Capstone Project – 1

Google Play Store App Review Analysis

Team Members

Tushar Natani

Anup Prasad

Introduction

Google Play Store or formerly Android Market, is a digital distribution service developed and operated by Google. It is an official apps store that provides variety content such as apps, books, magazines, music, movies and television programs. It serves an as platform to allow users with 'Google certified' Android operating system devices to download applications developed and published on the platform either with a charge or free of cost. With the rapidly growth of Android devices and apps, it would be interesting to perform data analysis on the data to obtain valuable insights.

Data Pipeline

- **Data Preparation:** In this section, we will be loading the Google Store Apps data stored in csv using pandas which is a fast and powerful python library for data analysis and easy data manipulation in pandas DataFrame object. It is usually used for working with tabular data (e.g data in spreadsheet) in various formats such as CSV, Excel spreadsheets, HTML tables, JSON etc.

- **Data Cleaning:** In this part, we have removed unnecessary features. Since there were nearly many columns with null values.

- **EDA:** In this part, we have done some exploratory data analysis.

Description of App Dataset Columns

1. **App :** The name of the app.
2. **Category :** The category of the app.
3. **Rating :** The rating of the app in the Play Store.
4. **Reviews :** The number of reviews of the app.
5. **Size :** The size of the app.
6. **Install :** The number of installs of the app.
7. **Type :** The type of the app (Free/Paid).
8. **Price :** The price of the app (0 if it is Free).
9. **Content Rating :** The appropriate target audience of the app.
10. **Genres :** The genre of the app.
11. **Last Updated :** The date when the app was last updated.
12. **Current Ver :** The current version of the app.
13. **Android Ver :** The minimum Android version required to run the app.

Diagnosing the App DataFrame

By diagnosing the DataFrame, we know that:

1. There are 13 columns of properties with 10841 rows of data.
2. Column 'Reviews', 'Size', 'Installs' and 'Price' are in the type of 'object'.
3. Values of column 'Size' are strings representing size in 'M' as Megabytes, 'k' as kilobytes and also 'Varies with devices'.
4. Values of column 'Installs' are strings representing install amount with symbols such as ',' and '+'.
5. Values of column 'Price' are strings representing price with symbol '$'.
6. There is only one 'float64' type column which is 'Rating'.

Description of User Dataset Columns

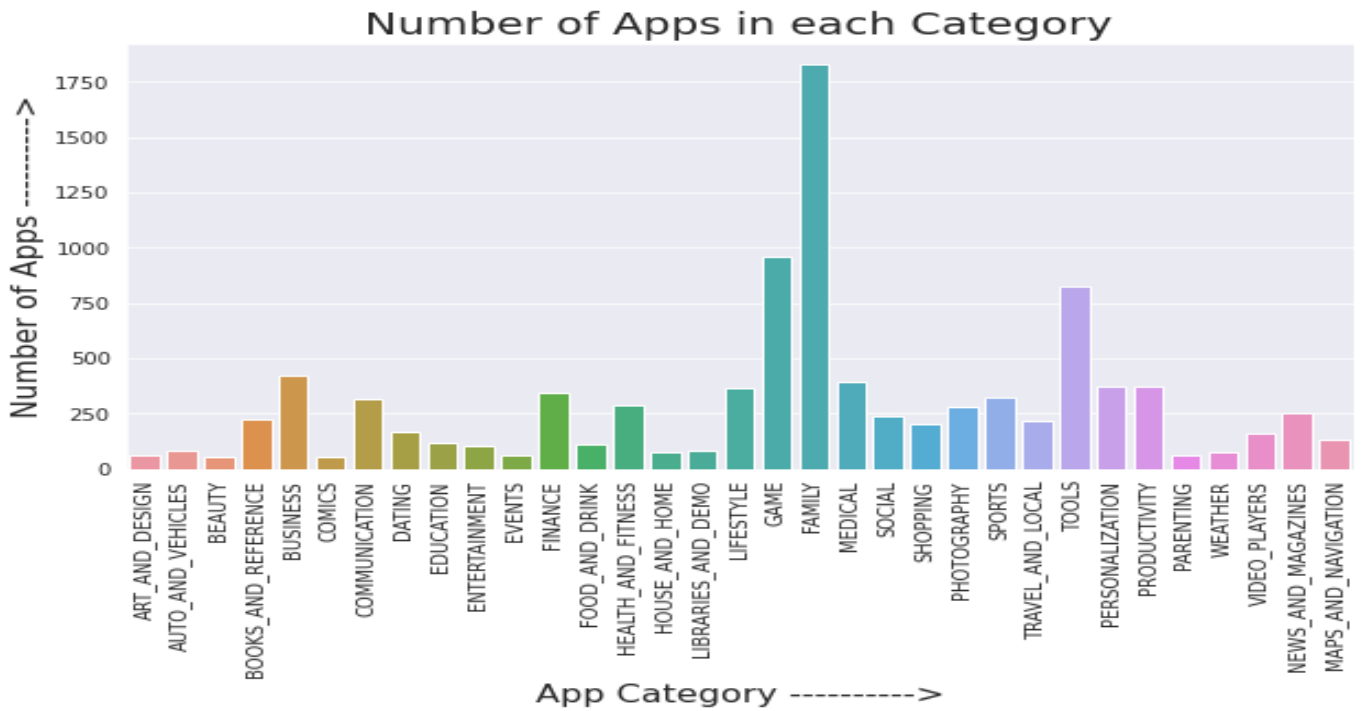
1. **App :** Category of Apps.
2. **Translated_Review :** Actual review in understandable manner.
3. **Sentiment :** Sentiment basically determines the attitude or the emotion of the writer, i.e., whether it is positive or negative or neutral.
4. **Sentiment_Polarity :** Sentiment Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement.
5. **Sentiment_Subjectivity :** Sentiment Subjectivity generally refer to personal opinion, emotion or judgment, which lies in the range of [0,1].

