📱 Case Study: App Store / Play Store Analytics

◆ ASK PHASE

Business Task

This project looks at mobile apps in the Google Play Store. I want to find out what makes some apps more popular than others. I will use data to see how things like rating, category, or price affect downloads. This can help app developers and marketers make better choices.

Stakeholders

- App developers and product managers
- Marketing teams who want more downloads
- Business owners or startup founders
- People who invest in apps

Key Objectives

- Find out what features help apps get more downloads or better ratings
- Compare free and paid apps
- Look for patterns in successful apps
- Give advice to improve app listings

Guiding Questions

1. Do better ratings mean more downloads?
2. What app categories are most popular?
3. Do free apps get more users than paid ones?
4. Do reviews or app size affect ratings?

◆ PREPARE PHASE

About the Dataset

I found a dataset on Kaggle. It is called "Google Play Store Apps." It has data about many apps on the Google Play Store. The dataset includes app name, category, rating, number of reviews, number of installs, price, and more.

Link to dataset: https://www.kaggle.com/datasets/lava18/google-play-store-apps

Why I chose this dataset

This dataset has the kind of information I need to answer my questions. It shows how users rate apps. It also shows if the app is free or paid. This helps me compare different types of apps. It is also popular and used in many beginner projects.

How I will check data quality

I will look for missing values and wrong entries. I will also check if some columns have strange symbols or extra spaces. I will fix these before using the data. I may also remove rows that do not have complete data.

Tools I plan to use

I plan to use Google Sheets or Excel to explore and clean the data. Later I may use R or Tableau for charts.

◆ PROCESS PHASE

Cleaning the Category column

During the pivot table creation, I found that the Category column had unexpected values like numbers (e.g., 1.9) or blanks. These caused errors in the pivot results. To fix this, I filtered the Category column using a custom formula `=ISNUMBER(B2)` to show only invalid entries. I then removed those rows, which allowed the pivot table to summarize the installs by category correctly. This helped ensure the charts were based on clean and meaningful data.

Cleaning the Type column

While preparing for the Free vs Paid app analysis, I noticed that the Type column had invalid entries like 'NaN'. These entries could interfere with the pivot table and chart results. To fix this, I used the filter tool in Google Sheets:

1. I clicked the filter icon on the Type column.
2. I chose "Filter by condition > Text contains" and typed in 'NaN'.
3. This showed only the rows with invalid values.
4. I deleted those rows to keep only valid values: 'Free' and 'Paid'.

This cleanup made sure the chart comparing Free vs Paid apps would be based on clean, accurate data.. These entries could interfere with the pivot table and chart results. To fix this, I filtered the Type column to show only rows with 'NaN' or blanks, then deleted those rows. This ensured that the column only included valid values: 'Free' and 'Paid'.

Cleaning the Installs column

The installs column (F) had values like "10,000+" and "1,000,000+". I used Find and Replace to remove the plus signs and commas. Then I used this formula to turn the text into a number:

Formula used:
```
=VALUE(SUBSTITUTE(SUBSTITUTE(F2, "+", ""), ",", ""))
```

Cleaning the Price column

The price column (J) had dollar signs like "$2.99". I used a formula to remove the dollar sign and change the text to a number. Free apps were set to 0.

Formula used:
```
=IF(J2="Free", 0, VALUE(SUBSTITUTE(J2, "$", "")))
```

Cleaning the Size column

The size column (E) had values like "14M" or "500k". I created a formula that changed "M" to megabytes and "k" to kilobytes. I also left out "Varies with device".

Formula used:
```
=IF(OR(E2="Varies with device", E2=""), "", IF(RIGHT(E2,1)="M",
VALUE(LEFT(E2,LEN(E2)-1))*1000000, IF(RIGHT(E2,1)="k",
VALUE(LEFT(E2,LEN(E2)-1))*1000, "")))
```

Cleaning the Reviews column

The reviews column (D) looked okay, but I used a formula to be sure all values were numeric and had no spaces.

Formula used:
```
=IF(ISNUMBER(D2), D2, VALUE(TRIM(D2)))
```

How I filled formulas quickly

Instead of dragging, I used Ctrl + Shift + Down and Ctrl + D to fill the formulas down the whole column. This saved time and worked great.

