Capstone Project-1

Play Store App Review

Team Members-

Md. Nawab Ali

Jahnavi Jaolekar

Kaustubh M.Amare

OVERVIEW OF DETAILED ANALYSIS ON PLAY STORE APPS

- Introduction
- Problem Statement
- Material
- Understanding of the Dataset
- Steps Involved
- Exploratory Data Analysis
- Conclusion

INTRODUCTION

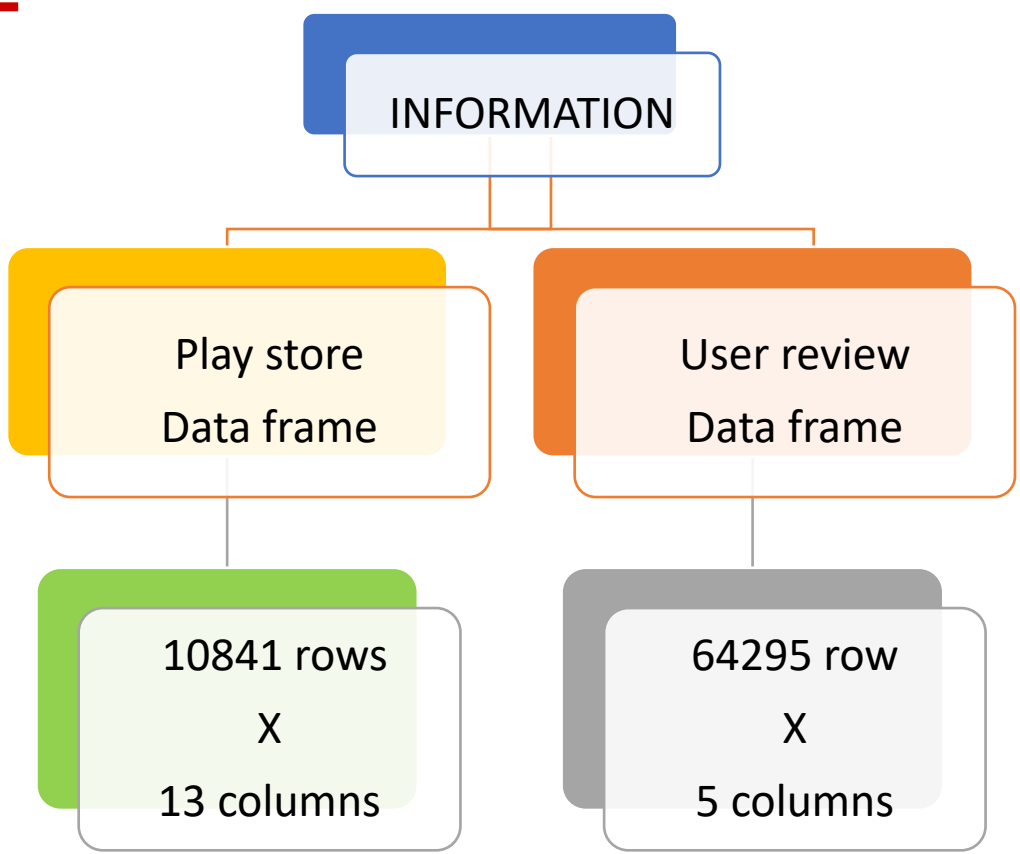
- An enormous range of mobile apps have been developed as a result of the rising popularity of mobile operating system-enabled devices like smartphones and tablets.
- Users may find and enjoy their preferred Android applications, games, movies, TV shows, books, and other stuff on Play Store.
- It contains more than 3.5 million Android Apps and it has a user base of more than 1 billion individuals.
- Actionable insights can be drawn for developers to work on and capture the Android market.

PROBLEM STATEMENT

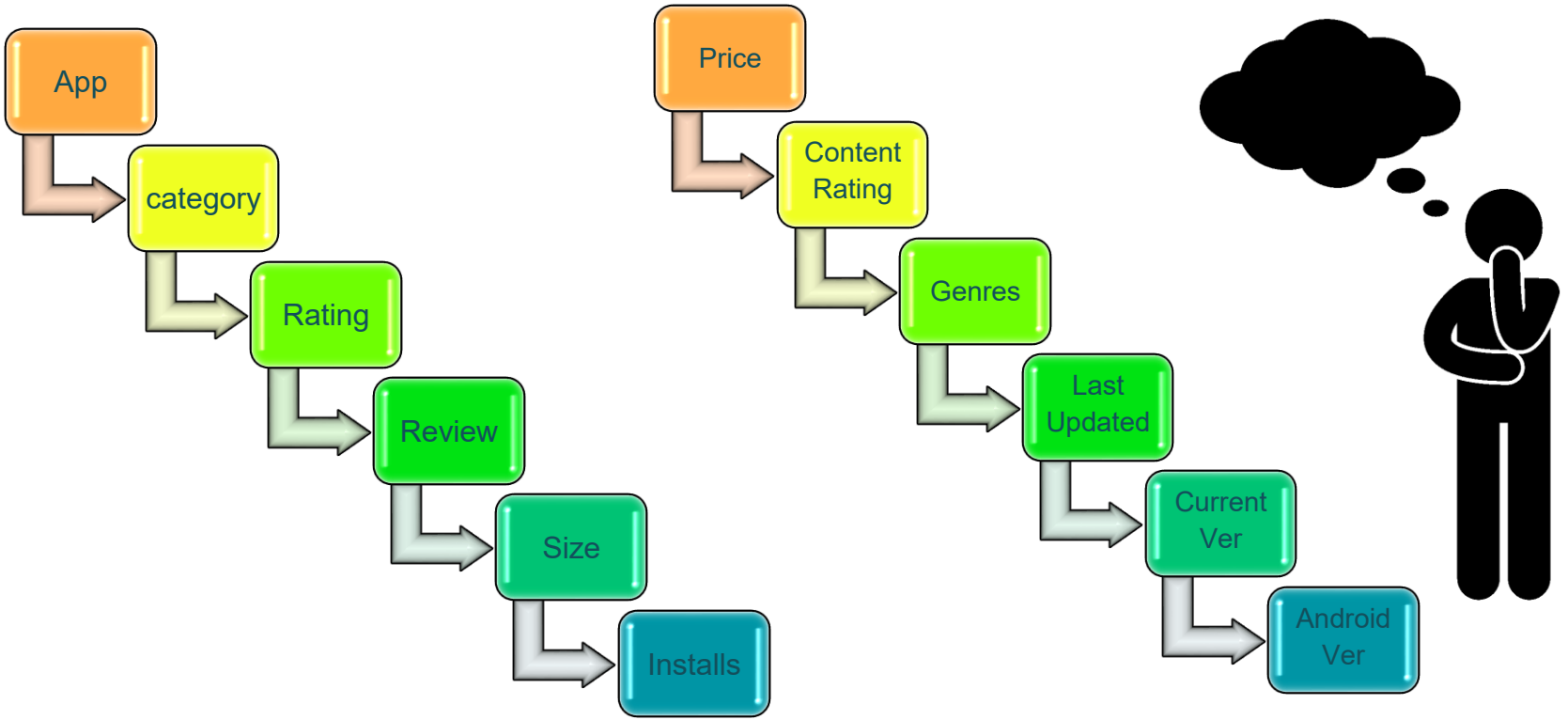
- To explore and analyze the data to discover key factors responsible for app engagement and success.