Diagnosing the User DataFrame

1. There are 5 columns with 64295 rows of data.
2. ‘App’, ‘Translated_Review’ and ‘Sentiment’ are of ‘object’ data type.
3. ‘Sentiment_Polarity’ and ‘Sentiment_Subjectivity’ are of ‘float64’ data type.
4. The ‘App’ column does not contain any null values.
5. Rest of the columns contains 26883 null values.

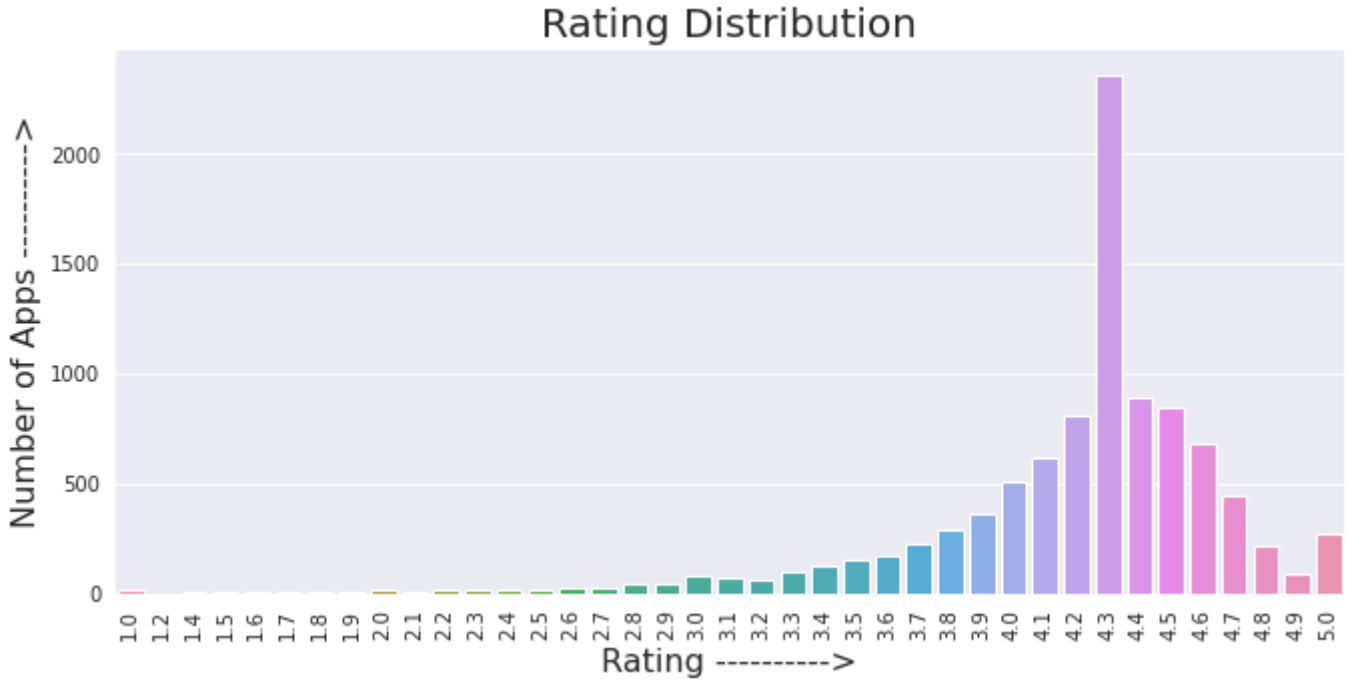
Exploratory Analysis and Visualization (EDA)

