Google Play Store Data Analysis

Detailed Project Report

ARYAN VARSHNEY

25 April 2025

Executive Summary

- Analyzed Google Play Store app data
- Used Power BI for data transformation & visualization
- Explored installs, ratings, reviews, app types
- Built a dashboard to identify top categories & trends
- Delivered key insights for datadriven decisions

Problem Statement

- The Google Play Store hosts thousands of apps across diverse categories.

- Users rely on ratings, reviews, and install counts to choose apps, but trends and insights are not always obvious.

- Businesses and developers need data-backed decisions to optimize app development, marketing, and pricing.

Objectives

- Analyze app performance based on key metrics: installs, reviews, ratings, and pricing.

- Compare how Free vs Paid apps perform across categories.

- Identify the most popular and highest-rated app genres.

- Use Power BI to visualize trends and uncover hidden patterns in user engagement and app growth.

Dataset Overview

Source: Public dataset (Kaggle)

Rows: 11,000+ app entries

Key Columns:

App, Category, Rating

Reviews, Installs, Type, Price

Last Updated, Android Version

Preprocessed to remove nulls, duplicates & outliers

Tools & Methodology

Tools Used:

- **Python**

- Power BI

- Excel (for cleanup)
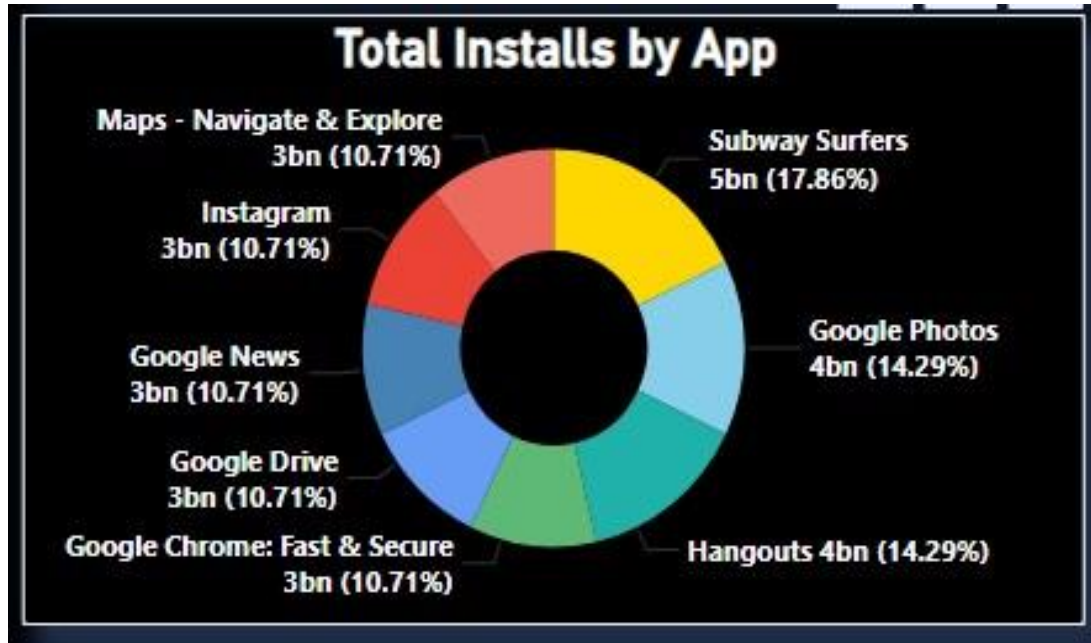
- Power Query Editor

Methodology (ETL):

- Import data (CSV)

