

Capstone Project - 1

Project Title : Play Store App Review Analysis

**Submitted by :
Mohd Azeem Ansari**

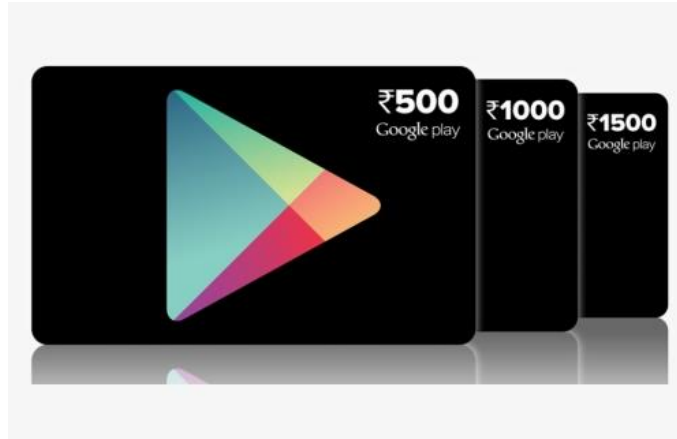
Tabular Array

1. Defining Problem statement.
2. Looking into the dataset.
3. Performing Data cleaning.
4. Removing Duplicates and Null Values.
5. Exploratory Data Analysis.
6. Visualization.
7. Summary and Conclusion



Google Play
Store

The Dilemma



- ❖ Google Play Store is an online store where people go to find their favourite apps, games, movies, TV shows, books, and more. Google Play is available in 190 countries and partners with developers from all over the world to distribute their apps.

Data Pipeline

1. Data cleaning process-1 : In the first part unnecessary features. Since, there were some columns with many null values.
2. Data cleaning process-2 : In this part, we manually go through each features selected from part 1. And encoded the categorical features and converting their datatype from object to float to perform different required operation in analysis.
3. EDA : In this part, we are performing some exploratory data analysis (EDA) on the features selected in part-1 and 2 to have a look on different trends.
4. Visualization : In this final step, we create different visualization graphs such a bar graphs, boxplot, pie chart and correlation matrix to have a big picture of our analysis.

Dataset : The dataset consists of 13 columns:- App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Genres, Last Updated, Current Ver, and Android Ver with 10841 Rows.

1. App - Application name
2. Category - Category the app belongs to
3. Rating - Overall user rating of the app
4. Reviews - Number of user reviews for the app
5. Size - Size of the app
6. Installs - Number of user downloads/installs for the app
7. Type - Paid or Free
8. Price - Price of the app
9. Content Rating - Age group the app is targeted at — Children / Mature 21+ / Adult
10. Genres - An app can belong to multiple genres (apart from its main category). For eg, a musical family game will belong to.
11. Last Updated - Date when was the application last updated.
12. Current ver - Current version of the app
13. Android ver - Which android version the app can supports.

Define Dependent Variable

- **ps_df1** : to store the given dataframe.
- **ps_df2** : to store the second dataframe of user reviews
- **cleaned_data** : list to store the different dataset after cleaning.
- **top_5_installed** : to store the data of top 5 apps installed by the user
- **top_5_reviewed** : to store the data of top 5 apps most reviewed by the users
- **top_5_exp** : to store the data of top 5 expensive apps available on the playstore
- **fin_low_rat** : to store 5 apps from the Finance category having the lowest rating
- **top_5_ins_game** : to store top 5 most installed apps in Game category

Major Outcomes of EDA

❑ The top 5 apps installed by the user are:

- Facebook,
- Google Photos,
- Google+,
- Gmail,
- Google Street View.

❑ The top 5 apps most reviewed by the users are:

- Facebook,
- WhatsApp Messenger,
- Instagram,
- Messenger – Text and Video Chat for Free,
- Clash of Clans.

