Key Insights from Google Play Store EDA

1. General Overview

- The dataset contains **2,312,944** records and **24** features.
- Data was collected in **June 2021** and includes app details like category, ratings, installs, pricing, and developer information.

2. Data Cleaning & Preparation

- **Missing values** were handled by imputing appropriate values or dropping irrelevant columns.
- **Feature engineering** was applied to convert size into MB, handle date-time values, and create an app age feature.
- No duplicate rows were found, but some apps had the same name with different package IDs.

3. Key Findings

3.1 App Ratings & Downloads

- Apps with higher downloads generally have better ratings.
- No strong correlation was found between ratings and downloads for lesser-known apps.

3.2 Free vs. Paid Apps

- **Free apps dominate** in terms of number and downloads.
- Paid apps do not necessarily have higher ratings.

3.3 Most Popular App Categories

- **Top 5 by Number of Apps**:
  - Education
  - Music & Audio
  - Tools
  - Business
  - Entertainment
- **Top 5 by Downloads**:
  - Tools
  - Communication
  - Productivity
  - Photography
  - Video Players & Editors

3.4 App Size & Popularity

- Apps **under 200MB** tend to have more downloads.
- Larger apps usually belong to specialized categories like gaming.

3.5 Content Ratings & Install Base

- Apps rated **"Everyone"** have the highest number of installs.
- The **"Teen"** category also has a significant share of downloads.

3.6 App Release Trends

- Most apps were released in **2020**.
- However, apps from **2015** have the highest number of total downloads.
- App age does **not strongly** influence download numbers.

4. Conclusion & Recommendations

- **Focus on free app models** for better market reach.
- **Optimize app size** to be under 200MB for higher adoption.
- **Target popular categories** like Tools and Communication for better downloads.
- **Consider user engagement strategies** for older apps to maintain relevance.

Further analysis with machine learning models can help predict app success based on historical data trends.

Author: Muhammad Waqas
Email: waqasliaqat630@gmail.com
GitHub: profile
LinkedIn: Profile