

# Capstone Project

## Play Store App Review Analysis

By

Ankita Gupta

WHY ANALYZE THE GOOGLE PLAY STORE?



Mobile App  
Market is  
set to grow  
20% by  
2023



how popular it's going to  
be?

Android Apps comprise  
90% of the Mobile App  
Market



What makes an  
App popular?  
Can we predict



What are some  
interesting patterns in  
user behavior related to

app usage & feedback

# Introduction

- Android is the most popular operating system in the world, with over 2.5 billion active users spanning over 190 countries.
- Google Play was launched on March 6, 2012, bringing together Android Market marking a shift in Google's digital distribution strategy .
- The Google Play Store is the largest and most popular Android app store.
- There are more than 3.04 million apps found on Google Play Store.
- The Play Store apps data has enormous potential to drive app-making businesses to success.

- Actionable insights can be drawn for developers to work on and capture the Android market. The main goal of our project is-

- 1) **The purpose of our project is to gather and analyze detailed information on apps in the Google Play Store in order to provide insights on app features and the current state of the Android app market.**
- 2) **The Objective of the project to Explore and analyze the data to discover key factors responsible for app engagement and success.**

## Problem Statement

- ❑ Two datasets are provided, one with **basic information** and the other with **user reviews** for the respective app.
- ❑ We must examine and evaluate the data in both datasets in order to identify the important characteristics that influence app engagement and success.

# So, what factors influence an app's success?

An app is said to be successful if it has:

- ❑ A high average user rating
- ❑ A good number of positive reviews
- ❑ A good number of monthly average users
- ❑ High revenue per customer and so on.

## Agenda

- ❑ Introduction
- ❑ Category wise play store apps installs
- ❑ Category wise most popular apps
- ❑ Top 10 apps in play store considering all the parameters
- ❑ Average installs, category wise
- ❑ Most installed apps in communication category
- ❑ Category wise percentage of paid apps

- ❑ Category wise top installed paid apps
- ❑ Average rating of paid apps
- ❑ Correlation between Rating ,Installs and Price
- ❑ Category wise installed apps with content rating
- ❑ Percentage reviews sentiment distribution

## Dataset Preparation

- **Loading the data sets:** Two datasets, First Play store app dataset and User Reviews dataset.
- **Import Libraries:** NumPy, Pandas, Seaborn and Matplotlib

- **Data cleaning:** Null values, Finding and removing Outliers, Removing duplicate data.
- **Data Imputation:** Filling the missing categorical values with mode and numerical values with median. Conversion of price, installs, reviews into numerical values.
- **Exploratory Data Analysis:** Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.

## Attributes in Google Play store Data

- 1.App :** This column Contains the name of the app for each observation.
- 2.Category :** This column Contains Category to which the app belongs.
- 3.Rating :** This column contains the average rating for the app.

- 4.Reviews :** This column contains the number of reviews that the app has received on the play store.
- 5.Size :** This column contains the amount of memory the app occupies on the device.
- 6.Installs :** This column contains the number of times that the app has been downloaded and installed from the play store.
- 7.Type :** This column contains the information whether the app is free or paid.
- 8.Price:** If the app is a paid app, this column contains the data about its price.
- 9.Content Rating:** This column contains the maturity rating of the app i.e. the age group of the audience for which it is suitable.
- 10.Genres:** This column contains the data about to which genre the app belongs. Genres can be considered as a further division of the group of Category.
- 11.Last Updated:** Contains the date on which the latest update of the app was released.
- 12.Current Version:** Contains information on the current version of the app available on the play store.
- 13.Android Version:** Contains information about the android versions on which the app is supported.



# Attributes in User reviews

1. **App-** Application name
2. **Translated Review-** User review
3. **Sentiment-** Positive/Negative/Neutral
4. **Sentiment Polarity-** Sentiment polarity score
5. **Sentiment Subjectivity-** Sentiment subjectivity score

