

EXPLORATORY DATA ANALYSIS OF GOOGLE PLAY STORE APPS

Market Trends, App
Performance & User Behavior



Google Play

TOOLS USED

Python, Pandas, Matplotlib, Seaborn & Plotly

PRESENTED BY

Tejas Jadhav

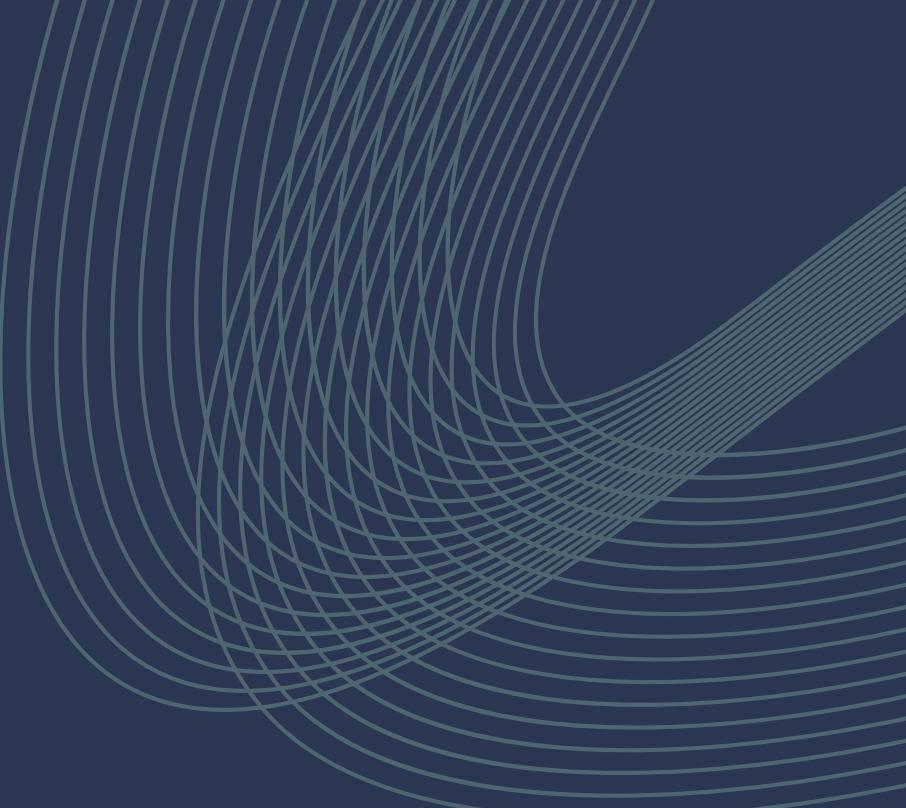


TABLE OF CONTENTS

| | |
|--|-----------|
| Problem Statement & Objective | 3 |
| EDA Workflow | 4 |
| Key Questions Explored | 6 |
| Data Overview | |
| Insights | 7 |
| Category Level Analysis | 8 |
| Correlations & Patterns | 9 |
| Business/Developer Takeaways | 11 |
| Conclusion | 12 |

PROBLEM STATEMENT & OBJECTIVE

Problem Statement:

The Google Play Store hosts millions of apps across different categories. It's challenging to understand trends, user preferences, and which types of apps are most successful.

Objective:

To explore the dataset of Google Play Store apps and uncover patterns, insights, and trends that can help app developers and businesses make better decisions.



EDA WORKFLOW

For this analysis, a structured workflow was followed, involving data collection, understanding, cleaning, exploration, and summarization of insights to allow a clear understanding of the dataset and its trends.



KEY QUESTIONS EXPLORED



- Which app categories are most popular ?
- Do free apps get more installs than paid apps ?
 - How does price affect ratings ?
- What factors influence high ratings ?
- Are more reviews linked to better ratings ?



AUSTEN
TECH

www.reallygreatsite.com
hello@reallygreatsite.com
123-456-7890

DATA OVERVIEW

The dataset provides key information about Google Play Store apps, including ratings, reviews, installs, price, category, and type. Additional details like content rating, genres, app size, last update, and Android version capture app characteristics and trends.

Data Source: Kaggle

| Dataset Features | Feature Description |
|------------------|---|
| App | Name of the application as listed on the Google Play Store |
| Category | Primary category under which the app is classified |
| Rating | Average user rating of the app on a scale of 1 to 5 |
| Reviews | Total number of user reviews submitted for the app |
| Size | Storage size of the app as reported on the Play Store |
| Installs | Approximate number of times the app has been installed by users |
| Type | Indicates whether the app is Free or Paid |
| Price | Price of the app (0 indicates a free app) |
| Content Rating | Target age group suitability (e.g., Everyone, Teen, Mature) |
| Genres | Specific genre or sub-genres associated with the app |
| Last Updated | Date when the app was most recently updated |
| Current Ver | Current version of the application |
| Android Ver | Minimum Android OS version required to run the app |

DATA QUALITY CHALLENGES & ANOMALIES

Several inconsistencies were found in the dataset, which could have affected the analysis if left unaddressed.

DATA ANOMALIES

- **Missing values** are mainly present in the Rating column, with very few missing entries in Current Ver, Android Ver, Content Rating, and Type.
- **Columns requiring cleaning or type conversion:** Reviews, Size, Installs, Type, Price, Last Updated, Android Ver
- **Columns with notable anomalies or invalid entries:**
 - Category (e.g., 1.9)
 - Rating (e.g., 19.)
 - Type (e.g., 0)
 - Installs (e.g., Free)
 - Price (e.g., Everyone)
 - Size and Android Ver (e.g., Varies with device)

DATA CLEANING & TREATMENT

Inconsistent and missing values were addressed, and key features were cleaned and standardized for analysis.

DATA CLEANING SUMMARY

- Missing values, mainly in **Rating** and a few version-related fields, were handled using median imputation.
- Features such as **Reviews**, **Size**, **Installs**, **Price**, **Type**, **Last Updated**, and **Android Ver** were cleaned and standardized.
- Inconsistent and invalid entries (e.g., incorrect categories, unrealistic ratings, mixed text values) were identified and corrected.
- A new feature, **Minimum Android Version Required**, was created from Android version data.
- **Current Ver** was removed as it is a technical detail with limited business relevance.



INSIGHTS