❑ The top 5 expensive apps available on the Playstore are:

- I'm Rich - Trump Edition,
- I'm Rich/Eu sou Rico//أنا غني 我很有錢,
- I am Rich Plus,
- I am rich (Most expensive app),
- I Am Rich Premium.

❑ The 5 apps from the Finance category having the lowest rating are :

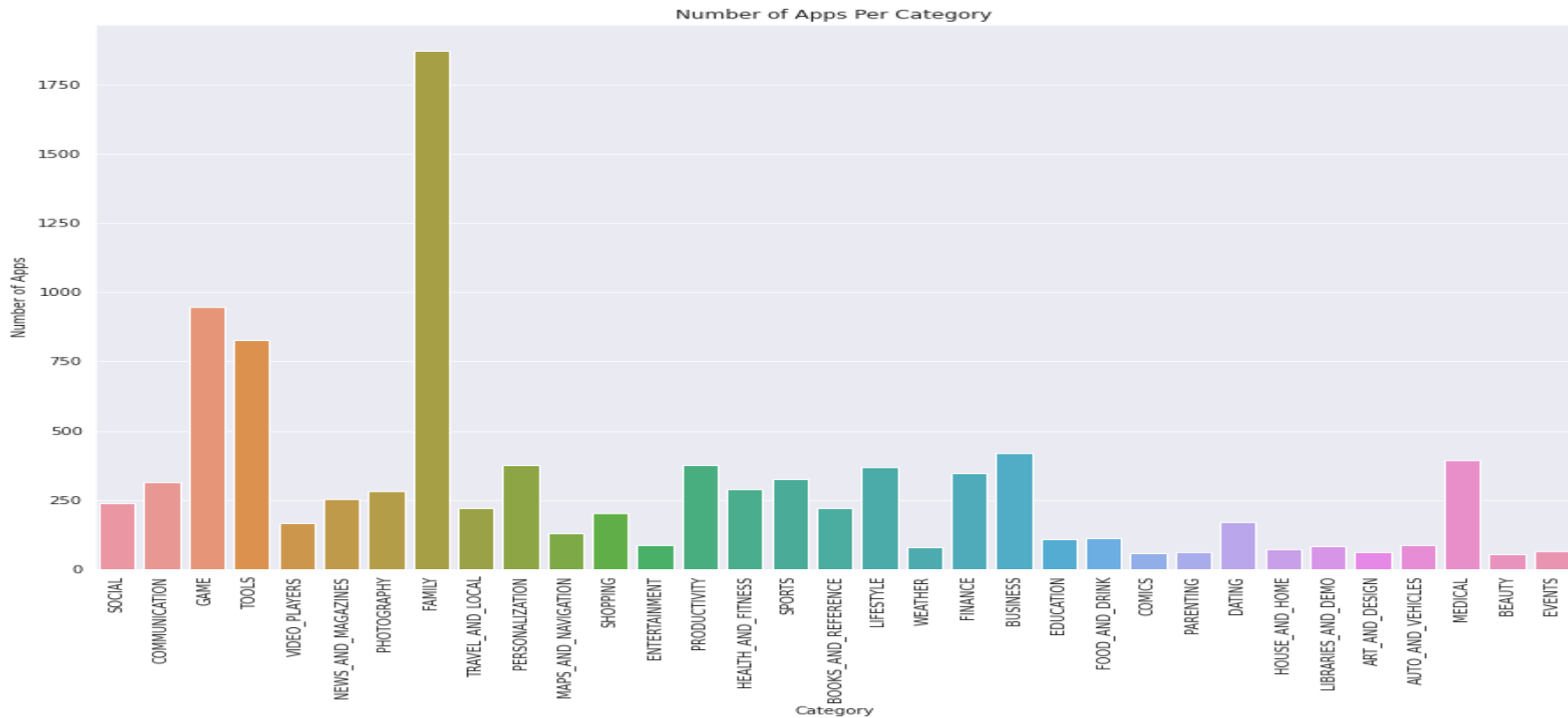
- CB Mobile Biz,
- Tech CU Card Manager,
- EY TaxChat,
- CB Mobile Access,
- Bitcoin BX Thailand PRO.

