Capstone Project -1

on

Play Store App Review Analysis

By

Shubhangi Dharmik
(Individual)

Data Analysis by following procedure

1. **Problems Statements**

2. **Data Summary**

3. **Data Cleaning**

4. **Data Manipulation**

5. **Data Visualization**

6. **Conclusion**

Problems Statements

- Analysis based on prices for the apps.

- Analysis based on Ratings of the apps.

- Analysis based on the category and types of apps.

- Analysis based on getting Reviews of the apps.

- Analysis based on installation of apps by user.

- Analysis on category of user using these apps.

Data Summary

Data set name – Play Store App Review Analysis including information on apps and its price, type, reviews, rating etc.

Shape –
- Rows - 10841
- Columns – 13

Important Columns - App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Current Version, Android Version

Data Cleaning

```
[ ] df1.isnull().sum()
```

```
App                0
Category           0
Rating          1474
Reviews            0
Size               0
Installs           0
Type               1
Price              0
Content Rating     1
Genres             0
Last Updated       0
Current Ver        8
Android Ver        3
dtype: int64
```

df1.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   App             10841 non-null   object
 1   Category        10841 non-null   object
 2   Rating          9367 non-null    float64
 3   Reviews         10841 non-null   object
 4   Size            10841 non-null   object
 5   Installs        10841 non-null   object
 6   Type            10840 non-null   object