- Our mission is to assist Android developers in understanding what is the motivating factor for people to download an app.
- For this reason, we want to examine the relationships between installs, prices, ratings, reviews, and categories.

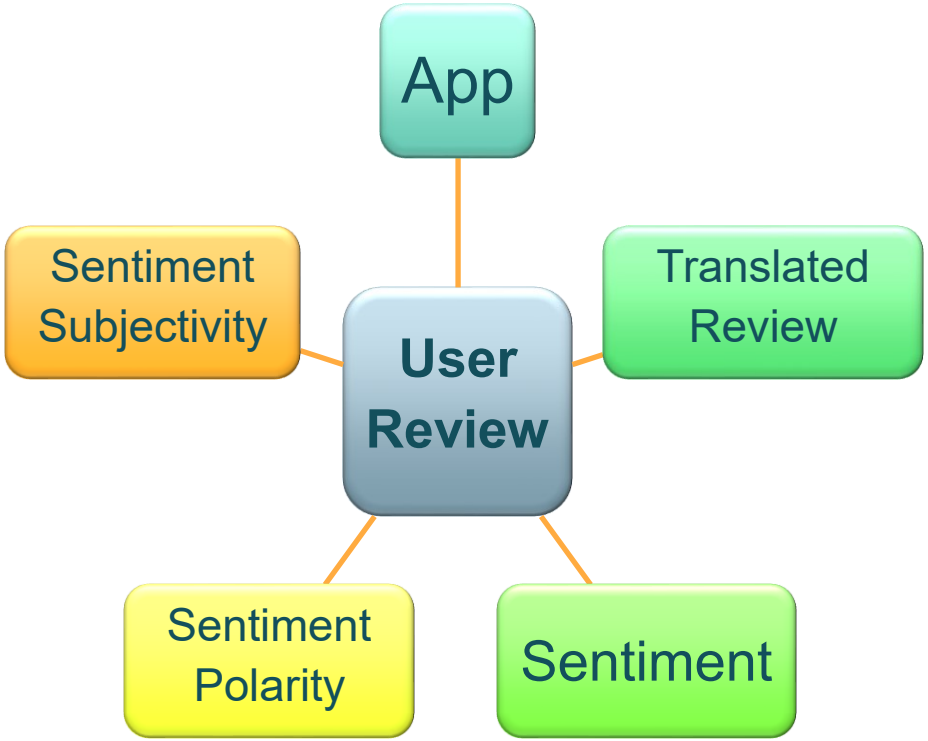
MATERIAL

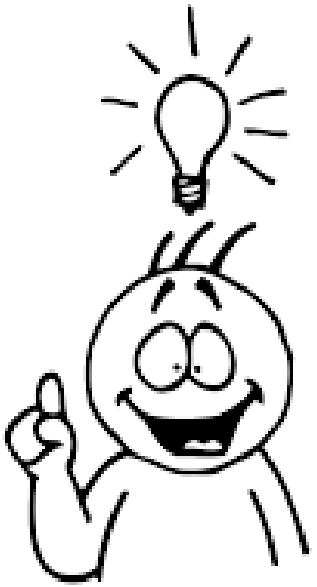

INFORMATION
Play store
Data frame
User review
Data frame
10841 rows
X
13 columns
64295 row
X
5 columns

UNDERSTANDING OF THE PLAY STORE DATA



UNDERSTANDING OF THE USER REVIEW DATA



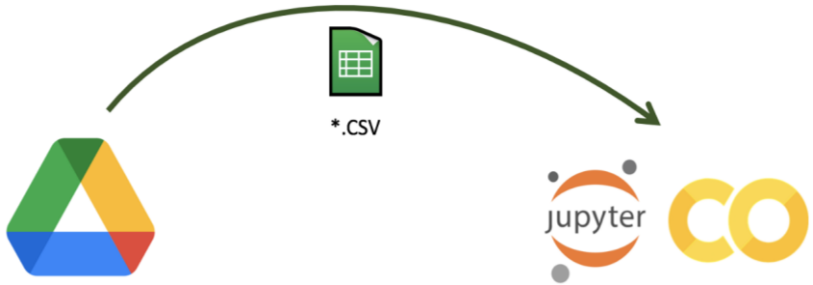


STEP 1: IMPLEMENTING THE DATASET

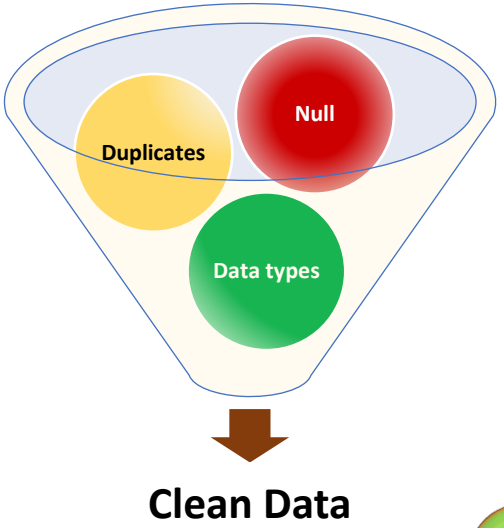
First we have imported the necessary libraries as:-

- Import numpy as np
- Import pandas as pd
- Import matplotlib.pyplot as plt
- Import seaborn as sns
- Import warning

- Then we have loaded the CSV files from google drive
- The given two dataset one is play store data review and second is user review dataset both were imported and defined as play_store_df & user_review



STEP 2: CLEANING THE DATA



- Removal of unwanted Observations
- Fixing Structural Errors
- Managing Unwanted Outliers
- Handling missing Data
- Removing data Duplication
- Final check

STEP 3: PREPARING THE DATA

Elimination of unnecessary details such as Last Updated, Current Ver, and Android version.

