

Capstone Project -01

Play Store App Review Analysis (EDA)

By

Vishal Sakore
Dhanashri Dewade
Atul Jadhav



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 as a platform to allow users with 'Google certified' Android operating system devices to download applications developed and published on the platform either with a paid 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.



Dataset Preparation

- **Import Libraries:** NumPy, Pandas, Seaborn, Matplotlib and Datetime.
- **Loading files:** Add both files Play store app dataset and User Reviews dataset.
- **Explore and Analyze the data:** Find useful information and describe it briefly.
- **Data cleaning:** Null values, Finding and removing Outliers, Removing duplicate data and Correcting Data Types.
- **Find Insights:** Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.

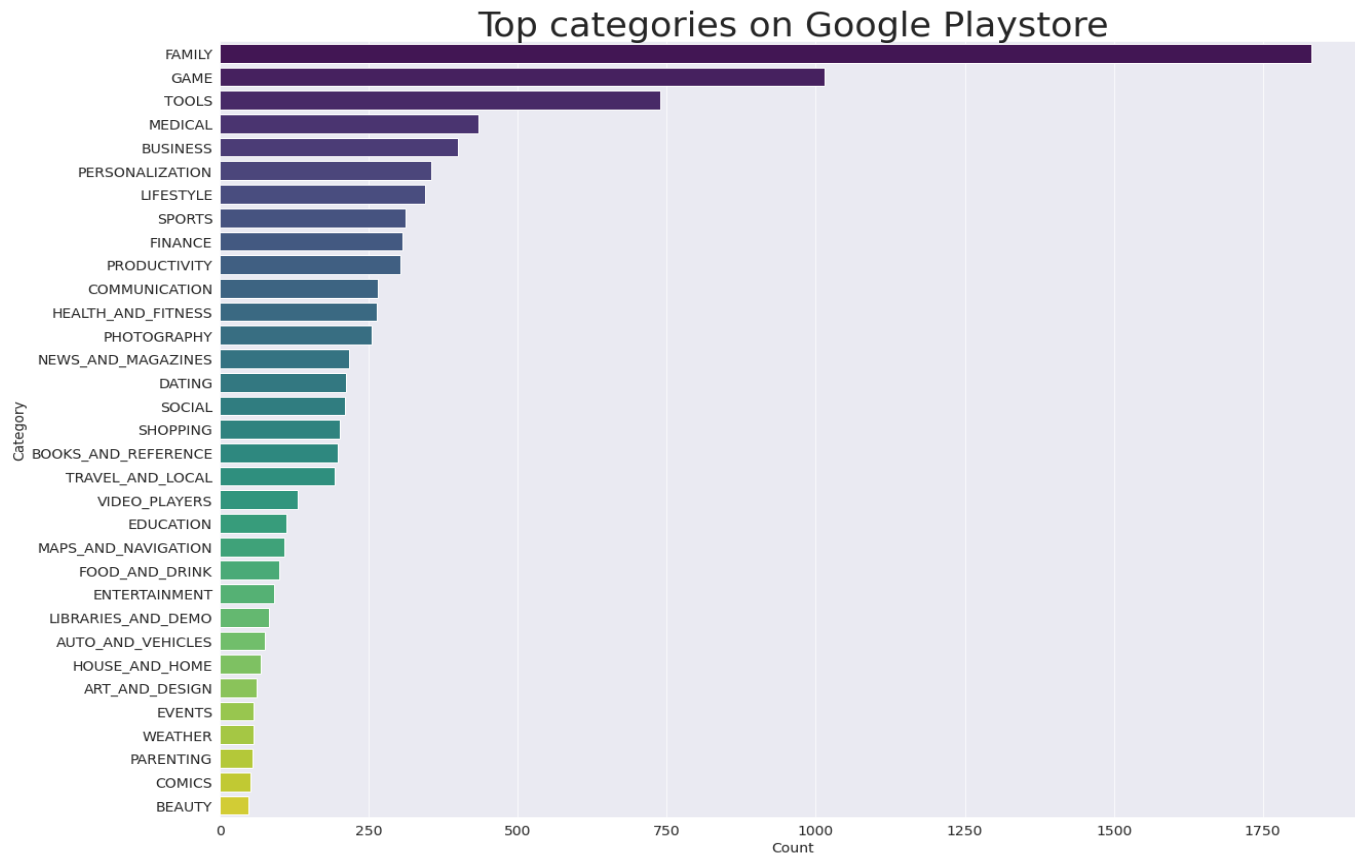
First Dataset : Play Store Data

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
8. Price : The price of the app
9. Content Rating : The rating given by user
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

