First Capstone Project

Alma Striker Team :- Play Store App Review Analysis

Team Members

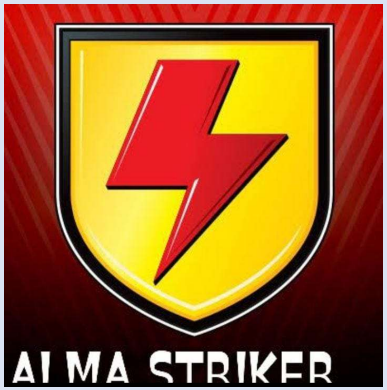
Adesh Dhole

Chetan Mangale

Amol Waghmare

Mohit Sonwane

Akash Soni



Points for discussion

1. Discuss the problem statement
2. Get familiar with the Bussiness logic
3. Connect With the Database
4. Perform analysis on the Data
5. Free apps VS Paid apps
6. Impact of ‘Rating’
7. Conclusion

Defining the problem statement

As we get the detail data of App store data which we have data in the plane csv format. so now first thing we have to be familiar with the database and try to understand the business logic .

after that we can actually start our Analyzation on the data..

Introduction to Play store Platform

- Play store is an Android Market serves as the official app store for certified devices running on the Android Operating system.
- Developed and Operated by Google, launched on 6th March, 2012.
- Also called as Google Play store.
- Approximately 3.48 million apps are in Play store.
- Each apps are having separate product page with user review and rating system.
- All the apps are separated into different categories based on the purpose.

Exploring the database

We have been provided with two databases

- User reviews database
- Shape of this database is (64295, 5)
- Here there are only two numeric values found, Sentiment Subjectivity, Sentiment Polarity.

- Play store app database
- Shape of this database is (10841, 13).
- Out of this thirteen columns we have numeric columns like Rating, Reviews, Size, Installs.

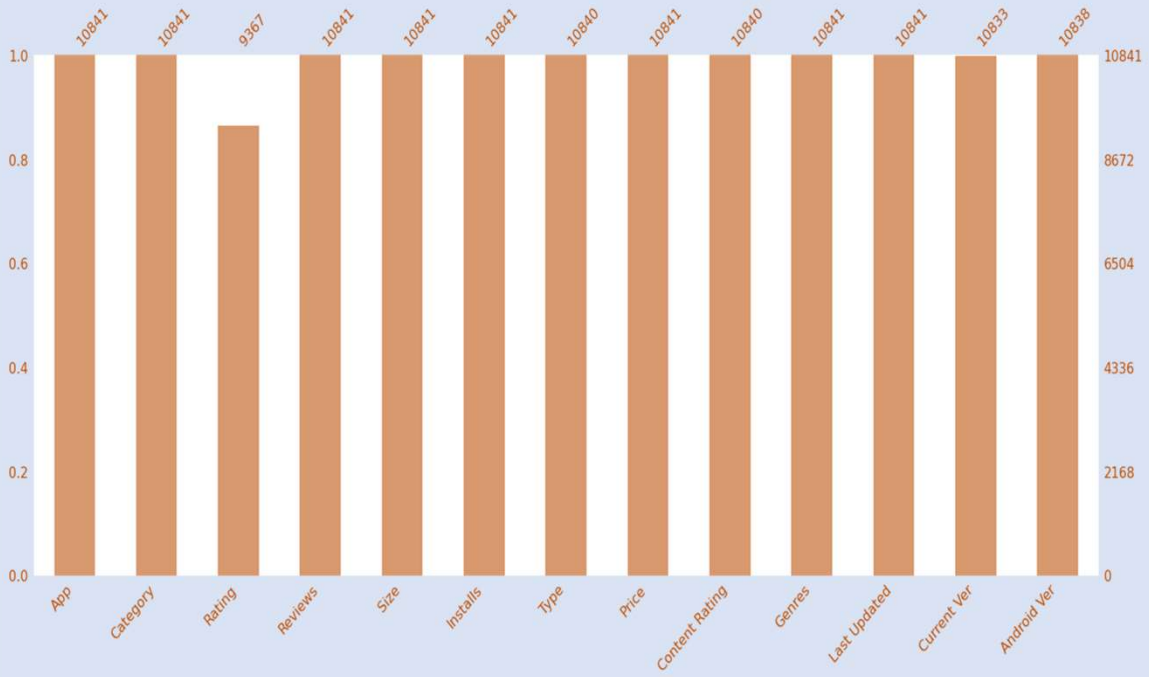
```
#viewing the database info

user_reviews_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 64295 entries, 0 to 64294
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   App                     64295 non-null  object
 1   Translated_Review       37427 non-null  object
 2   Sentiment               37432 non-null  object
 3   Sentiment_Polarity      37432 non-null  float64
 4   Sentiment_Subjectivity  37432 non-null  float64
dtypes: float64(2), object(3)
memory usage: 2.5+ MB
```

```
#viewing info of playstore dataframe
playstore_df.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
```

Presents of Null Value of Columns



As we all know ,once we get the data , the very first step will be find the null values , get the null values ,and try to fix . And if it is not affecting the databses so much then we can clear the rows containing the null values .