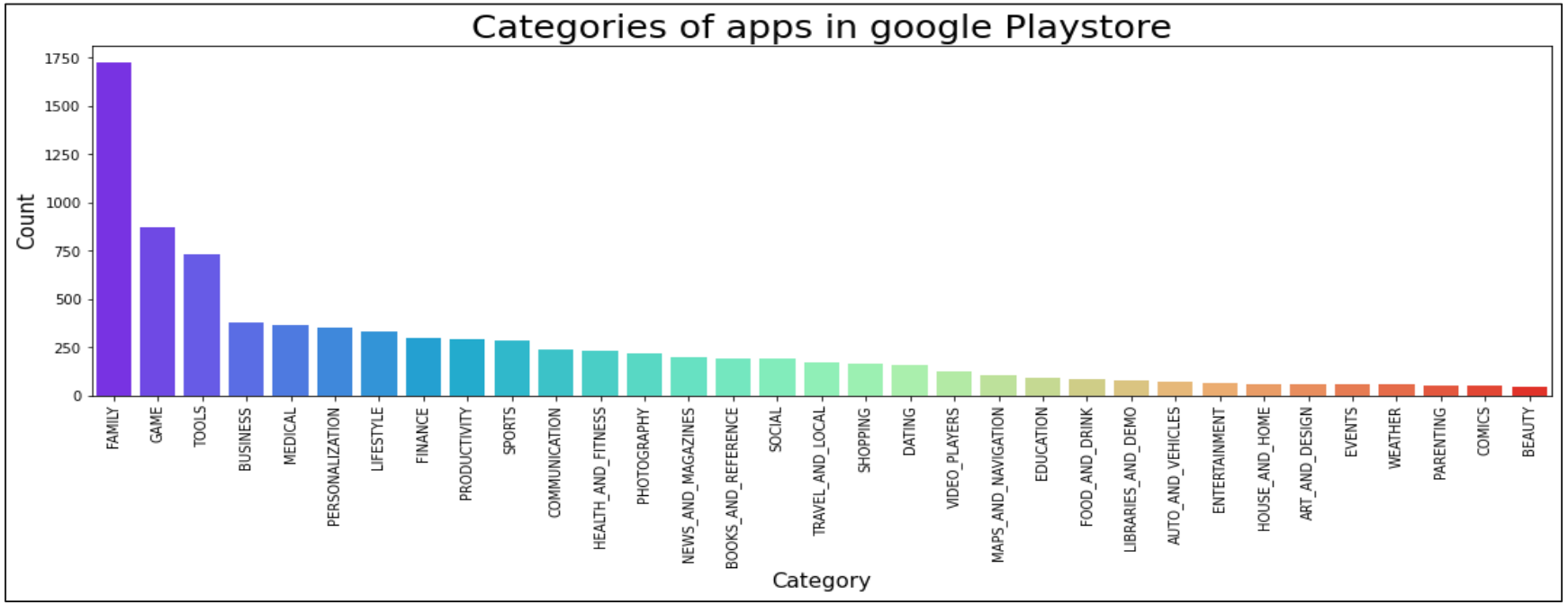
Convert the data type of variables such as Reviews, Size, Installs, and pricing, data imputations.

Conversion from type object to int or float.

STEP 4: EXPLORATORY ANALYSIS AND VISUALIZATION

Volume of apps

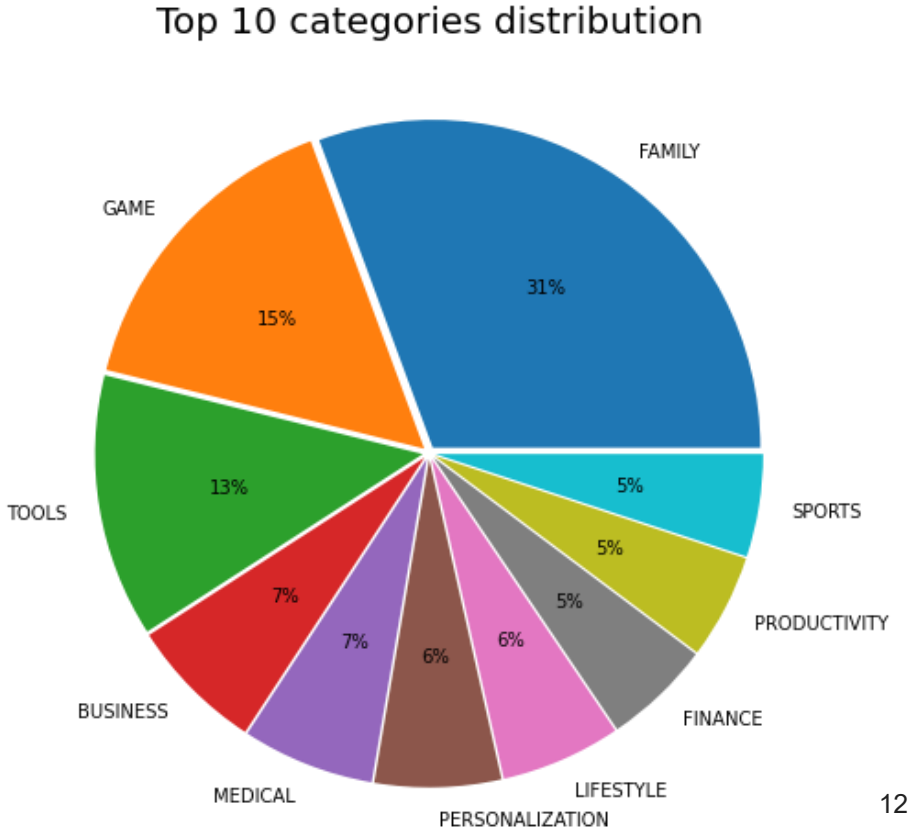
Total 33 unique categories are observed. This graph represents the volume of apps falling under each category



Top 10 Categories distribution

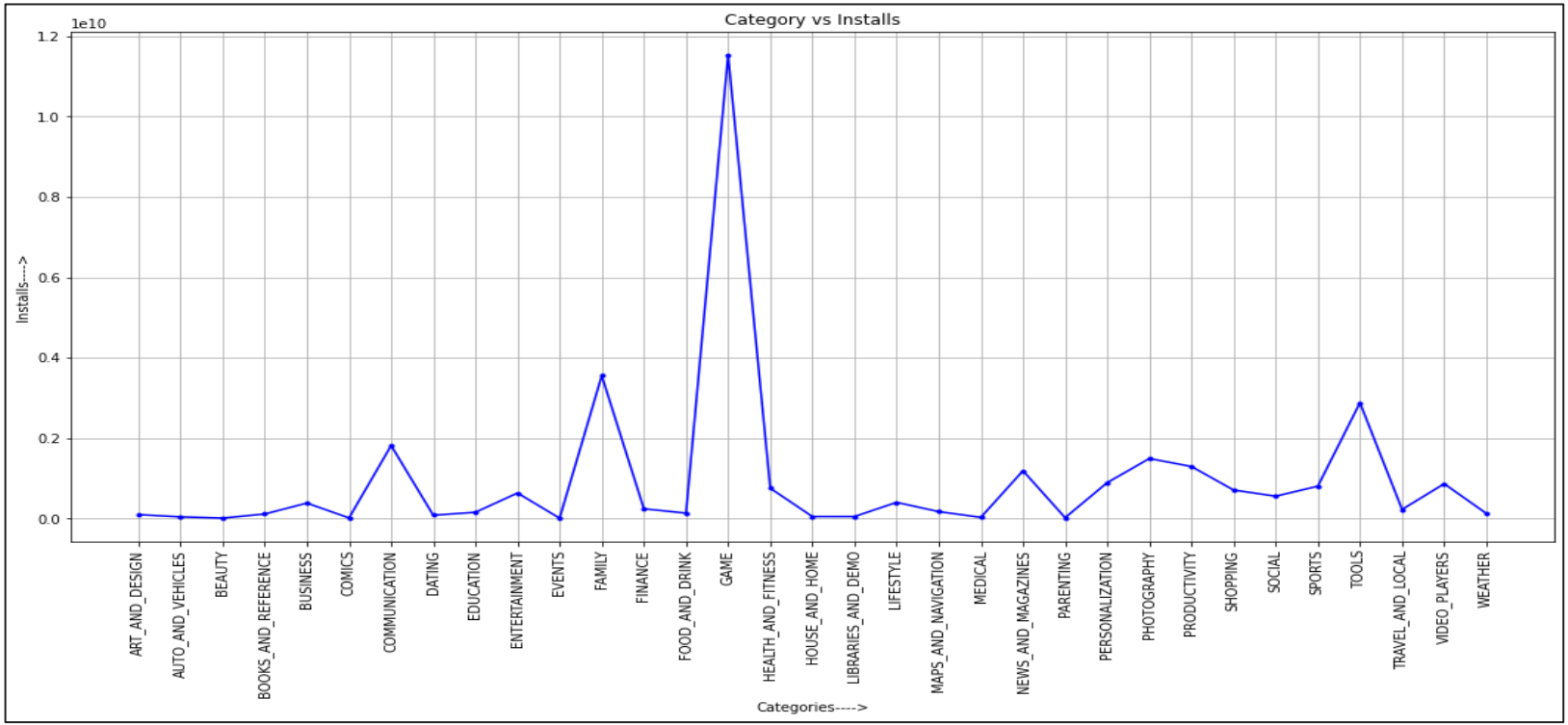
This pie chart represents the volume distribution of apps on google playstore in top 10 categories.