1. **Number of Apps in each category**



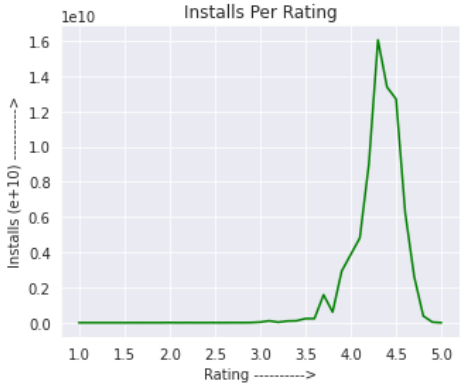
EDA (continued)

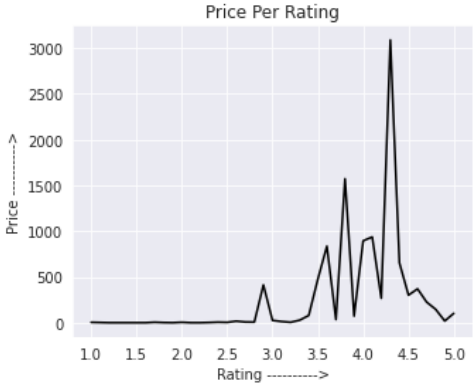
2. **Rating Distribution**:



EDA (continued)

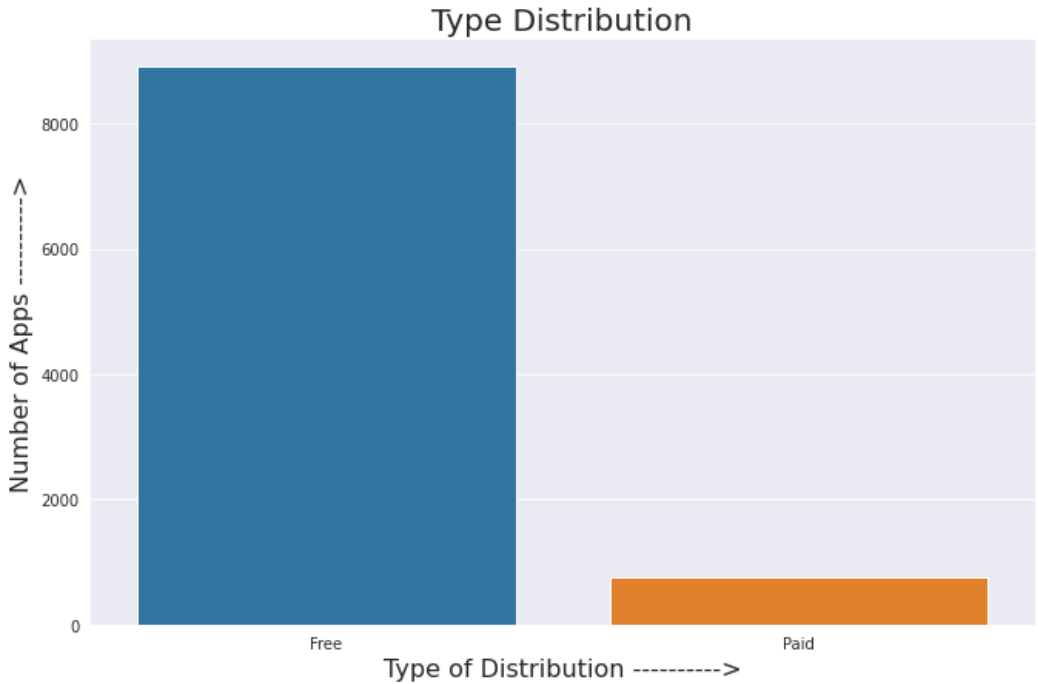
3. Reviews, Size, Installs and Price per rating:





