Internship Report

Advanced Google Play Store Data Analytics - Python

Name: Vijayalakshmi
Internship Organization: NullClass
Date: 06.01.2025 – 06.02.2025

---

1. Introduction

This project focuses on advanced data analytics of Google Play Store apps using Python. The goal is to extract meaningful insights from user reviews, app ratings, and installs through visualizations. The analysis involves filtering and segmenting data based on various criteria to understand user behavior and app performance.

The internship involved three major tasks:

1. **Visualizing sentiment distribution** of user reviews using a stacked bar chart.

2. **Comparing app category performance** through a grouped bar chart.

3. **Analyzing app size vs. rating trends** using a bubble chart.

---

2. Task Descriptions

Task 1: Sentiment Distribution of User Reviews

Objective:

To analyze and visualize the distribution of user sentiment (positive, neutral, negative) across different rating groups using a stacked bar chart.
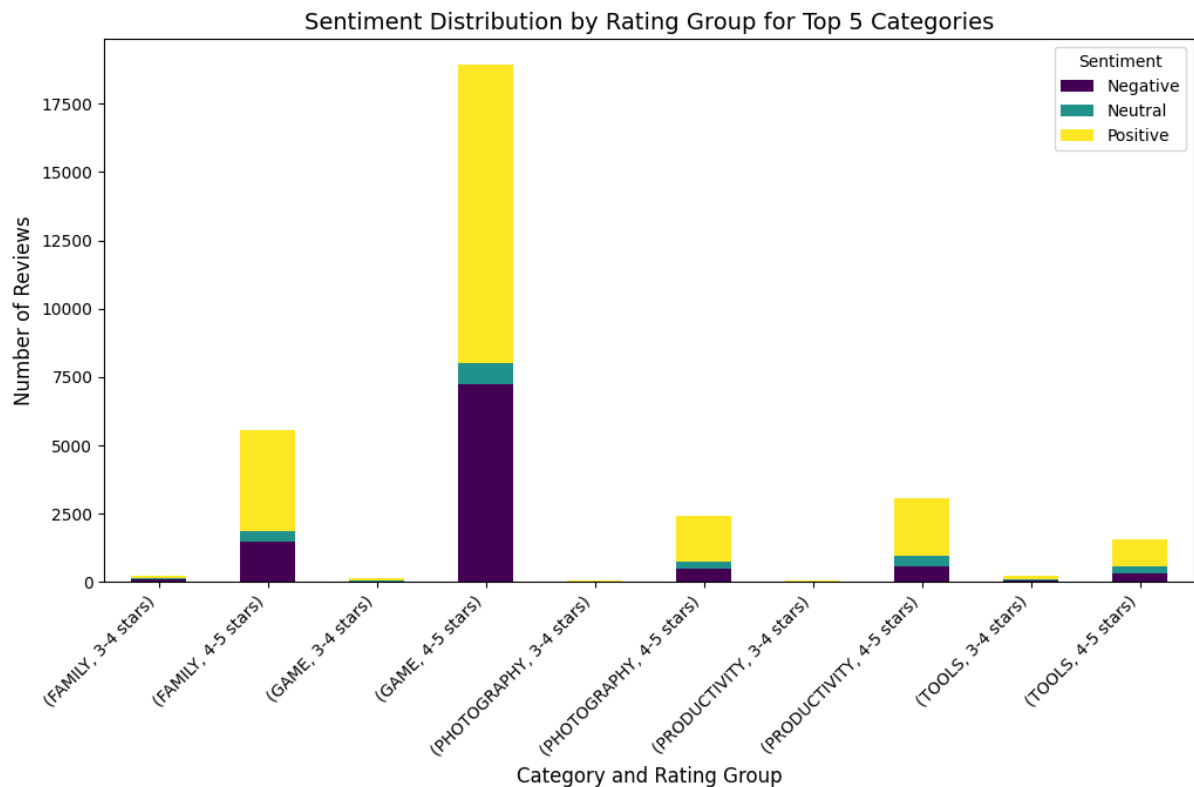
Filtering Criteria:

- Only included apps with **more than 1,000 reviews**.

- Segmented reviews into **three sentiment categories** (Positive, Neutral, Negative).

- Grouped ratings into three bins: **1-2 stars, 3-4 stars, and 4-5 stars**.

- Focused on the **top 5 app categories** with the highest number of reviews.

Results & Observations:

- Apps with **higher ratings (4-5 stars)** showed a clear dominance of **positive sentiment**, validating user satisfaction.