Among Top 10 Categories Family(31%), Games (15%) And Tools(13%) Contribute The Most



Category v/s Installs

‘Game’ category have maximum number of downloads, followed by ‘Family', 'Tools’ and ‘Communication’ respectively.

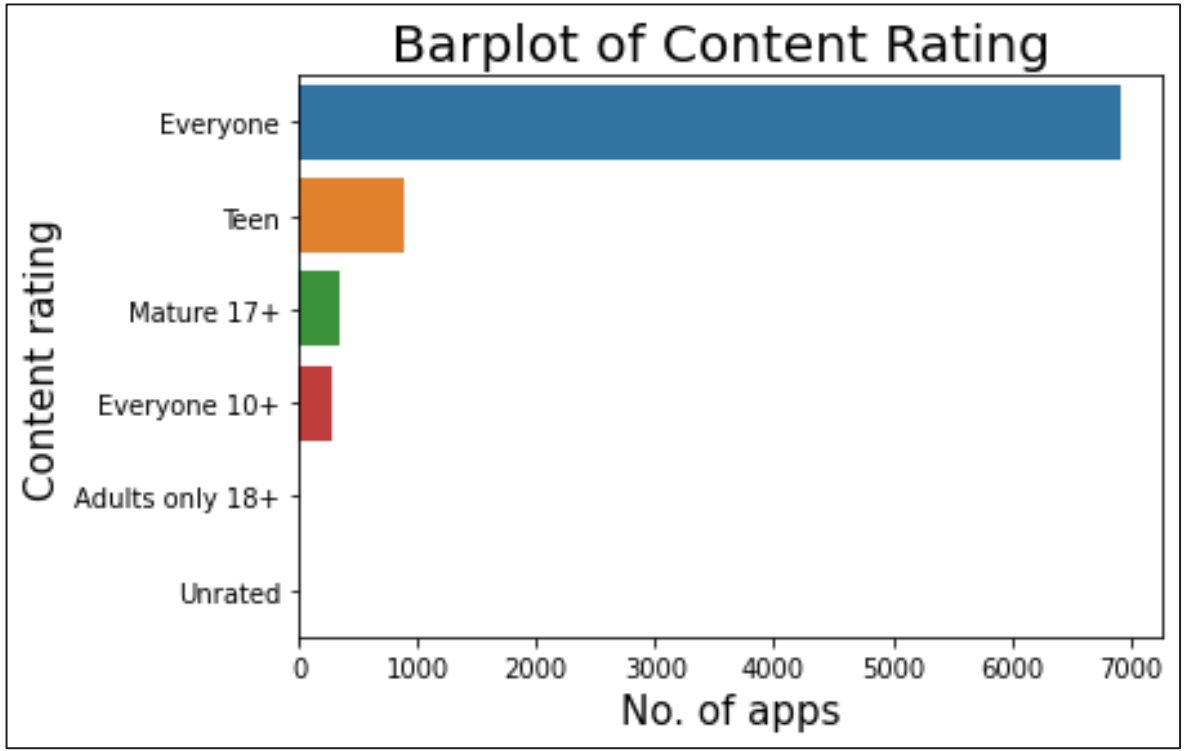


Top 10 5-star rated Categories

Total 266 in the data frame have 5 star rating, Amongst which 204 falls under top 10 category

Rank	Category	App Count
1	Family	65
2	Lifestyle	29
3	Medical	25
4	Business	18
5	Tools	17
6	Game	12
7	Health and fitness	12
8	Personalization	10
9	Social	8
10	Productivity	8
Total		204

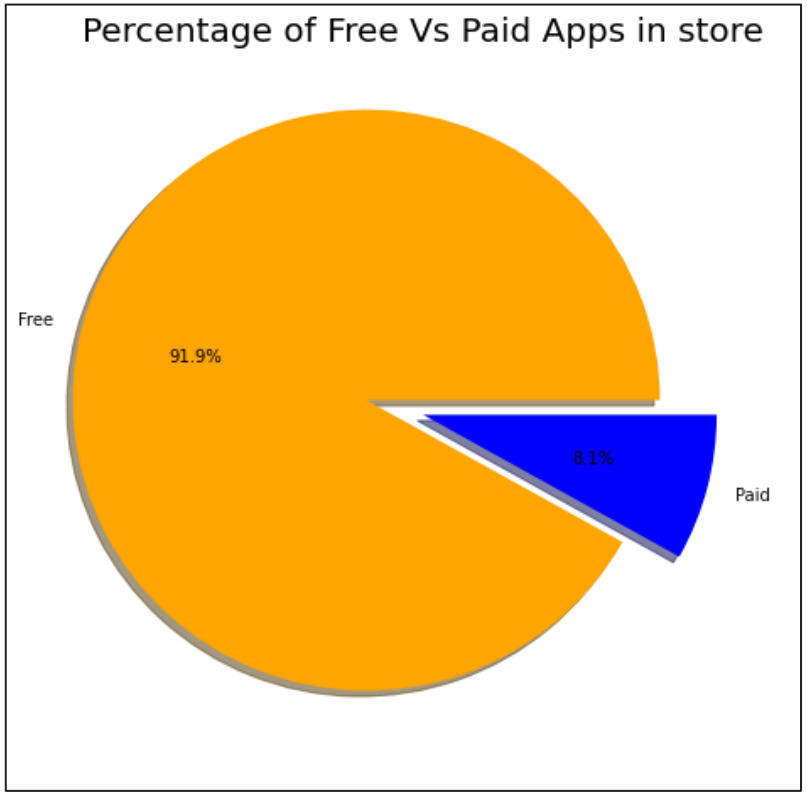
Content Rating



The Everyone category has the most apps, as can be seen from the plot above.