Data Cleaning

Data Exploration

Predictive Modeling

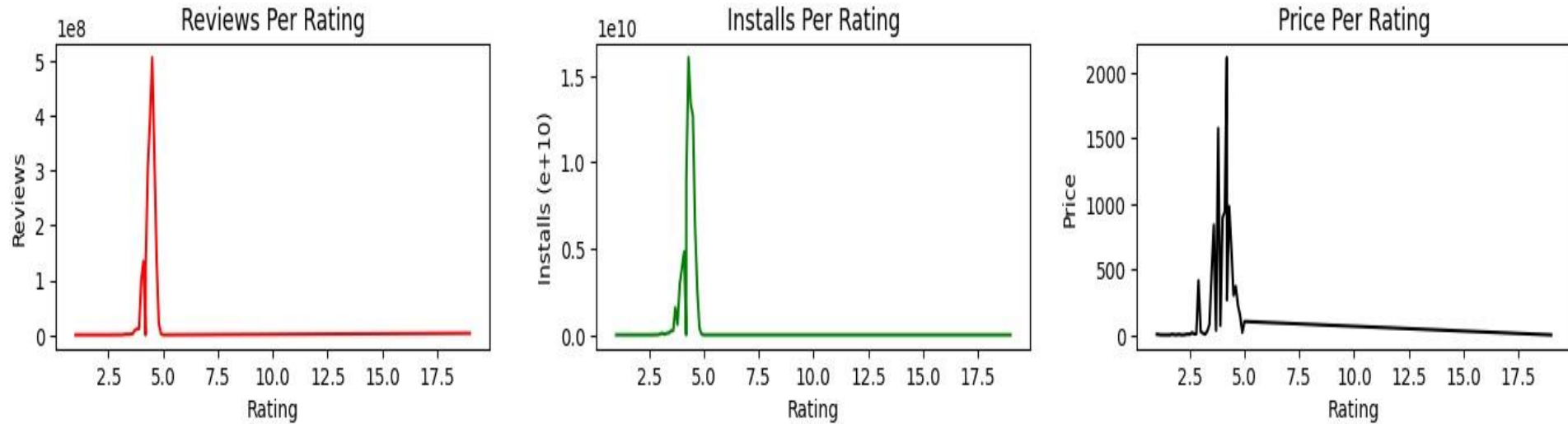


## OVERVIEW OF ANALYSIS

Understand the structure of the dataset and clean data before analysis

Uncover initial patterns, characteristics, and points of interest using visual exploration

Formulate a statistical model to forecast an outcome using relevant predictors



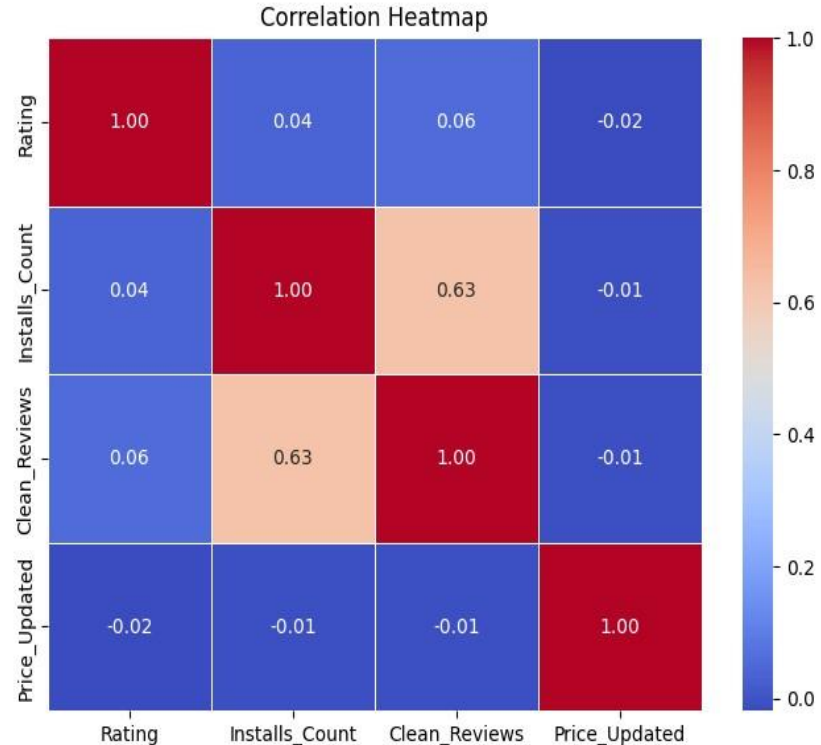
## Pairwise Plot- Reviews, Installs, Price

From the above plottings, we can imply that most of the apps with higher rating range of 4.0 - 4.7 are having high amount of reviews and installs. In terms of price, it doesn't reflect a direct relationship with rating, as we could see a fluctuation in term of pricing even at the range of high rating.