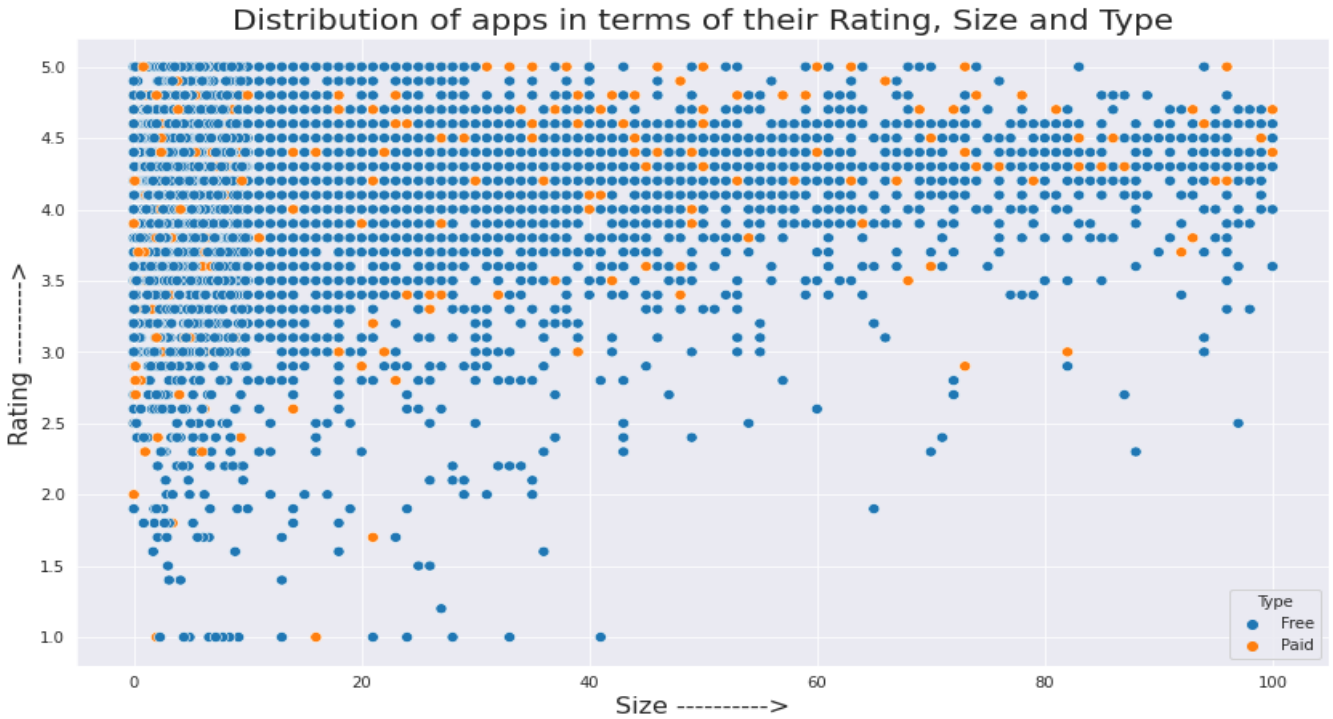
EDA (continued)

4. Application Type:



EDA (cont.)