 7   Price           10841 non-null   object
 8   Content Rating  10840 non-null   object
 9   Genres          10841 non-null   object
 10  Last Updated    10841 non-null   object
 11  Current Ver     10833 non-null   object
 12  Android Ver     10838 non-null   object
dtypes: float64(1), object(12)
memory usage: 1.1+ MB
```
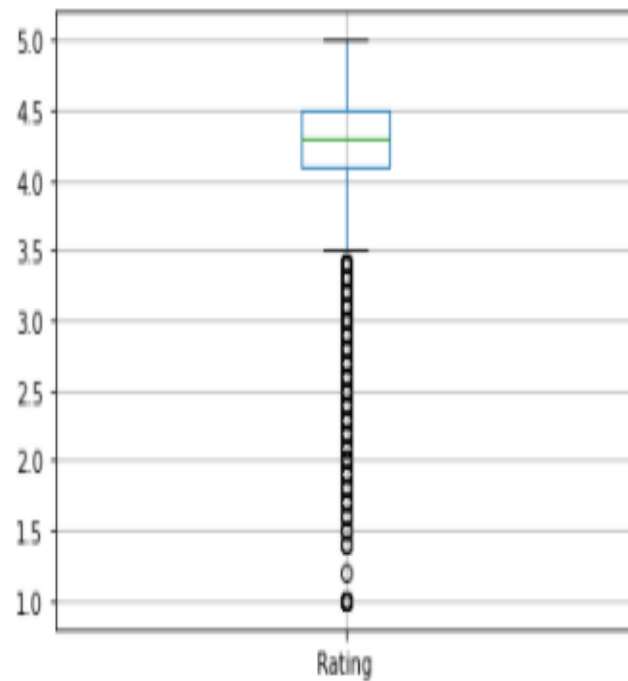
```
df1.boxplot('Rating');
```
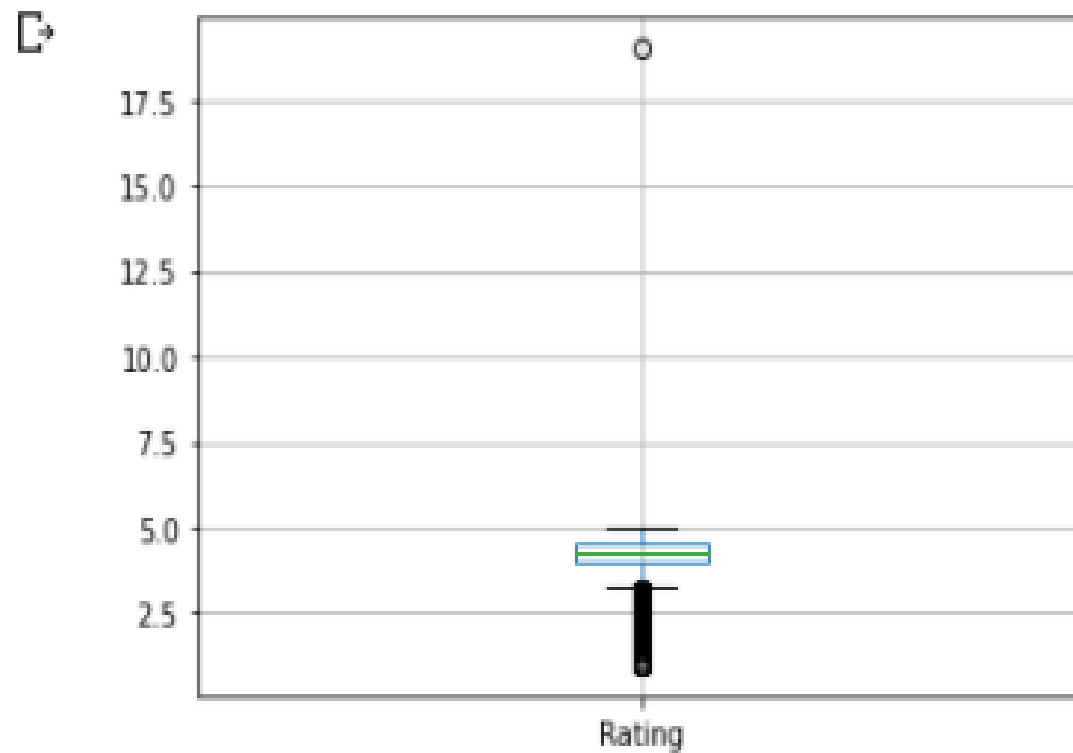
In data cleaning process, some null value are to be remove by dropping and some are removed by filling.
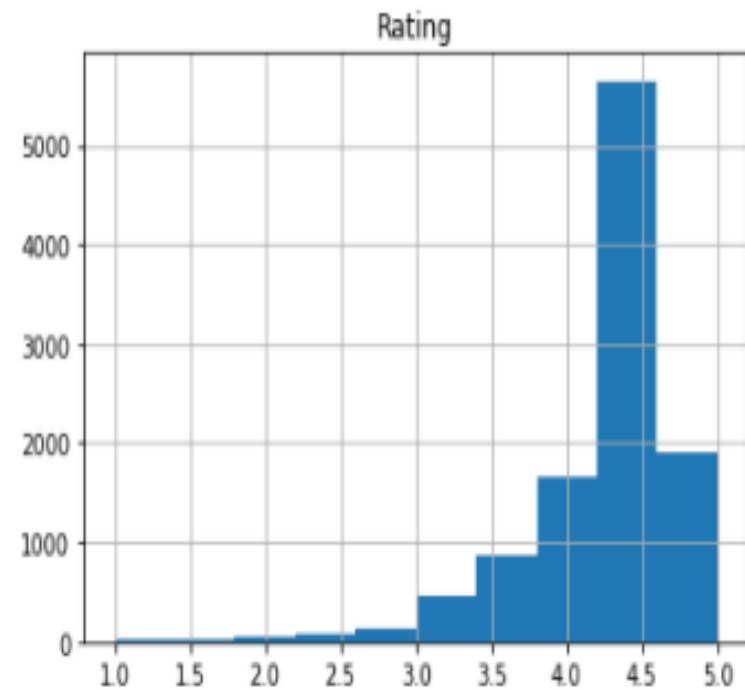
1. Remove NAN value from content Rating.

2. Replace NAN value from Rating.

3. Replace NAN value from Type.

4. Replace NAN value from android version.

5. Replace NAN value from current version.

