# Abstract:

Play store App Data Review Analysis

### Adesh Dhole

### Data Science Aspirant

**AlmaBetter, Bangalore.**

* + The play store is a leading App market for all kind of operating system.
  + Our feature analysis will be based on the number of columns given and the data associated with it and try to find the meaning full pattern and inside instincts.

# Problem Statement

* + Here , data are provided are such a way that , how many downloads , what kind of apps are there, what people prefer to used and installs on that basis we have to do further analysis..
  + App name: This is the column which contains the name of the app.
  + Category: The Category is a column by which the apps got separated based on the application and purpose.
  + Rating: Rating plays a huge role in finding the correct apps. It was manually given by the users.
  + Review: Review are here to show how many numbers of review it got, regardless its value
  + Size: Size is a factor which is not the same for all the devices, and it fully dependent.
  + Installs: Installs gives us the data of total installed users count.
  + Type: Type defines whether the app is free or paid.
  + Price: Here price is mentioned for the paid apps, for free apps 0 will be given.
  + Content Rating: Here, on content rating, we got the idea that which age group would follow which kind app and app category
  + Genres: Genres are like tags, an app can be coming under more than one Genres, based on the usage.
  + Android Version: This feature gives us the supported device versions of android.

The main agenda of our data analysis to find that what kind of app is on the verge to be trend, and what people will actually choose to used

# Exploring the database

### We have provided with two databases

**Play Store Database**

* + Shape of this database is (10841, 13).

**User reviews database**

* + Shape of this database is (64295, 5).
  + Here there are only two numeric values found,

Here, the first step would be clean the data, deal with the missing data, and locate and delete them.

Then, the actual analyzation would start.

**The Key Points**

**No of apps per categories:**

The categories to, what kind of app are actually preferred.

**Top categories of downloads:** It holds the top categories amongst the download

**Distribution of apps per categories:**

here distribution of apps amongst the categories are found and further used for graph as well.

**Distribution of Ratings per categories:** as like above, the distribution of ratings would also been finds.

**Application categories with their average rating:**

Here we got to know what will be the average rating of each app categories

### Count of current version:

here we got the actual count of apps require which are support the current version

**Count of Android Version:**

here we got the actual count of apps require which are require what kind of android version

**Number of installs type wise genres:**

this holds the no of installs type according to genres; this is the one and only inspection we do on the merged data of two tables

different aged people wise ratings count.

# Number of apps per categories analysis

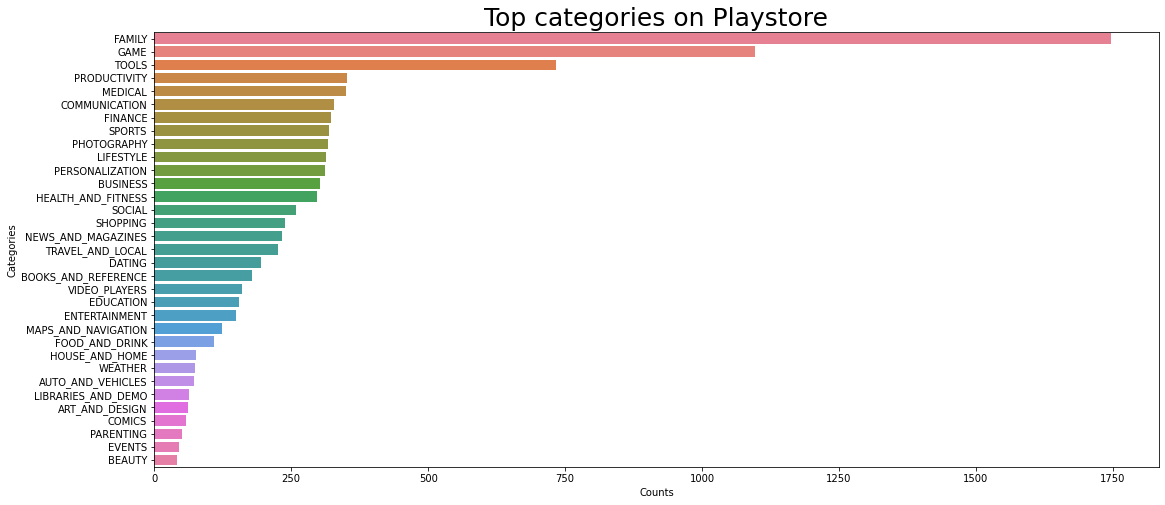
Here to analyze, first we have to find unique values of apps, rating and categories