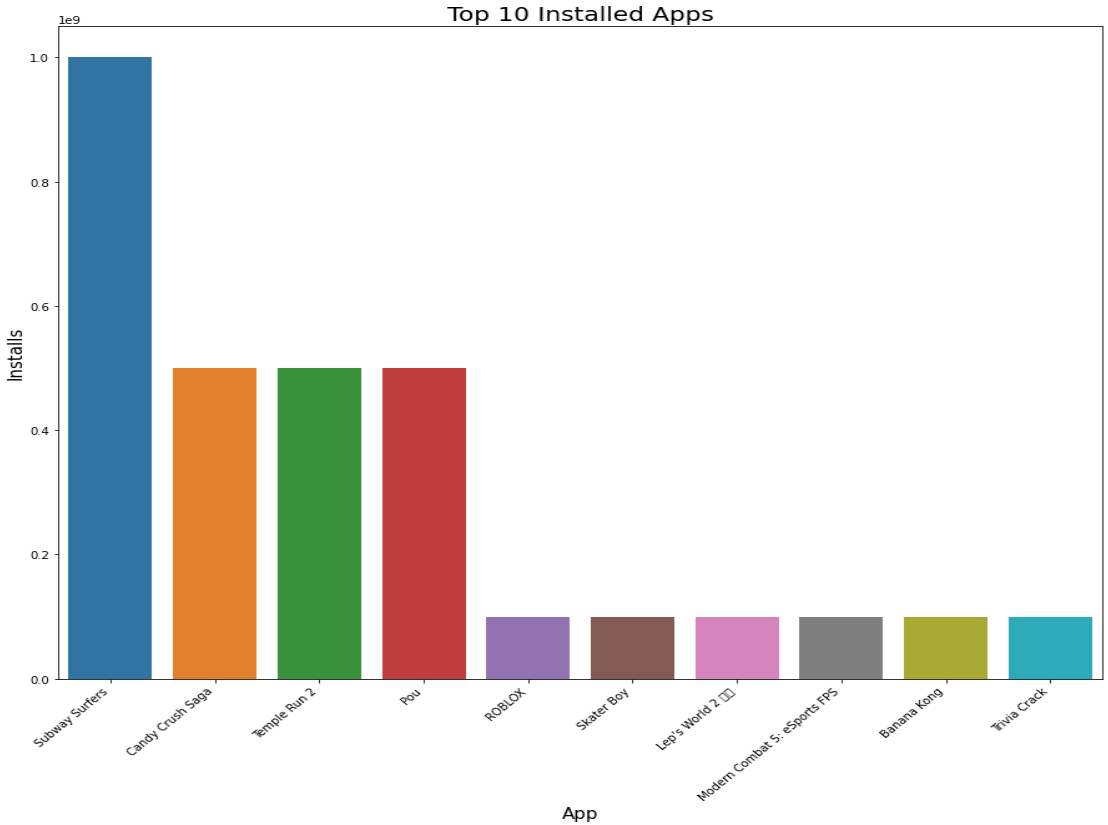
Free vs Paid apps

On an average 90% of apps on play store are freely available
Thus attracting larger audience base.



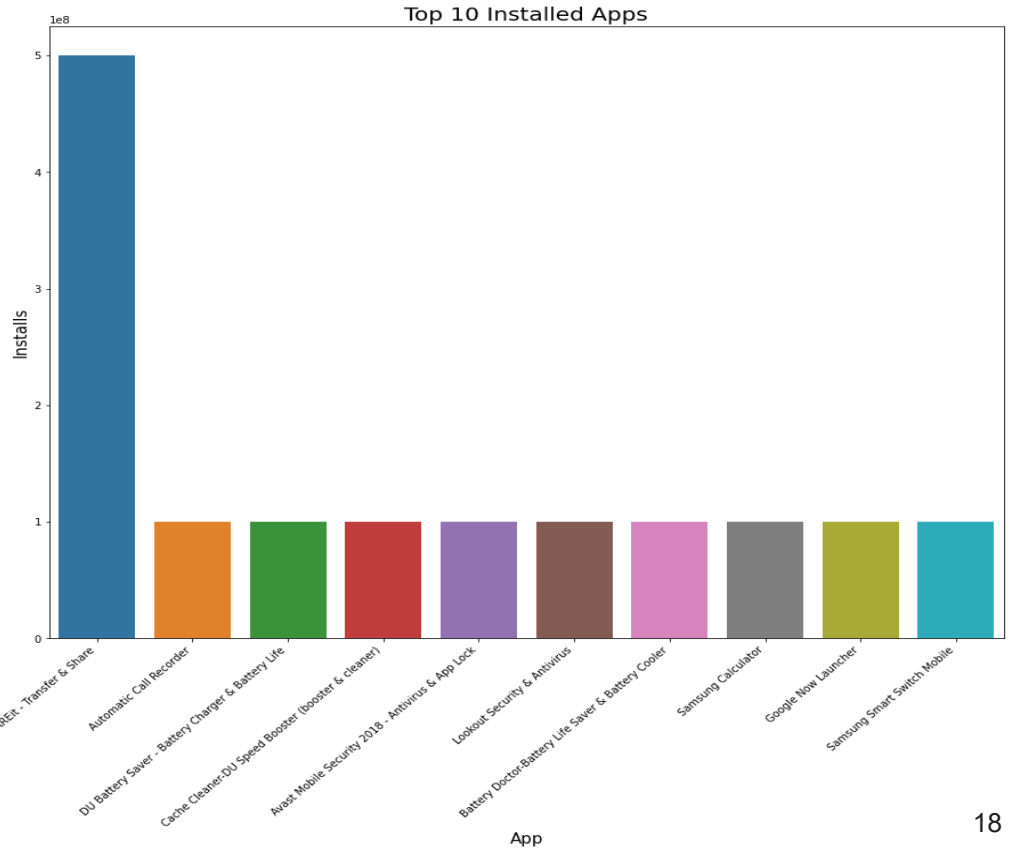
Top 10 installs in Game

Thus,
Subway surfers, Candy crush, temple run2, Skater bot, Modern combat 5 & Banana Kong etc. are some of the apps having highest installs in ‘Game’ category.



