Capstone Project - 1

Individual : Play Store App Review analysis

Name

Azeem Husain

OBJECTIVE:

The play store apps data has enormous potential to drive app making business to success. Actionable insights can be drawn for developers to work on and capture the Android market. We have two files to analysis and we need to find some results for developers

Process to analysis over a data:

1. **Paste useful Libraries**
2. **Paste all ‘csv’ files**
3. **Data Cleaning**
   a. **Check for null values**
   b. **Check for Outliers**
   c. **Remove Column for less Data Entries**
4. **Data Manipulation**
5. **Data Visualization**

Questions on which all analysis are focus:

1. **Which type of apps have most chances to be installs by user.**
2. **Which type of apps can be easily sold.**
3. **Which type of apps are more reliable.**
4. **Which type of audience are mostly targeted by Play Store apps.**
5. **How Size, Reviews, Installs and Price are correlated.**

‘CSV’ Files

1.Play Store Data

2.User Reviews Data

Libraries, which will be helpful during our analysis

1. Numpy
2. Pandas
3. Seaborn
4. Matplotlib

Data Cleaning

a. Check for null values

Isnull()

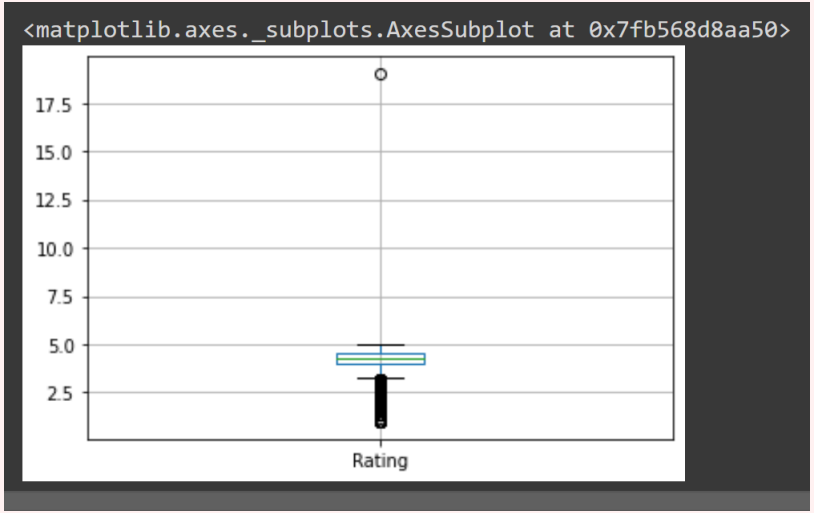
Isnull().sum()

b. Remove Column for less Data Entries

dropna()

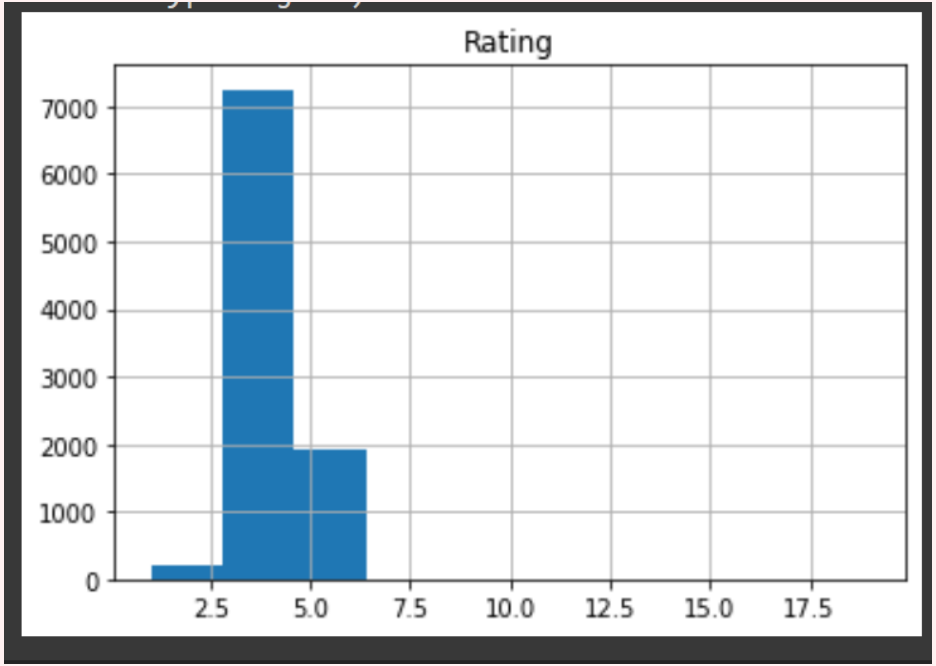
Check for Outlier in Play Store Data by Boxplot

It shows an outlier present in column ‘Rating’.