❑ The top 5 most installed apps in Game category are:

- Subway Surfers,
- Candy Crush Saga,
- My Talking Tom,
- Temple Run 2,
- Pou.

Visualizing Different Trends

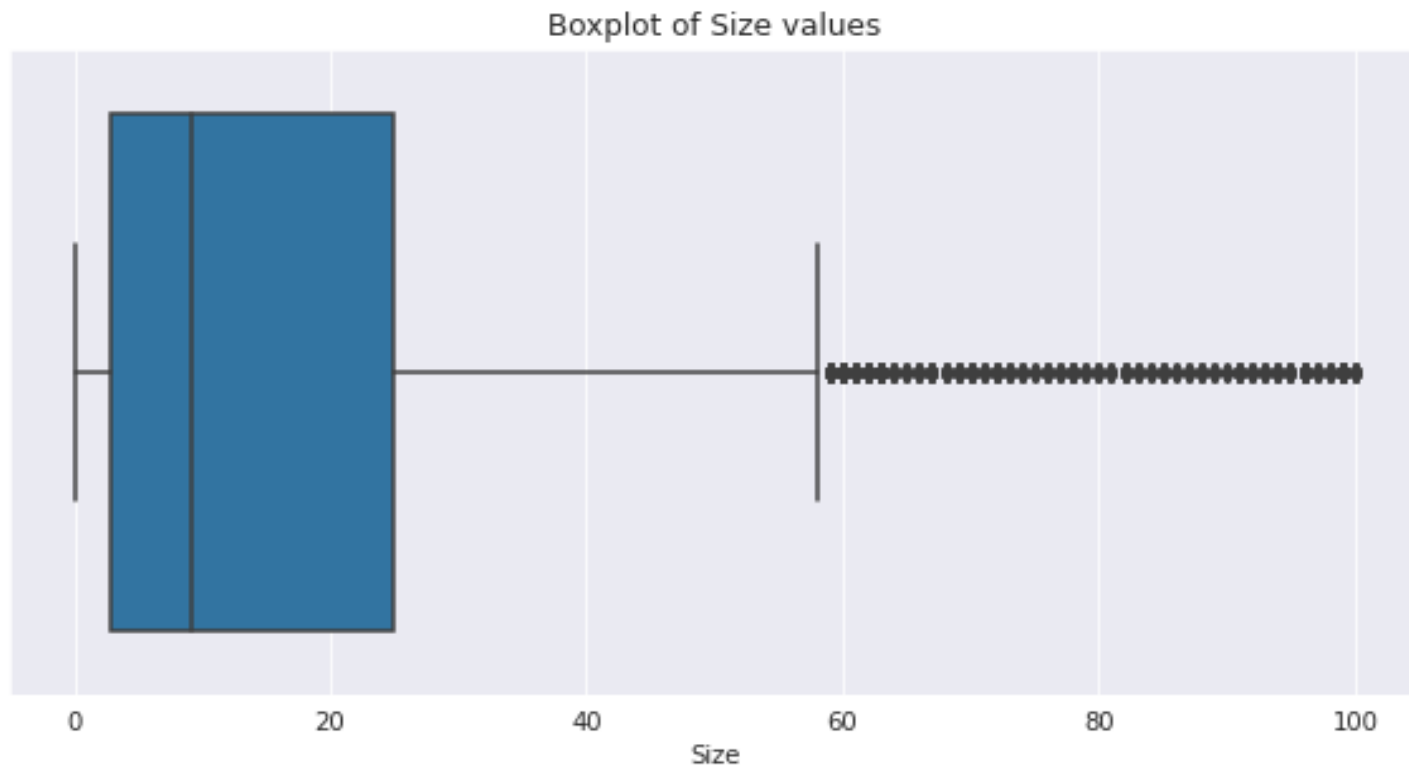
❖ Category of Apps



✓ Most of the apps available on the play store are from the categories of 'FAMILY', 'GAME' and also 'TOOLS'.