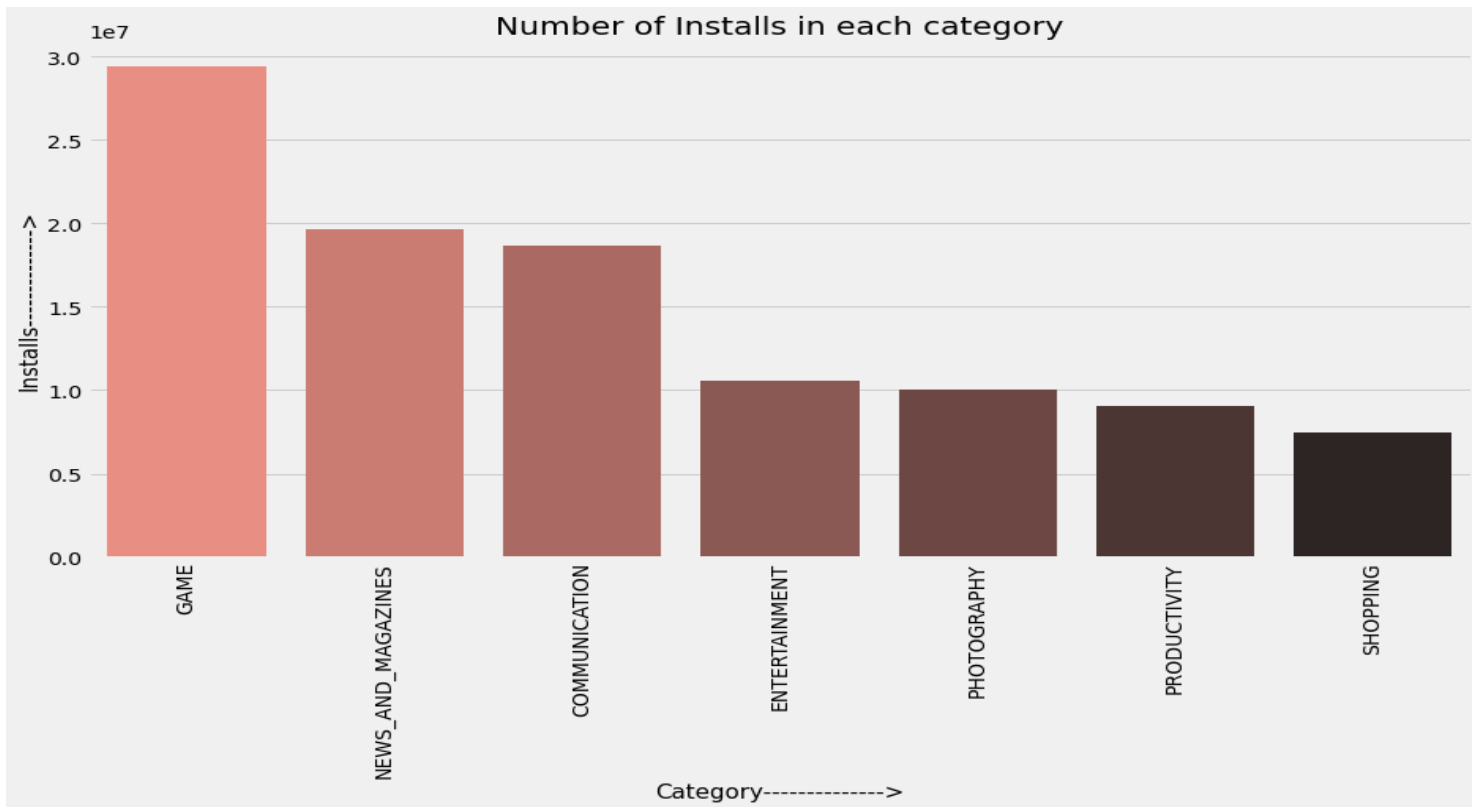
Second Dataset: User Reviews

1. App : The name of the application
2. Translated_Review : User comment
3. Sentiment : User view or opinion
4. Sentiment_Polarity : Strongness of Sentiment polarity
5. Sentiment_Subjectivity : It is a public opinion and not a factual information