## Correlation Heatmap

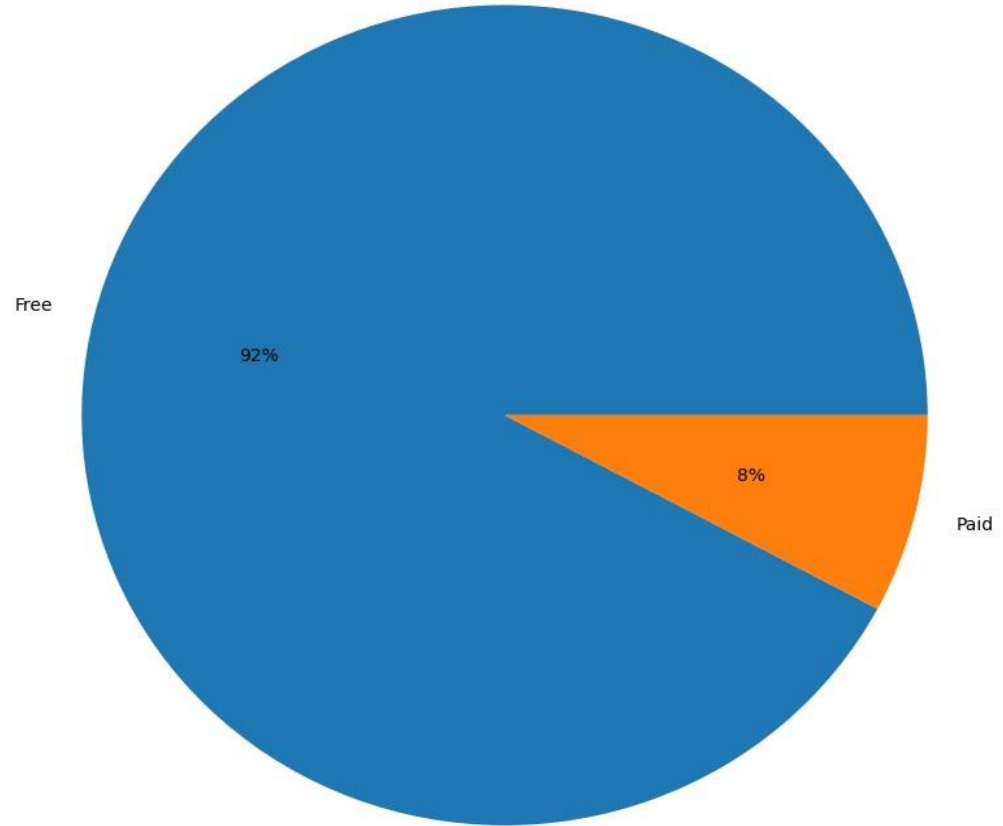
- There is a strong **positive** correlation between the **Reviews** and **Installs**.
- The Price is slightly **negatively** correlated with the Rating, Reviews and Installs.
- **Rating** is slightly **positively** correlated with the **Installs** and **Reviews**.





## Percentage of Paid apps v/s Free apps

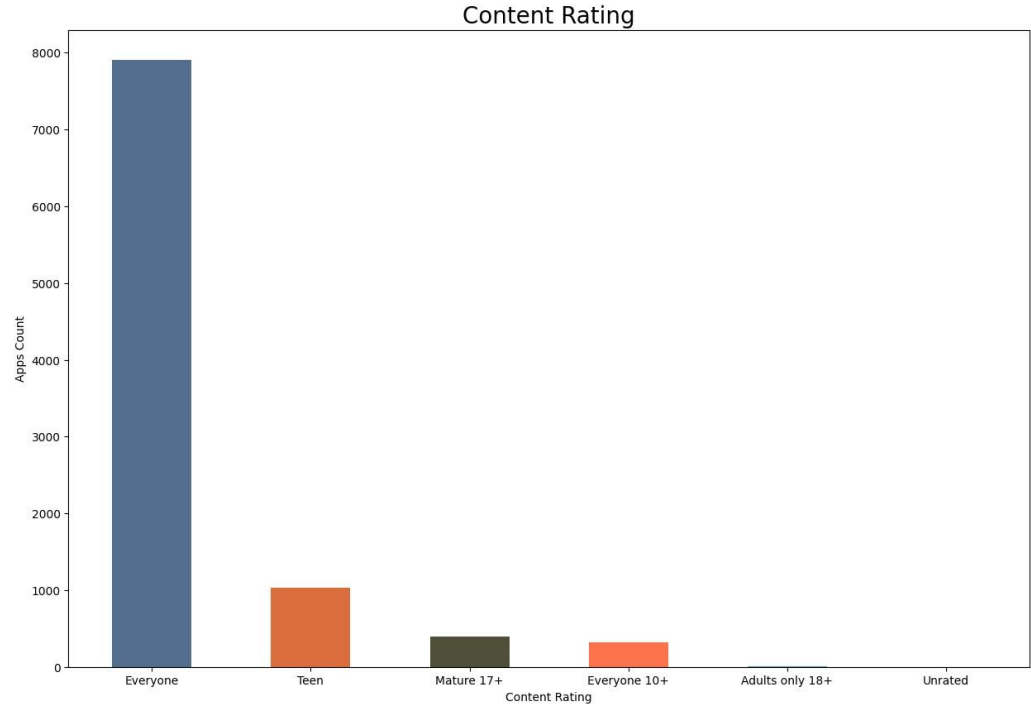
We Observed that **92% of Apps are free** and only **8% of Apps are paid** in Play store.