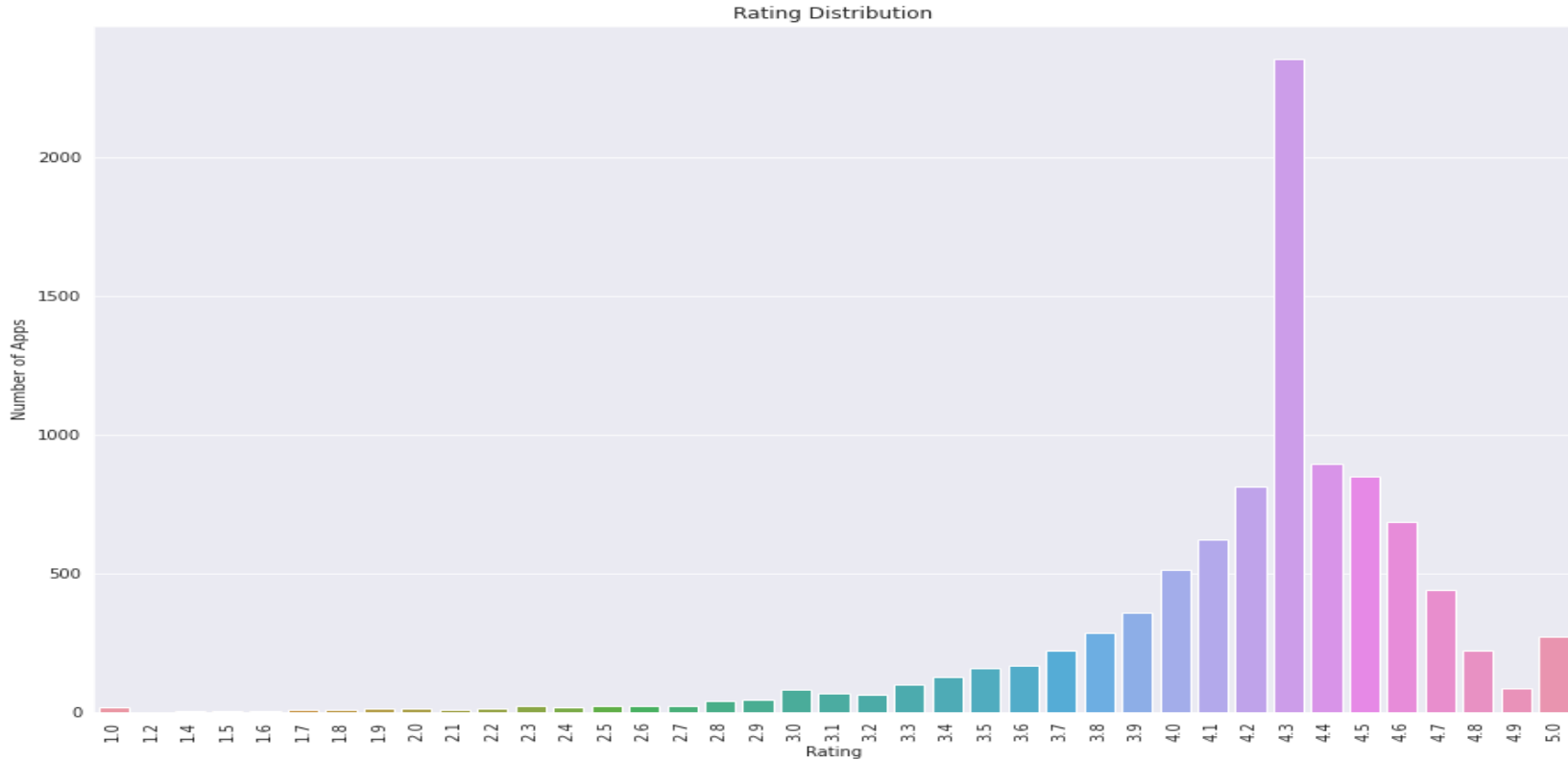


Sizes of Application



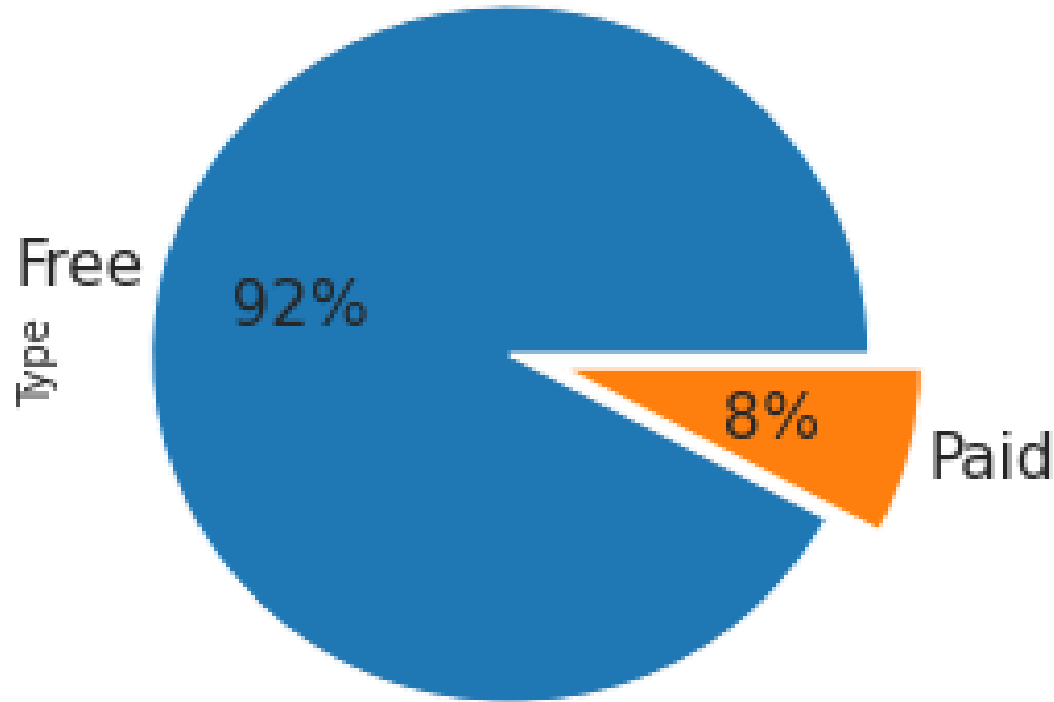
- ✓ Majority of the apps available are within the range of size 5Mb to 25Mb and their median size is around 9Mb.

❖ Rating Trend



✓ Most of the apps in the Play Store are having rating higher than 4 or in the range of 4 to 4.7.