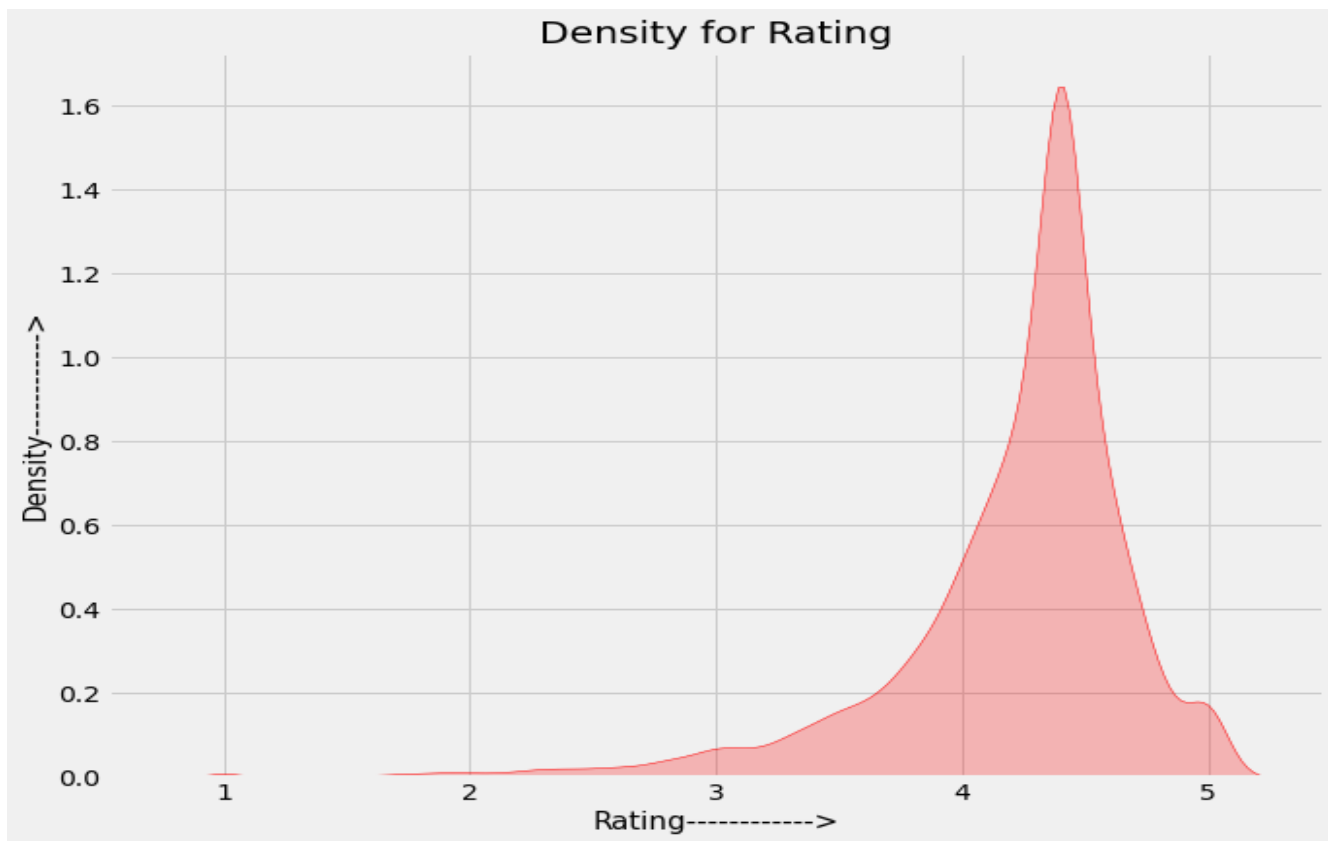
Top categories



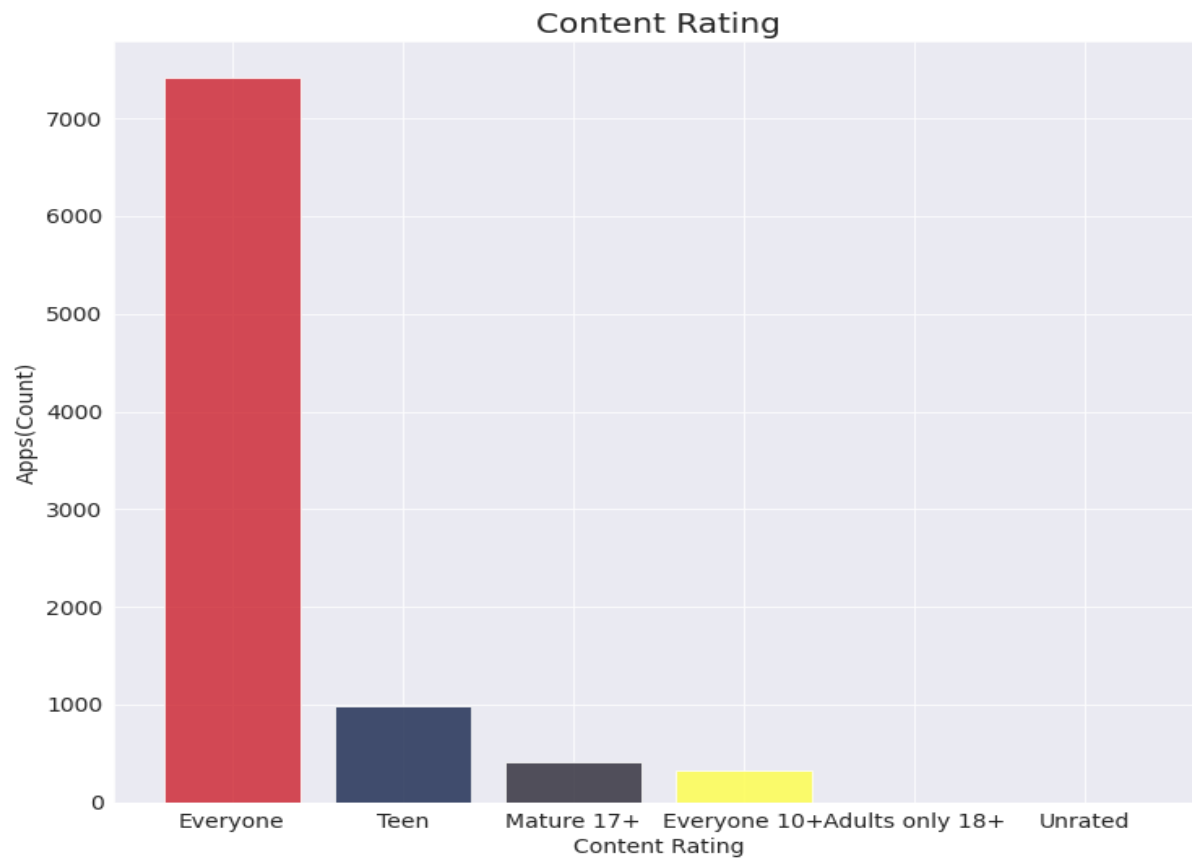
Number of Installs in each category



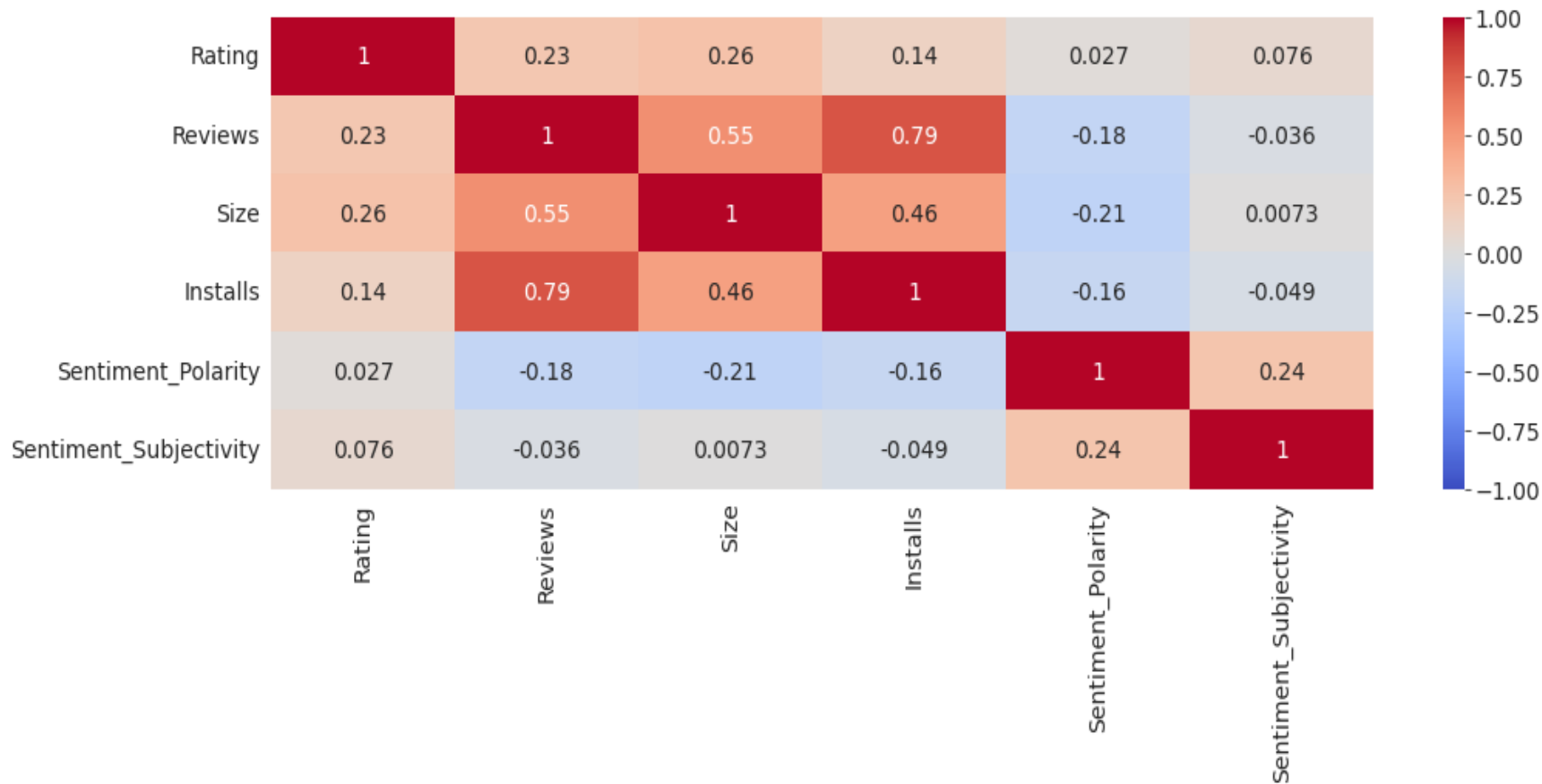
Density for Rating



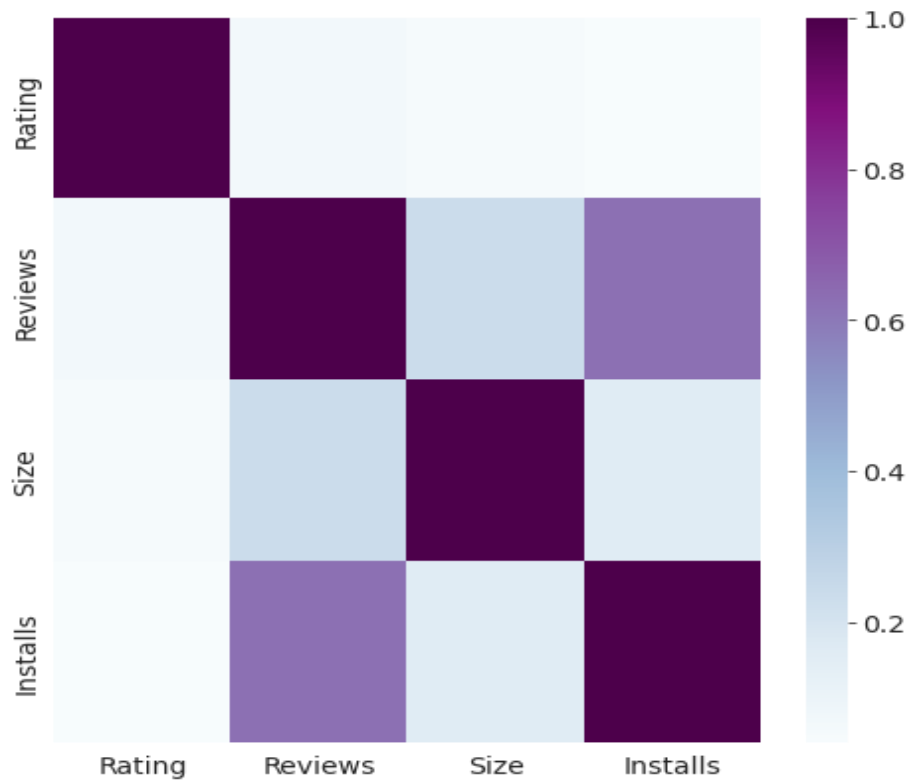
Content Rating