◆ ANALYZE PHASE

Project Reflection

I didn't clean all the columns at the beginning of the project. Instead, I cleaned the columns as I needed them. This helped me stay focused on the questions I wanted to answer and saved time. In real-world projects, analysts often clean data step by step, that is, discovering issues during the analysis. This method allowed me to handle each task in a clear and manageable way.

Trends I explored

I looked at the relationship between installs and other variables. I focused on rating, category, price, reviews, and size. I wanted to see what factors might affect the number of installs.

Charts and Visual Analysis

To support the analysis, I created several charts based on the cleaned data. I used Google Sheets to generate each chart. These visuals helped to answer the main business questions and uncover patterns in the data.

Note: Due to the size and scale of the dataset, some scatter plots such as Rating vs Installs, Reviews vs Installs, and Size vs Installs were created using filtered samples (up to 500 rows) to improve readability. The charts were inserted into the project to show these insights visually.

1. Top 10 Categories by Total Installs
To avoid crowding the chart with too many app categories, I used a pivot table to summarize installs by category. Then I sorted the results and used the top 10 categories for the bar chart.
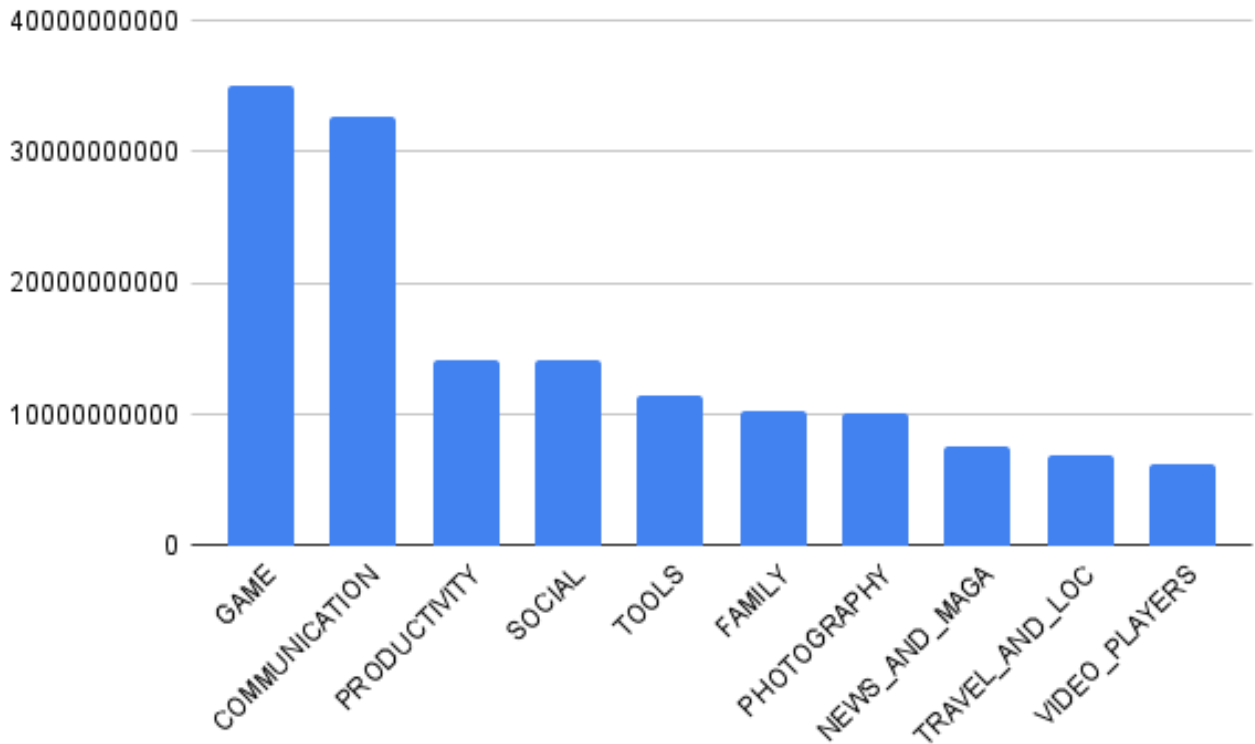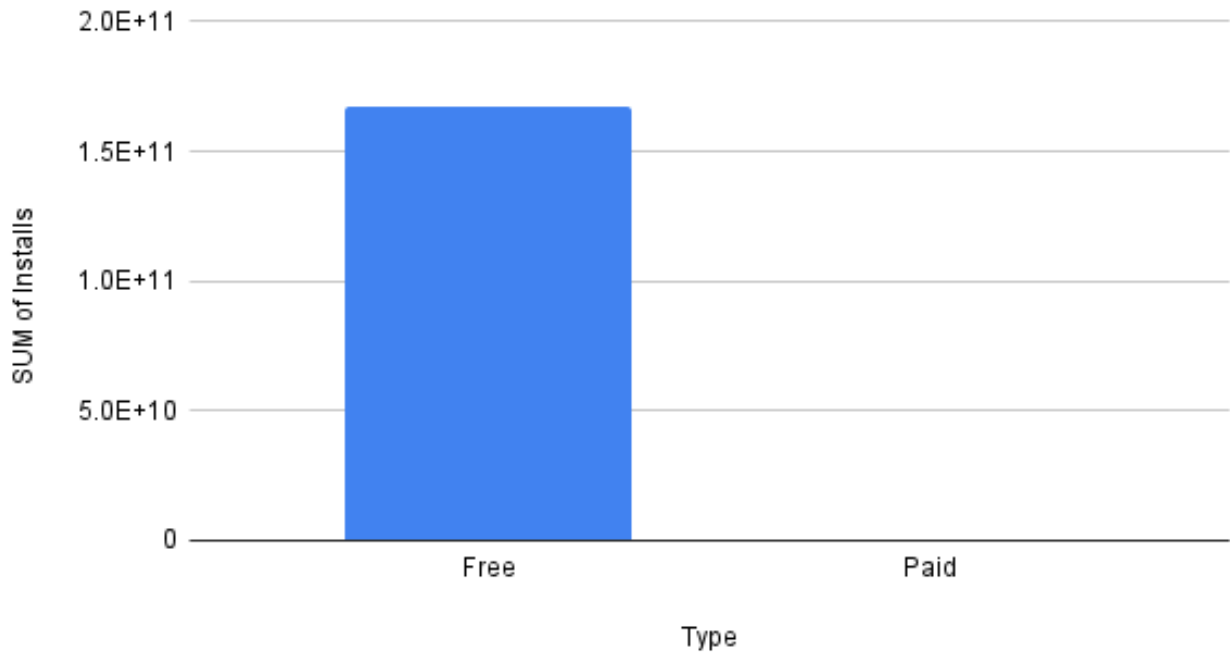

