Play store Data Analysis Report

Background:

The dataset contains historical information about mobile apps from the Google Play Store, covering app details such as ratings, reviews, size, installs, and Android version compatibility. It provides insights into trends and user behavior from June 2012 to February 2019, helping businesses understand market dynamics.

Problem Statement:

The goal of this analysis is to gain insights into the mobile app industry to guide business decisions and app development. Key areas of focus include identifying popular app categories, analyzing ratings, reviews, and installs, and understanding trends in Android version usage, which will help businesses make informed strategies for app development and user engagement.

Data Overview:

The dataset consists of 10,838 rows and 12 columns, with the following attributes for each mobile app:

1. **App**: The name of the mobile app.
2. **Category**: The app's classification or genre (e.g., Games, Social, Education).
3. **Rating**: The app's rating, reflecting user satisfaction.
4. **Reviews**: The total number of reviews submitted for the app.
5. **Size**: The size of the app file in megabytes (MB).
6. **Installs**: The number of times the app has been installed.
7. **Price**: The price of the app (free or paid).
8. **Content Rating**: The app's content rating (e.g., Everyone, Teen).
9. **Genres**: The specific genre of the app (e.g., Action, Puzzle).
10. **Last Updated**: The date when the app was last updated.
11. **Current Version**: The version of the app available in the Play Store.
12. **Android Version**: The minimum Android version required for the app to run.

Methodology:

1. **Data Collection:**
    - The dataset was collected from the Google Play Store, containing mobile app details from June 2012 to February 2019. The data includes information such as app names, categories, ratings, reviews, installs, app size, and more.
2. **Data Cleaning:**
    - The data was cleaned to remove inconsistencies, missing values, and duplicates. Rows with missing values were handled either by removing or imputing with relevant values where necessary. Any outliers were analyzed and dealt with appropriately.
3. **Data Exploration:**
    - Initial exploration of the data was conducted to understand the distribution of categories, ratings, installs, and other key variables.
    - Statistical summaries were performed to understand data distribution.
4. **Data Transformation:**

- Data types were adjusted (e.g., converting ratings and installs to numeric values for analysis).
- Categorical columns like Category, Genres, and Android Version were encoded as needed for analysis.
- The Price column was transformed to indicate whether the app is paid or free.

5. **Data Analysis:**
   - Most Used Android Version: Identified the most commonly used Android version across the dataset by analyzing the Android Version column.
   - Categories of Mobile Apps: Analyzed the distribution of apps across different categories to identify trends and user preferences.
   - Most Installed Categories: Aggregated the total installs by category to identify the most popular app types.
   - Top Rated Apps: Identified apps with 5-star ratings and the top 10 apps based on ratings and reviews.
   - Insights on Size and Installs: Correlated app size with the number of installs to determine if larger apps tend to get more or fewer installs.
6. **Data Visualization:**
   - Used various visual tools (charts, graphs, and pivot tables) to represent trends, such as:
     - Pie charts for the distribution of app categories.
     - Bar charts for top-rated and most installed apps.
     - Histograms to display ratings distribution.

Technical Processes

- Use pivot tables for summarizing data.
- Calculate averages, variances, and growth rates.
- Create charts and graphs for visual representation.
- Apply filters and sorting for specific analyses.
- Use functions like Count and Sum IF for data aggregation.

Key Findings

Most Used Android Version:

- The most widely used Android version across the dataset is **4.1 and up**, which suggests that apps should be compatible with this version and higher for maximum reach.

App Categories:

- The dataset includes **33 distinct app categories**, with the following key categories:
  - **Games**
  - **Social**
  - **Sports**
  - **Education**
  - **Productivity**
  - **Entertainment**

    - **Health and Fitness**
    - **Lifestyle**
- This reflects the diverse nature of mobile apps available on the Google Play Store, catering to a wide range of user needs.

Most Installed Category:

- The **Games** category stands out with the highest number of installs, totaling **35,086.02 million** installs. This indicates that gaming apps dominate the mobile app landscape in terms of user engagement.

Total Number of Categories:

- The dataset covers a total of **33 categories**, showcasing the extensive range of mobile apps across various industries, including entertainment, education, lifestyle, and more.

Apps with 5-Star Ratings:

- There are **271 apps** that have received a perfect 5-star rating, indicating high user satisfaction and the potential for strong user engagement and loyalty.

Top 10 Apps Based on Ratings:

- The top 10 apps based on user ratings are:
    1. **ROBLOX**
    2. **CBS SPORTS APP - SCORES, NEWS, STATS & WATCH LIVE**
    3. **DUOLINGO: LEARN LANGUAGES FREE**
    4. **8 BALL POOL**
    5. **CANDY CRUSH SAGA**
    6. **ESPN**
    7. **ZOMBIE CATCHERS**
    8. **BOTMASTERS**
    9. **SNIPER 3D GUN SHOOTER: FREE SHOOTING GAMES – FPS**
    10. **SUBWAY SURFERS**

Recommended Analysis

Q1: - Which is the most used Android version?
Q2: - Which all categories of mob apps are existing from this dataset?
Q3: - Which is the most installed category of mob app from this dataset?
Q4: - How many total categories are there?
Q5: - Which apps have got 5-star ratings
Q6: - Which are the top 10 mob apps based on ratings?

Conclusion

The Mobile Apps Playstore Data Analysis provides valuable insights into the mobile app landscape, highlighting trends in Android version usage, popular app categories, and top-rated apps. The analysis reveals that the most used Android version is 4.1 and up, with the Games category having the highest number of installs. Additionally, the study identified 33 app categories and found that 271 apps have earned 5-star ratings. The top-rated apps like ROBLOX and Duolingo offer guidance

for developers to focus on user satisfaction and high engagement. These insights are essential for strategic decision-making in app development and marketing.

Project Owner

Name: Vikant Kumar

Date: 21-08-2023