5. **Distribution of apps in terms of their Rating, Size and Type**:

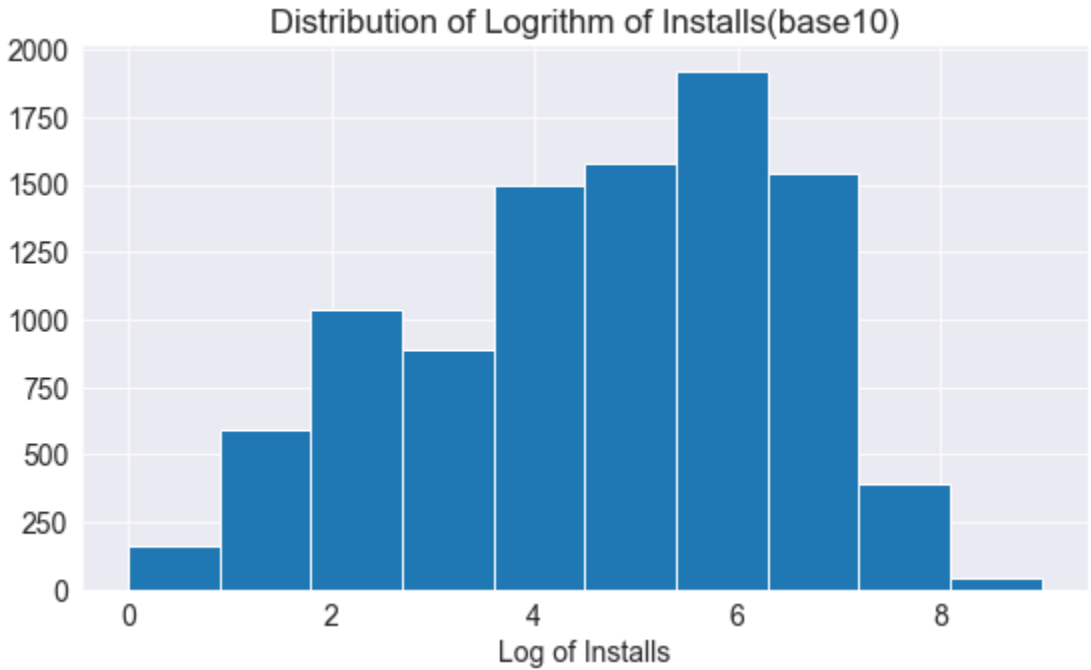


EDA (cont.)

6. Correlation between Rating, Reviews, Size, Install and Price:

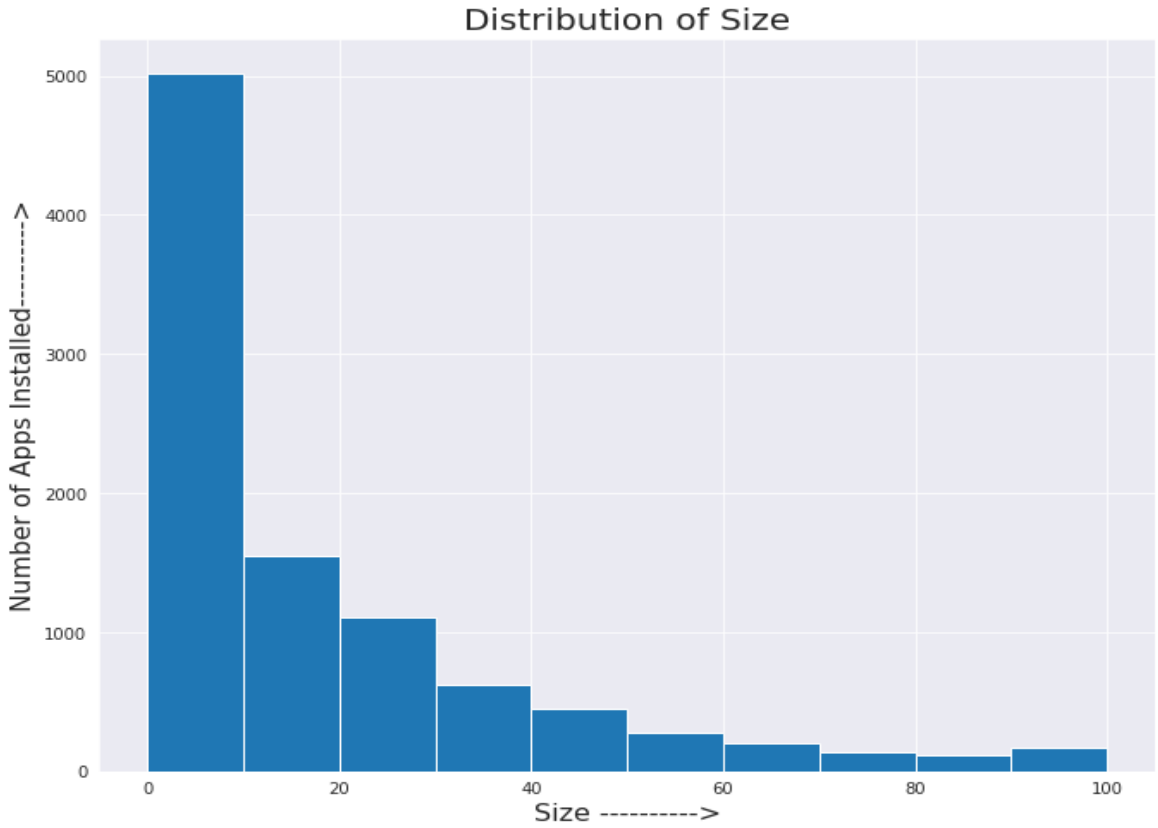
EDA (cont.)