This chart shows that apps in the Games, Communication, and Tools categories are the most downloaded.

2. Free vs Paid Apps – Total Installs

I compared the total installs for Free and Paid apps using a bar chart. This helps show user behavior and preference.
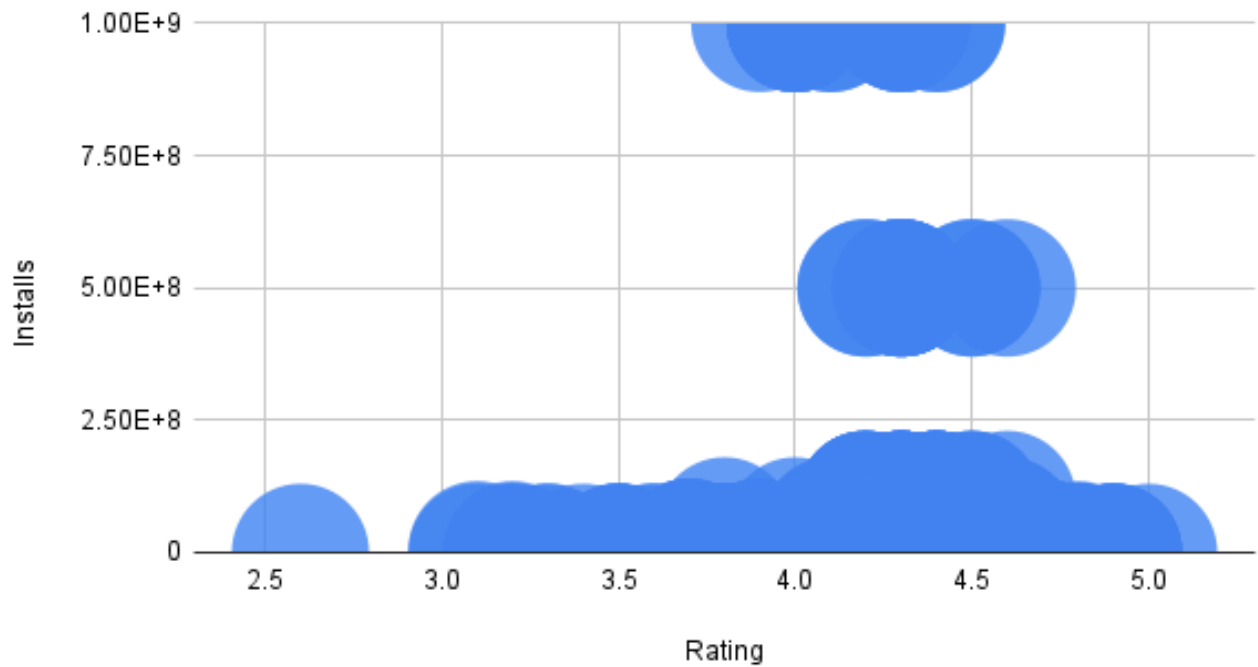