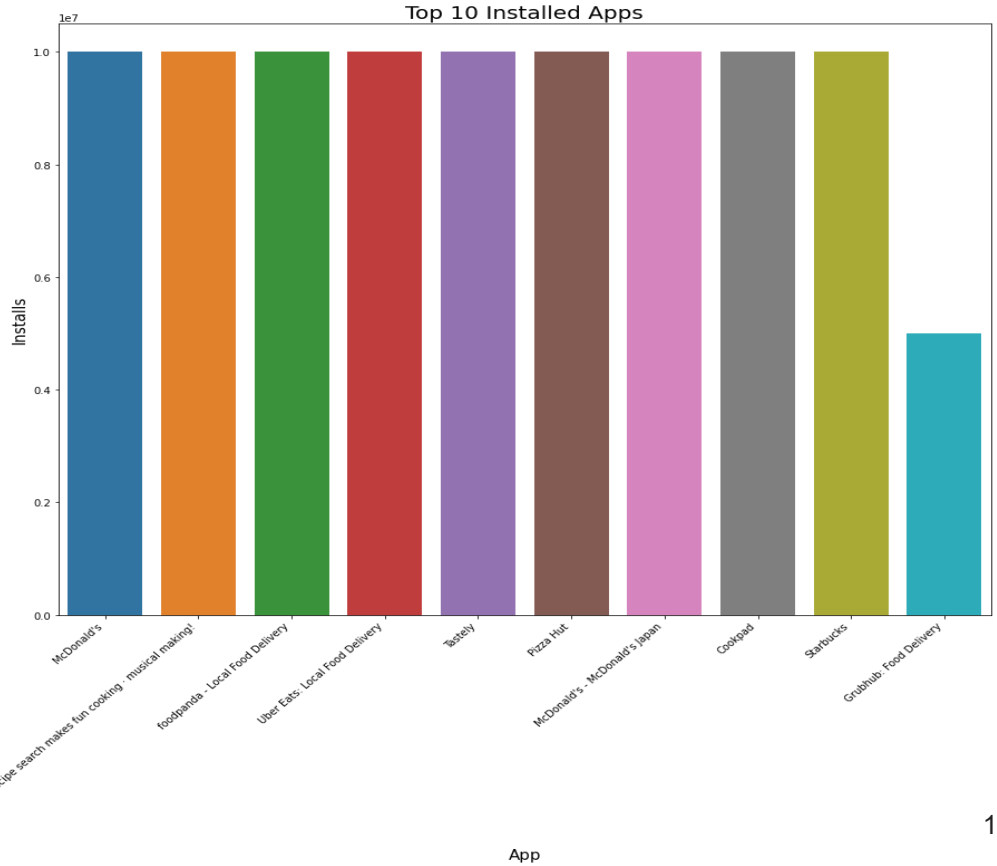
Top 10 installs in Tools

Thus,
Shareit, Automatic call recorder, Du battery saver, Cache cleaner Du, Avast mobile security are some of the top 10 apps installed from ‘Tools’ category.



Top 10 installs in Food & drink

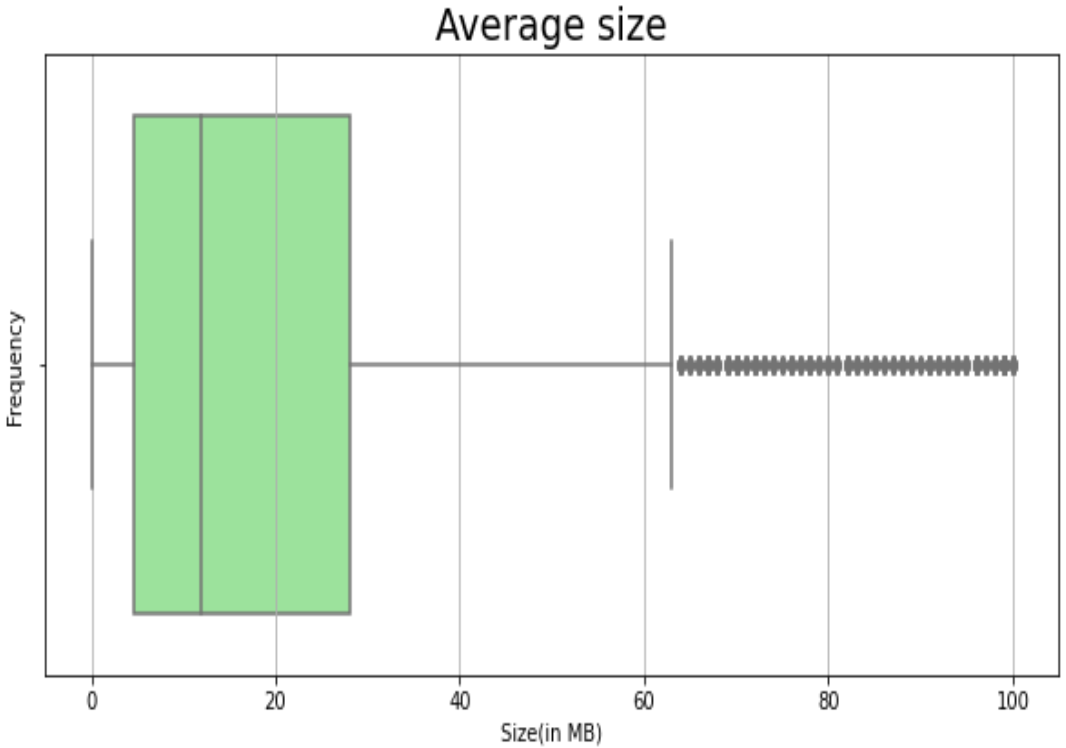
Thus, McDonald's, food panda, UberEATS, Pizza hut and Starbucks are some of the highly downloaded apps in "Food and drink category"



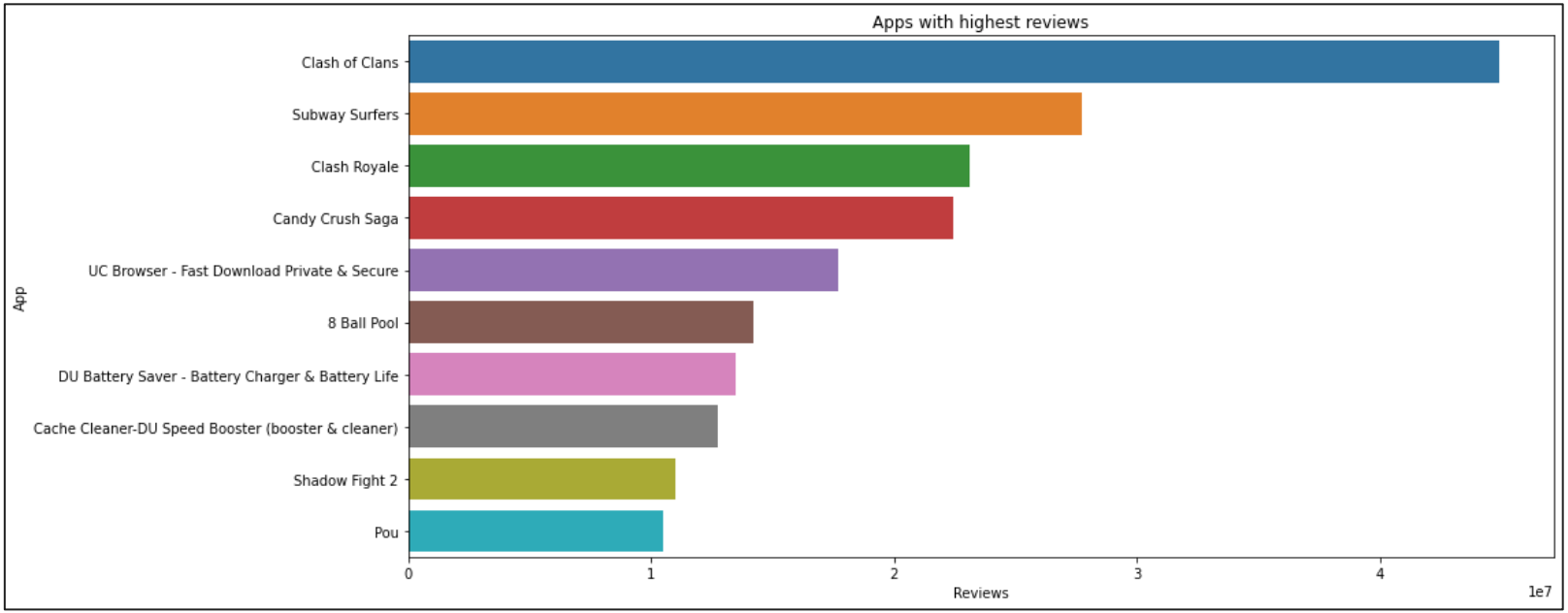
Size of apps

As we can see from the box plot above,

- 25% of apps are under 5 MB
- 25% of apps are between 30 Mb to 63 Mb
- Majority i.e. 50% lies between 5 Mb to 30 Mb.
- Many outliers are all the way upto 100 Mb
- Median is 12 Mb

