CAPSTONE PROJECT PLAY STORE APP REVIEW ANALYSIS

BY SARATH LONAPPAN

Points for Discussion

- ~ Data Summary
- ~Category Review Comparison
- ~ Category Install Comparison
- ~Content Rating
- ~Rating for each genre
- ~Apps in each category
- ~Translated Review in each Sentiment
- ~Size and Price relation
- ~Correlation heatmap analysis
- ~Possible pairplots
- ~Conclusion

We have two dataframes with shape (10841,13) and (64295,5)

First dataframe - consists mostly of numerical data like price,installs,size and textual data like category,genres,Android version.

| [] | [] dfl.head() | | | | | | | | | | | | | |
|----|----------------|--|----------------|--------|---------|------|-------------|------|-------|-------------------|------------------------------|---------------------|--------------------|----------------|
| | | Арр | Category | Rating | Reviews | Size | Installs | Туре | Price | Content Rating | Genres | Last Updated | Current Ver | Android Ver |
| | 0 | Photo Editor & Candy Camera & Grid & ScrapBook | ART_AND_DESIGN | 4.1 | 159 | 19M | 10,000+ | Free | 0 | Everyone | Art & Design | January 7, 2018 | 1.0.0 | 4.0.3 and up |
| | 1 | Coloring book moana | ART_AND_DESIGN | 3.9 | 967 | 14M | 500,000+ | Free | 0 | Everyone | Art & Design;Pretend Play | January 15, 2018 | 2.0.0 | 4.0.3 and up |
| | 2 | U Launcher Lite – FREE Live Cool Themes, Hide | ART_AND_DESIGN | 4.7 | 87510 | 8.7M | 5,000,000+ | Free | 0 | Everyone | Art & Design | August 1, 2018 | 1.2.4 | 4.0.3 and up |
| | 3 | Sketch - Draw & Paint | ART_AND_DESIGN | 4.5 | 215644 | 25M | 50,000,000+ | Free | 0 | Teen | Art & Design | June 8, 2018 | Varies with device | 4.2 and up |
| | 4 | Pixel Draw - Number Art Coloring Book | ART_AND_DESIGN | 4.3 | 967 | 2.8M | 100,000+ | Free | 0 | Everyone | Art & Design;Creativity | June 20, 2018 | 1.1 | 4.4 and up |
| | % | | | | | | | | | | | | | |

Second dataframe - Mainly regards with the user feedback information like reviews, sentiments.

[] df2.head()

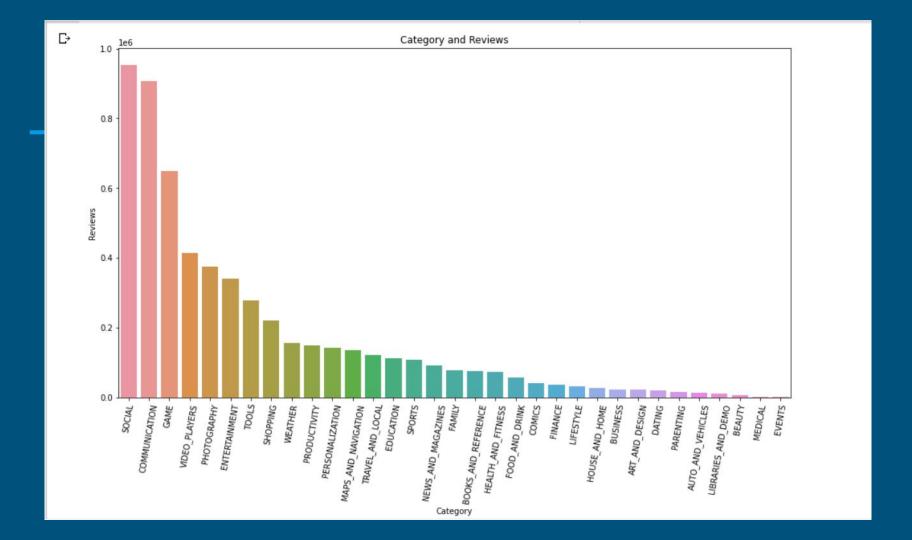
| 10 Best Foods for YouI like eat delicious food. That's I'm cookingPositive1.000.53333310 Best Foods for YouThis help eating healthy exercise regular basisPositive0.250.28846210 Best Foods for YouNaNNaNNaNNaN10 Best Foods for YouWorks great especially going grocery storePositive0.400.87500010 Best Foods for YouBest idea usPositive1.000.300000 | Арр | Translated_Review | Sentiment | Sentiment_Polarity | Sentiment_Subjectivity |
|---|-----------------------|---|-----------|--------------------|------------------------|
| 10 Best Foods for You NaN NaN NaN NaN NaN NaN NaN 10 Best Foods for You Works great especially going grocery store Positive 0.40 0.875000 | 10 Best Foods for You | I like eat delicious food. That's I'm cooking | Positive | 1.00 | 0.533333 |
| 10 Best Foods for You Works great especially going grocery store Positive 0.40 0.875000 | 10 Best Foods for You | This help eating healthy exercise regular basis | Positive | 0.25 | 0.288462 |
| | 10 Best Foods for You | NaN | NaN | NaN | NaN |
| 10 Best Foods for You Best idea us Positive 1.00 0.300000 | 10 Best Foods for You | Works great especially going grocery store | Positive | 0.40 | 0.875000 |
| | 10 Best Foods for You | Best idea us | Positive | 1.00 | 0.300000 |