The next is the missing value , we also have to deal with that

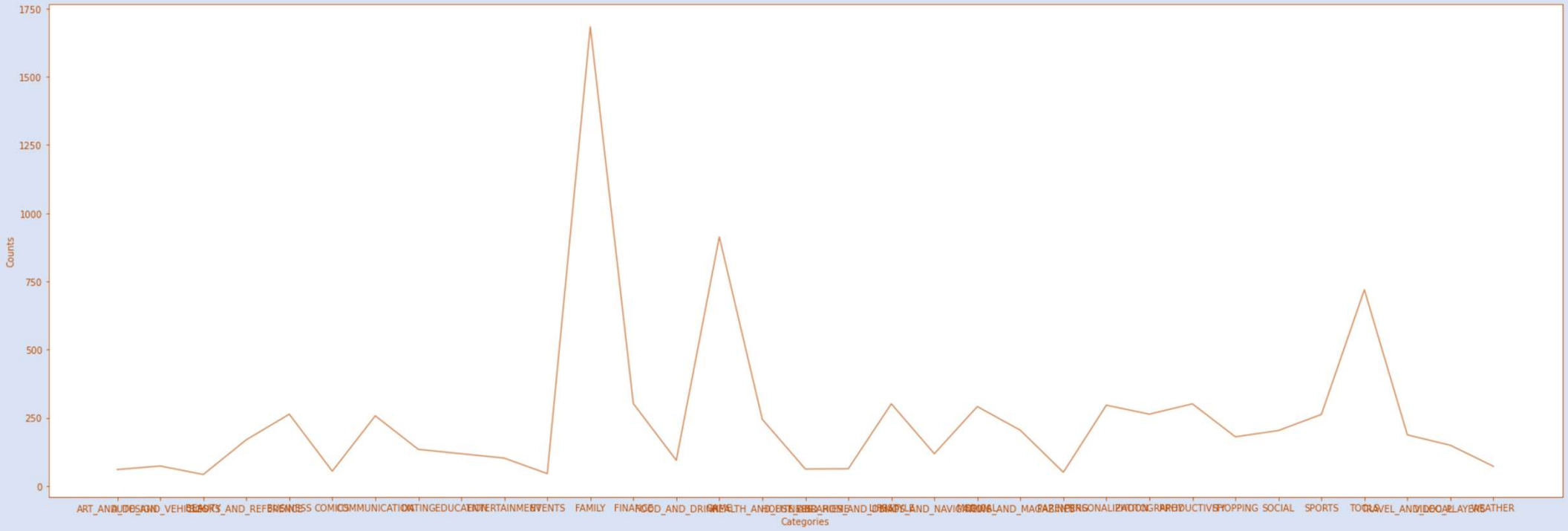
this two thing will affect whole Data Set if they are not settled.

so after the removing the data which are creating a problem to do analyze , what we do here , we just to print of a graph of two columns , i.e “Categories” and “Apps” .

So on that graph , we got that how many Apps are there with respective each Apps categories.

the graph between categories and apps are shown on next slide , you can check for reference