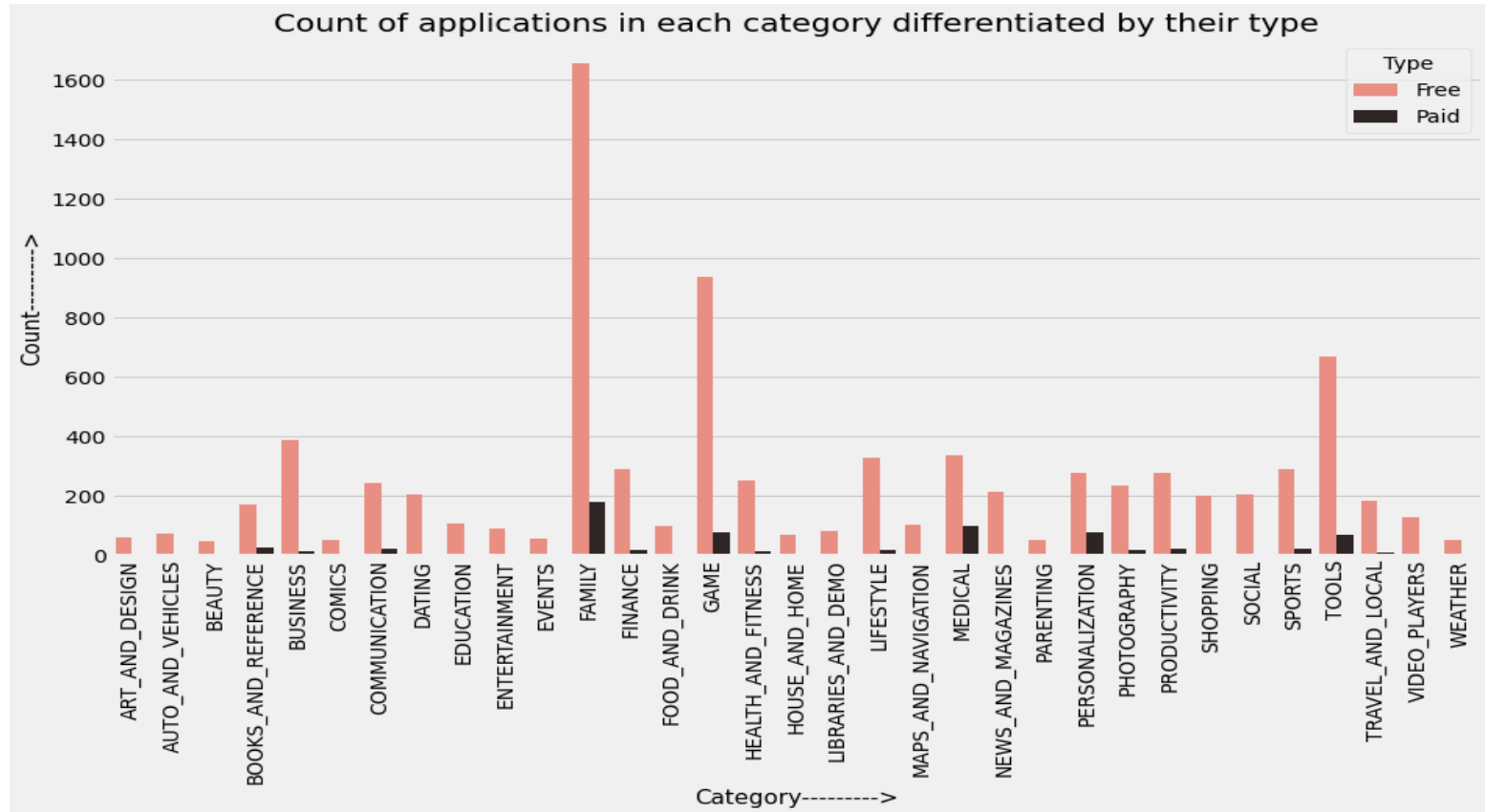
Correlation Heatmap



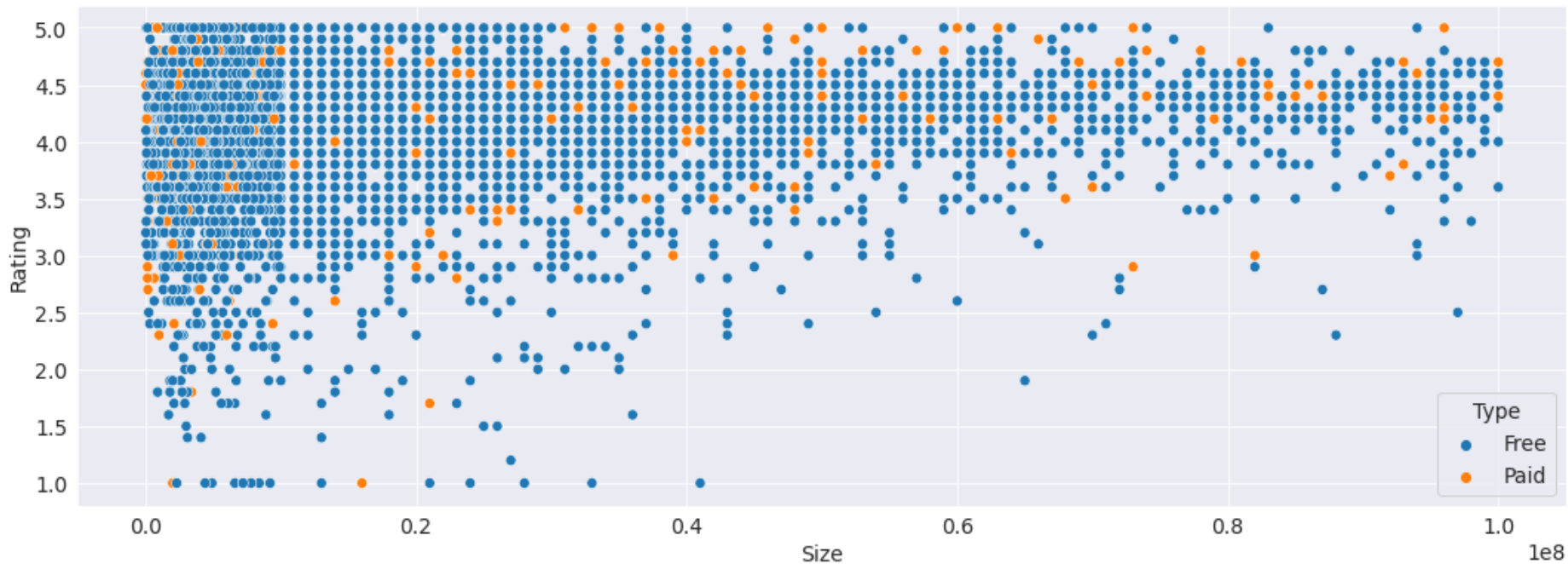
Correlation map



Count of applications in each category differentiated by their type



Distribution of apps in term of their rating, size and type



Top 5 expensive apps

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
4367	I'm Rich - Trump Edition	LIFESTYLE	3.6	275	7300000.0	10000	Paid	400.00	Everyone	Lifestyle
5354	I am Rich Plus	FAMILY	4.0	856	8700000.0	10000	Paid	399.99	Everyone	Entertainment
5358	I am Rich!	FINANCE	3.8	93	22000000.0	1000	Paid	399.99	Everyone	Finance
5369	I am Rich	FINANCE	4.3	180	3800000.0	5000	Paid	399.99	Everyone	Finance
5373	I AM RICH PRO PLUS	FINANCE	4.0	36	41000000.0	1000	Paid	399.99	Everyone	Finance

Top 5 Reviewed apps

url: head(5)