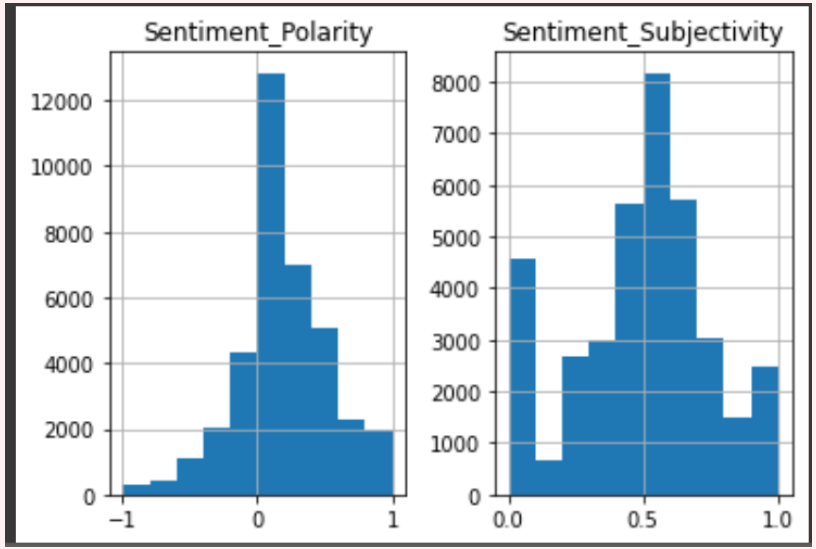
Check for Outlier in Play Store Data by Histogram

It shows an outlier present in column ‘Rating’.



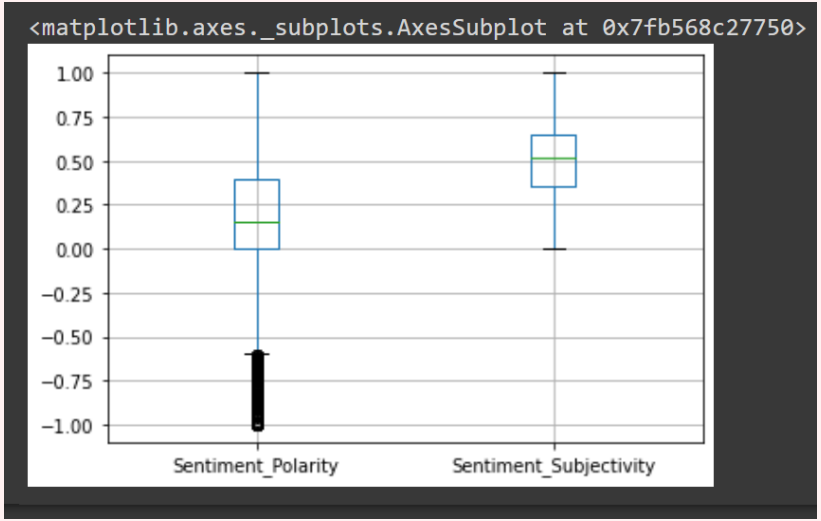
Check for Outlier in User Reviews Data by Histogram

It shows no outlier present in User Reviews Data.



Check for Outlier in User Reviews Data by Boxplot

It shows no outlier present in User Reviews Data.

