

Capstone Project-1 Play Store App review analysis

Member
Shubham Awale



Data Summary

We will complete this project by using following steps-

- Here we have two csv files. We will upload both the files to read the data
- After reading the data we will do the Data Cleaning operation.
- We will perform the some statistical operation on our dataset.
- We will do the exploratory analysis and data visualization.
- After all the analysis we will come to the conclusion.



DATA

SUMMARY
Play Store apps
data

CATEGORY

APP

RATING

REVIEWS

INSTALLS

SIZE

TYPE

PRICE

GENRES

LAST UPDATED

CURRENT VER

ANDROID VERSION





SENTIMENT

SENTIMENT POLARITY

SENTIMENT SUBJECTIVITY



Objective

- Installation of application by users according to the categories.
- What are the mostly demanded applications in playstore.
- Factors that affect the installation of application by the user.



Dependent Variables

Play store csv file -

- Category
- Ratings
- Reviews
- Size
- Type
- Price



Dependent Variables

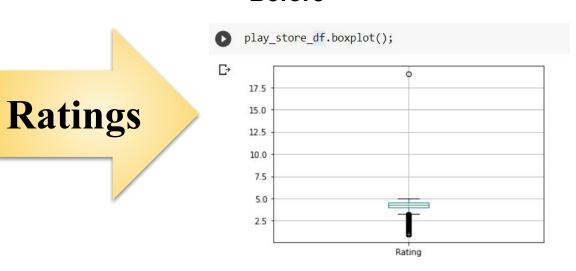
User review csv file -

- Sentiment
- Sentiment Subjectivity

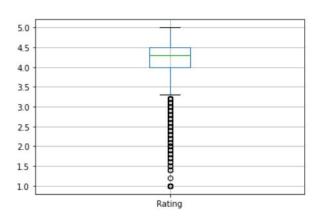


Data Cleaning

Before



After removing outliers





Data Cleaning

Before

play_store_df.isnull().sum()

Category 0 Rating 1474 Reviews 0 Size 0 Installs 0 Type 1 Price 0 Content Rating 0 Genres 0 Last Updated 0 Current Ver 8 Android Ver 2 dtype: int64

After

play_store_df.isnull().sum()

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

Null Values



Data Cleaning

Before

play_store_df.info()

class 'pandas.core.frame.DataFrame'> RangeIndex: 10841 entries, 0 to 10840 Data columns (total 13 columns):

Duca	COTAMINO (COCAT	15 COLUMNIS / 1	
#	Column	Non-Null Count	Dtype
	22222		
0	Арр	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	Туре	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
dtype	es: float64(1),	object(12)	
memoi	ry usage: 1.1+ M	В	

After

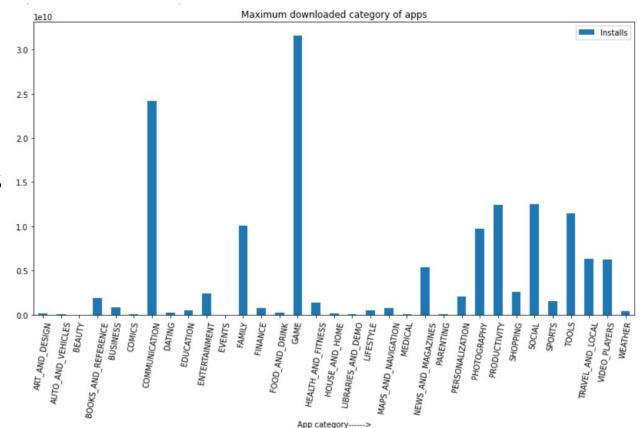
- play_store_df.info()
- <class 'pandas.core.frame.DataFrame'> Int64Index: 10357 entries, 0 to 10840 Data columns (total 13 columns):

#	Column	Non-Null Count	Dtype
0	Арр	10357 non-null	object
1	Category	10357 non-null	object
2	Rating	10357 non-null	float64
3	Reviews	10357 non-null	int64
4	Size	10357 non-null	float64
5	Installs	10357 non-null	int64
6	Туре	10357 non-null	object
7	Price	10357 non-null	float64
8	Content Rating	10357 non-null	object
9	Genres	10357 non-null	object
10	Last Updated	10357 non-null	object
11	Current Ver	10357 non-null	object
12	Android Ver	10357 non-null	object
	es: float64(3), ry usage: 1.1+ M	int64(2), object B	(8)

