Capstone Project

Play store App review Analysis

By

Ankur Nain

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 .
- Android is the dominant mobile operating system today more than 85% of all mobile devices running Google’s OS. 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

❖WHY ANALYZE THE GOOGLE PLAY STORE?

- Mobile App Market is set to grow 20% by 2023
- Android Apps comprise 90% of the Mobile App Market
- What makes an App popular? Can we predict how popular it’s going to be?
- What are some interesting patterns in user behavior related to app usage & feedback

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
- Average sizes of apps in each category
- Category wise percentage of paid apps
- Category wise top installed paid apps
- Average rating 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.

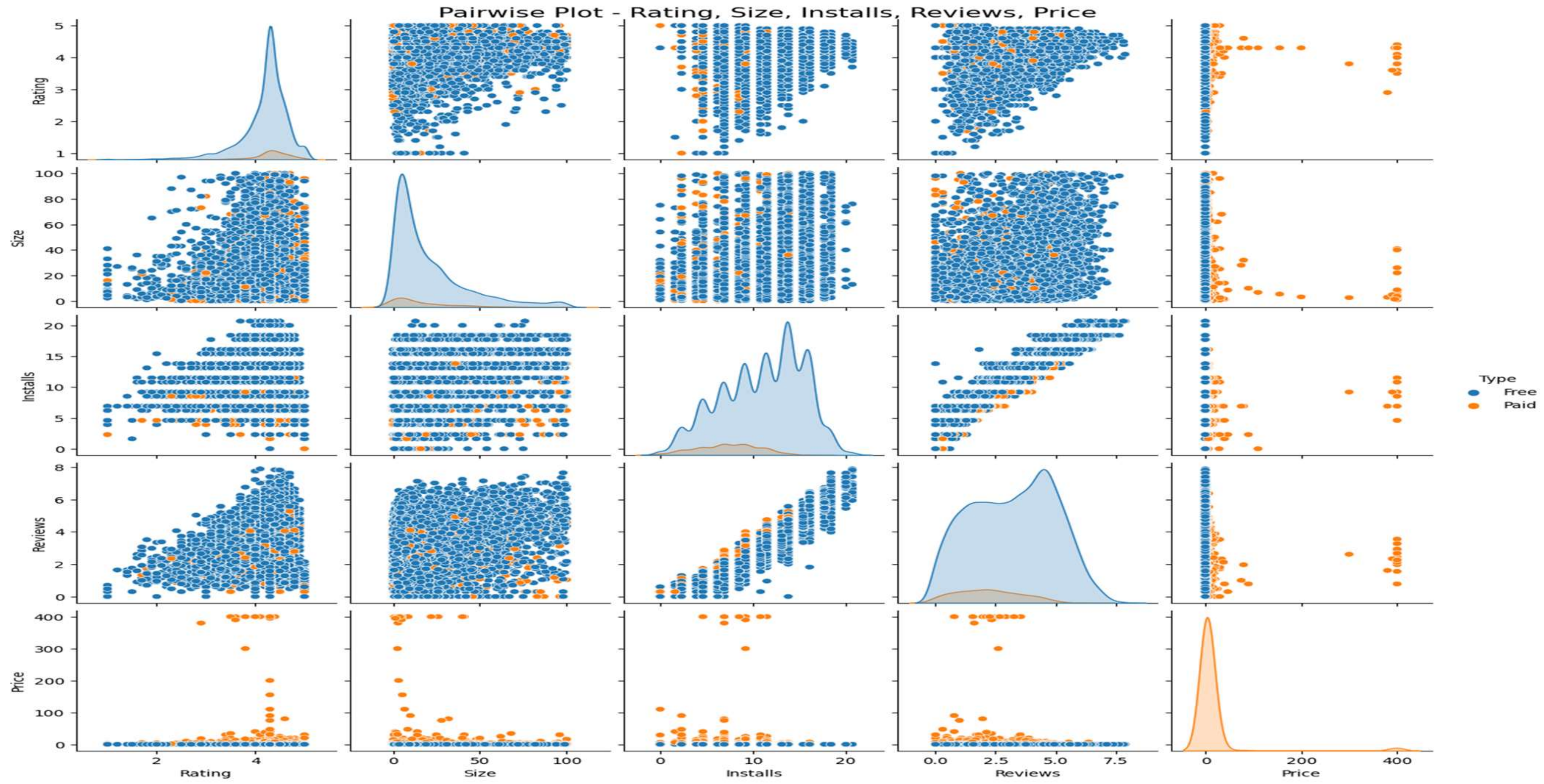
❑Attributes in User reviews

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

❑ 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- Ratings, Size, Installs, Reviews, Price