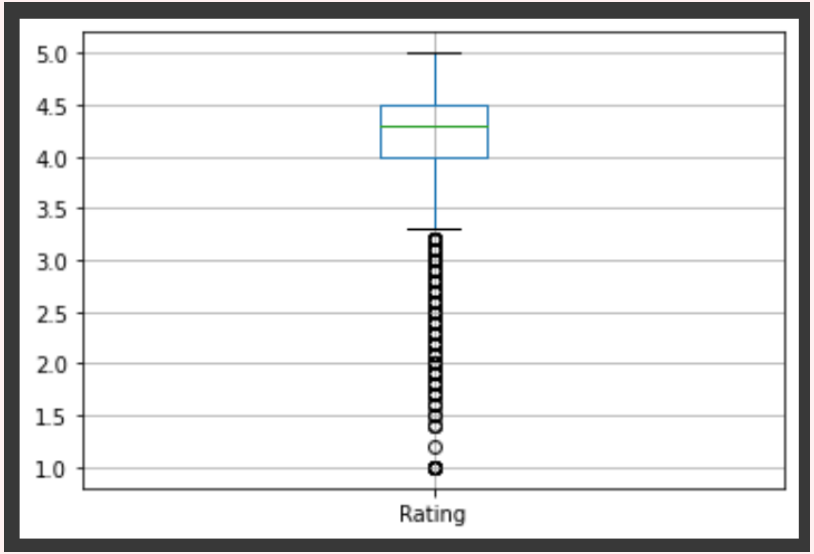


Outlier removing

We can remove outlier by drop() method.

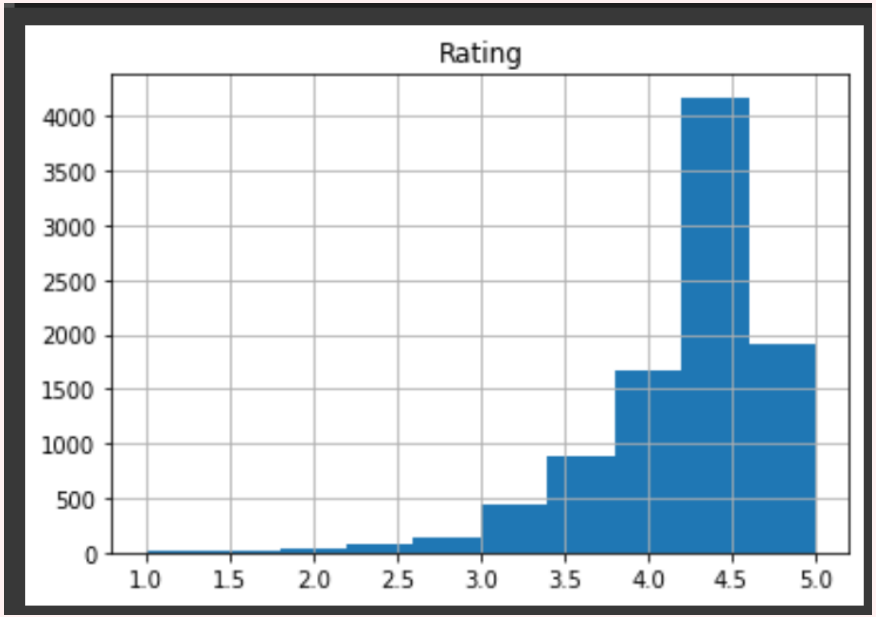
Boxplot after outlier removing

This Boxplot shows that Outlier is removed in Rating.



Histogram after outlier removing

This Boxplot shows that Outlier is removed in Rating.



Data Manipulation

fillna()

This method comes into Data Manipulation we can fill all null values by this method.