Data Manipulation

In data manipulation process, data type of the values has to be change

1. Replace value of price from $ to blank.

2. Replace + and , from Installs.

3. Convert object data type of price into float.

4. Convert object data type of installs into float.

5. Convert object data type of reviews into float.

Data Visualization
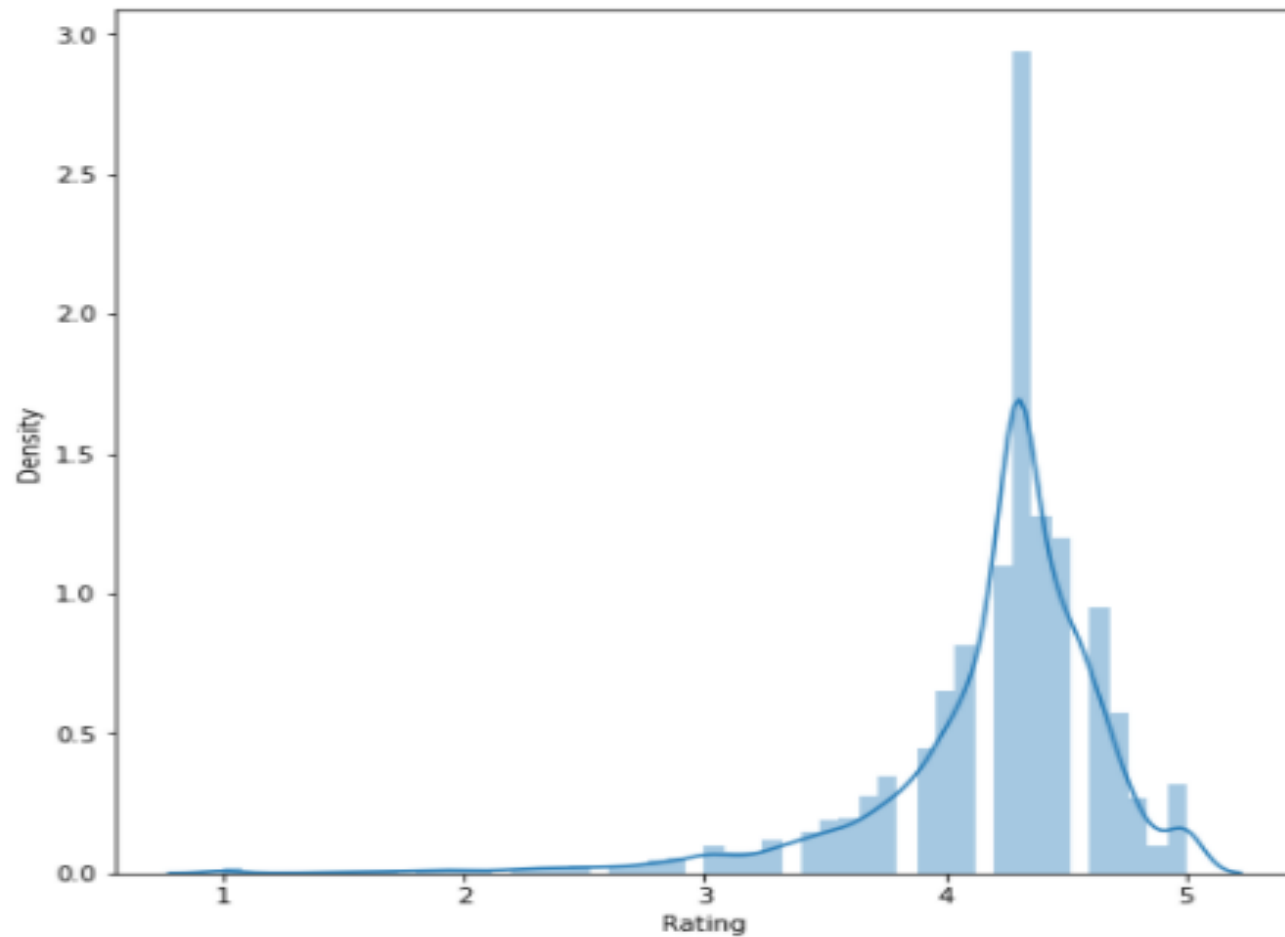
Number of apps in each category

Total apps are divided into 33 categories
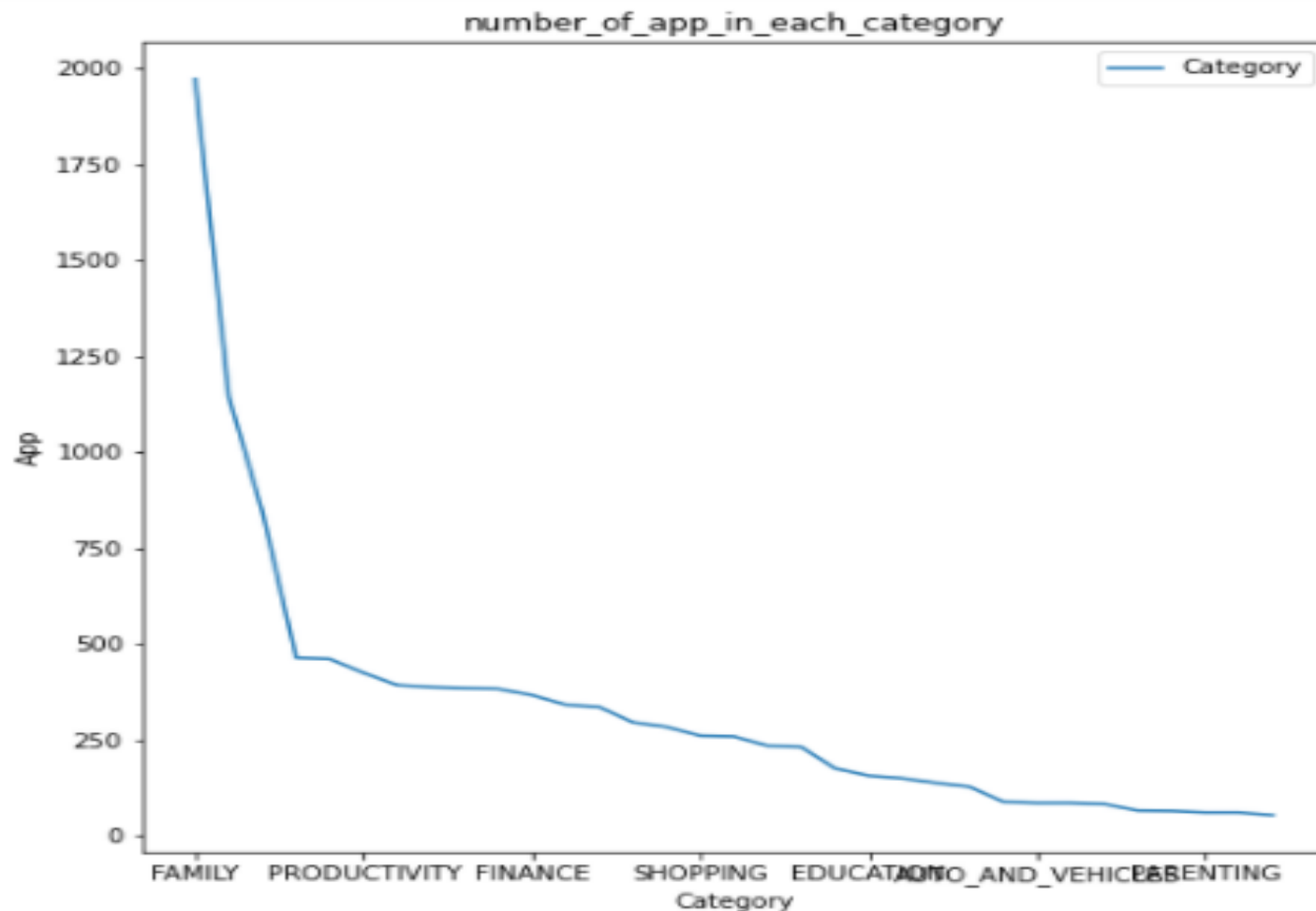
Category	Number	Category	Number
GAME	1144	DATING	234
TOOLS	843	BOOKS_AND_REFERENCE	231
MEDICAL	463	VIDEO_PLAYERS	175
BUSINESS	460	EDUCATION	156
PRODUCTIVITY	424	ENTERTAINMENT	149
PERSONALIZATION	392	MAPS_AND_NAVIGATION	137
COMMUNICATION	387	FOOD_AND_DRINK	127