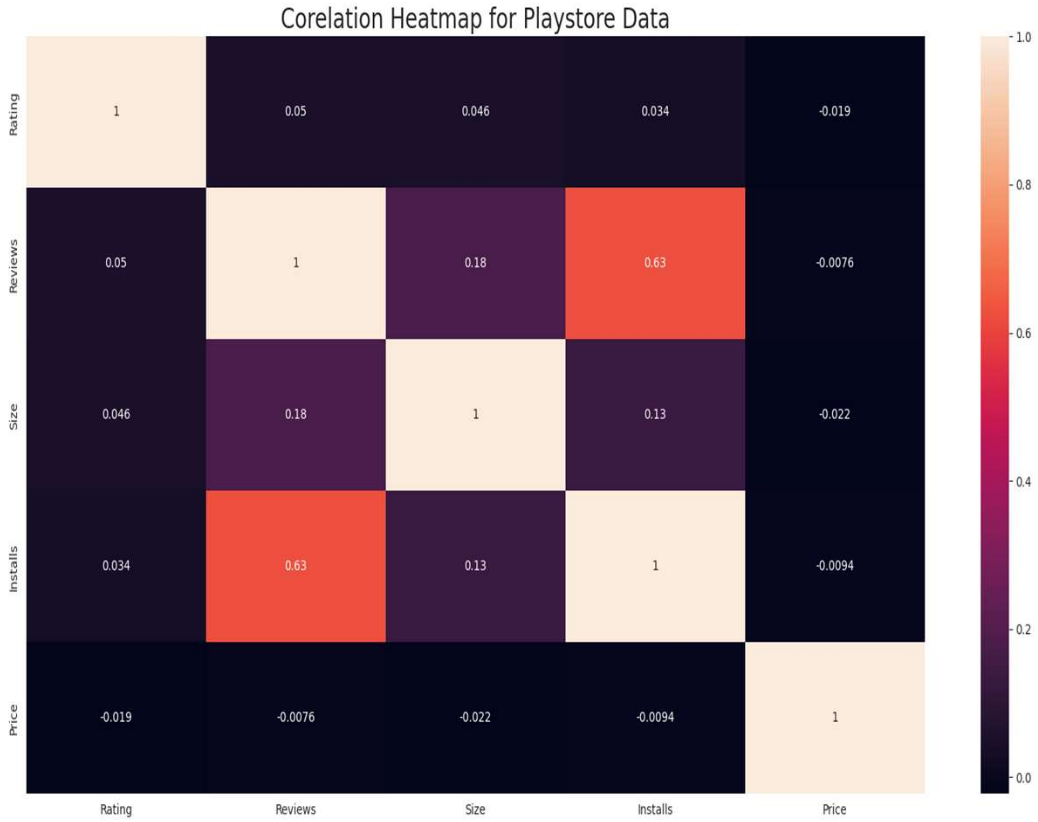


Correlation Heatmap

- There is a strong **positive** correlation between the **Reviews** and **Installs**

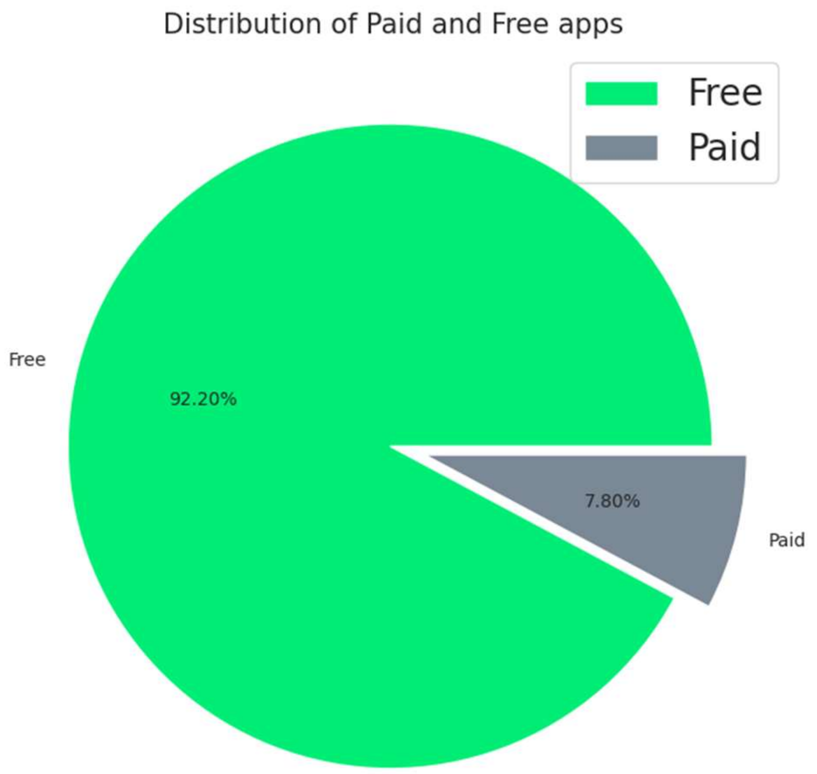
- The price is slightly negatively correlated with rating,review,and installs.