Category Review Comparison

Review is the formal assessment with the intention of instituting change if necessary.

Reviews make the app more efficient.

From the graph in the next slide, we get that Social and Communication category are reviewed more while Medical and Events are reviewed less.

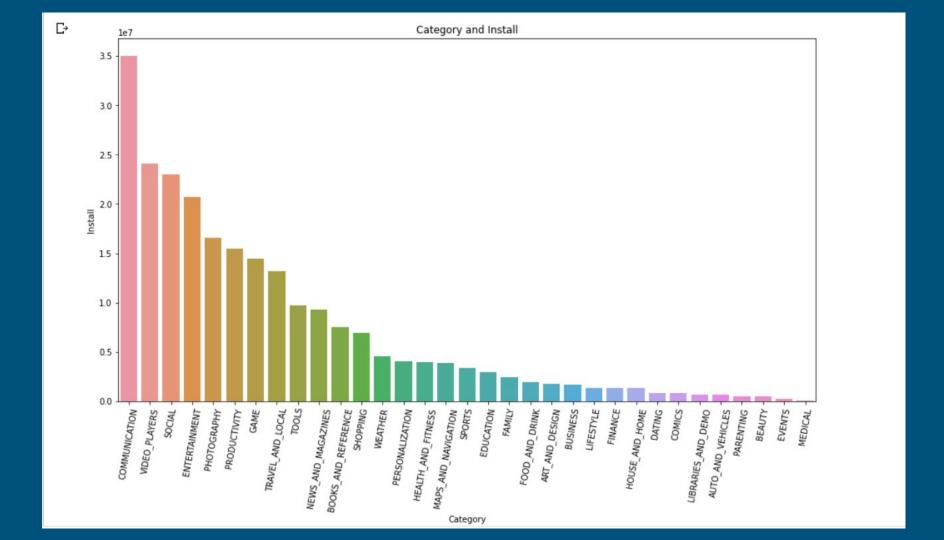


Category Install Comparison

Downloading of apps is the only way of earning for the app developers.

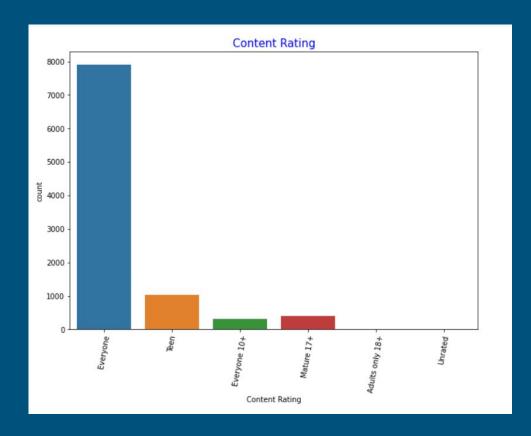
So they have to ensure that app released should make an impact among users and get a high count of downloads.

From graph, Communication category has the highest number of downloads while Medical and Events are having least downloads.



Content Rating

- Content rating is essential for every app as some apps will be prevented from the usage of kids. Some games and dating apps will be some of the examples because of its matured content.
- From graph, apps that can be used by everyone are more.
- Adults only and unrated apps are released least.



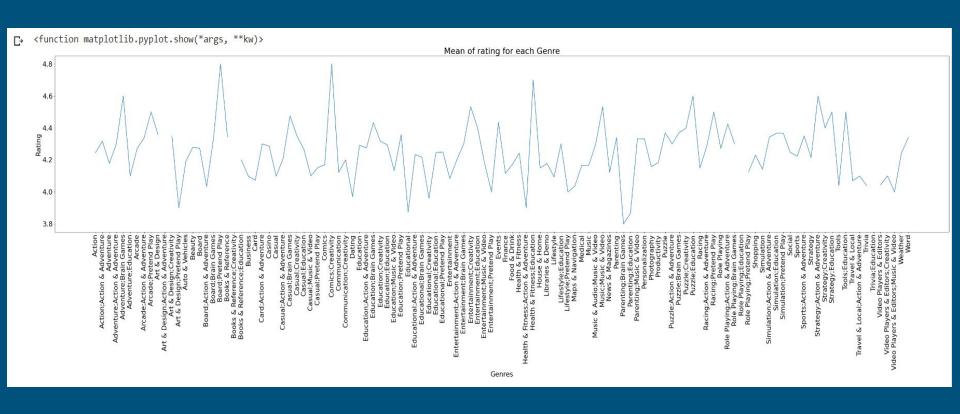
Rating for each Genre

The mean rating for each genre is compared

 Comics and Creativity, Board and brain games are two genres that has high mean rating.