SPORTS	384	HOUSE_AND_HOME	88
LIFESTYLE	382	AUTO_AND_VEHICLES	85
FINANCE	366	LIBRARIES_AND_DEMO	85
HEALTH_AND_FITNESS	341	WEATHER	82
PHOTOGRAPHY	335	ART_AND_DESIGN	65
SOCIAL	295	EVENTS	64
NEWS_AND_MAGAZINES	283	PARENTING	60
SHOPPING	260	COMICS	60
TRAVEL_AND_LOCAL	258	BEAUTY	53

Analysis based on ratings

Analysis based on Ratings for each category of apps

There are 274 apps having 5 Star Ratings

ratings for the category
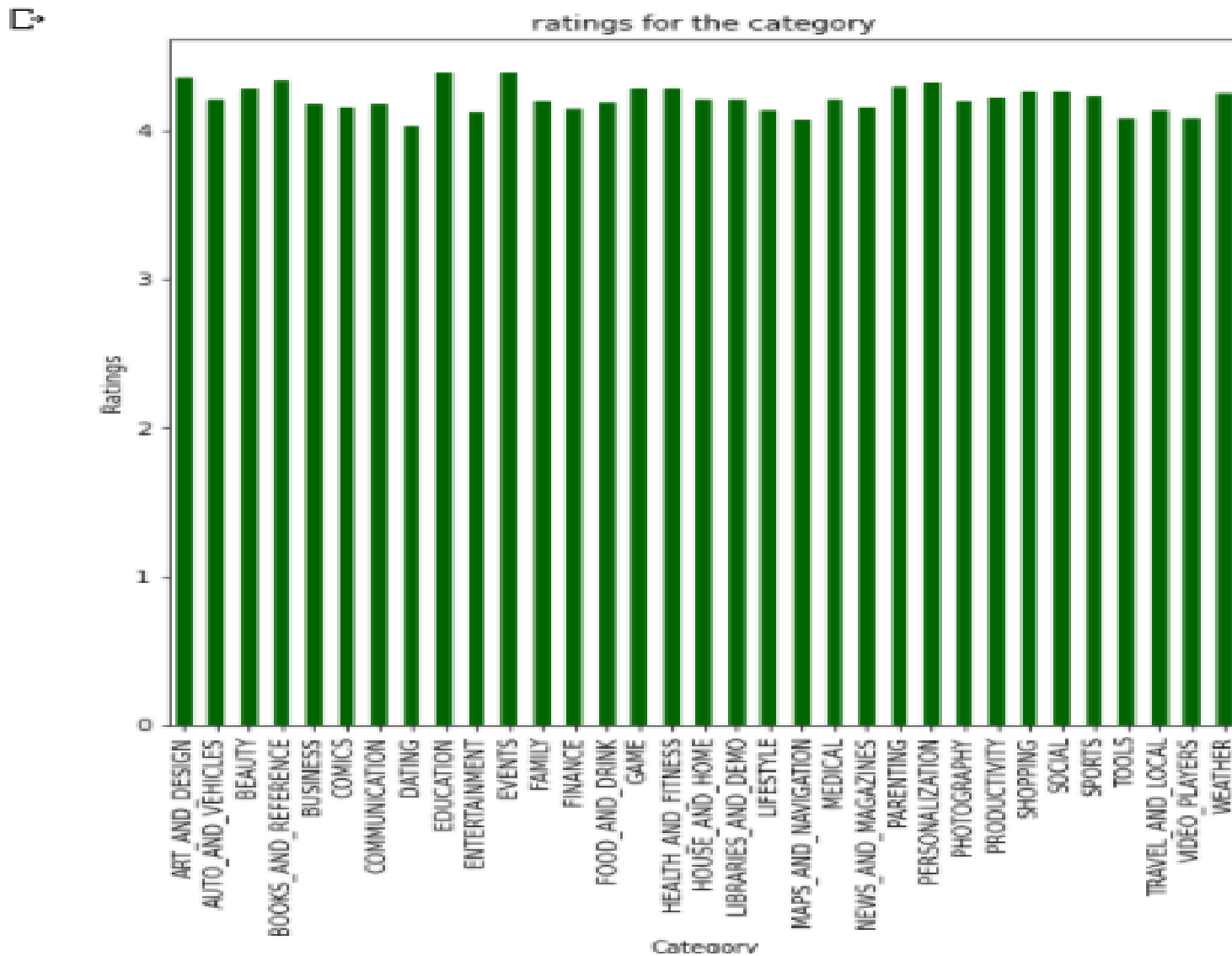
Analysis based on rating for the types of apps
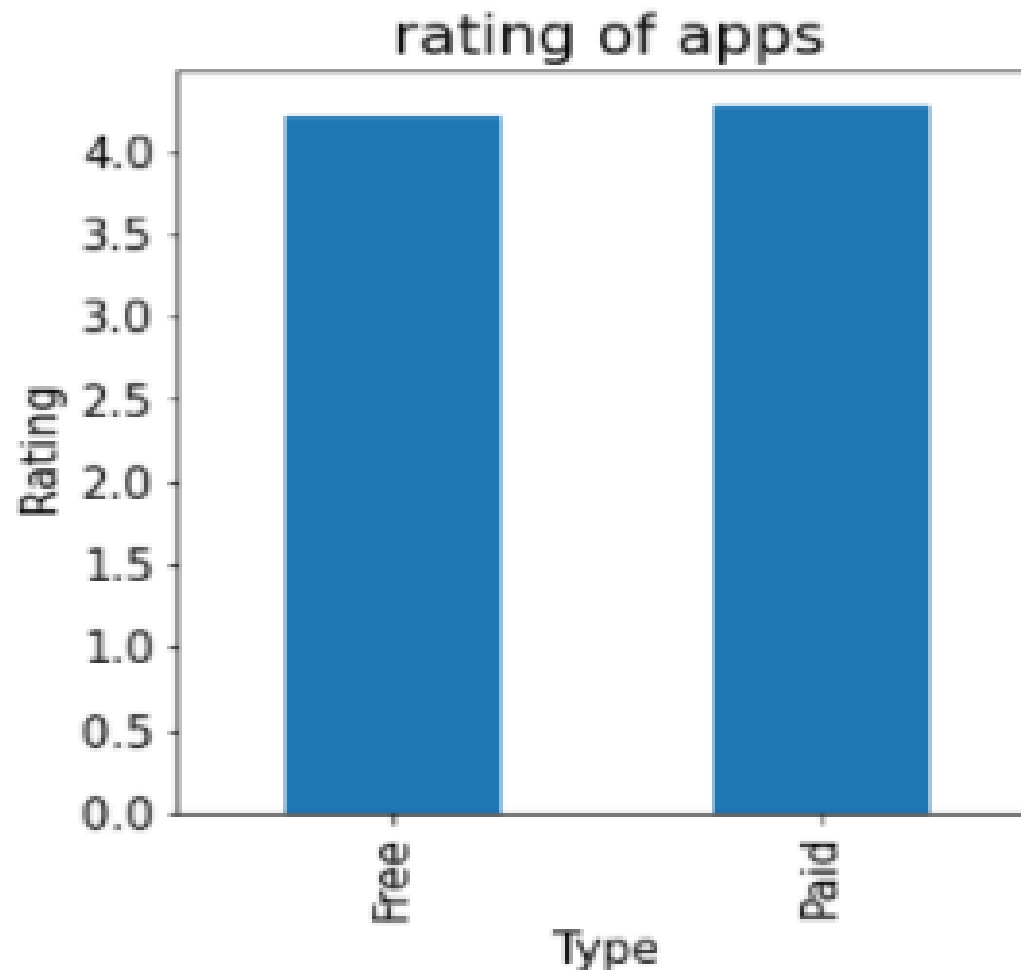
Total apps are divided into two types

| SR NO	TYPES	NUMBER	RATING
1.	FREE APPS	10040	4.20
2.	PAID APPS	800	4.27

Analysis based on Installs
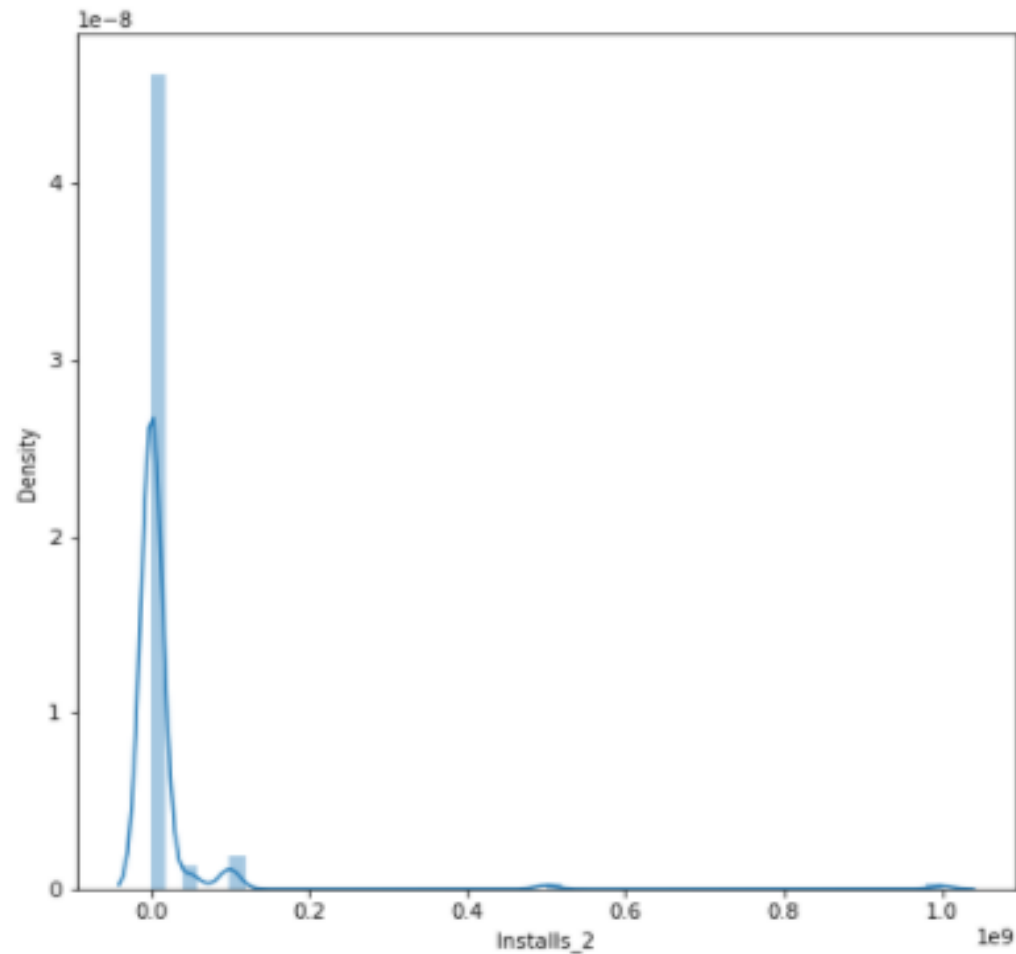
Top 5 apps which has maximum installation