- The **Rating** is slightly **positively** correlated with the **Installs** and **Reviews**

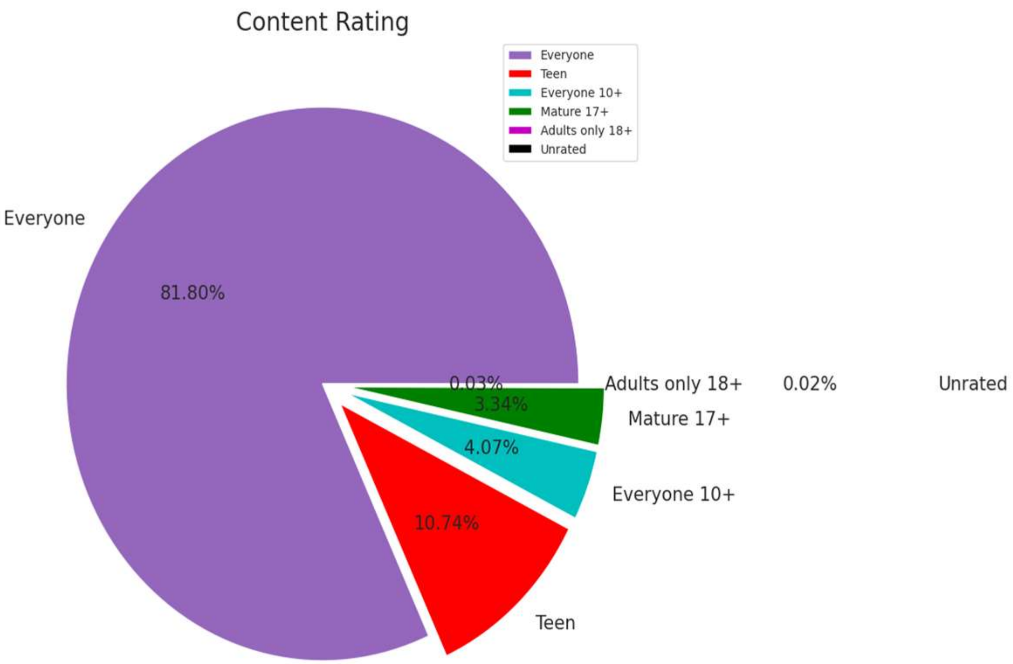


❑Percentage of Paid apps v/s Free apps

We Observed that **92.20% of Apps are free** and only **7.80% of Apps are paid** in Play store