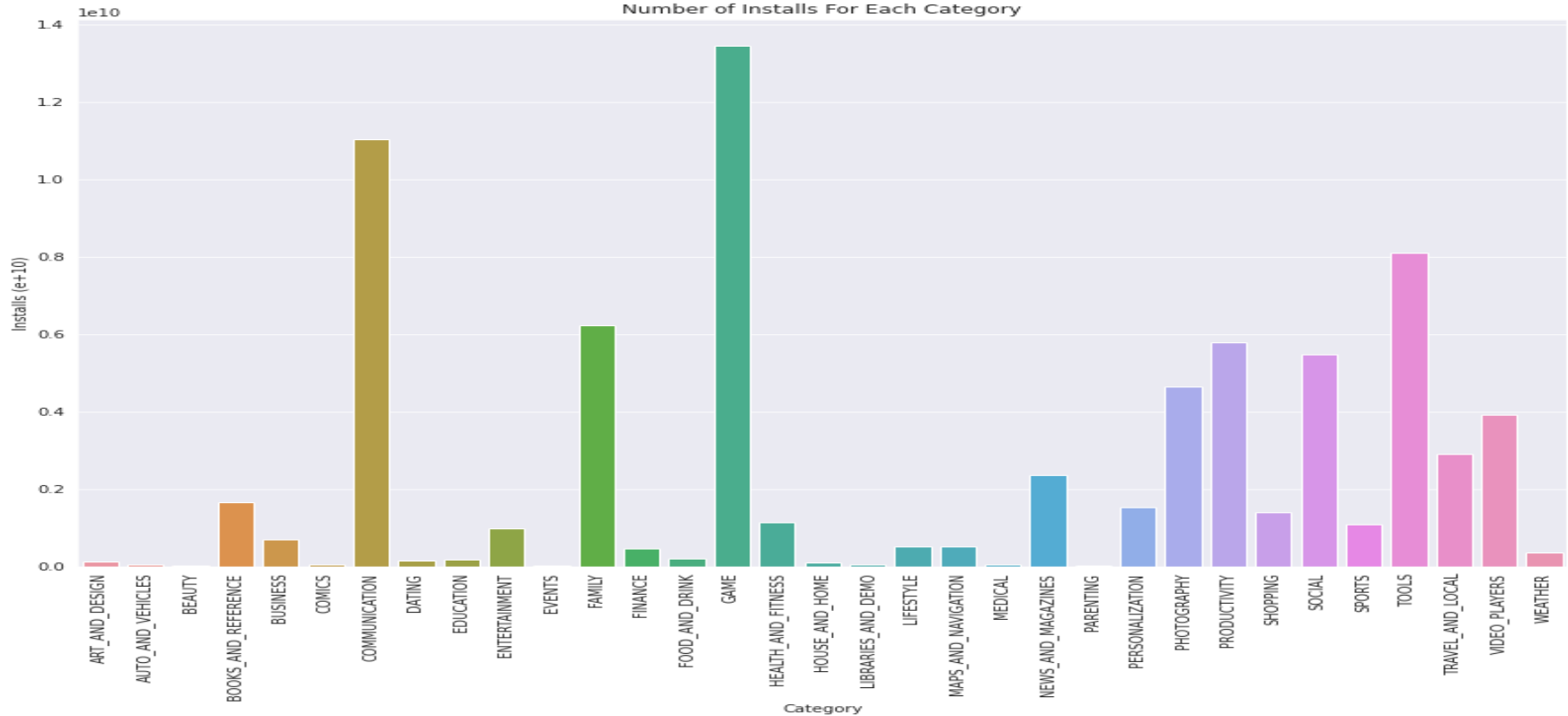
❖ Types of Applications



- ✓ Majority of the apps available on the Play Store are Free apps(92%) and only few apps are paid (8%).