* + Then we find data of apps per categories
  + Each categories have number of installs, particular type and so on the further things would be calculated

## p2

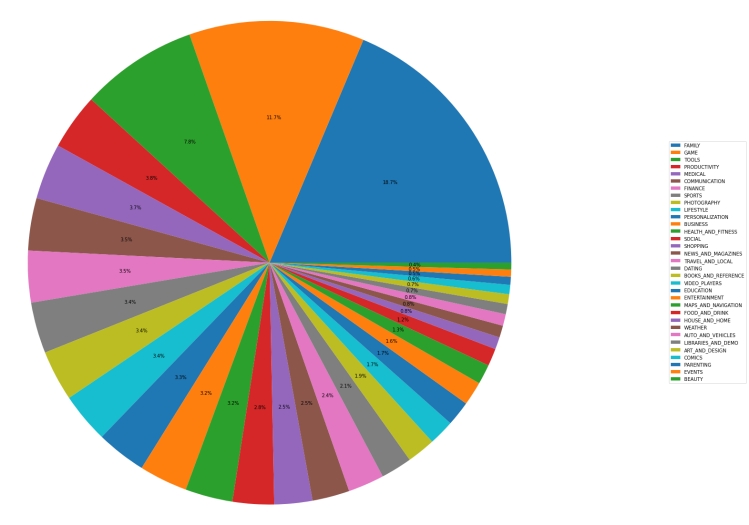
## Top Categories on play store

* + Here we got to know that what kind of categories of apps are in most downloads, so that we can get to know that what most people are in look for!!

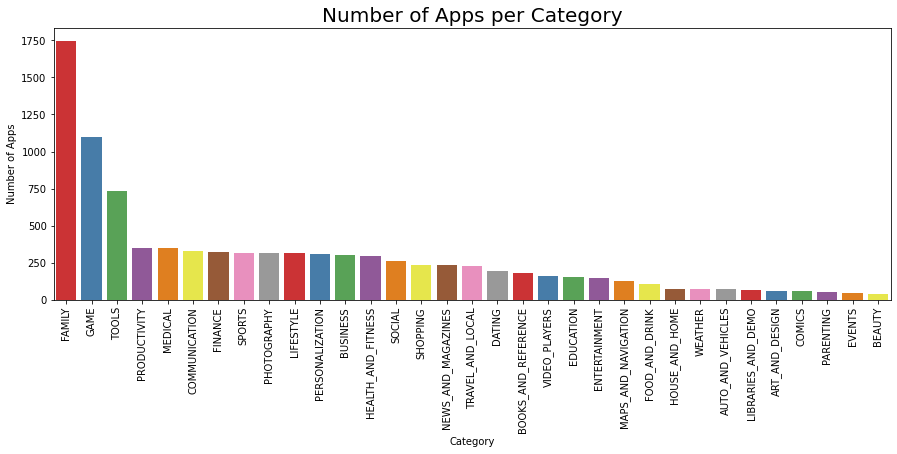


## Distribution of apps across the categories

Now we are here also providing what number of apps of, of certain categories, to get the flow of downloads.