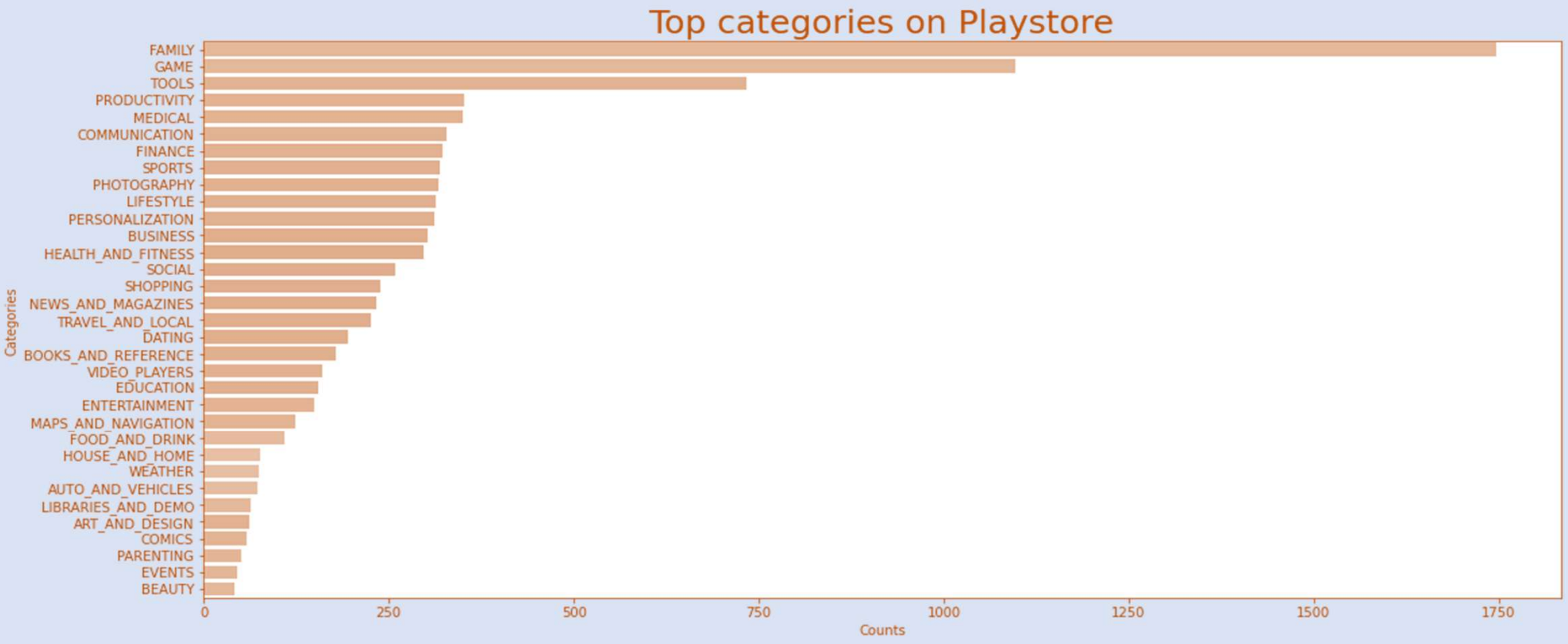
Counts of Categories



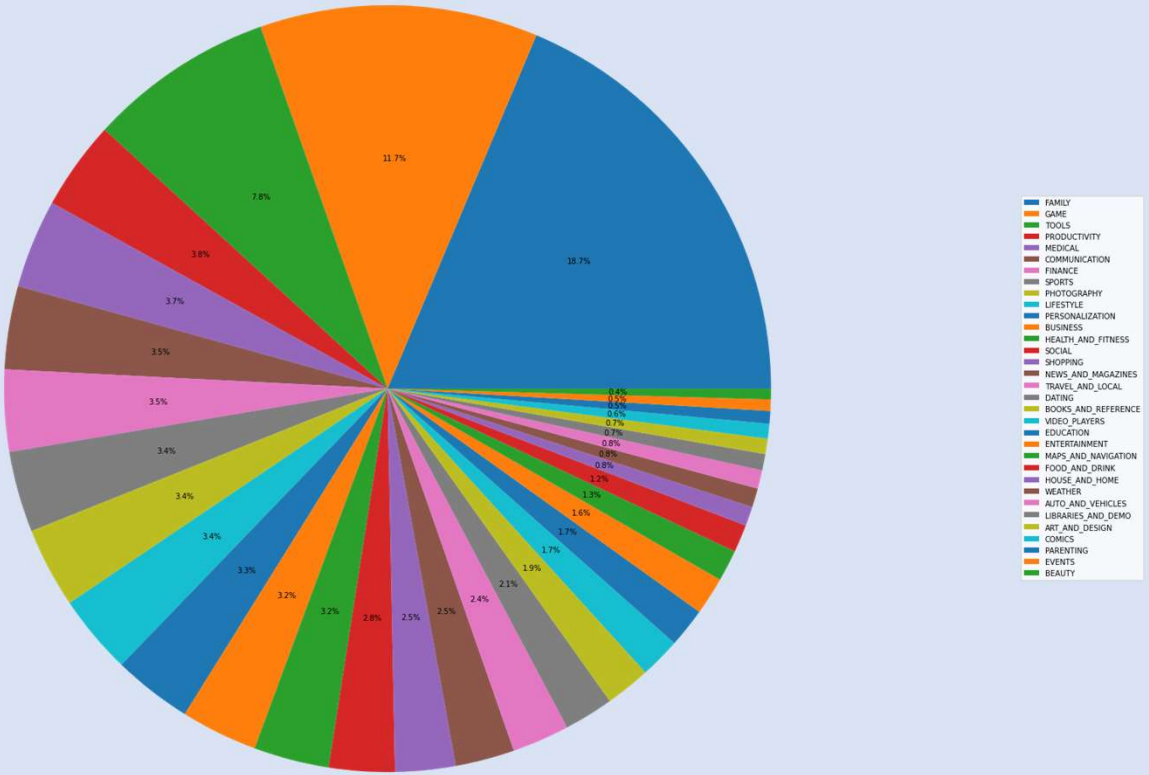
Next thing is that we find the top categories of apps with count which are available on play store . so we can get the idea what kind of application of we have in bulk in App store

As the graphs are lill bit big in size , we have to use show it on other page , because , if we minimize the graph , the data will not visible properly.