Data Info

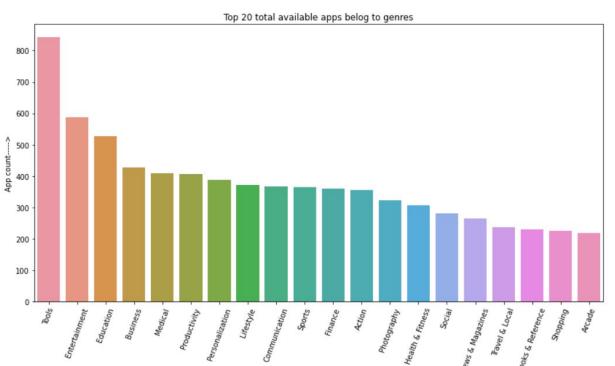


Apps having maximum installs by category



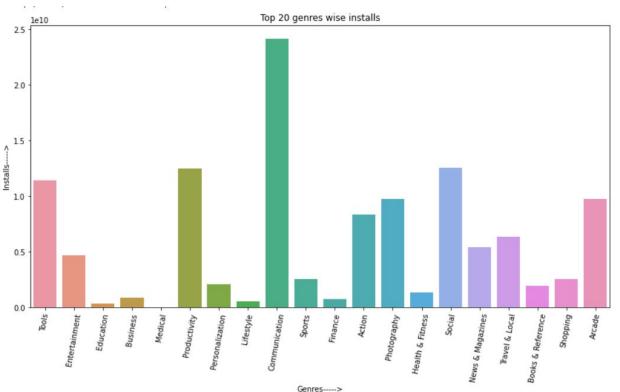


Top 20 genres and their apps present in data



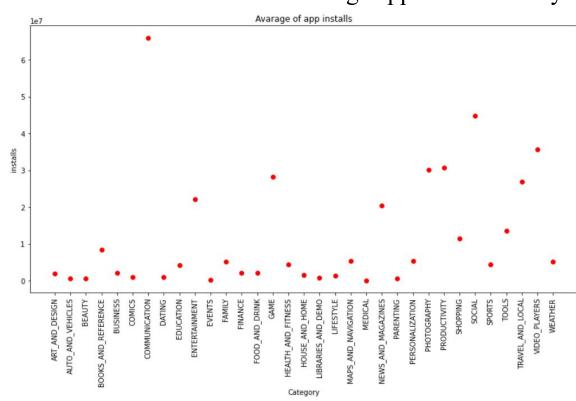


The most popular applications are under communication genres





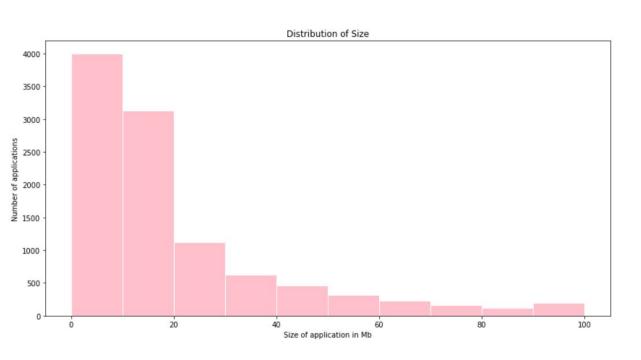
Average app installation by category



The communication apps having maximum average installs



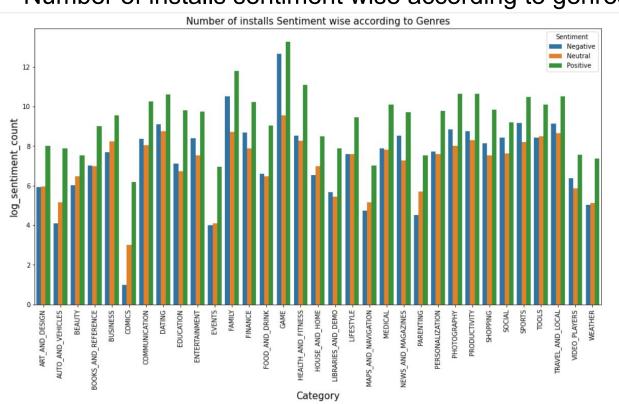
Apps present on play store with respect to their size



In the range of 0 Mb to 20 Mb their are maximum apps present on the play store

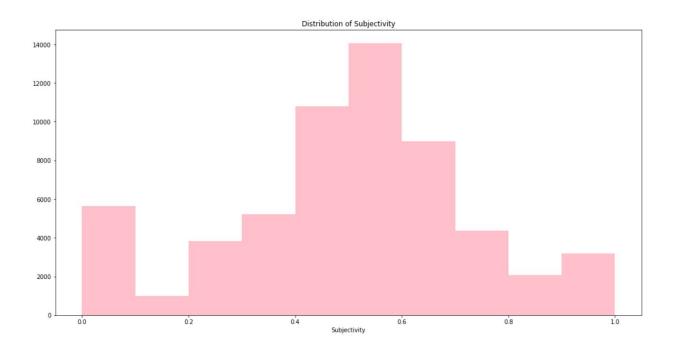


Number of installs sentiment wise according to genres





Subjectivity - Personal opinion of the user about the application





Conclusion

- Communication category apps are in high demand for all type of users.
- Game apps and communication apps have maximum downloads
- Users installs the application depends according to previous rating and reviews
- There are many tools app present in play store
- The social apps have maximum average installs
- Users installs the apps depends on previous rating and review



Challenges

Understanding the sentiments of the user age wise.

Difference between installing the application and using the application.



References

https://seaborn.pydata.org/examples/index.html

https://matplotlib.org/3.1.1/index.html



Q & A