7. **Histogram of Log Install:**



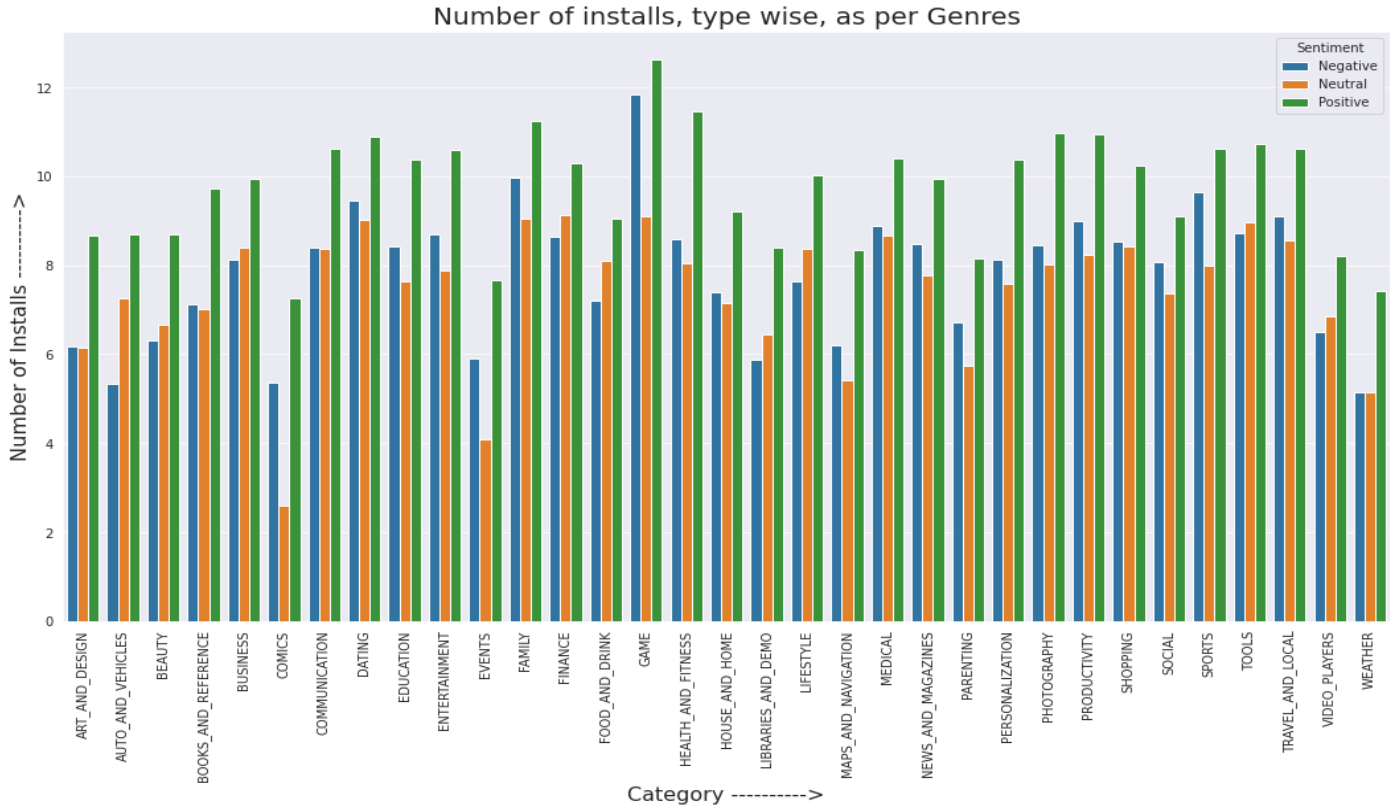
EDA (cont.)