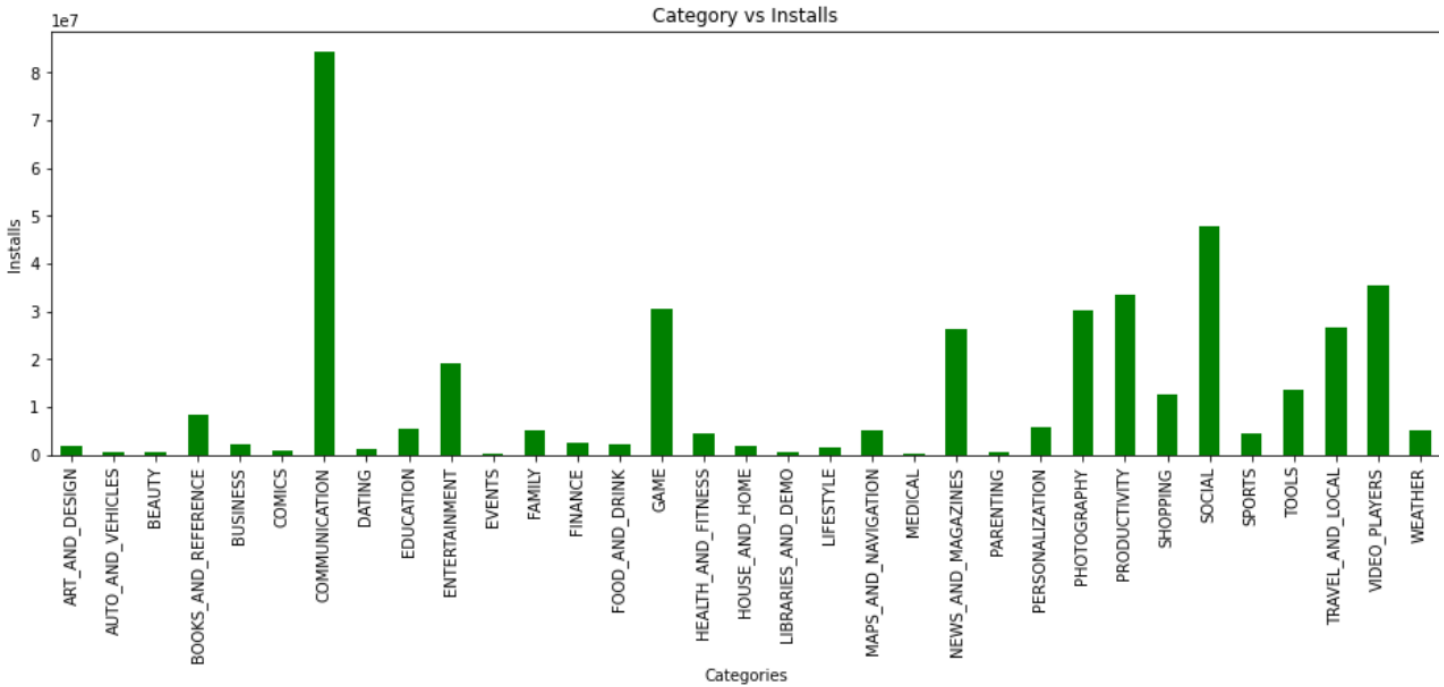
Data Visualization

We will visualize our data to solve these problems.

1. **Which type of app have most chances to be installs by user.**
2. **Which type of apps can be easily sold.**
3. **Which type of apps are more reliable.**
4. **Which type of audience are mostly targeted by Play Store apps.**
5. **How Size, Reviews, Installs and Price are correlated.**

