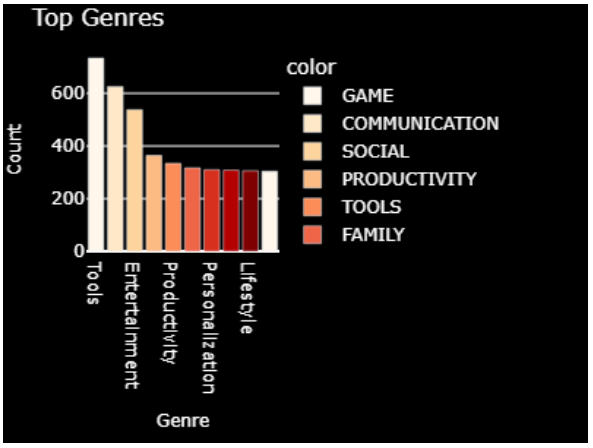


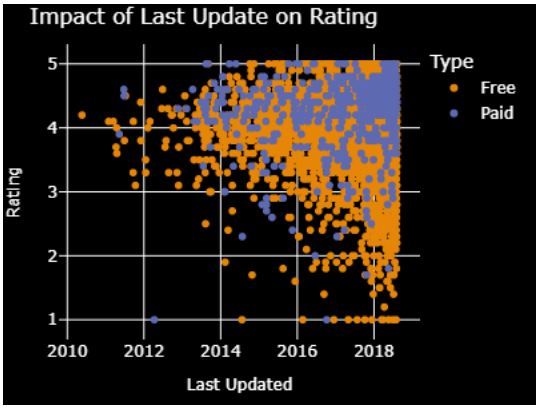
fig7 **fig8**

fig7 - Revenue by Category:

- "Games" generate the highest revenue, followed by "Social" and "Tools" categories.
- This aligns with installation trends, reinforcing the profitability of entertainment apps.

fig8 - Top Genres:

- "Productivity", "Tools", and "Family" are the most popular genres in the store.
- This suggests a balanced demand for both entertainment and practical applications.



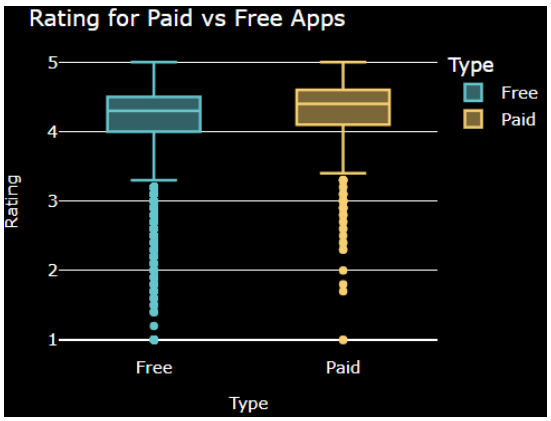


fig9 **fig10**

fig9 - Impact of Last Update on Rating:

- Apps updated recently tend to have higher ratings, indicating the importance of regular updates.
- Free apps show more variation in ratings compared to paid apps.

fig10 - Rating for Paid vs. Free Apps:

- Paid apps generally receive higher ratings than free apps.
- This could indicate better quality control and user expectations for paid services.

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)