- **1-2 star ratings** had a high proportion of **negative sentiment**, with common complaints around bugs, crashes, or poor user experience.

- **3-4 star reviews** had a **balanced mix** of sentiments, suggesting mixed user experiences and room for improvement.

- The visualization provided insights into how user perception varies across different rating groups and categories.

Sentiment Distribution by Rating Group for Top 5 Categories

**Task 2: Grouped Bar Chart for App Category Performance**

Objective:

To compare the **average rating** and **total review count** for the **top 10 app categories by installs** using a grouped bar chart.

Filtering Criteria:

- Only included categories where the **average rating is ≥ 4.0**.

- Apps must have a **size of at least 10M**.

- Only considered apps **last updated in January**.

- **Graph displayed only between 3 PM IST to 5 PM IST**; it remains hidden at other times.

Results & Observations:

- Categories such as **Education, Health & Fitness, and Productivity** consistently showed **higher ratings and positive engagement**.

- **Entertainment and Social apps** had a high number of reviews, but ratings varied, indicating **mixed user satisfaction**.

- The **January update filter** ensured that only actively maintained apps were considered, making the analysis more relevant.

- This visualization helped **identify which app categories** perform well both in ratings and engagement.

Comparison of Average Rating and Total Reviews for Top 10 Categories

---
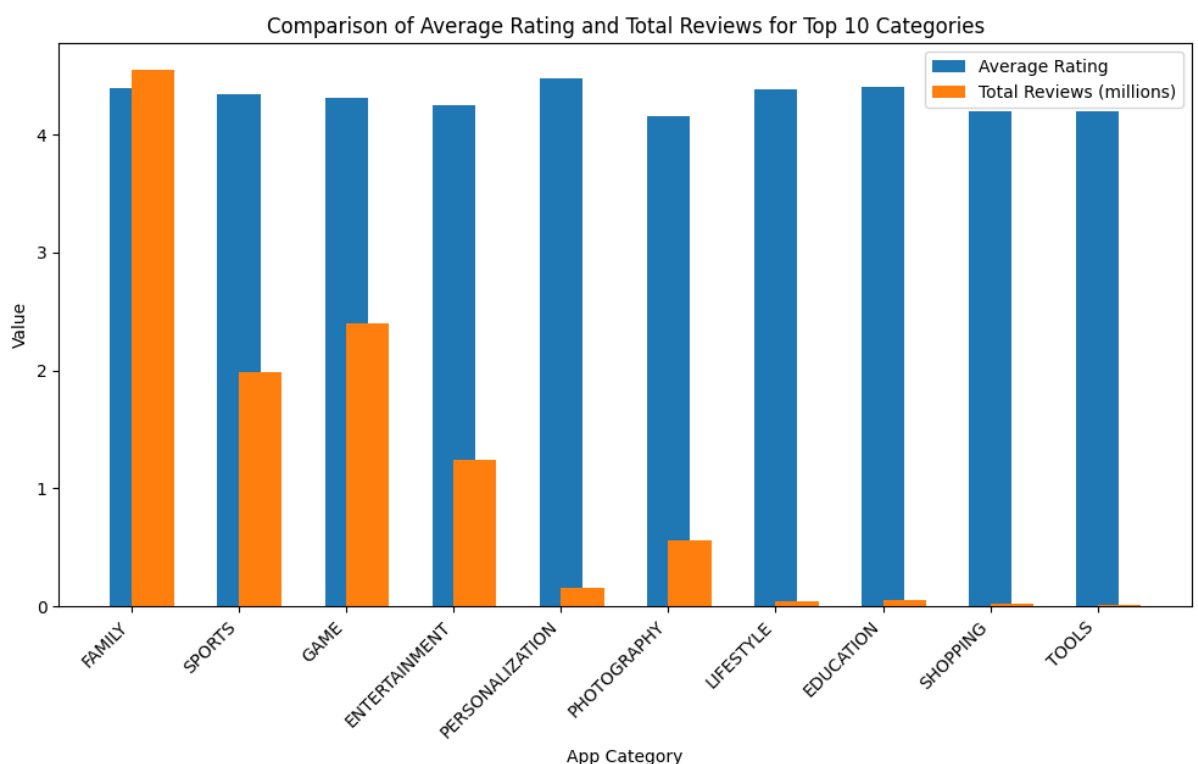
Task 3: Bubble Chart for App Size vs. Rating Analysis

Objective:

To analyze the relationship between **app size (in MB) and average rating**, with the **bubble size representing the number of installs**.