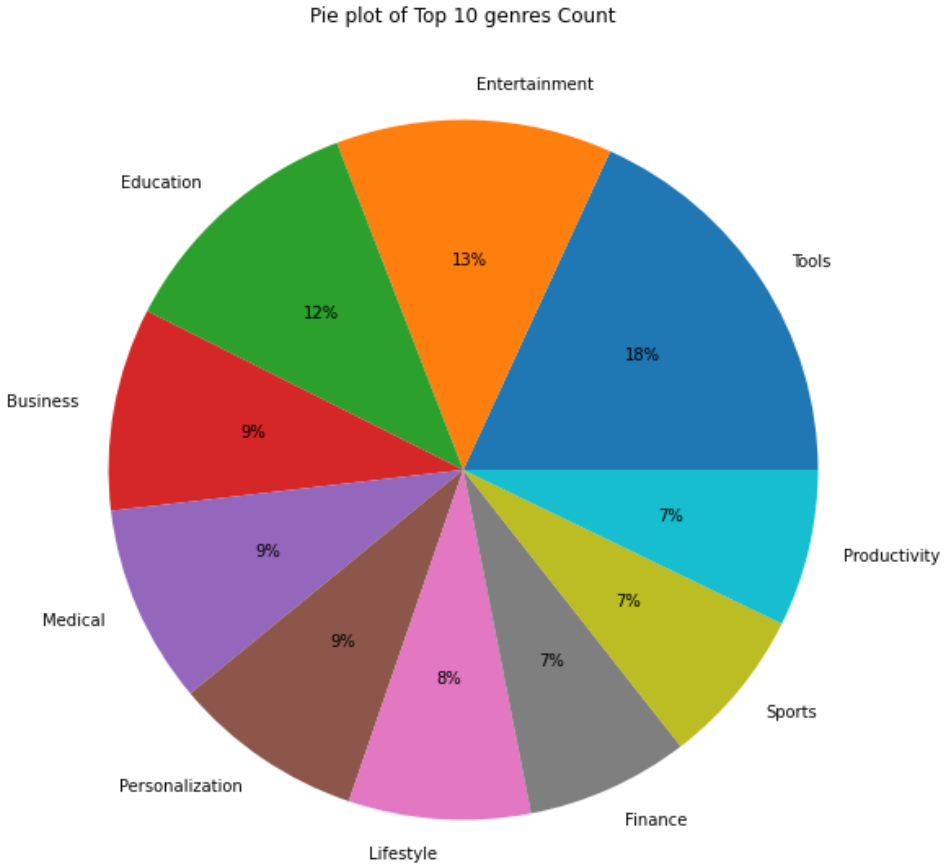


Highly reviewed apps



As we can see that Clash of clans , Subway surfers, Candy crush, UC browser , Shadow fight 2 etc. are some of the highest reviewed apps

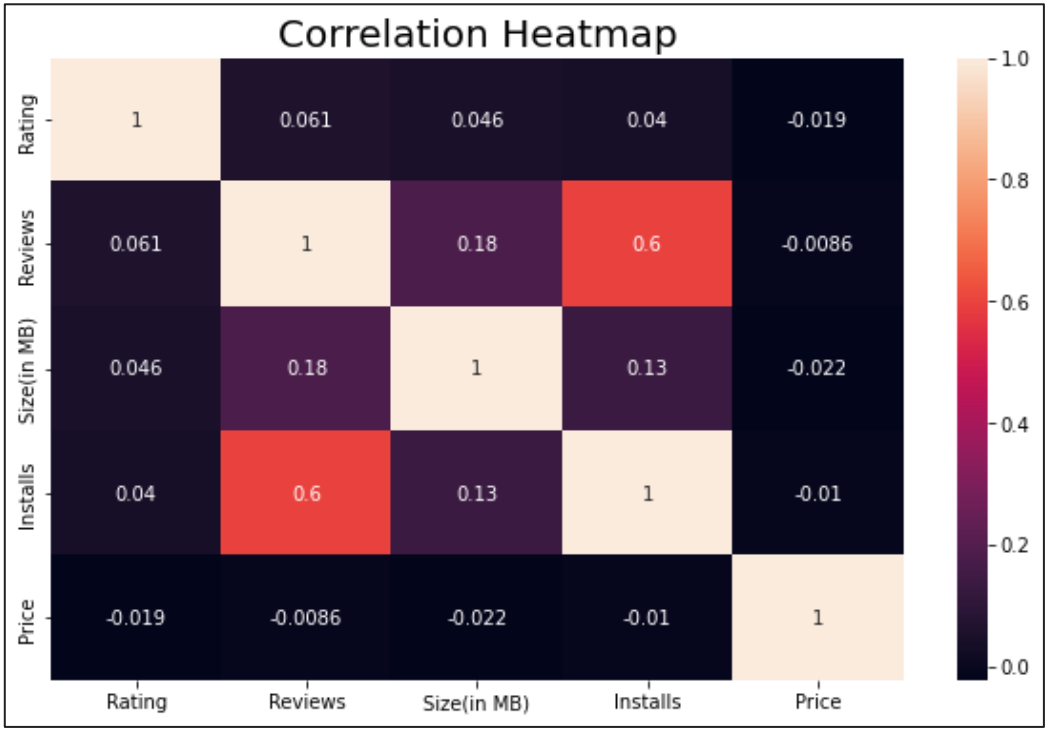
Top 10 Genres



Correlation heat map

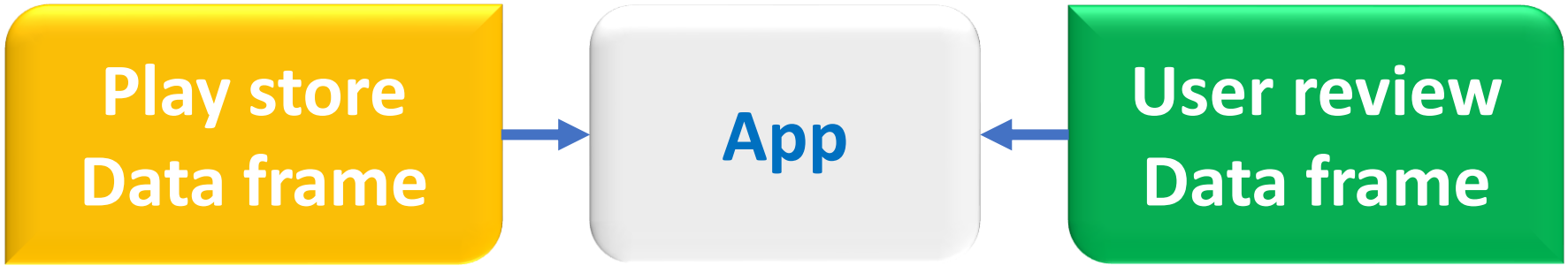
Thus from the heat map we can observe as follows: -