8. **Distribution of Size:**



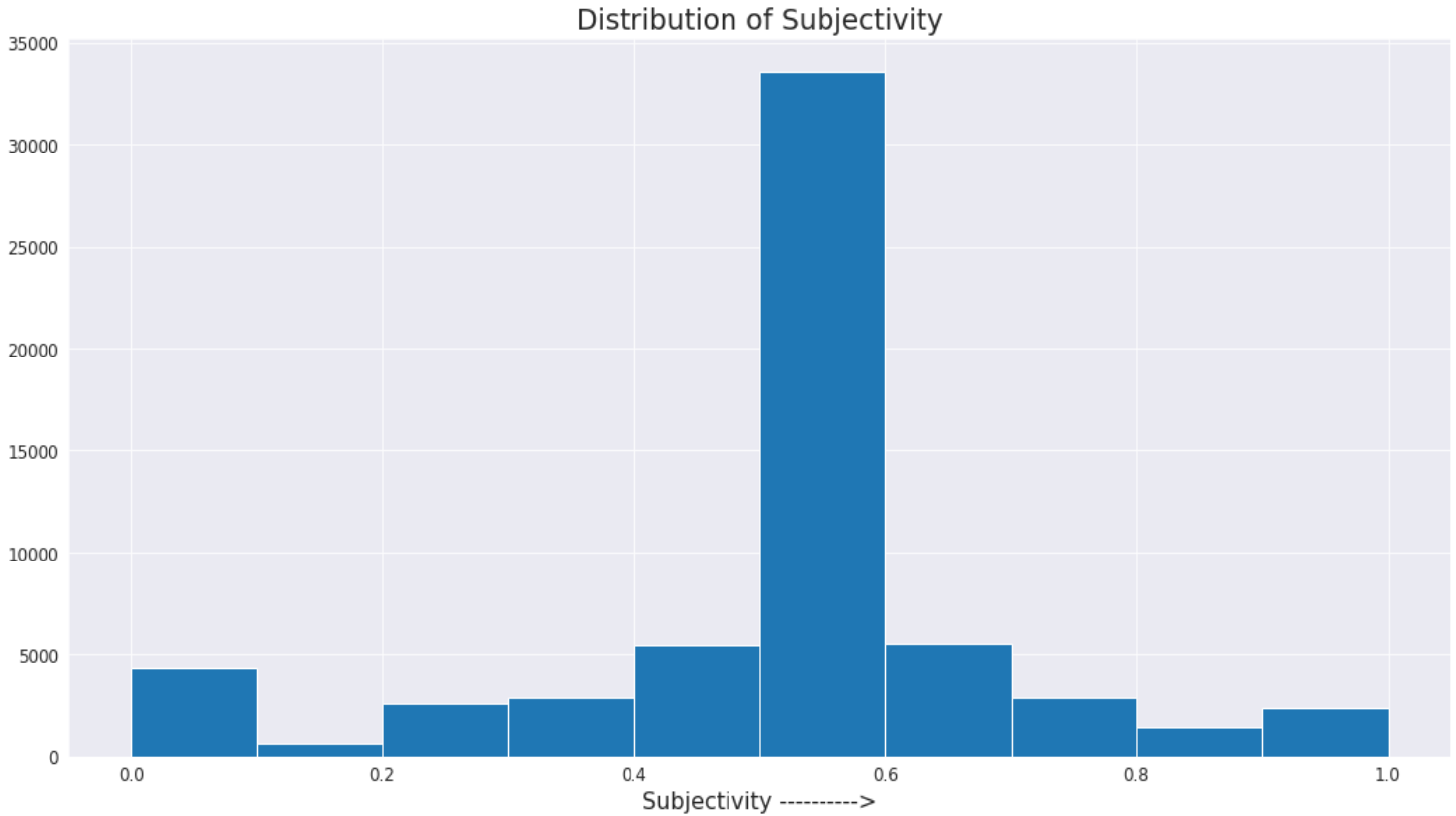
EDA (cont.)