# Content Rating

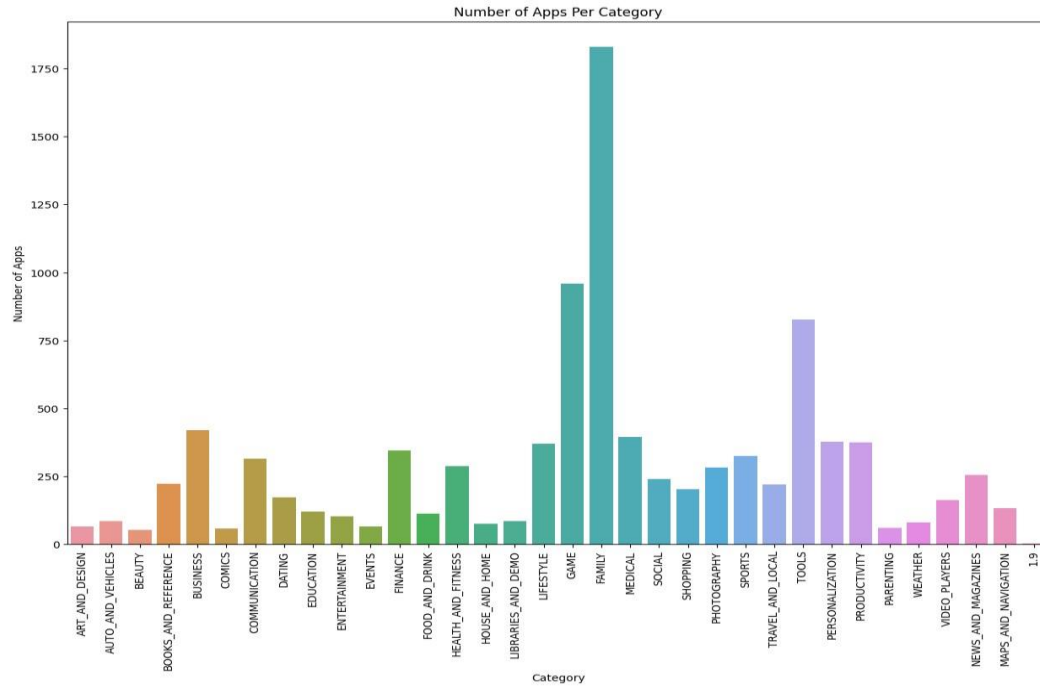
From the above plot we can see that Everyone category having majority of apps count.

A majority of the apps (81.80%) in the play store are can be used by everyone. The remaining apps have various age restrictions to use it.



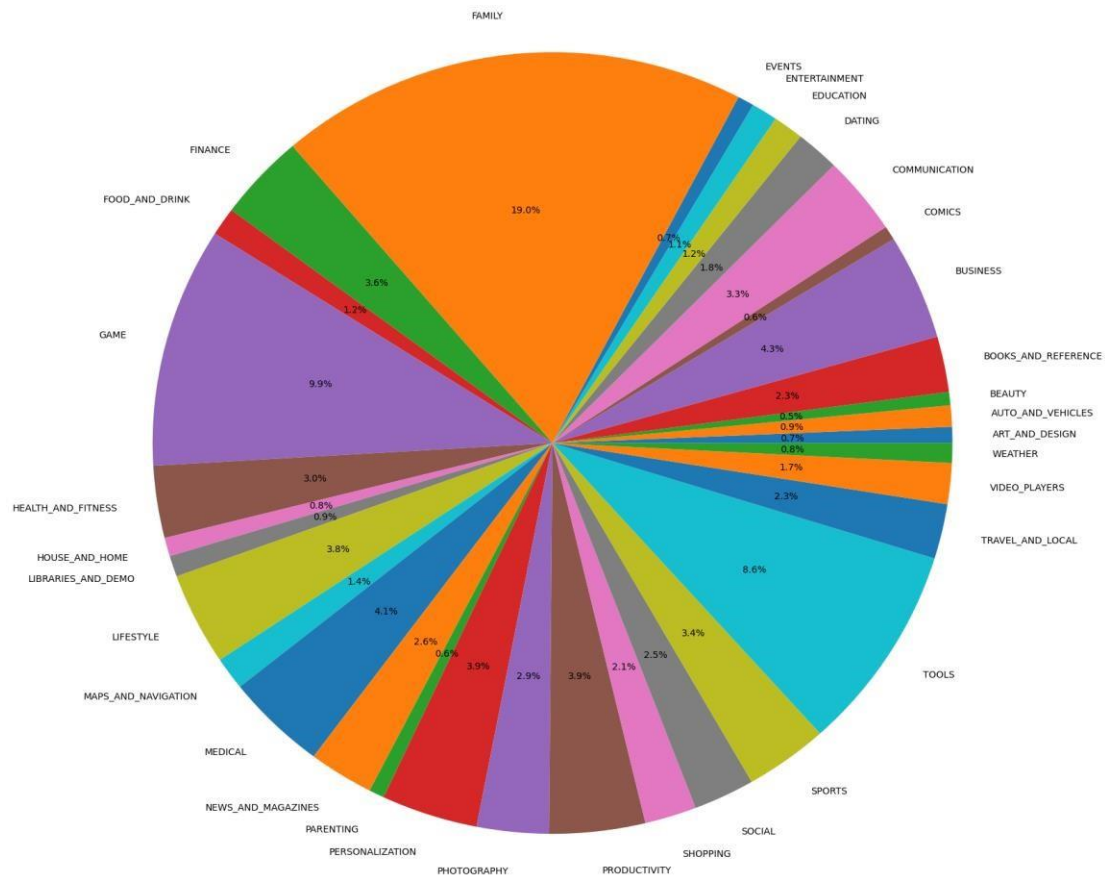
## Count of Applications in each category