```
[ ]   updated_df.nlargest(5,['Installs_2'])
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Price_1	Installs_1	Installs_2
152	Google Play Books	BOOKS_AND_REFERENCE	3.9	1433233.0	Varies with device	1,000,000,000+	Free	0	Teen	Books & Reference	August 3, 2018	Varies with device	Varies with device	0.0	1,000,000,000	1.000000e+09
335	Messenger – Text and Video Chat for Free	COMMUNICATION	4.0	56642847.0	Varies with device	1,000,000,000+	Free	0	Everyone	Communication	August 1, 2018	Varies with device	Varies with device	0.0	1,000,000,000	1.000000e+09
336	WhatsApp Messenger	COMMUNICATION	4.4	69119316.0	Varies with device	1,000,000,000+	Free	0	Everyone	Communication	August 3, 2018	Varies with device	Varies with device	0.0	1,000,000,000	1.000000e+09
338	Google Chrome: Fast & Secure	COMMUNICATION	4.3	9642995.0	Varies with device	1,000,000,000+	Free	0	Everyone	Communication	August 1, 2018	Varies with device	Varies with device	0.0	1,000,000,000	1.000000e+09
340	Gmail	COMMUNICATION	4.3	4604324.0	Varies with device	1,000,000,000+	Free	0	Everyone	Communication	August 2, 2018	Varies with device	Varies with device	0.0	1,000,000,000	1.000000e+09