Parenting and brain game genre has the lowest rating.

The graph has 4 non joining points. It shows the presence of null values.



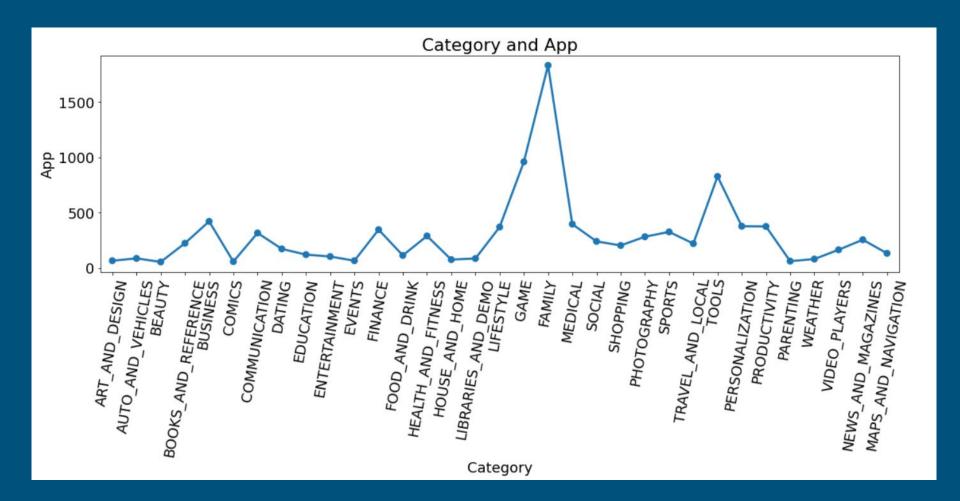
Number of apps in each Category

The apps developed will be more in the category which is common to all age and gender classifications.

Dating apps will be used by matured people.

Beauty apps will be more used by females.

From graph Beauty, Comics and Parenting has the minimum number of apps while Family has the most number of apps.

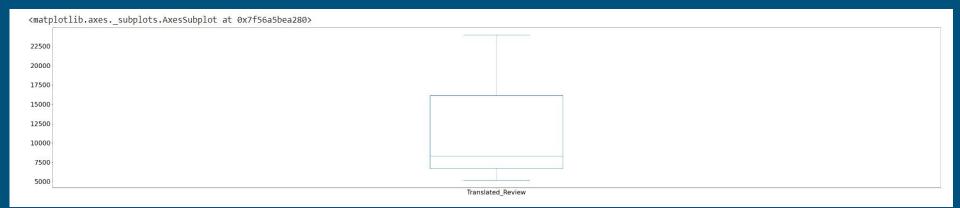


Translated Review for each sentiment

Positive, Negative and Neutral are the different sentiments.

From the boxplot, we can see that the minimum value is the neutral sentiment.

The maximum value is the positive sentiment while middle line is the negative sentiment



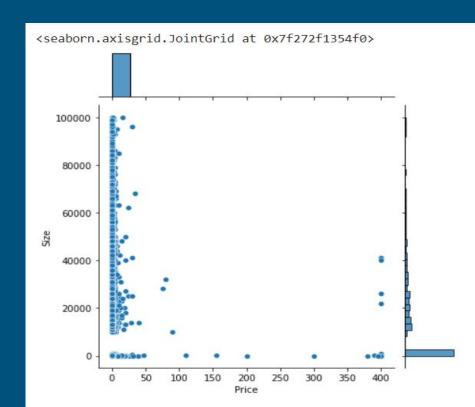
Size and Price

Jointplot consist of 3 plots where outer bars represent the distribution of independent variables x and y.Inside is a bivariate graph which shows the how y varies with x.

Size and price are not proportional to each other.

From graph ,the paid apps have comparatively low size.

Apps that have high size are free.



Correlation heatmap analysis

Correlation is used to find the pairwise correlation of all columns in the Pandas Dataframe.

The value of most positively correlated value is 1 while most negatively correlated value is -1.

From heatmap, we see that reviews and installs are positively correlated which is denoted in orange squares.