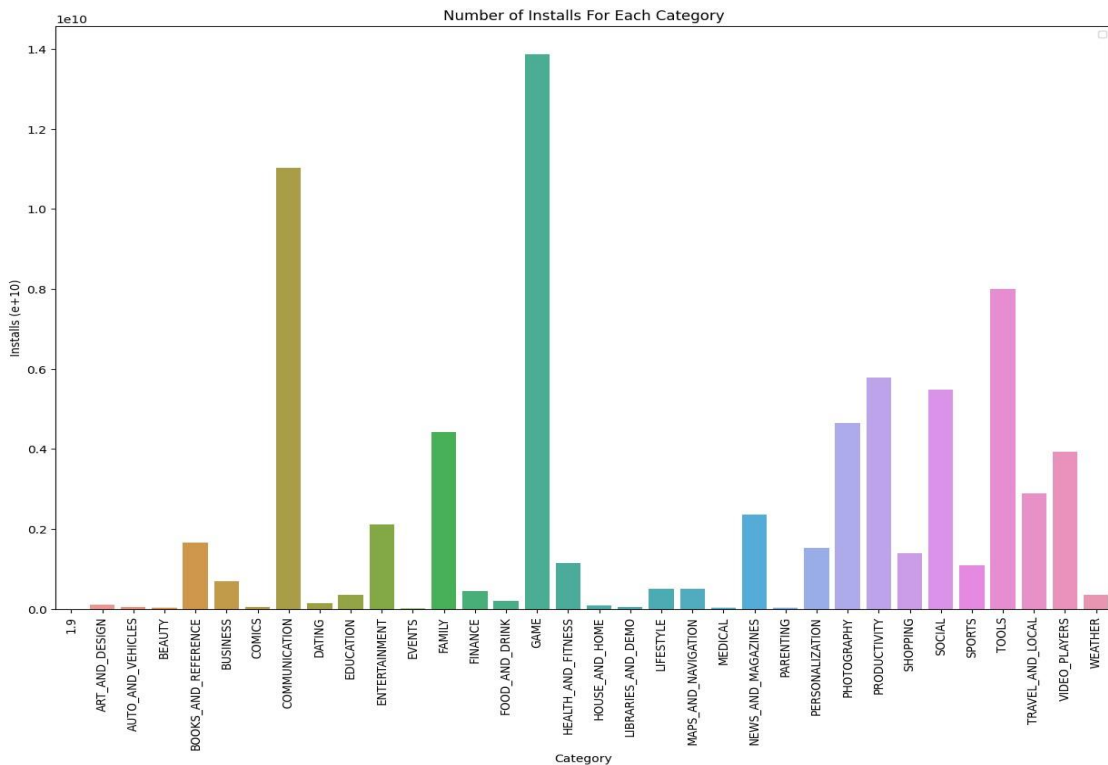
Family and Game apps have the highest market prevalence. Surprising Tools, Business and Medical apps are also at the Top Count of applications.