- Clean & transform in Power Query

- Create calculated columns

- Build visuals & publish dashboard

Data Modeling & Transformation

- ➢ **Removed duplicates and null values**

- ➢ **Cleaned incorrect data types (e.g. Ratings > 5)**

- ➢ **Extracted year from "Last Updated"**

- ➢ **Created new columns:**

  - ✓ **Install Range (binned installs)**

  - ✓ **App Age (optional)**

- ➢ **Simple model — flat table**

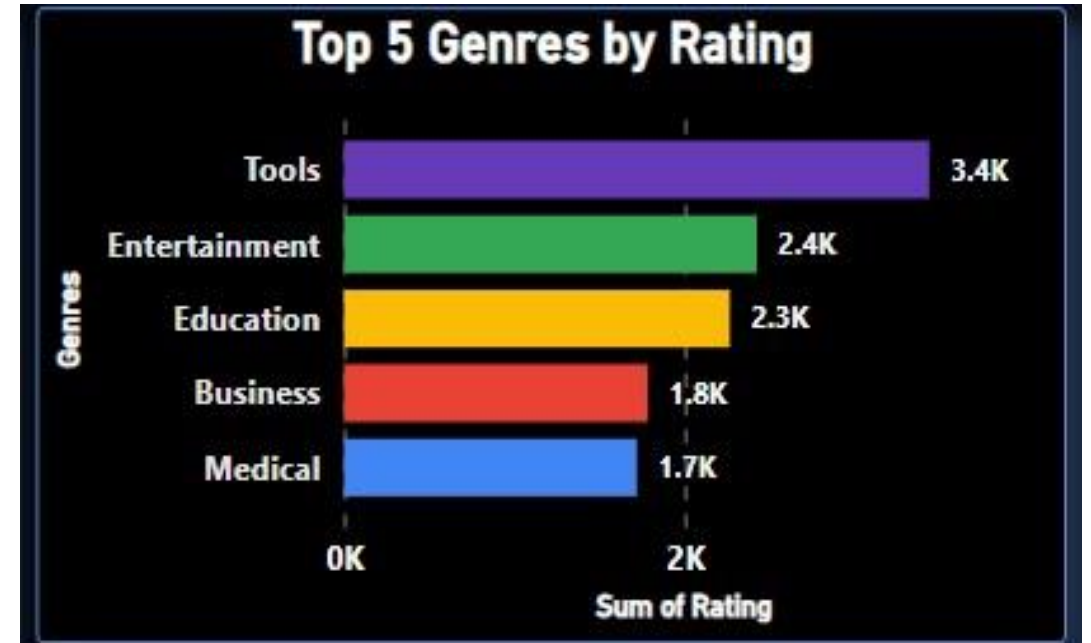- ➢ **Added Date Table (optional) for time analysis**

Total Installs by App vs Top 5 Genres by Rating

➔ *Insight:* "Subway Surfers and Google Photos are
Entertainment apps receive the among the most

➔ *Insight:* "Tools and
installed apps."highest user

ratings."