This chart confirms that Free apps are downloaded far more often than Paid apps.

3. Rating vs Installs

I made a scatter plot to see if higher app ratings lead to more installs.
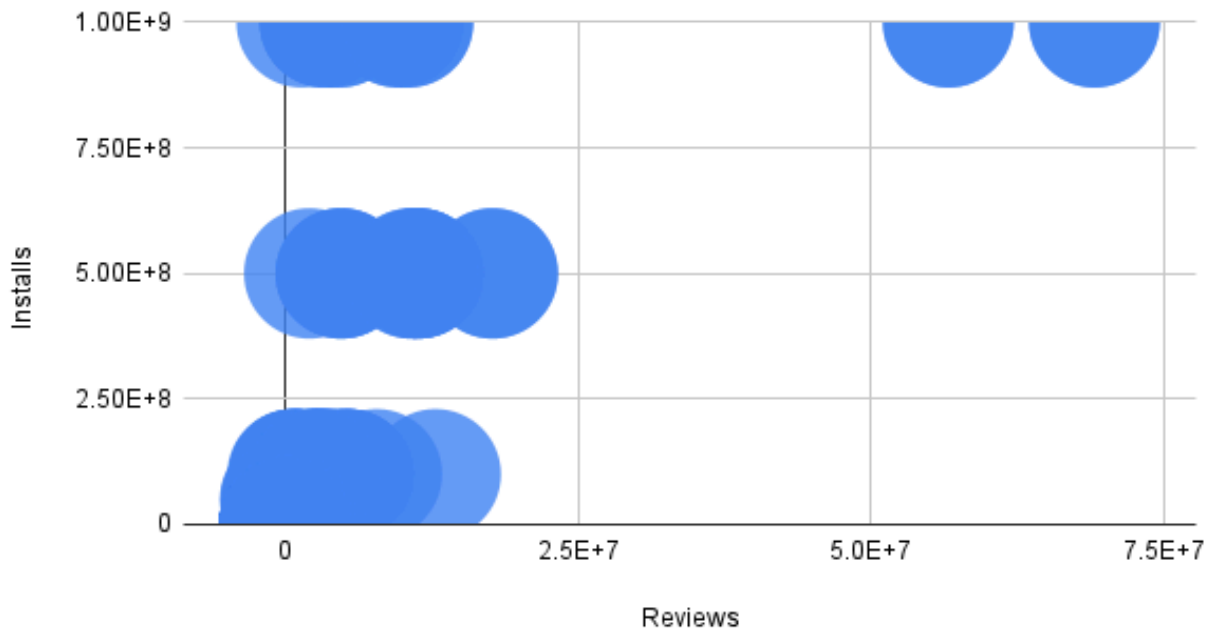

Installs vs. Rating

This chart shows a general trend that apps with higher ratings also have higher installs.