## Percentage of Applications in each category

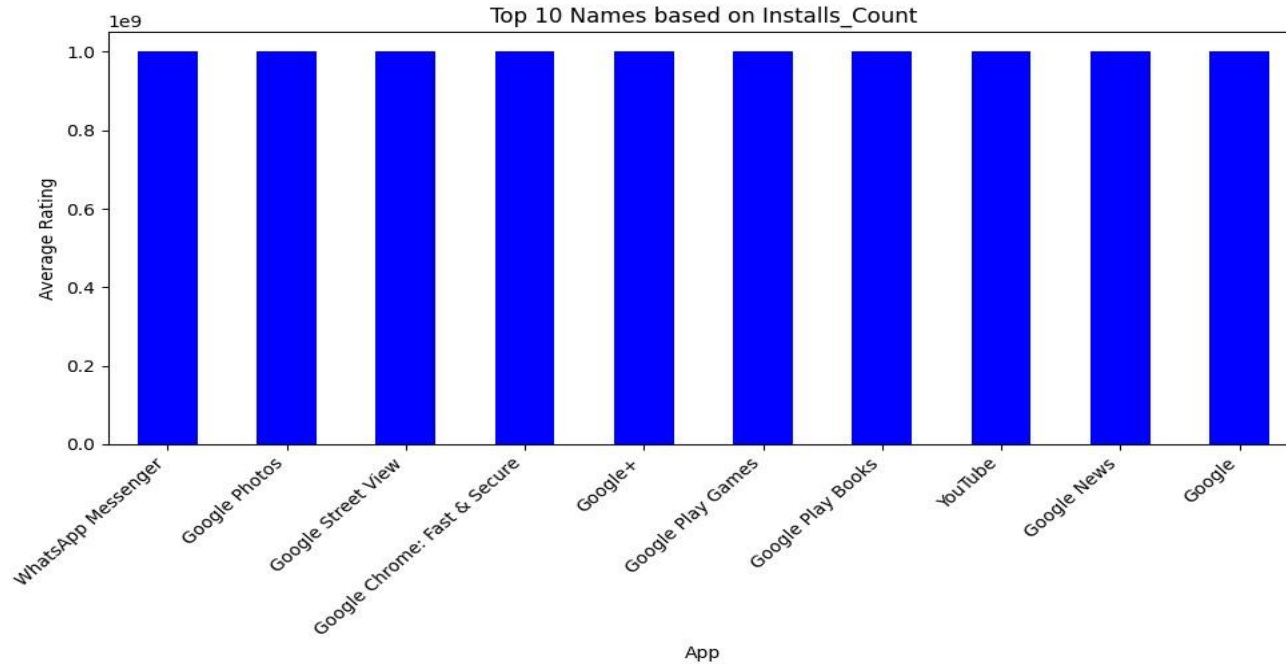


# Category App's have most number of installs

The Game, Communication and Tools categories has the highest number of installs compared to other categories of apps.



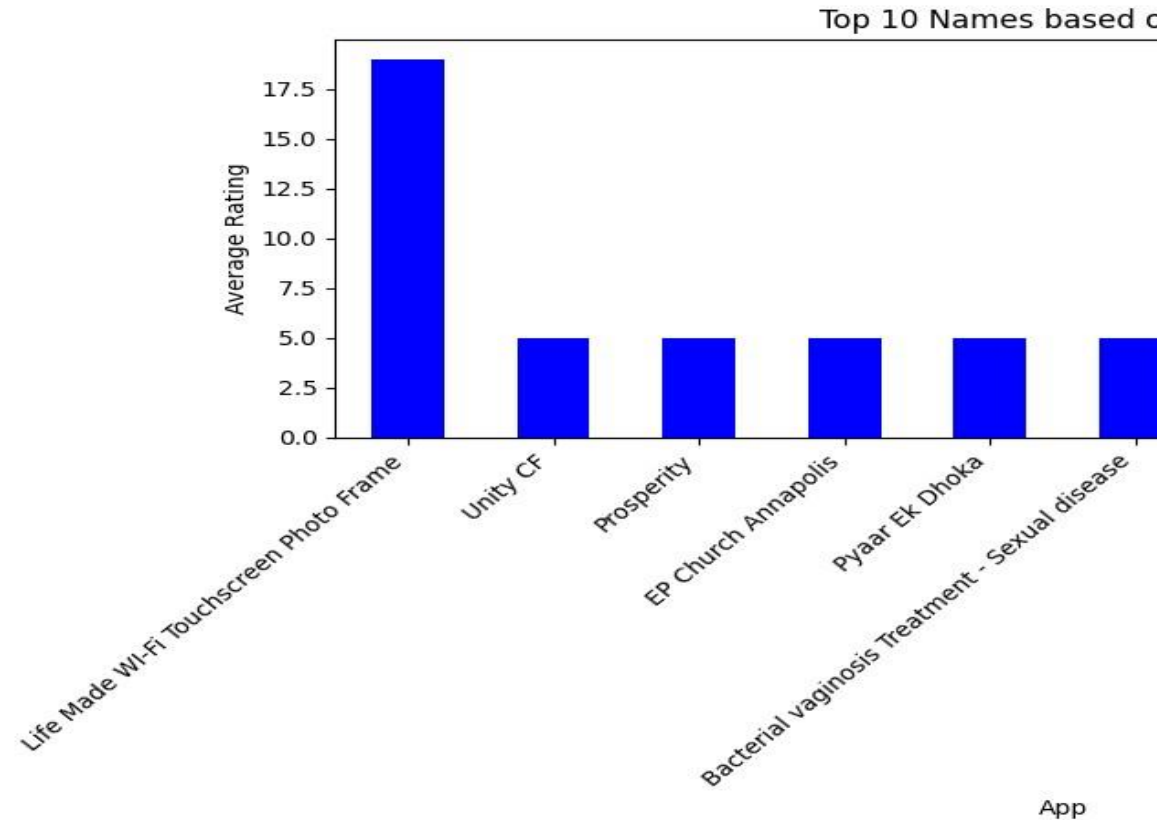
# Top 10 most installed apps in any category