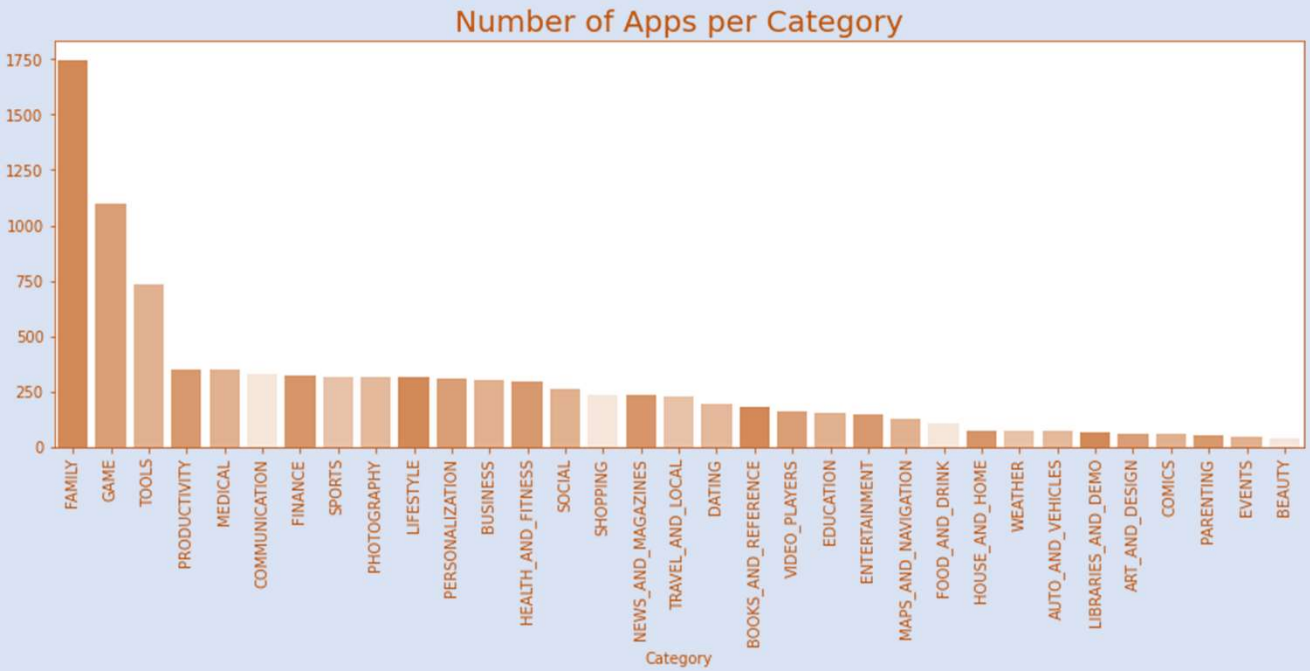
Top Categories On Playstore



Distribution of Apps across Various Categories



Number of Apps per Category



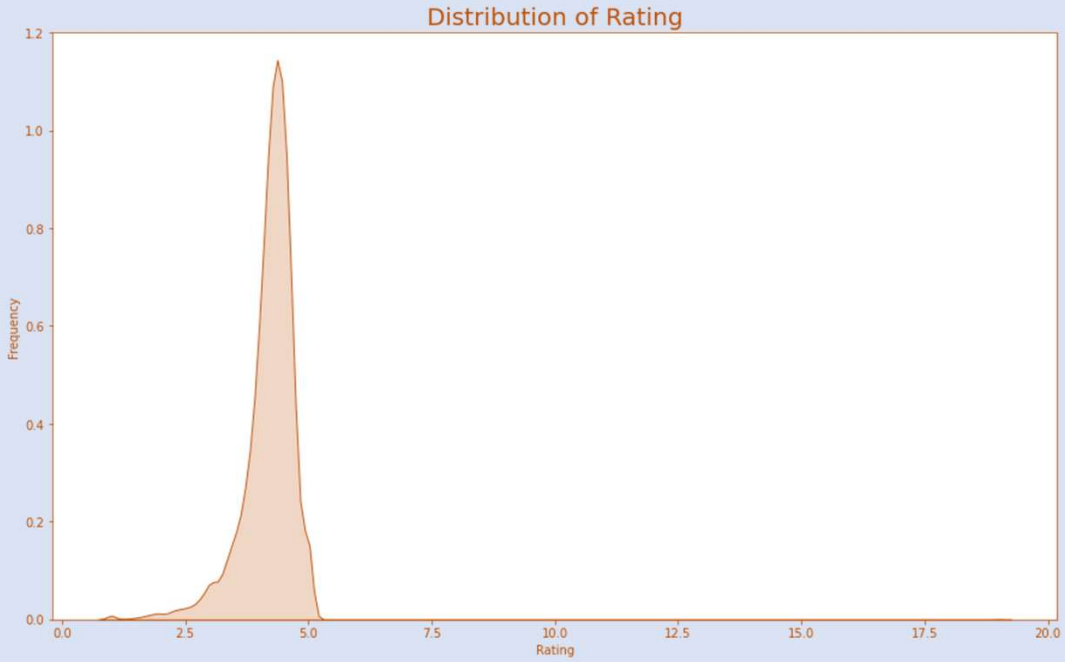
here as previosly we find the number of apps available on store , so aside that we plot this plot to get the total number of installed apps .