4. Reviews vs Installs
This scatter plot shows how the number of reviews relates to the number of installs.
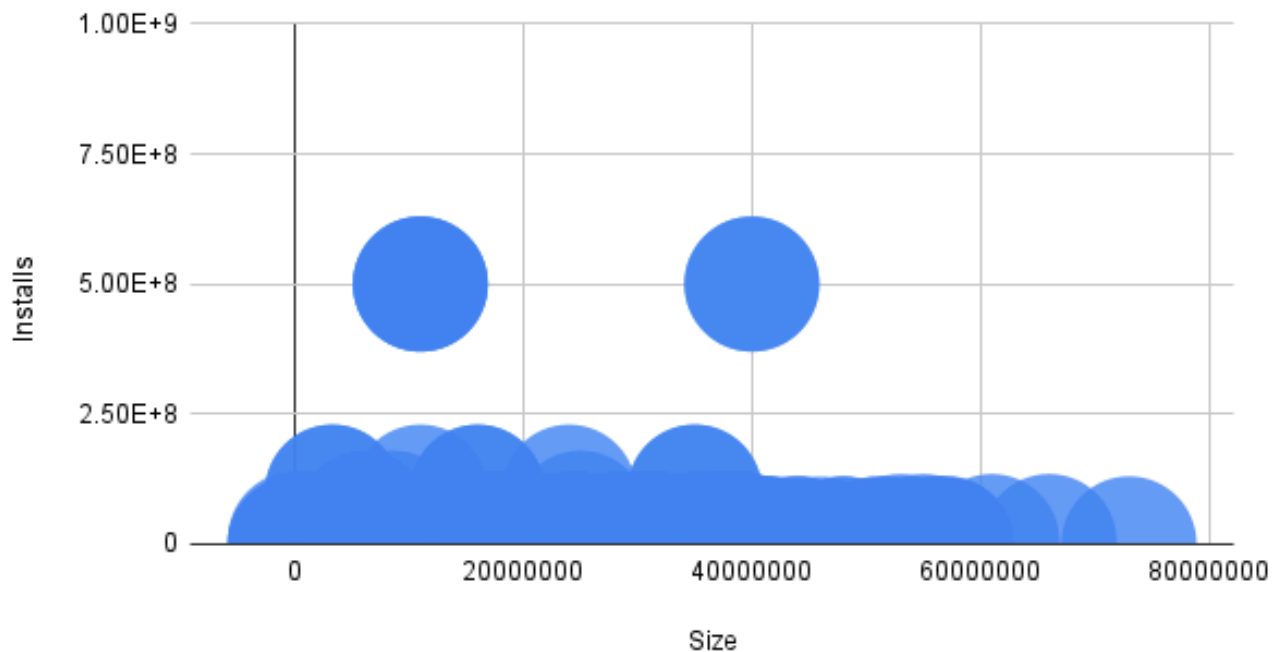

Installs vs. Reviews

Apps with more reviews usually have more installs. This shows user engagement and popularity.

5. Size vs Installs

I wanted to test if app size affects popularity. I used a scatter plot to compare cleaned size data with install counts.

Installs vs. Size

This chart shows no strong link between app size and installs. Small and large apps both performed well.

Findings

- Apps with higher ratings usually had more installs. Most top-rated apps had 4.0 stars or more.
- Free apps had more installs than paid apps. People downloaded free apps more often.
- Some categories like "Games" and "Communication" had more downloads than others like "Lifestyle" or "Beauty".
- Apps with many reviews often had high install numbers. This shows popularity and engagement.
- App size did not show a strong pattern with installs. Both small and large apps had high downloads.

What I used to find these insights

I used **Google Sheets** for the analysis. I sorted and filtered the data to explore different relationships. I created **bar charts** to compare app categories and price types. I also used

scatter plots to see how installs relate to ratings, reviews, and size. These visual tools helped me find simple patterns and answer the questions I started with.

◆ **SHARE PHASE**

Executive Summary

In this project, I explored a dataset of Google Play Store apps to find out what makes apps more successful. I looked at how factors like rating, category, price, reviews, and size affect the number of installs. I cleaned the data, visualized the results, and used those insights to answer key business questions.

Key Takeaways

- Apps with higher ratings tend to have more installs.
- Free apps are downloaded much more than paid ones.
- Categories like Games and Communication are among the most popular.
- Apps with more reviews also had more downloads, showing engagement.
- App size does not clearly affect the number of installs.

Recommendations

- Focus on creating apps in high-demand categories like Games and Communication.
- Keep the app free (at least initially) to attract more users.
- Encourage users to leave reviews to boost engagement and trust.
- Focus on app quality and user satisfaction to increase ratings and visibility.

These insights can help app developers, marketers, and product teams make smarter choices when launching or updating their apps in the market.