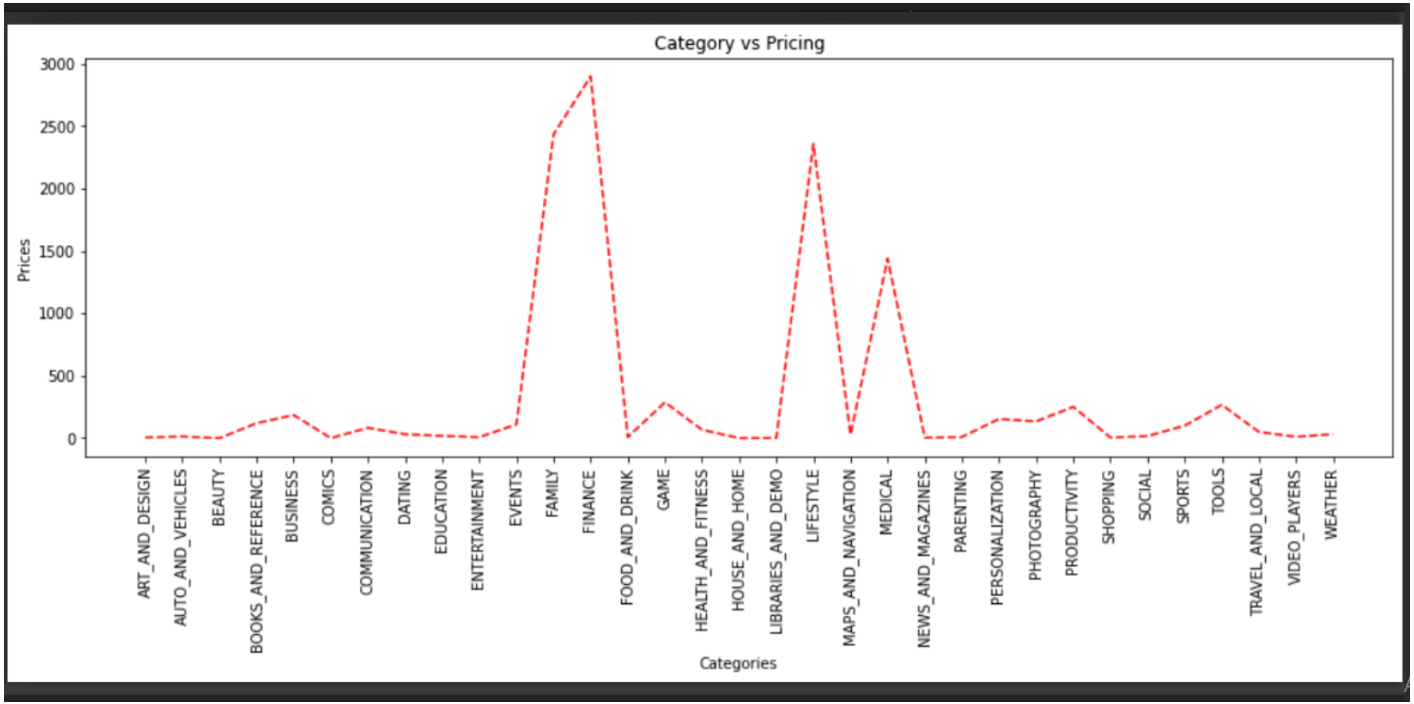
Category vs Installs

It shows mostly user prefer Communication apps.