as we saw , these two graphs are not so much different ,

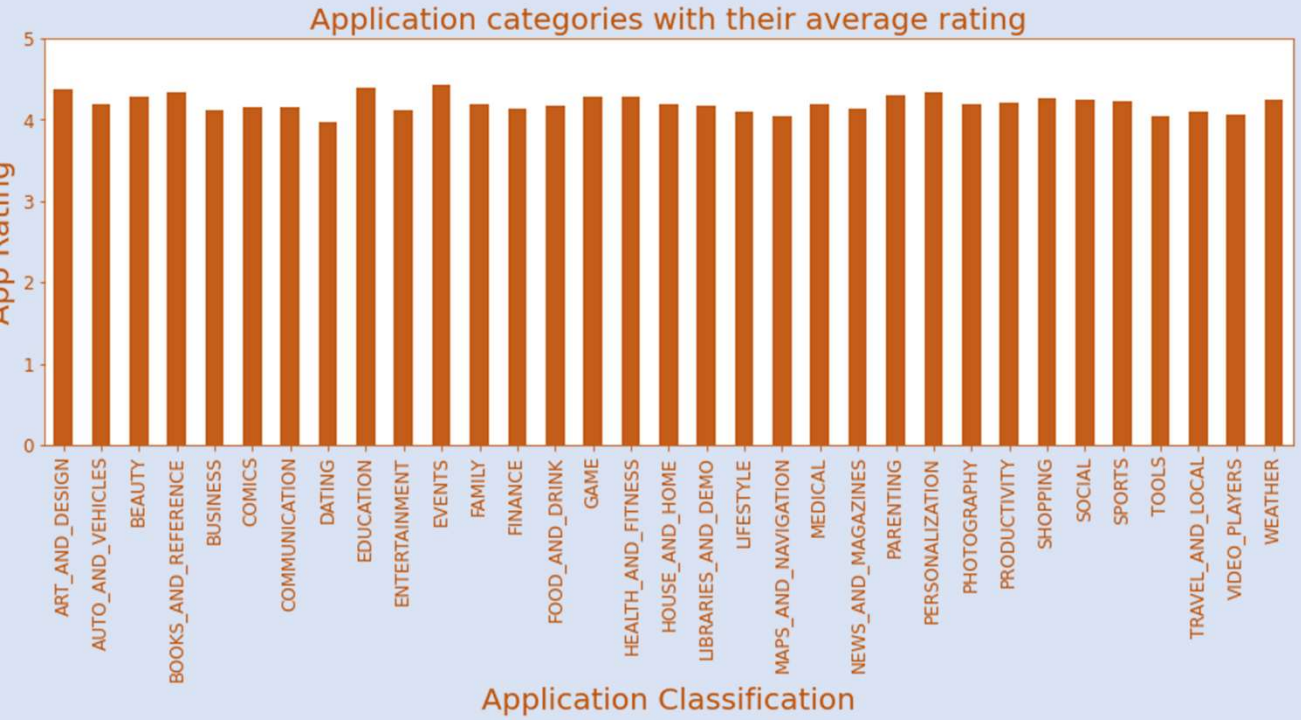
Distribution of ‘Rating’

here we found that the distribution of rating amongst the apps .

by using this chart ,we can found out the multiple algebric value for a further usage