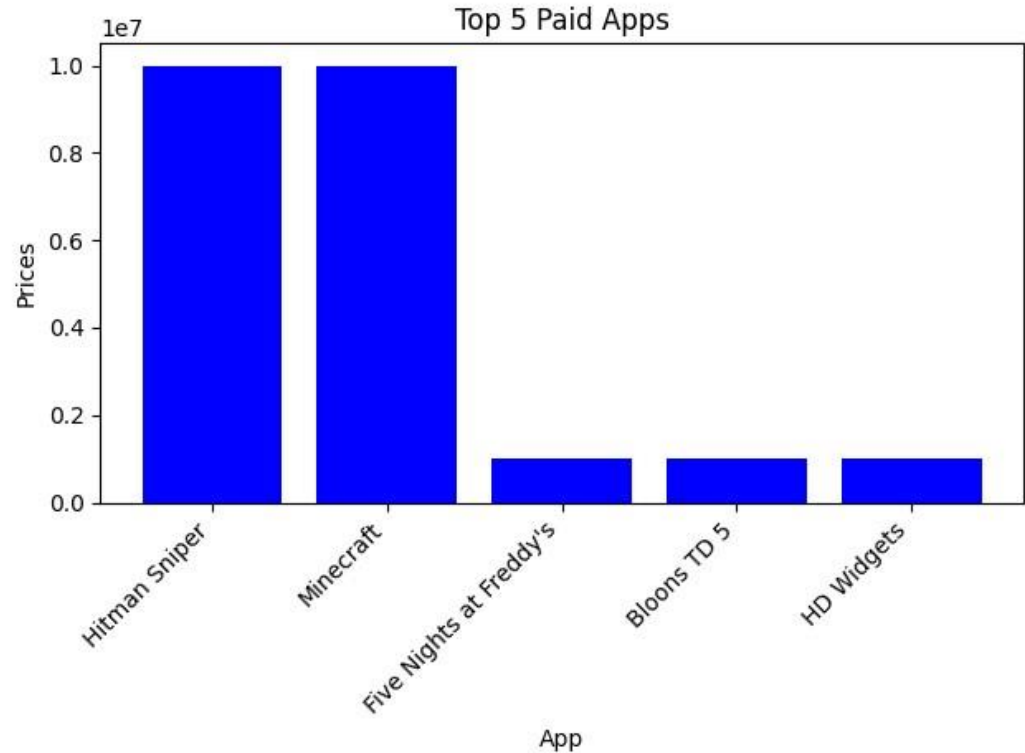
This graph shows the top 10 most installed apps are WhatsApp Messenger, Google Play Books, Subway Surfers, Maps - Navigate & Explore, Gmail, Instagram, Facebook, Google Chrome: Fast & Secure, Google+, YouTube.

- Hitman Sniper, Minecraft, Five Nights at Freddy's, Bloons TD 5, HD Widgets are the top paid most installed apps based on revenue generated.

## Top 10 apps based on Ratings



## Top 5 paid most installed app

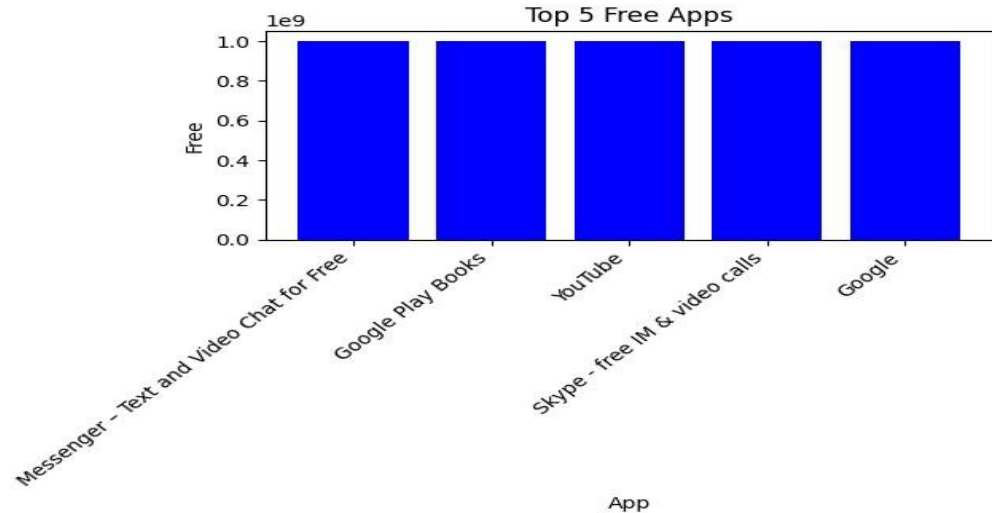




## Top 5 free most installed

1. Messenger – Text and Video Chat for Free
2. Google Play Books,
3. YouTube
4. Skype - free IM & video calls,
5. Google

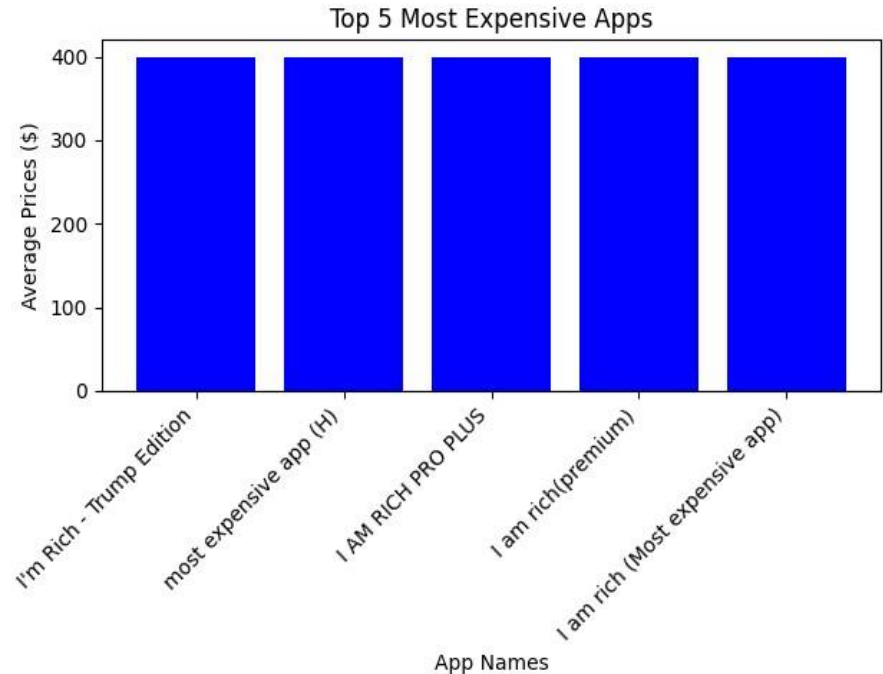
**top 3 most installed  
application category**