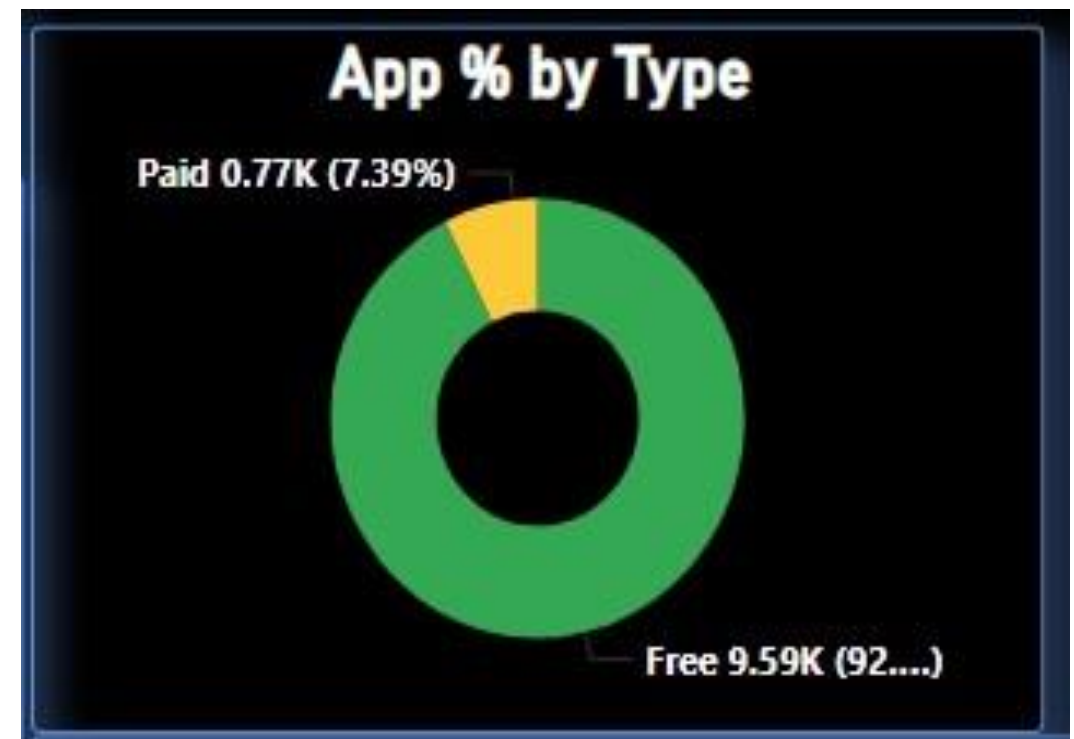# Top 5 Apps by ReviewsApp % by Type

➔ *Insight:* "Instagram and Facebook lead in total user reviews."➔ *Insight:* "More than 90% of the apps are free on the Play Store."