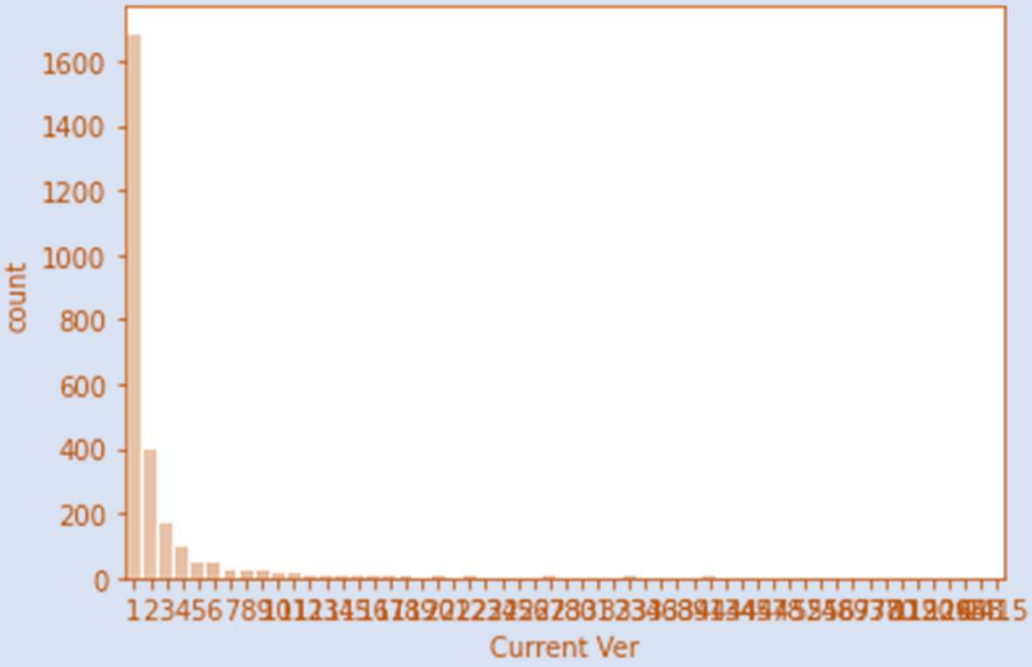


Application Categories with their Average Rating



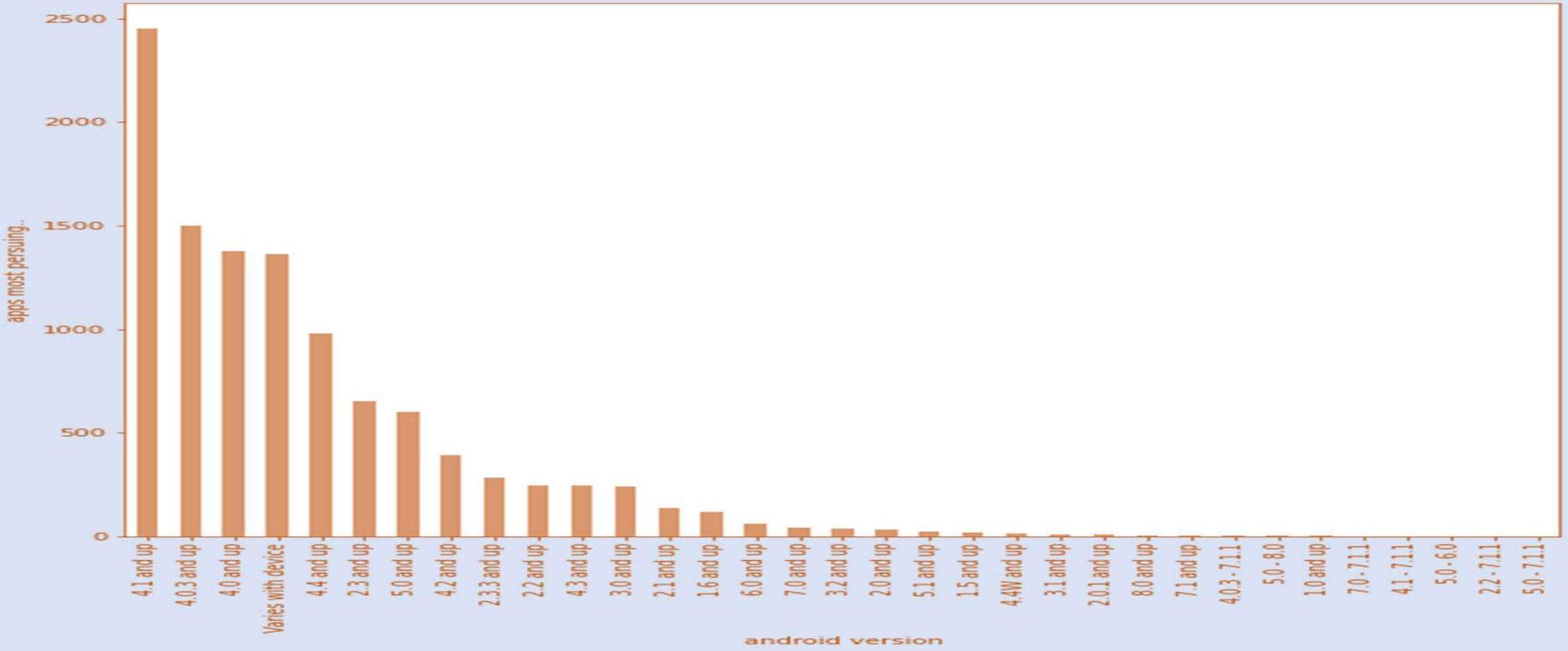
- The average rating Application per application categories is calculated here , so we can get that how the app category is able to delivering result or failed

Count of Current Versions on There Basis of Android Version and Current Version

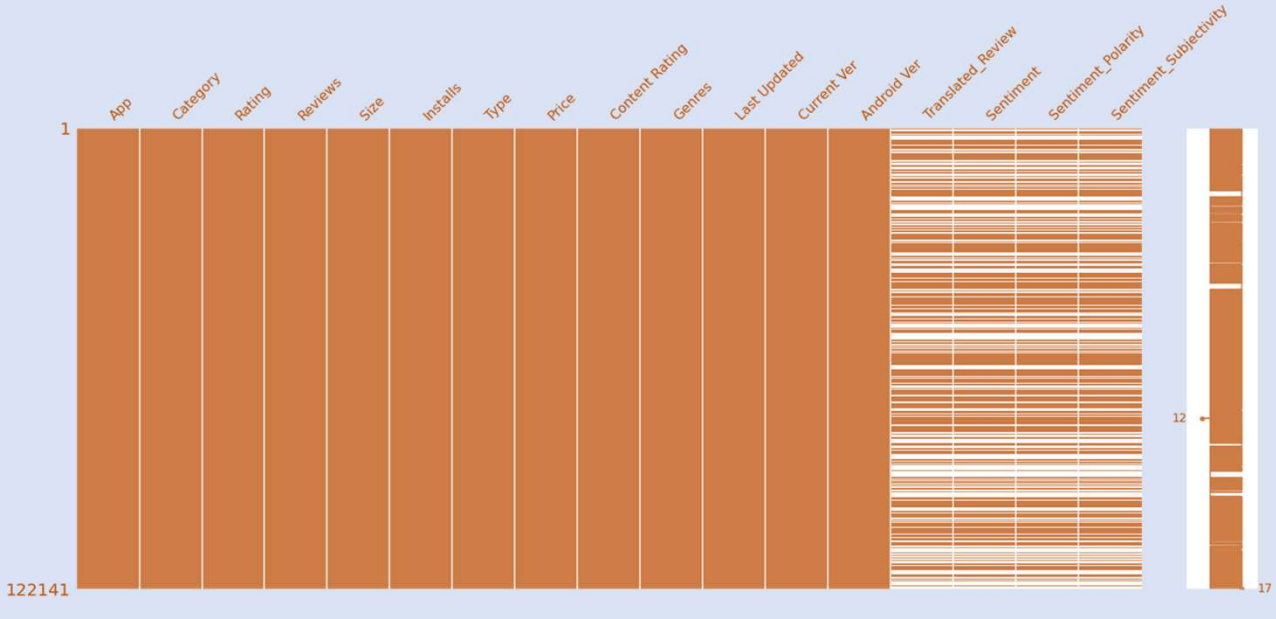


- here we just found the Count of each version of app installed in a system
- the detail description of number of apps are installed versus the required android version are shown in next slide , in detail