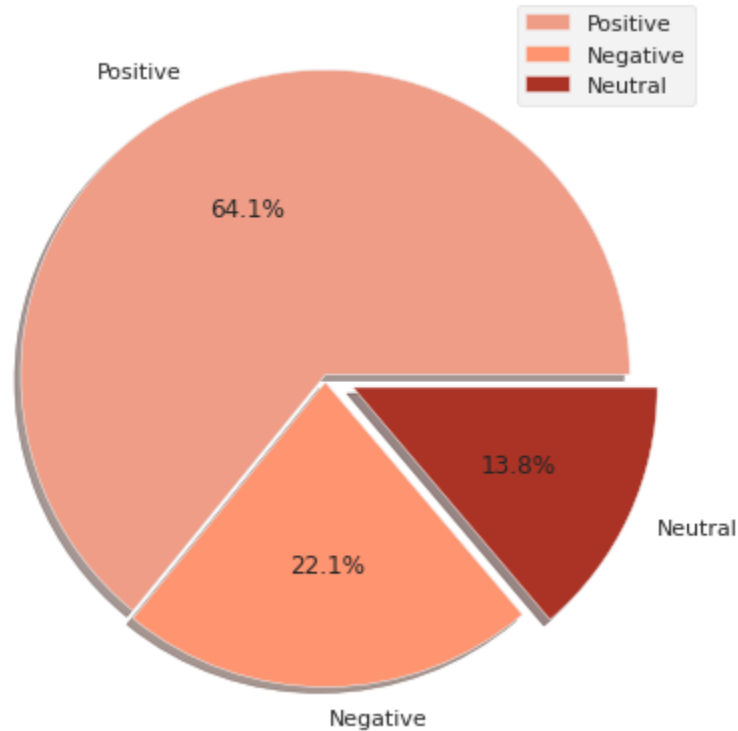


	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
2989	GollerCepte Live Score	SPORTS	4.2	9992	31M	1,000,000+	Free	0	Everyone	Sports	May 23, 2018	6.5	4.1 and up
4970	Ad Block REMOVER - NEED ROOT	TOOLS	3.3	999	91k	100,000+	Free	0	Everyone	Tools	December 17, 2013	3.2	2.2 and up
2723	SnipSnap Coupon App	SHOPPING	4.2	9975	18M	1,000,000+	Free	0	Everyone	Shopping	January 22, 2018	1.4	4.3 and up
2705	SnipSnap Coupon App	SHOPPING	4.2	9975	18M	1,000,000+	Free	0	Everyone	Shopping	January 22, 2018	1.4	4.3 and up
3079	US Open Tennis Championships 2018	SPORTS	4.0	9971	33M	1,000,000+	Free	0	Everyone	Sports	June 5, 2018	7.1	5.0 and up