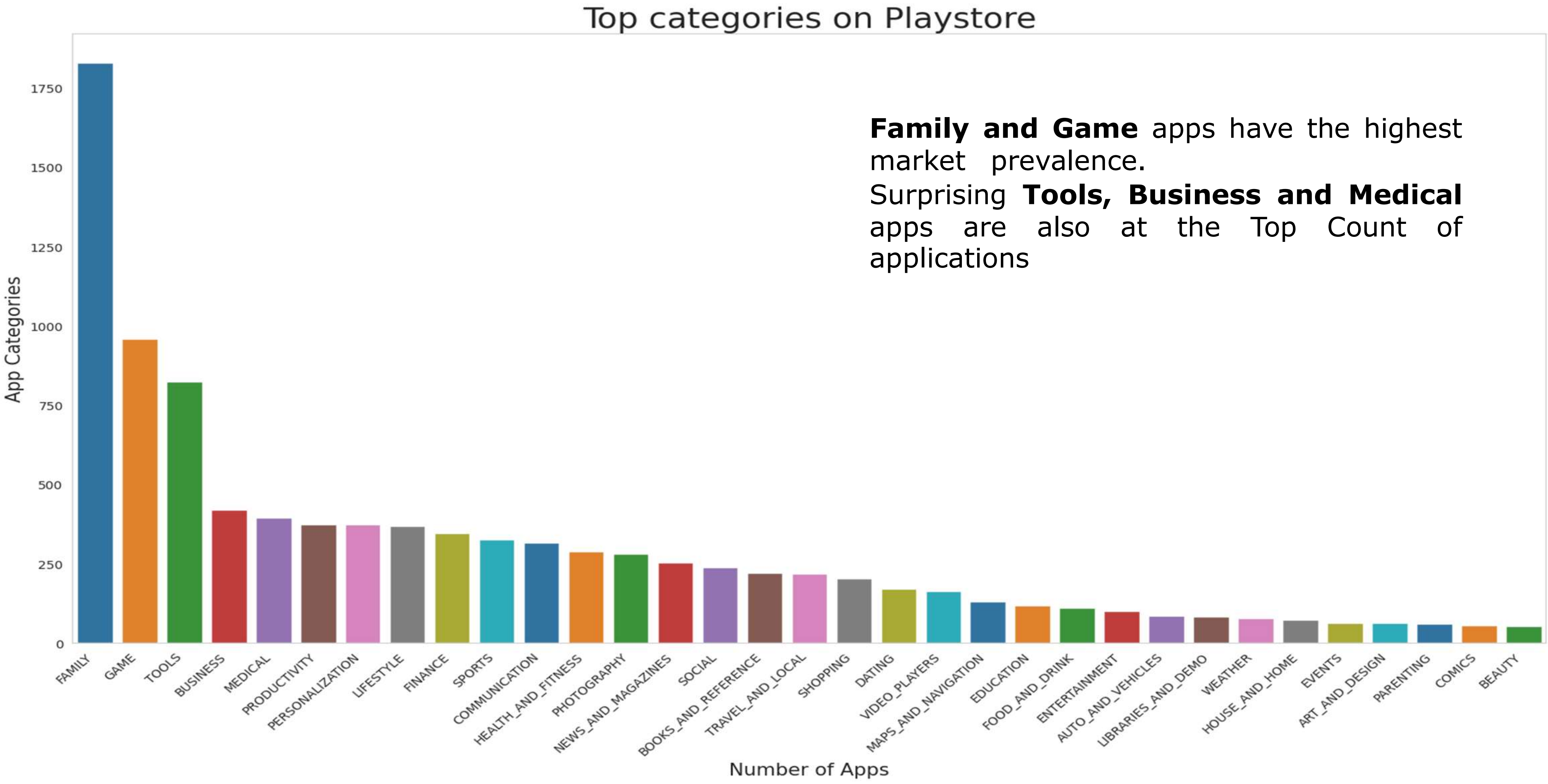
❑ Content Rating



- 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.