Distribution of Android

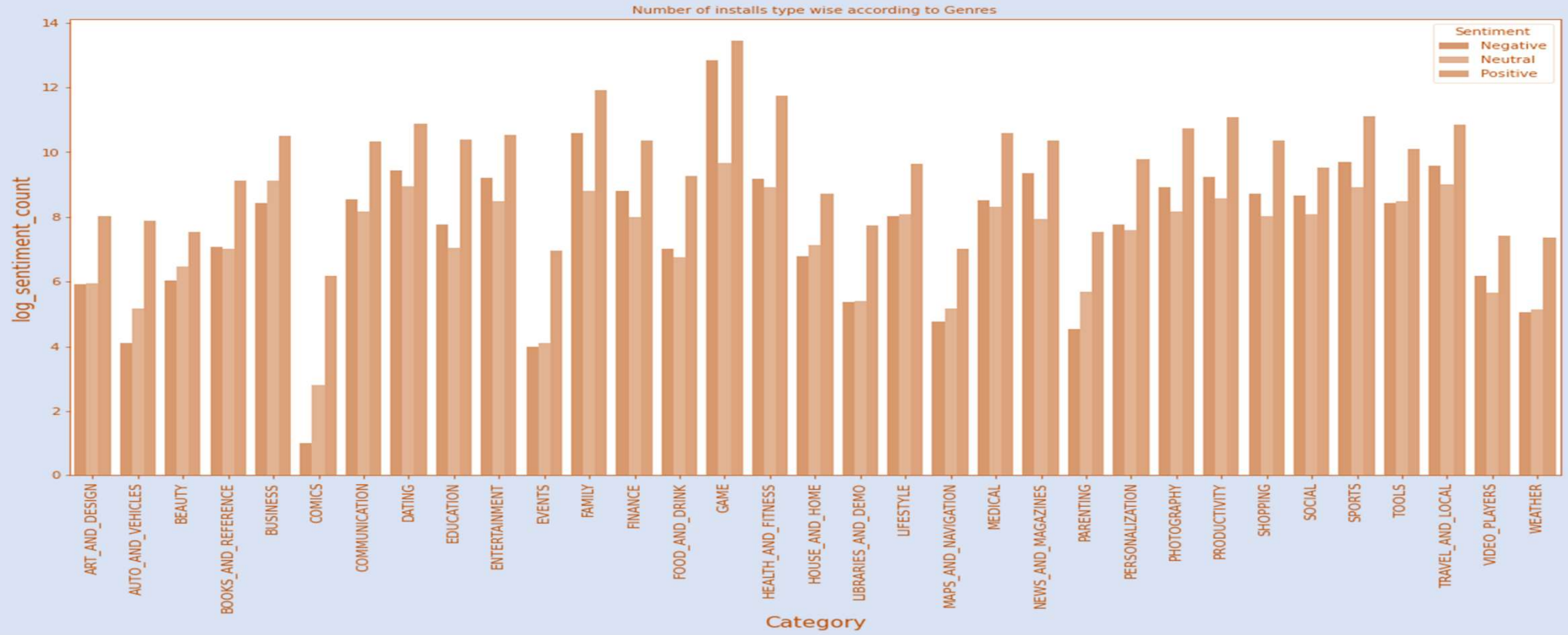


Analysis After Merginig Two Table

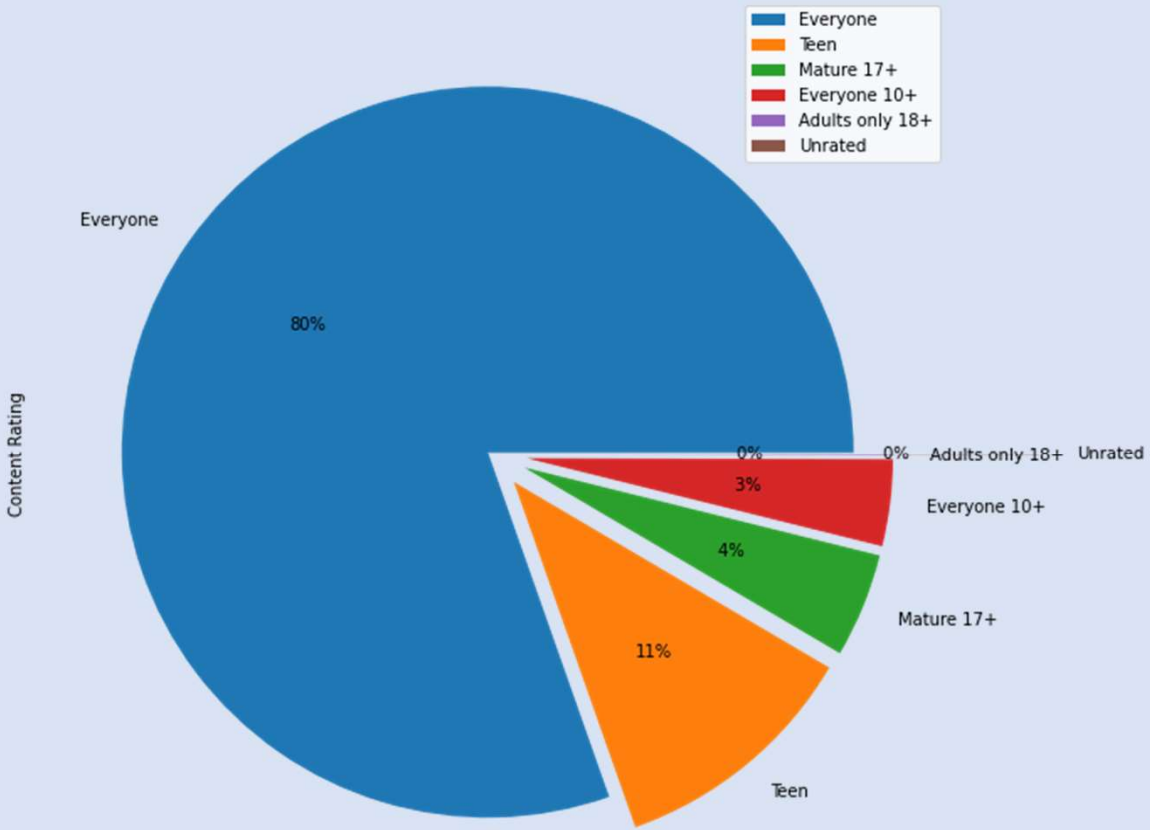


- this is we got the number of null values in a table where we merged the user reviews and play store data set
- then after that we do no of install type wise according to generes
- here we shows the sentiments level as on positive , neutral , negetive
- that is shown on next slide, in detail.

Number of Installs Type wise according to Genres

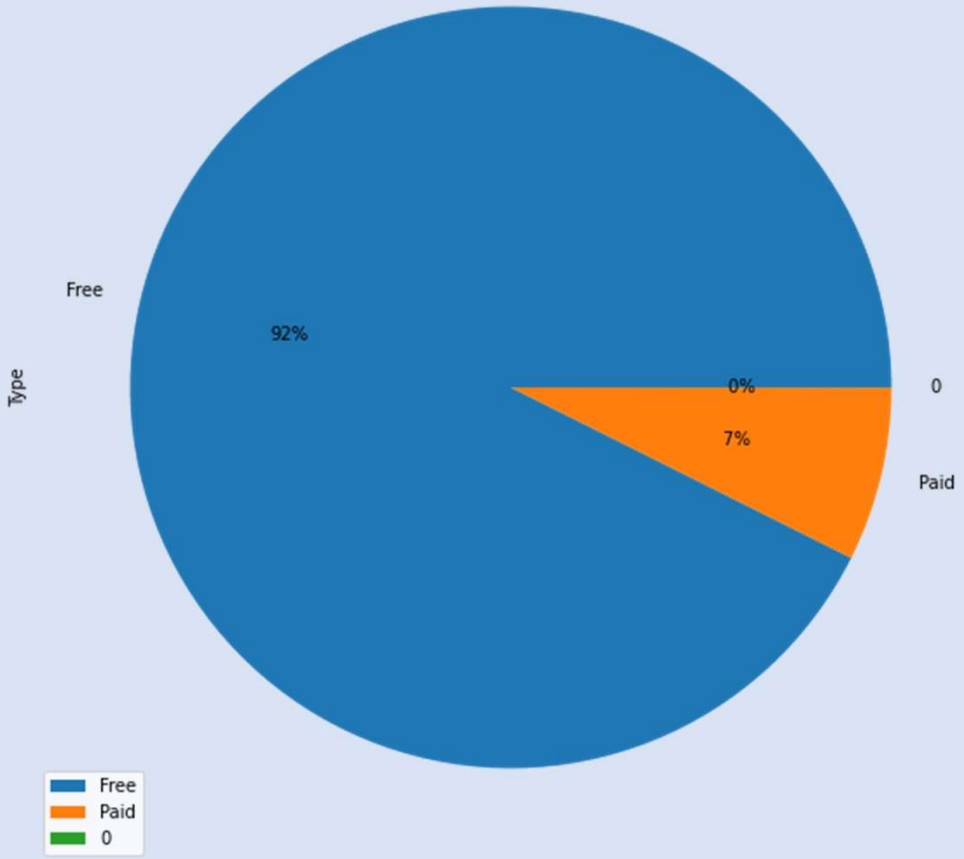


The Count of data of content for perticular age group



this is the most important data analysis that we have find what will the restricted age or for what age categories , the App is made for . we can understand what age group will be our customer for what kin of app

Free apps VS Paid apps Installs (Top genres based)



- as the ast we just show the comparison between paid and free apps , so occupation of kinds of app available in market , we get

Challenges

- The first we face the challenge like to understand the business logic and data set and correrelation between them.
- Then to deal the improper data.
- Dependence of one data column on another.
- between to find the meaning full insights, we do several steps , several methods we performs and from that we learned that how to think when we anaalyze the data ,what mindset we have to keep while doing data analyzation

Conclusion

At the end we come to conclusion that the basic mantra of apps , that if we want to spread our app , make it free , if we want to make it rare , just make them paid.

It’s a factual and jokes apart

We solve interesting analysis which will helps bussiness to get the data flow and trends of current time!!

Thank You