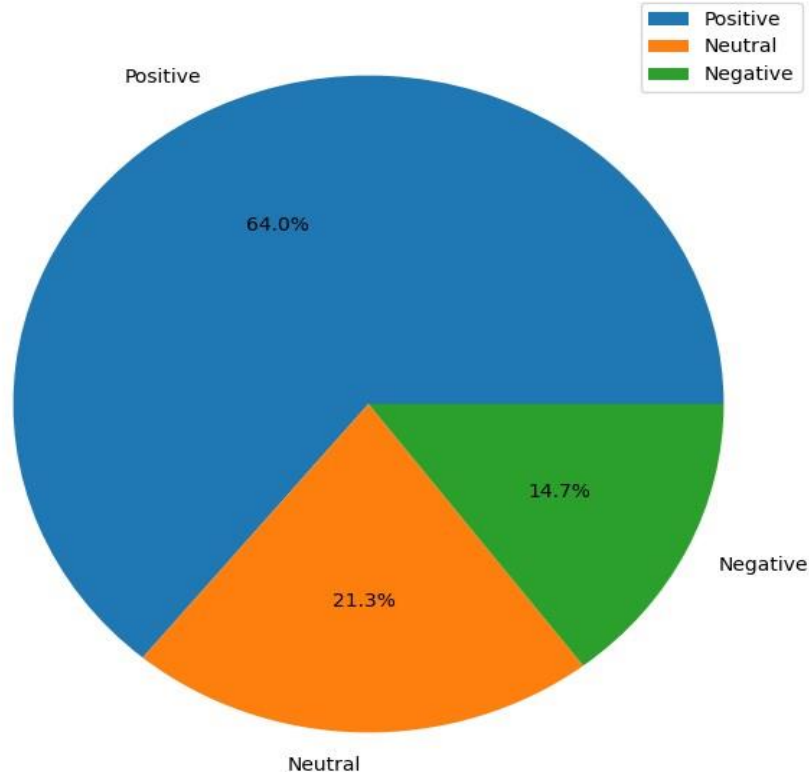
COMMUNICATION,BOOKS\_AND\_REFERENCE and TOOLS are the type of top 3 most installed applications.

## Top 5 Most Expensive Apps

- 1.I'm Rich - Trump Edition,
  - 2.most expensive app,
  - 3.I am rich(premium),
  - 4.I am Rich!, 5.I AM RICH PRO PLUS
- are the top 5 most expensive apps.



# Percentage of Review Sentiments



The number of Unique Apps from Play store and User reviews merged dataset are 816.

From Sentiment column, 64% are Positive, 21.3% are Negative and 14.7% are Neutral values.

## Challenges Faced Challenges Faced

- Reading the dataset and comprehending the problem statement.
- Examining the business KPIs for app development and devising a solution to the problem.
- Handling the error, duplicate and NaN values in the dataset.
- Designing multiple visualizations to summarize the information in the dataset and successfully communicate the results and trends to the reader.

## Conclusion

- 92.19% apps are Free and 7.81% apps are paid in type.
- **There are various categories amongst which there are multiple applications- Three major apps are Family, Tools and Games.**

- Applications receive both Ratings and Reviews and can be respectively graded too.
- Applications are installed by various users and mostly installed applications can be found, along with the information of their category, rating, type etc.
- Business, Game, Family and Tools have the highest number of applications.
- Free applications make 98% of the total of Applications, with a total count of more than 8000.
- Communication, Games and Tools are the most installed Application category.

## Conclusion[Contd.]

- Communication, Games and Tools are the most installed Application category.

- **Most of the apps with higher rating range of 4.0-4.7 are having high amount of reviews and installs. In terms of price, it doesn't reflect a direct relationship with rating, as we could see a fluctuation in term of pricing even at the range of high.**
- **Install Counts and Reviews are the most correlated and influence a lot of the downloads.**

**Thus a lot of factors play an important role in the user usage and popularity of any application. It can be because of the price of the application, based of the rating and reviews, also because of the number of installs, which in turn also affects the decision making of any app.**