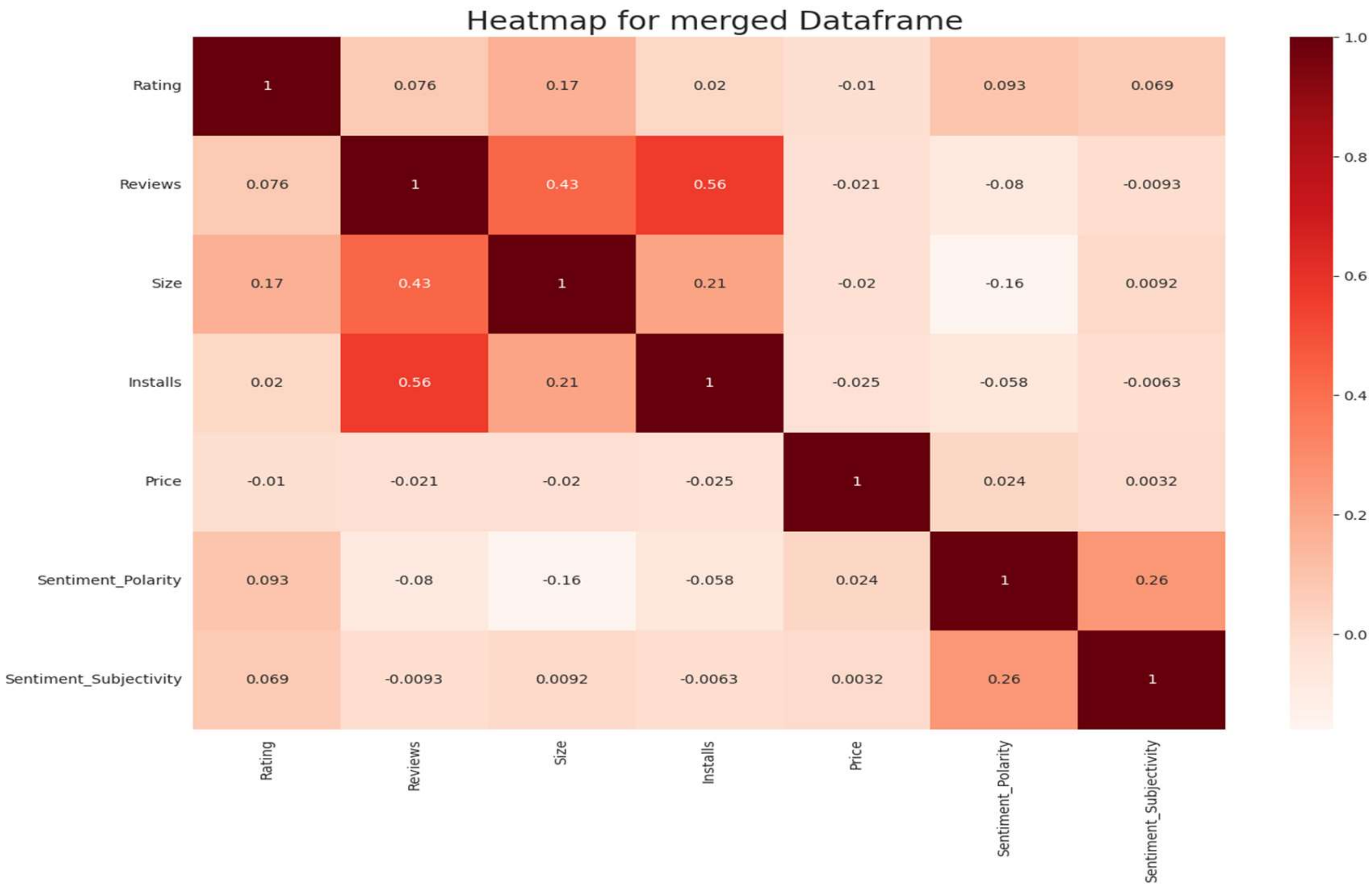
Count of Applications in each category



Family and Game apps have the highest market prevalence.

Surprising **Tools, Business and Medical** apps are also at the Top Count of applications

Co-Relation in merged data frame



	Rating	Reviews	Size	Installs	Price	Sentiment_Polarity	Sentiment_Subjectivity
Rating	1	0.076	0.17	0.02	-0.01	0.093	0.069
Reviews	0.076	1	0.43	0.56	-0.021	-0.08	-0.0093
Size	0.17	0.43	1	0.21	-0.02	-0.16	0.0092
Installs	0.02	0.56	0.21	1	-0.025	-0.058	-0.0063
Price	-0.01	-0.021	-0.02	-0.025	1	0.024	0.0032
Sentiment_Polarity	0.093	-0.08	-0.16	-0.058	0.024	1	0.26
Sentiment_Subjectivity	0.069	-0.0093	0.0092	-0.0063	0.0032	0.26	1

In this correlation matrix, There is not a significant relationship between Rating, Reviews, Size and Installs with respect to the Sentiment polarity and Sentiment subjectivity.

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.

shutterstock.com · 518802193

Conclusion’s

92.19% apps are **Free** and 7.81% apps are paid in type.

81.80% apps have **Everyone** content rating.

Events category has a **highest mean rating of 4.39** and Dating category has lowest 4.05 rating.

Family, Game and Tools are top three categories having 1906, 926 and 829 app count.

Most competitive category: **Family**

Category with the highest number of installs: **Game**

Tools, Entertainment, Education, Business and Medical are top Genres.

8783 Apps are having size less than or equal to **50 MB.**

7749 Apps has rating **more than 4.0** including both type of app.

Overall sentiment count of merged dataset in which **Positive sentiment count is 64%, Negative 22% and Neutral 14%.**