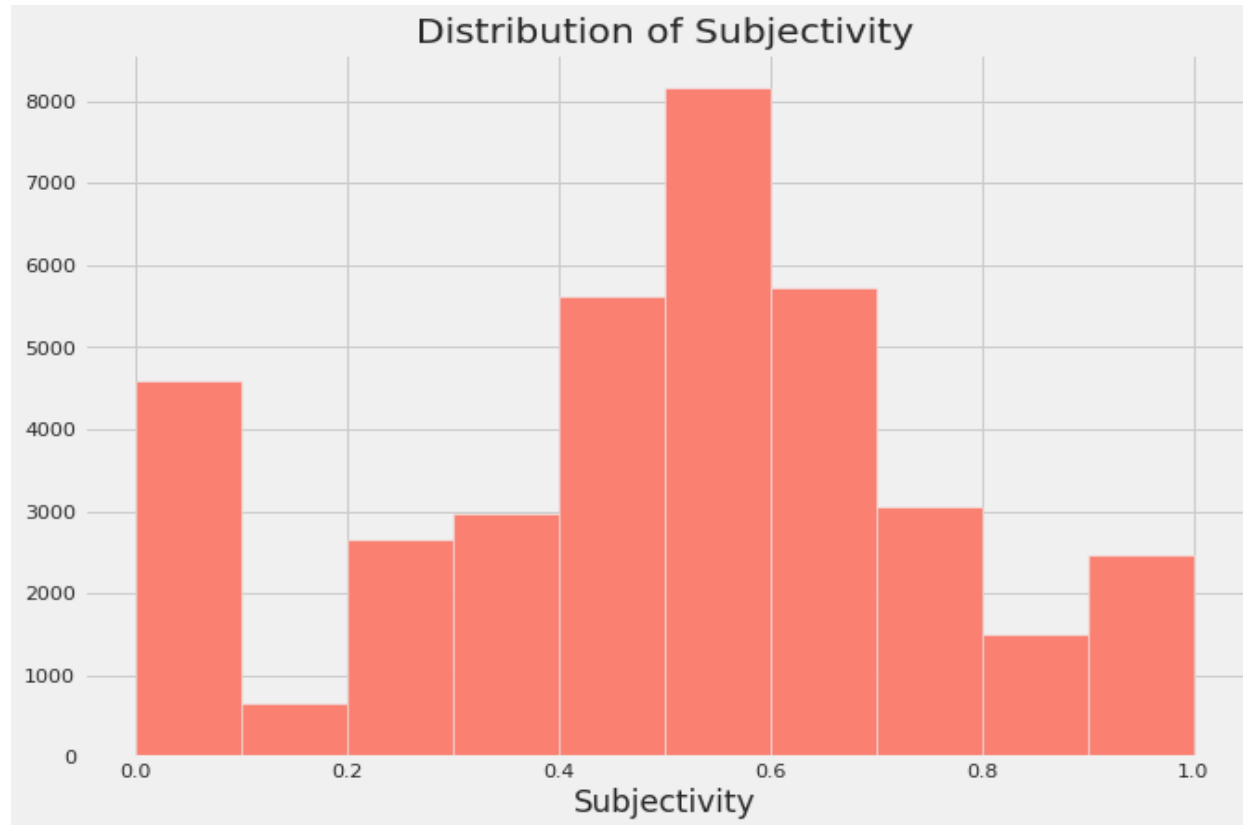


Percentage of review sentiments

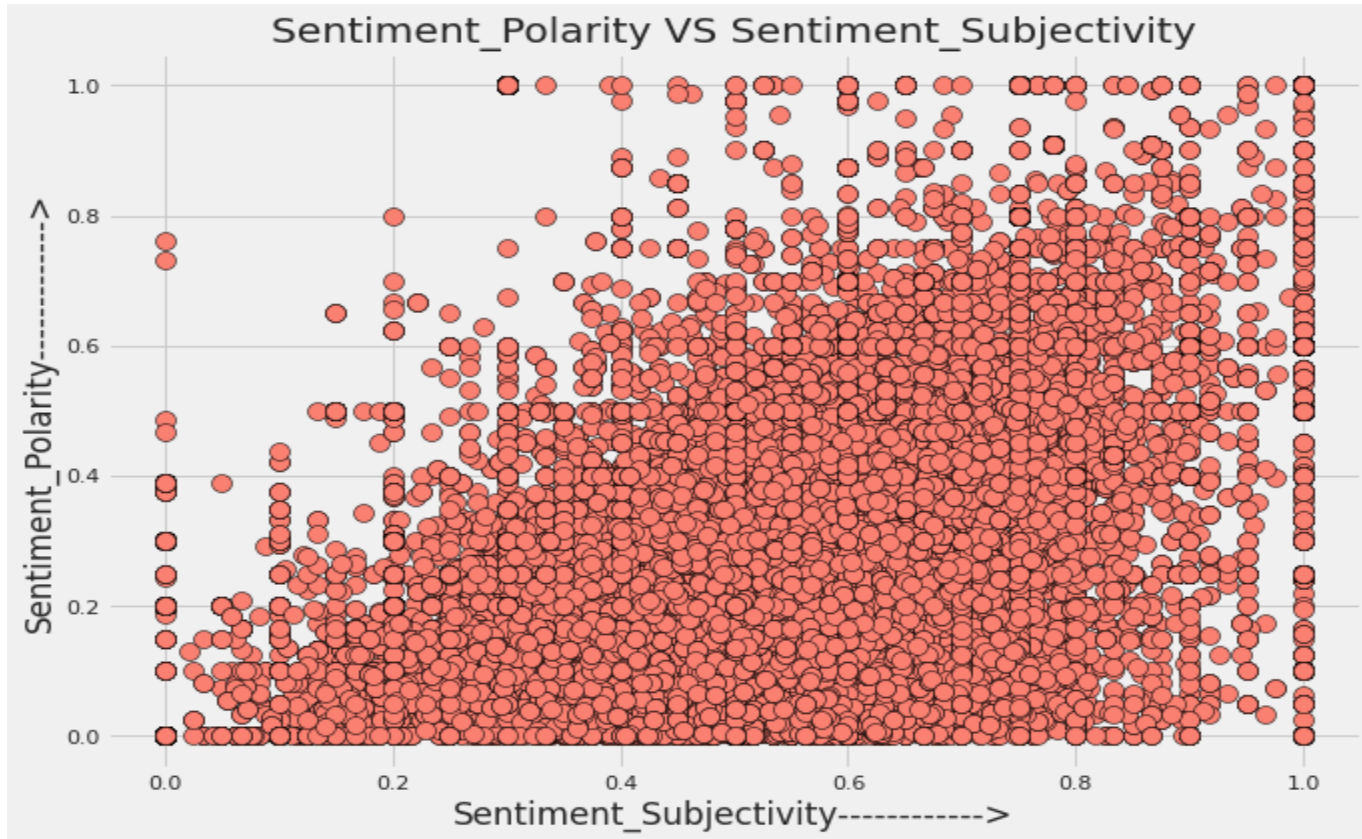
A Pie Chart Representing Percentage of Review Sentiments



Distribution of subjectivity



Sentiment_polarity VS Sentiment_Subjectivity



Conclusion

- Rating is very important factor for installation of apps as user mostly like to watch rating before using app, so developer should also work on updating their content as per the ratings.
- From all above we analyze rating and installation are related, so owners should encourage to write review of their app.
- Gaming have maximum number of installed app so anyone want to become developer can join this category.
- App category like events and beauty have not much reach, so one can also keep this consideration.
- Most of the apps are downloaded by teens, so users of other age category, must also be encouraged to install the apps

Conclusion

- Maximum number of apps present in FAMILY category then GAME category and then TOOLS
- 81.8% app can use every age group person.
- Adults only 18+ and unrated have very less number of apps.
- Gaming have maximum number of installed app.
- Around 92.2% app are freely available on play store. • 4.3 to 4.5 have maximum average rating on play store.

Challenges

- Computation time.
- Reading the dataset and comprehending the problem statement. Our major challenge was data cleaning.
- Also dataset contains duplicate data and irrelevant data, need to clear that first was also problematic