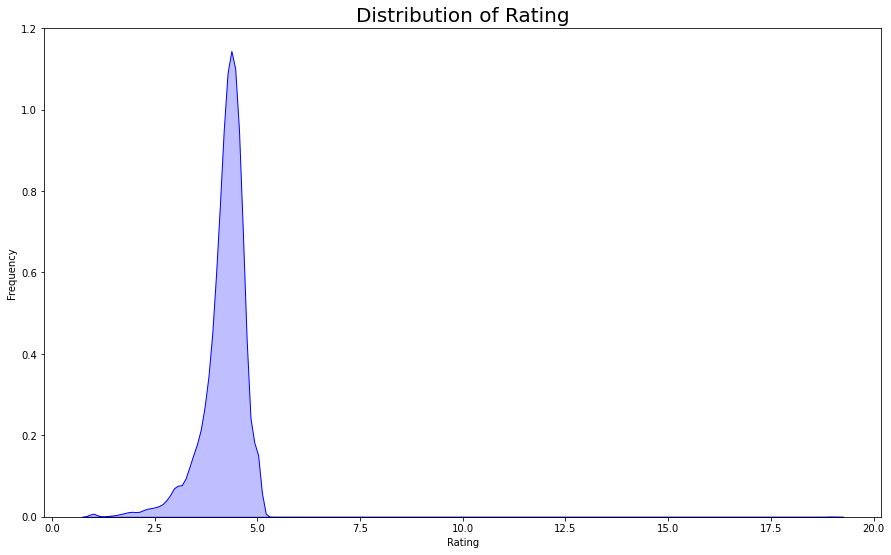


1. **Number of apps to be download** 
   * Here, no of apps been download on basis of categories we got here, but here we got actual number not a level.



# Rating-Distribution:

* + - This is the most important point to be noted while analysis on rating, we have to find what is the distribution of rating so we can get all the algebraic terms of rating columns

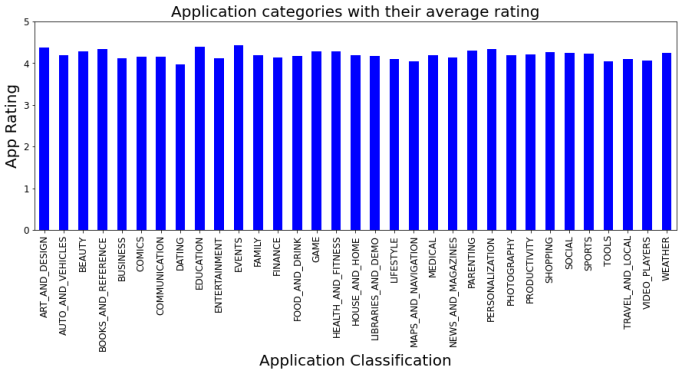


# Application categories with average Categories:

Here we got the thing like what are the

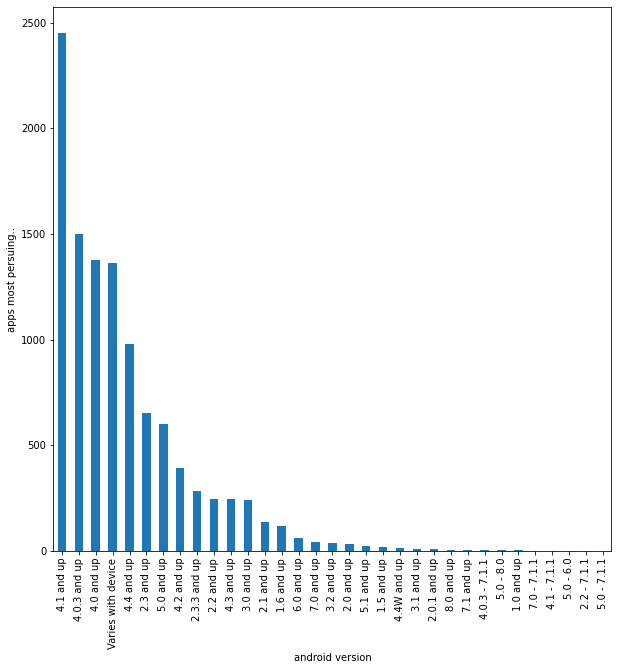
Average rating of each Application categories