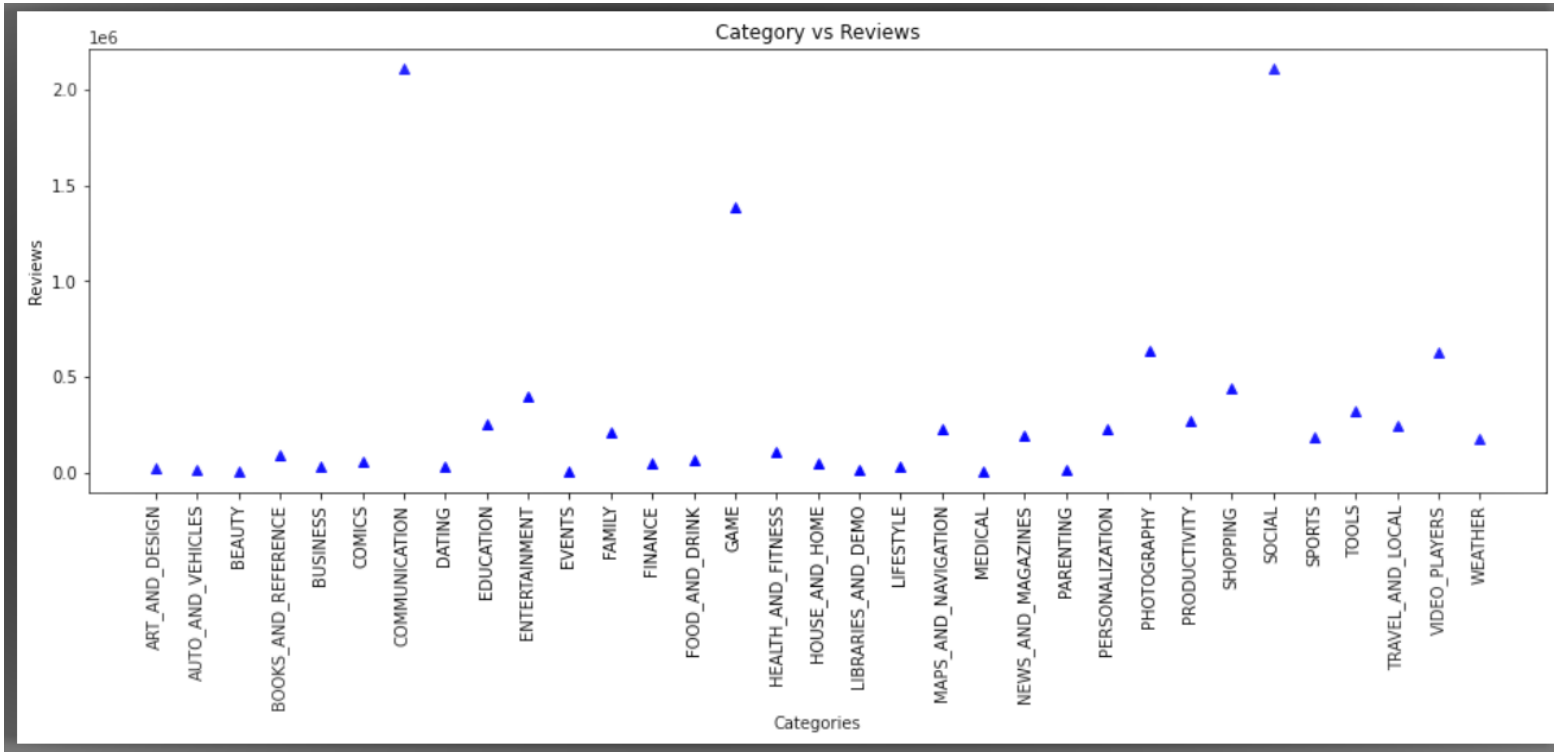
Category vs Price

It shows that Family, Finance, Lifestyle are those categories which can be easily sold out.



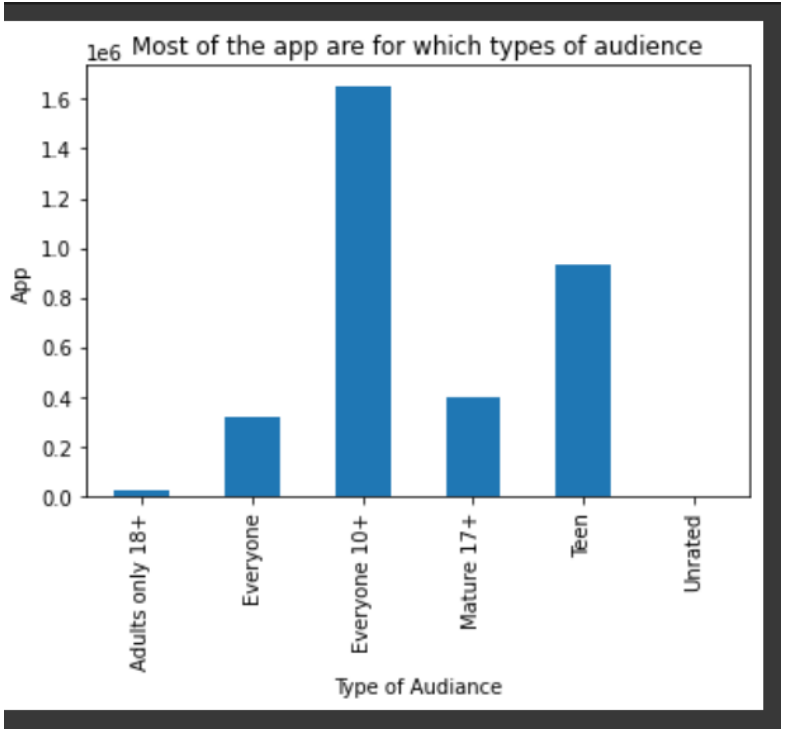
Category vs Reviews

It shows that communication, Game, Social categories app are more reliable than others.



Audience targeted by Play Store

- In this chart we can see that
- mostly apps are created for
- teen agers and other people
- who are over 10 year old.



Correlation between Size, Reviews, Installs and Price

By this graph we can see how a app effect with his price, reviews, installs and his size.

Conclusion:

There are some conclusion of our data analysis:-

1. You can believe on Communication, Social apps to install because they have lot of reviews.
2. If you want to make an app you need to make only Games, Social and Communication app because it can be easily install by users.
3. People buy only Finance, Lifestyle and Family app so other apps should be free of cost.
4. You should target everyone as an audience.
5. Size, Reviews and Price affect the installation of app.

Challenges during Analysis:

1. Lot of chunk need to handle.
2. Indentation need focus.
3. Lot of Practice need during analysis.
4. Selection of method also a challenge it can increase or decrease time.

Q & A

ThankYou …