- Reviews and installs have positive correlation
- Size and install have week correlation
- Rest of the factors are not affecting each other directly



Joining Data frames

Starting with data cleaning of User reviews data frame
Then joining User review data frame to Playstore data frame on common column ‘App’

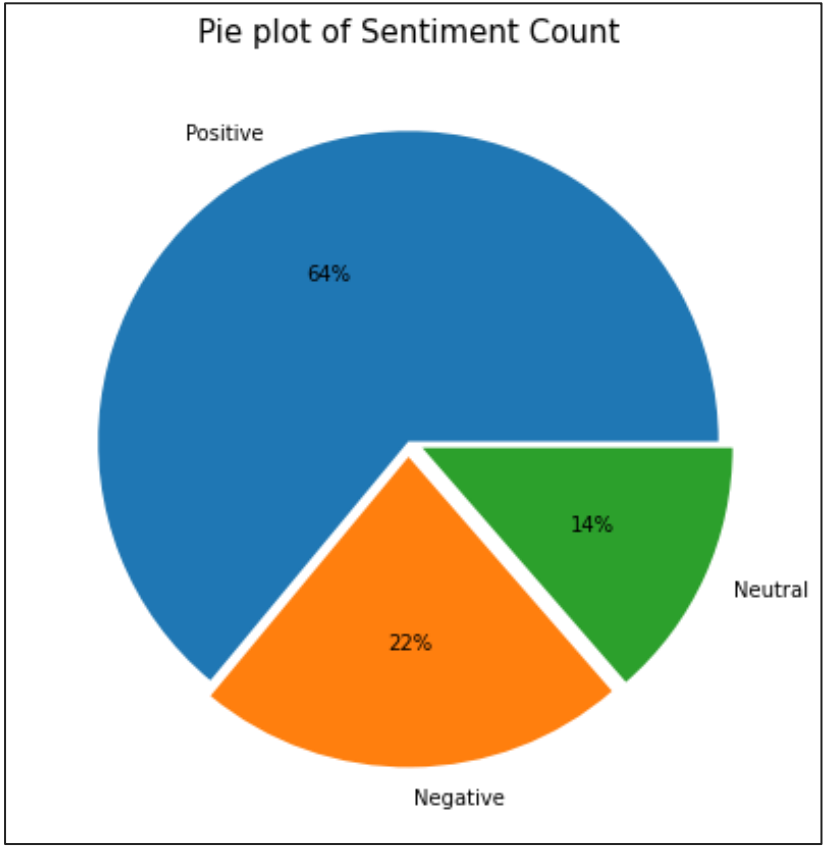


Sentimental Analysis

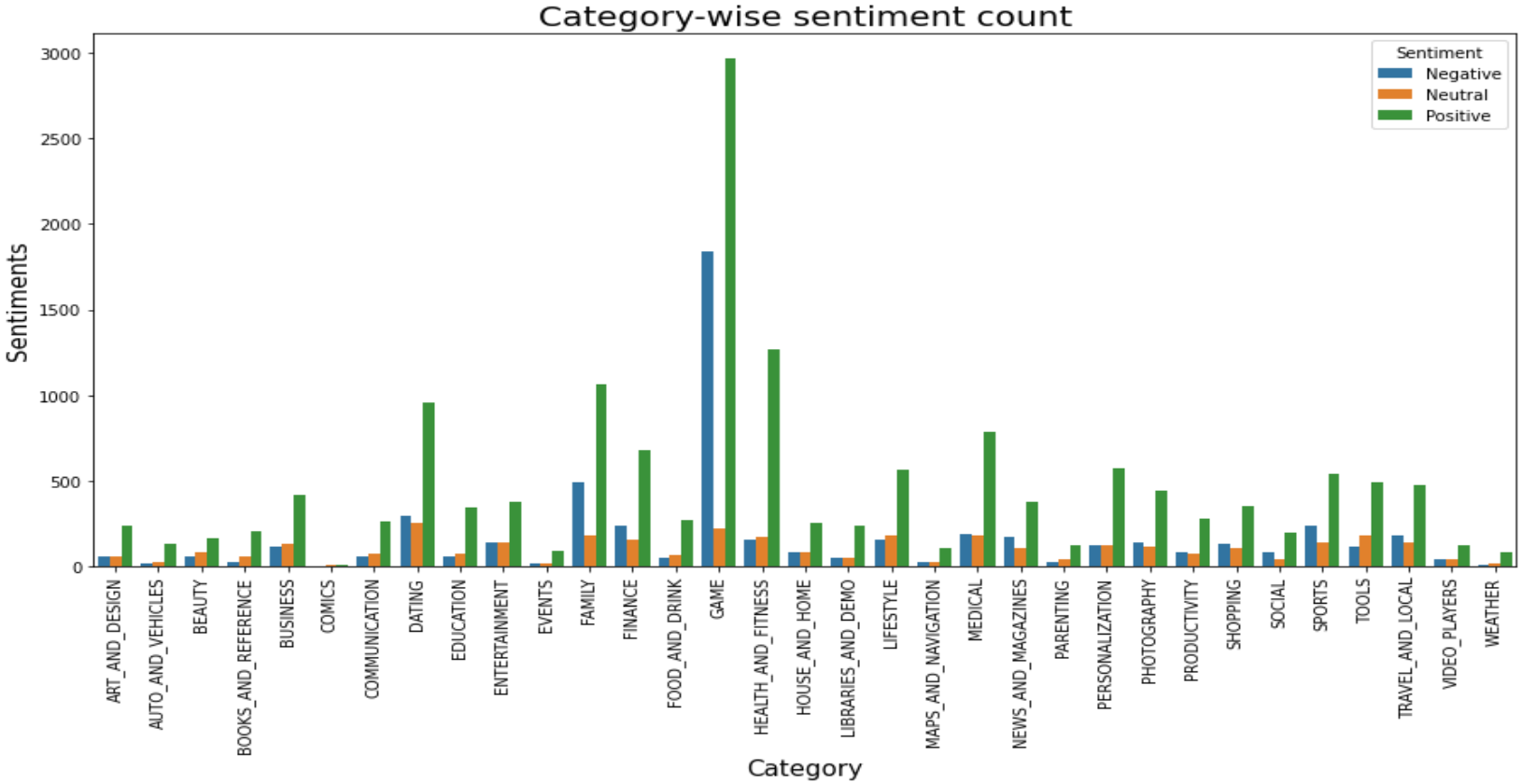
This chart interprets , as follows:-

- 64% positive reviews
- 14% neutral reviews
- And only 22% negative reviews.

Considering the large data base of apps we can conclude the maximum apps on play store have good reviews thus making the store trust worthy and helpful.



Sentimental Analysis vs category



CONCLUSION

- Joining the dots from given datasets, we came across a lot of information.
- After analysis and visualization, information can be classified into beneficial for customers and beneficial for developers.
- We discovered the top app categories, the most popular app categories, the proportion of free and paid apps, and the average app size. This information enables customers and users to make informed download decisions.
- Given that games receive the most installs, developers may want to consider gamifying their material (Educational Apps).

THANK YOU...