9. Distribution of type of reviews:



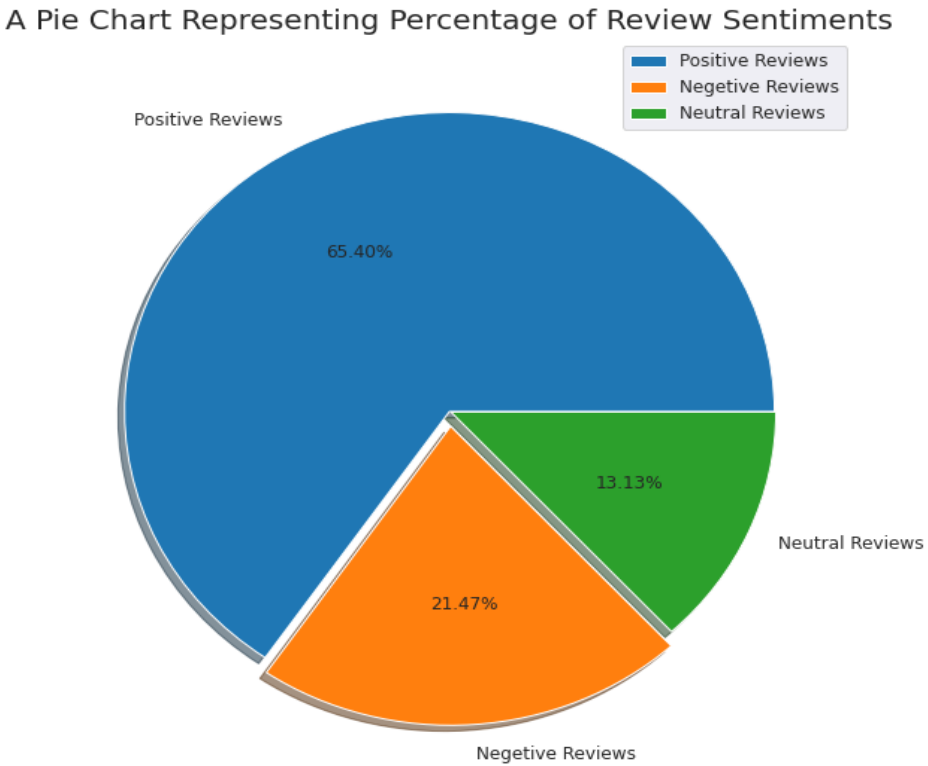
EDA (cont.)

10. Distribution of Subjectivity:



EDA (cont.)

11. Percentage of Review Sentiment:



Conclusion

- The Google Play Store Apps report 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. The trending 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.
- Other than that, the charts shown above actually implies that most of the apps having good ratings of above 4.0 are mostly confirmed to have high amount of reviews and user installs. There are some spikes in term of size and price but it shouldn't reflect that apps with high rating are mostly big in size and pricy as by looking at the graphs they are most probably are due to some minority. Furthermore, most of the apps that are having high amount of reviews are from the categories of SOCIAL, COMMUNICATION and GAME like Facebook, WhatsApp Messenger, Instagram, Messenger – Text and Video Chat for Free, Clash of Clans etc.
- Even though apps from the categories like GAME, SOCIAL, COMMUNICATION and TOOL of having the highest amount of installs, rating and reviews are reflecting the current trend of Android users, they are not even appearing as category in the top 5 most expensive apps in the store (which are mostly from FINANCE and LIFESTYLE). As a conclusion, we learnt that the current trend in the Android market are mostly from these categories which either assisting, communicating or entertaining apps.

Thank You