Top 5 Apps by Reviews

App % by Type

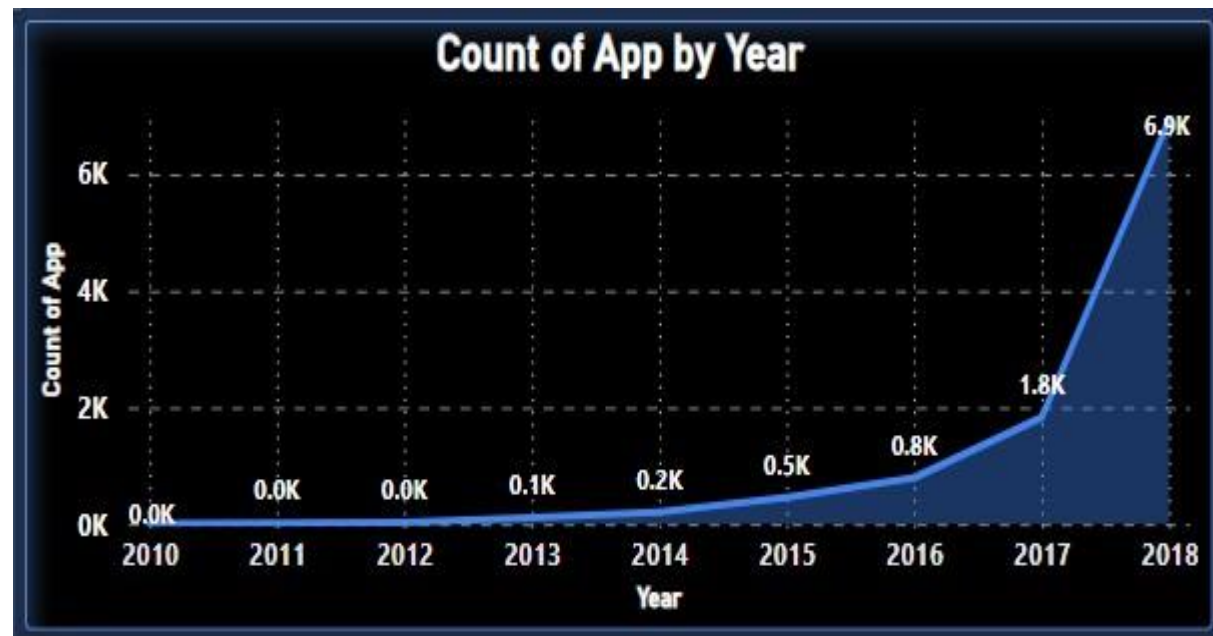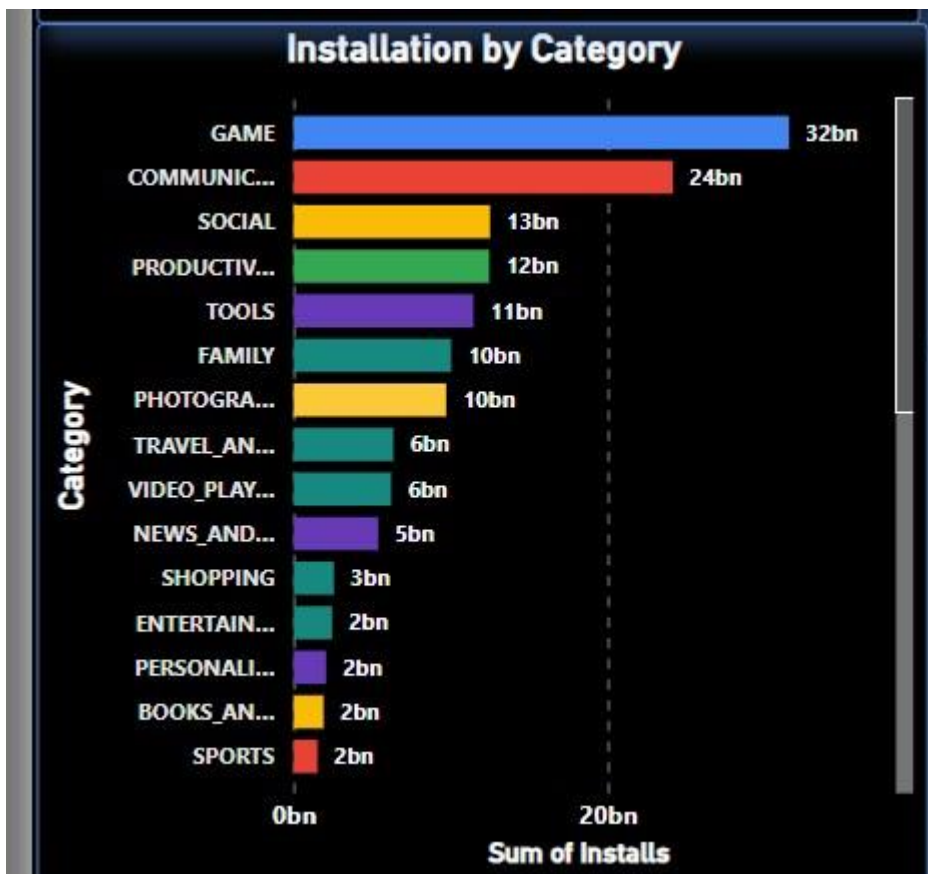## Installation by Category

Count of Apps by Year

➔ *Insight:* "Games are most Installed Category followed by Communication."
➔ *Insight:* number of apps surged rapidly after 2016."

Installation by Category

Count of App by Year

🔍 Key Findings:

➢ Most apps on Play Store are free (~92%).

➢ Categories like **Family** and **Games** dominate app counts and installs.

➢ Apps targeting **Everyone** and **Teen** age groups have the highest installations.

➢ **Subway Surfers** and **Google Photos** are among the top installed apps.

➢ **Instagram** and **Facebook** received the most user reviews.

➢ App launches surged sharply after 2016.

Recommendations:

➢ **Focus app development on popular categories (e.g., Games, Family).**

➢ **Consider offering free apps with in-app purchases to attract users.**

➢ **Optimize apps for broader content ratings (Everyone, Teen) for wider reach.**

➢ **Maintain high ratings and reviews to boost app visibility.**

🏁 Conclusion:

- Successfully analyzed Google Play Store data using Power BI.

- Identified key trends in app installs, ratings, pricing, and content ratings.

- Built an interactive dashboard for better business understanding and decision-making.

⚡ Challenges Faced:

➢ **Handling missing or inconsistent data entries (e.g., missing ratings).**

➢ **Cleaning text-based data types (e.g., "Varies with device" in size/version).**

➢ **Choosing the most impactful visuals for dashboard storytelling.**