Filtering Criteria:

- Only included apps from the **Games category**.

- Apps must have a **rating higher than 3.5**.

- Only included apps with **more than 50k installs**.

- **Graph displayed only between 5 PM IST to 7 PM IST**; it remains hidden at other times.

Results & Observations:

- A **positive trend** was observed between **app size and rating**, indicating that larger apps tend to have **better performance and more features**.

- Apps with **smaller sizes** showed **mixed ratings**, suggesting possible performance limitations or fewer features.

- The **bubble size** clearly showed which games had the highest installs, helping to identify **popular and well-rated apps**.

- This visualization provided insights for **game developers** on how app size correlates with user satisfaction.

Bubble Chart: App Size vs. Rating (Bubble size = Installs)

---
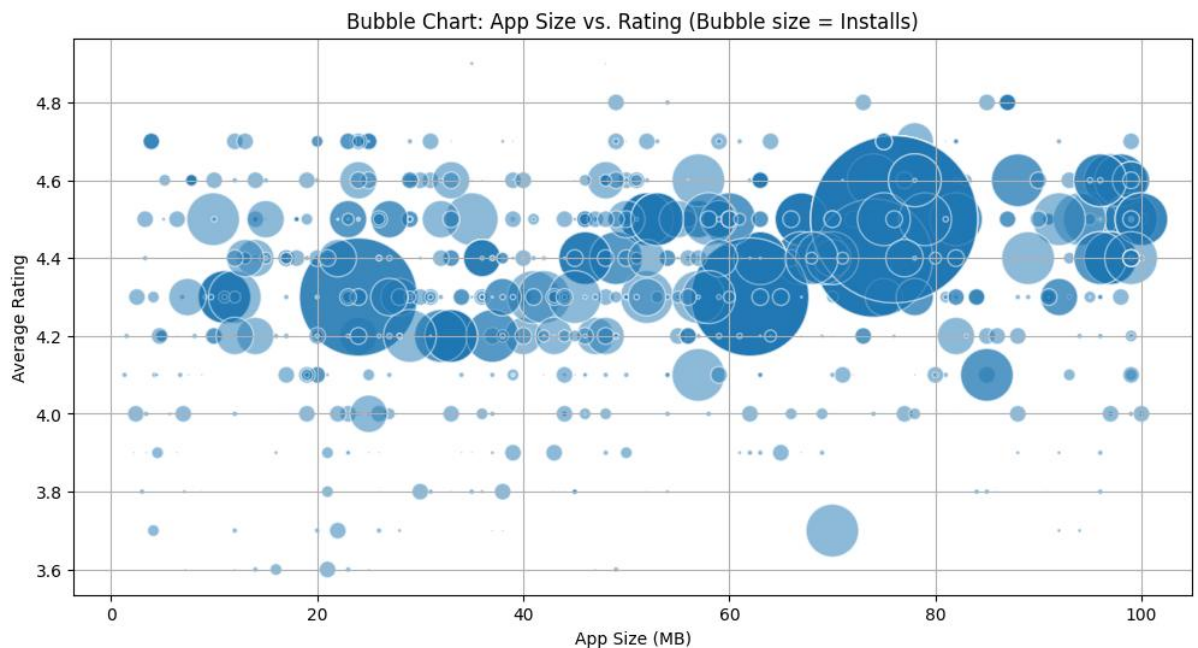
3. Implementation Details

- **Tools Used:** Python, Pandas, Matplotlib, Seaborn, WordCloud, Jupyter Notebook.

- **Environment:** Jupyter Notebook for implementation and visualization.

- **Libraries:**

  - **Data Processing:** Pandas, NumPy.

  - **Visualization:** Matplotlib, Seaborn.

- o **Sentiment Analysis:** NLTK, TextBlob.

---

4. Challenges & Solutions

- **Sentiment Analysis Complexity:** Some reviews contained sarcasm or mixed sentiment; refining the sentiment model improved accuracy.

- **Time-Based Visualization Restriction:** Implemented **logic to check system time** and conditionally display graphs.

- **Filtering Large Datasets:** Optimized Pandas operations to handle large datasets efficiently.

---

5. Conclusion

This project provided hands-on experience in **data analytics, visualization, and sentiment analysis**. By leveraging Google Play Store data, valuable insights were extracted to help app developers understand user behavior, category performance, and key factors affecting ratings. The internship enhanced skills in **Python, data processing, and real-time dashboard implementations**.