Installs Trend



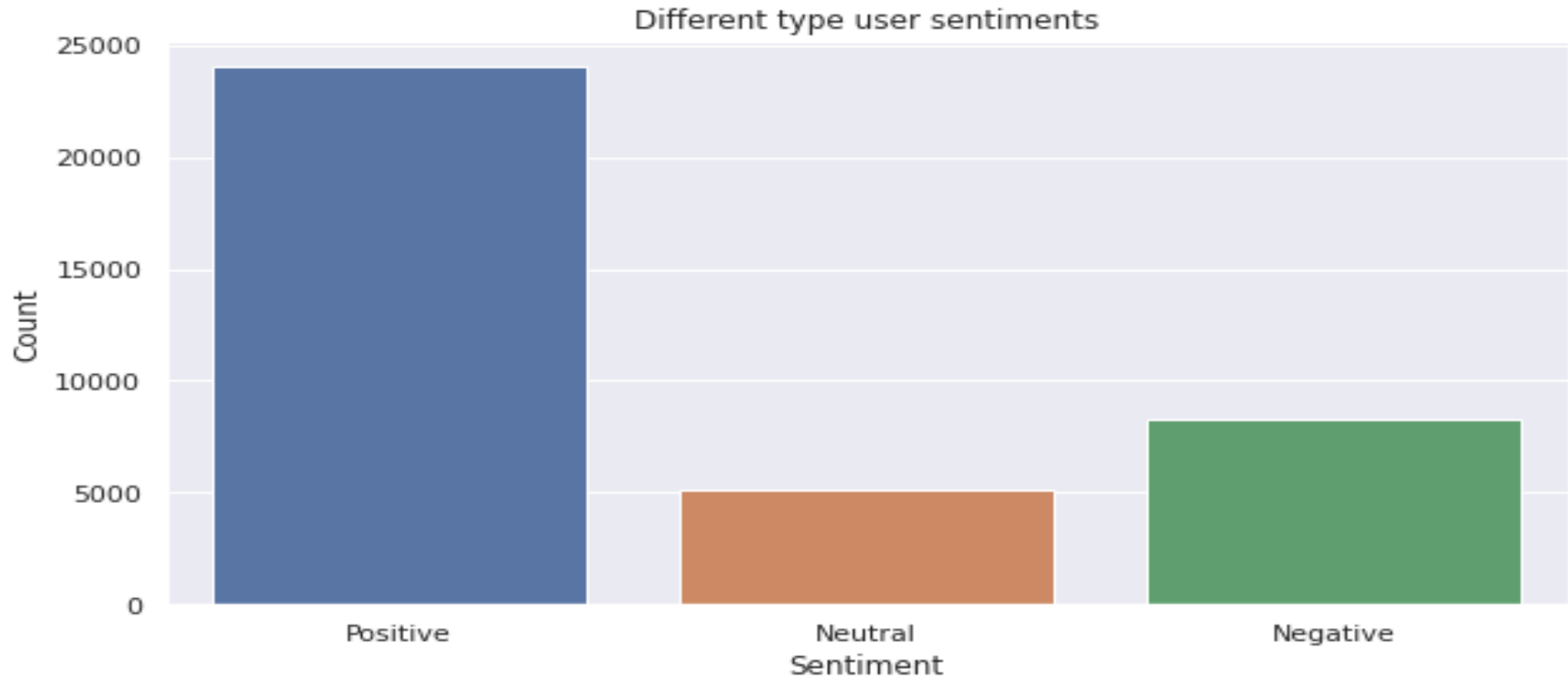
✓ Most of the apps being downloaded and installed are from categories of 'GAME' and 'COMMUNICATION'.

❖ Correlation Matrix



- ✓ Reviews and Installs features are more correlated and the value is 0.63.

❖ Sentiments of users towards the app store



- ✓ Most of the user have the positive attitude towards the apps available on the app store.

❖ Conclusion

- ✓ The Google Play Store Apps data analysis provides some useful insights regarding the trending of the apps in the play store.
- ✓ As per the graphs visualizations shown above, most of the trending apps (in terms of users' installs) are from the categories like GAME, COMMUNICATION, and TOOL even though the amount of available apps from these categories are twice as much lesser than the category FAMILY.

❖ Conclusion(contd.)

- ✓ The demands of these apps are most probably due to their nature of being able to entertain or assist the user.
- ✓ Besides, it also shows a good trend where we can see that developers from these categories are focusing on the quality instead of the quantity of the apps as most of them are getting positive sentimental support from the users.

❖ Conclusion(contd.)

- ✓ At the end, we can conclude that the current trend in the Android market are mostly from the categories which are either assisting, communicating or entertaining apps.

❖ Challenges

- ✓ Huge Chunks of data was to be handled keeping in mind not to miss anything which is even of the little relevance.
- ✓ Feature selection was quite a challenge as our database had many hidden facts and trends for EDA of Playstore Apps data analysis.

Q & A

THANK YOU