Price is negatively correlated with all the columns thus having negative values.

| | Rating | Reviews | Size | Installs | Price |
|----------|-----------|-----------|-----------|-----------|-----------|
| Rating | 1.000000 | 0.055087 | 0.061619 | 0.040181 | -0.021144 |
| Reviews | 0.055087 | 1.000000 | 0.176260 | 0.625165 | -0.007597 |
| Size | 0.061619 | 0.176260 | 1.000000 | 0.133997 | -0.022134 |
| Installs | 0.040181 | 0.625165 | 0.133997 | 1.000000 | -0.009404 |
| Price | -0.021144 | -0.007597 | -0.022134 | -0.009404 | 1.000000 |
| | | | | | |



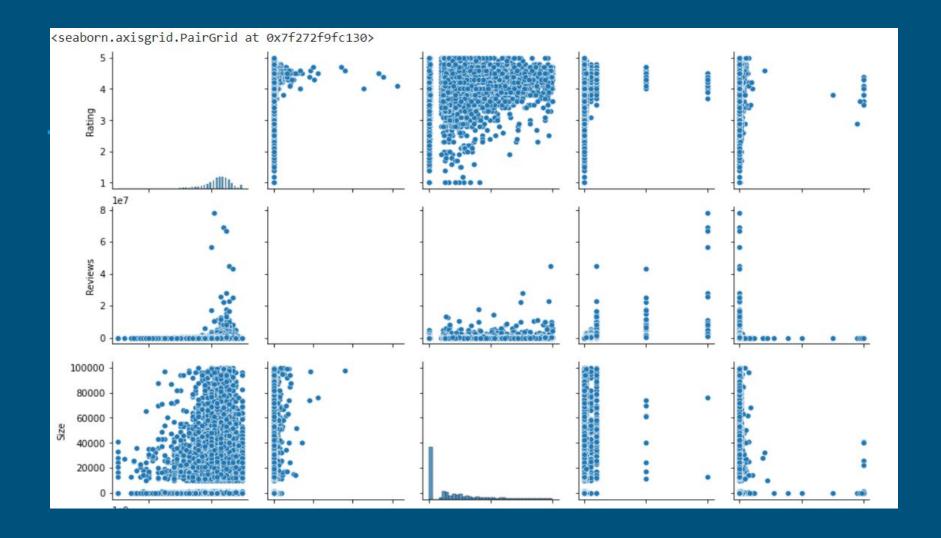
Possible pairplots

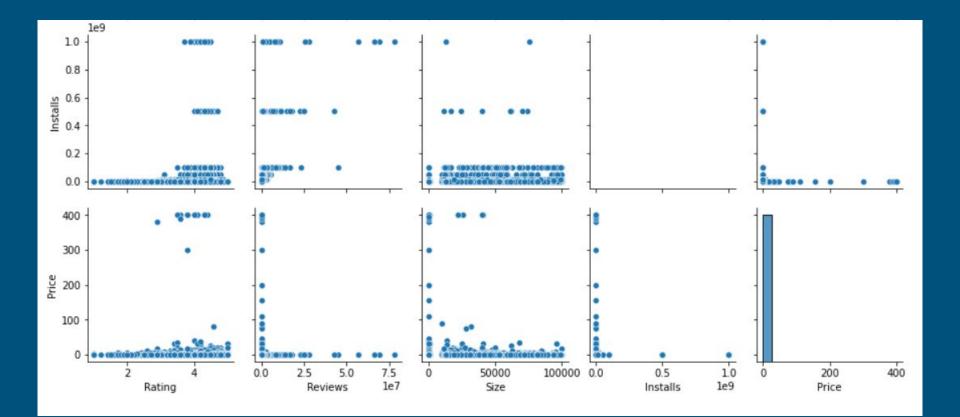
Pairplot is used to plot multiple bivariate distributions in a dataset.

The diagonal plots are the univariate plots. They will be histograms.

Pair plot is used to understand the best set of features to explain a relationship between two variables.

This also helps to form some classification models by drawing simple lines or making linear separations in our data set.





Conclusion

- The categories that have high reviews are Social and Communication
- That categories that have high installs are Communication and video player.
- But as video players have no much reviews ,users can't select the best video player app.
- Beauty,Medical and Events apps are low in number of reviews and installs.So developing a more user friendly and added features in these apps may increase the number of installs thus increasing the reviews.
- Apps used by everyone is more. Age restricted apps will be used by people in that age category thus reducing the installation of those apps.
- Comics and creativity, Board and brain games has the highest rating while parenting has the lowest rating. So improving the parenting apps features can result in more downloads.

- Family has the most number of apps while beauty,comics and parenting have low number of apps.Developing more apps in beauty,comics and parenting will suggest the users to installs and rate the new apps.
- High sized apps are free.Paid apps doesn't have much size as compared to free apps.

Thank you!!