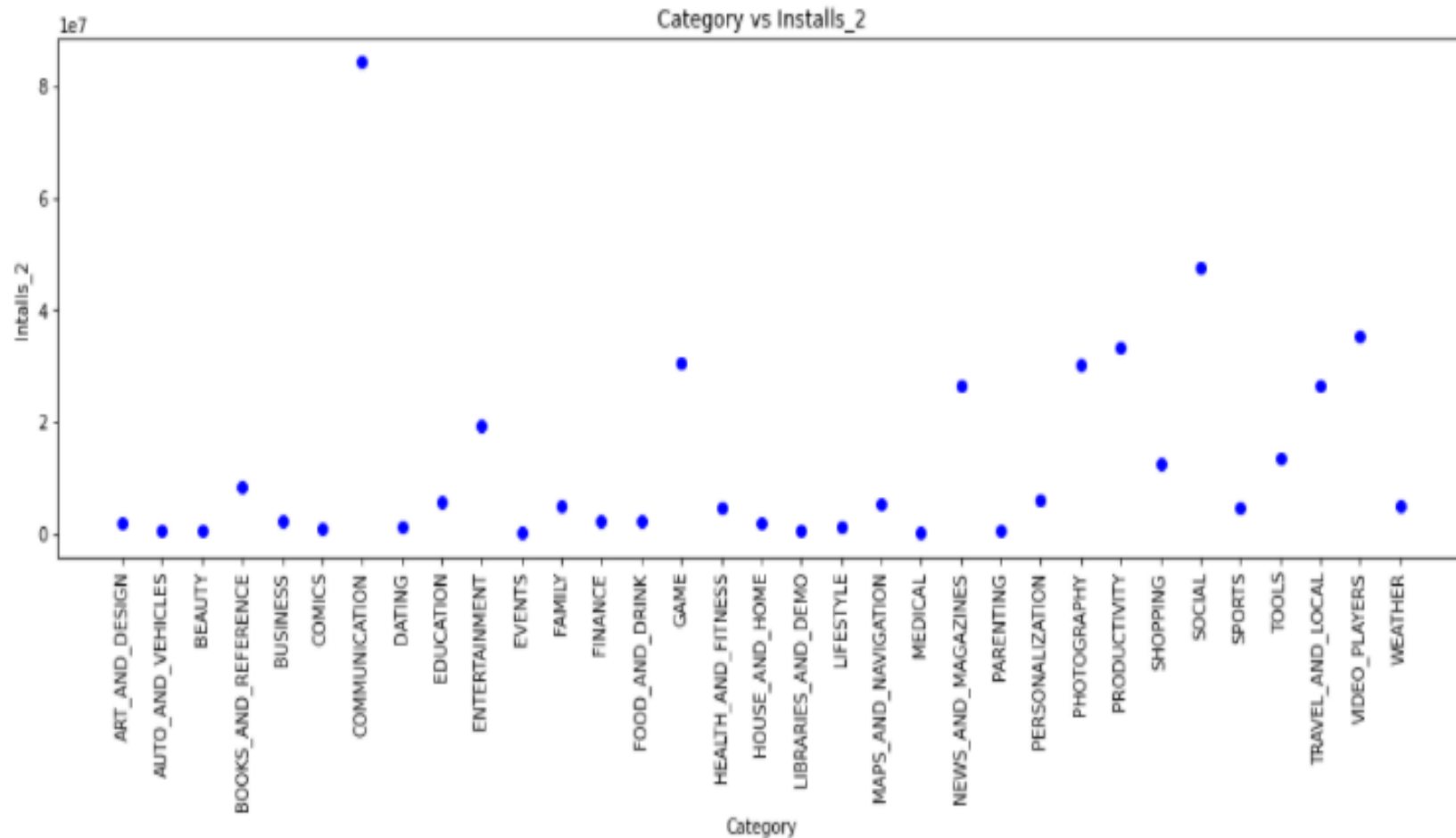# Analysis based on user those using apps

```
[ ]  updated_df.groupby(['Content Rating'])['Rating'].count().plot.bar(color="darkgreen")
     plt.show()
```
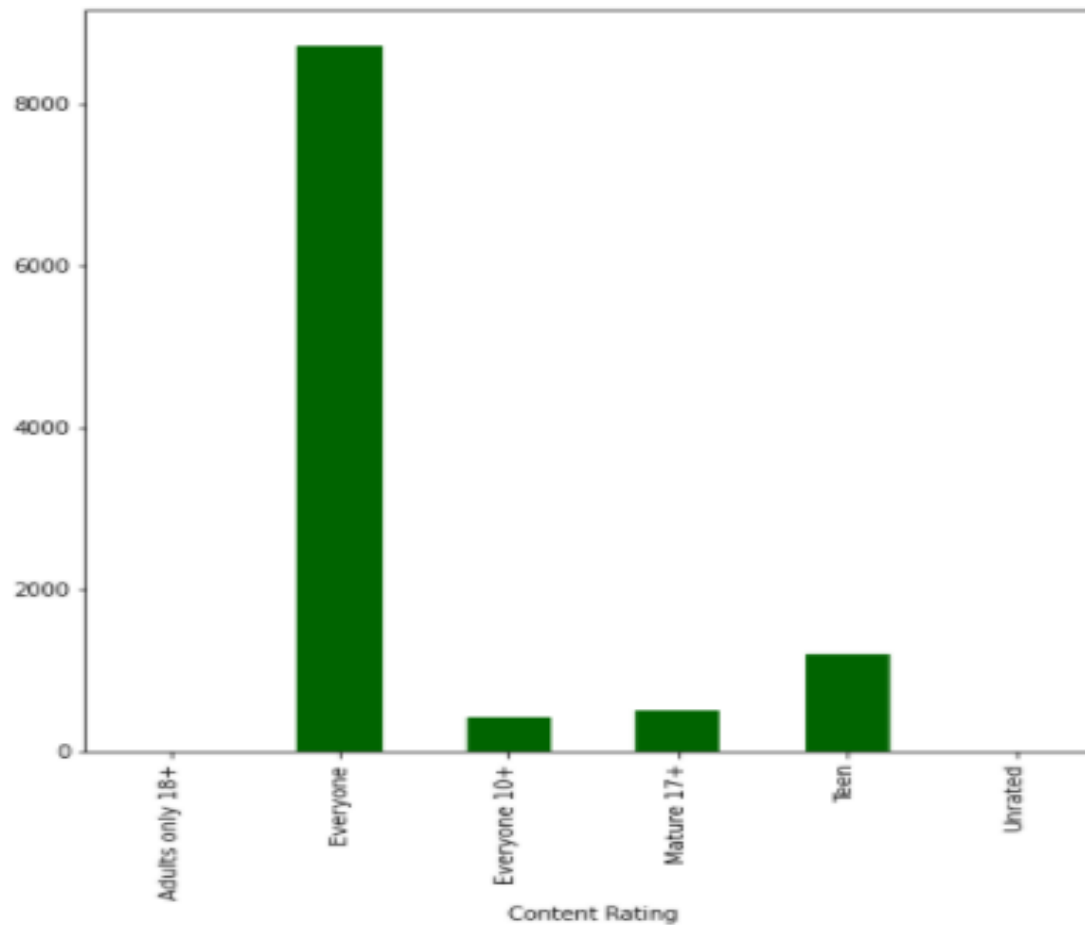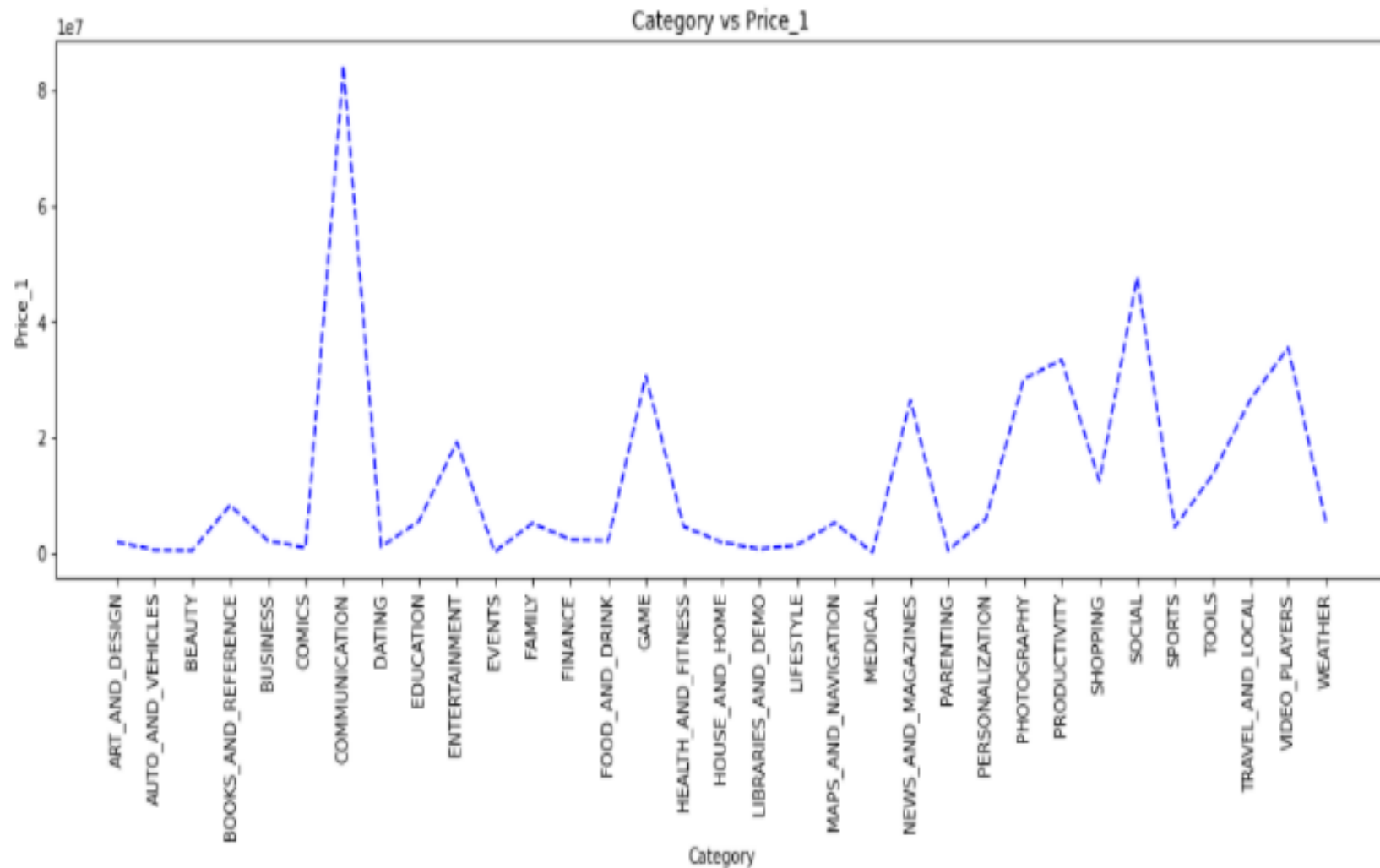
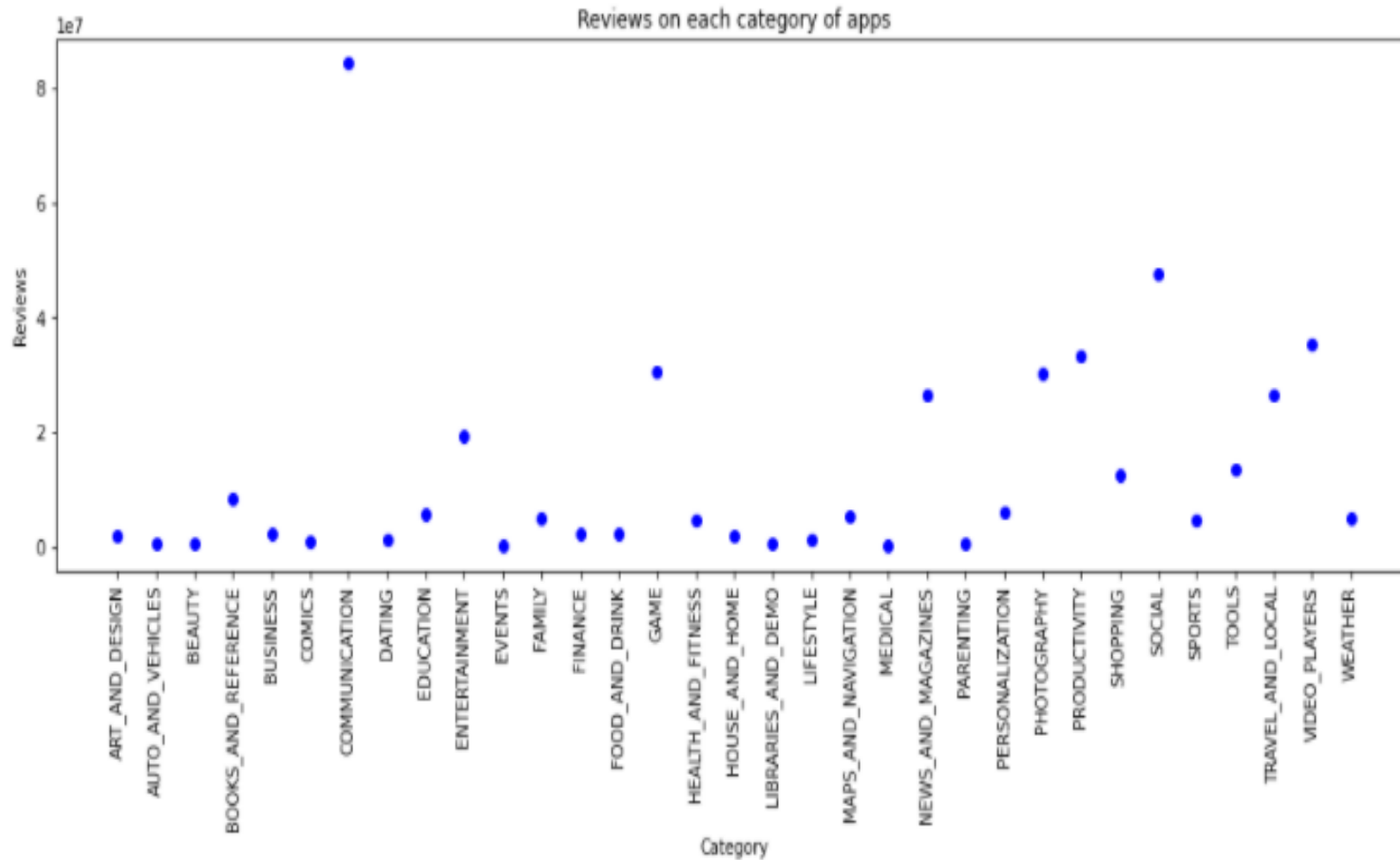# Analysis based on the category of apps and on installation

Analysis on the category of apps based on prices

Reviews on each category of apps

Text(0, 0.5, 'Reviews')

Reviews on each category of apps

Conclusions

1. Total number of apps are divided into 33 categories.

2. Apps are divided into two types i) Free Apps ii) Paid Apps

3. There are 10040 Free apps and 800 Paid apps in the database which has 4.2 and 4.27 rating respectively.

4. Communication Category apps has maximum installs.

5. Events and Education Category apps has maximum ratings

6. Communication category apps has maximum price too.

7. Facebook and WhatsApp Messenger app has maximum Reviews.

8. Maximum price of the app is I'm Rich-Trump Edition.

9. Top 5 maximum installs apps are: i) Google play Books ii) Messengers-Text and Video for Free iii) WhatsApp Messengers iv) Google Chrome: Fast and Secure v) Gmail

10. There are 274 apps which has 5 star ratings.

Thank You