So, we can understand the flow of likes and dislikes



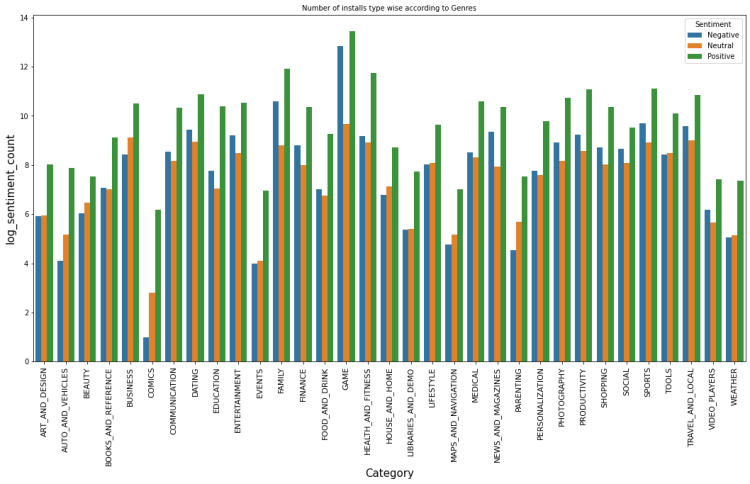
# Count of current version:

Here we have to find what are the count of current version on which Apps are supported.



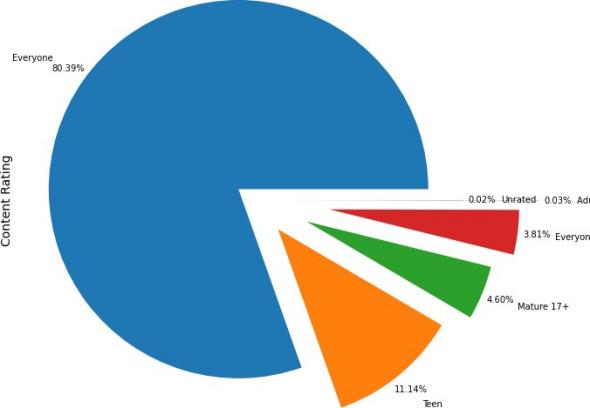
# Number of installs type wise according to genres:

Here we have to check first number of type vs installs and then on hues, genres are analyzed.



x

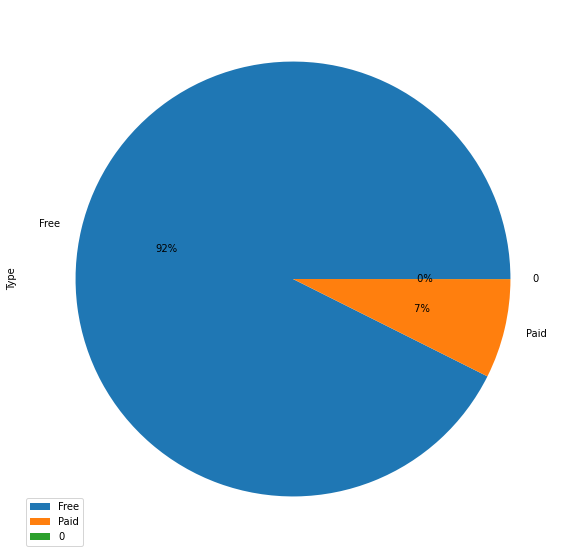
# Impact of ‘Rating’:



* + - The content rating shows the results for general contents as high.
    - The content rating type ‘Everyone’ has the most percentage value of 80.39%.
    - ‘Teen’ contents are second in the order with the percentage of 11.14%.
    - Adult’s only and unrated contents are least in this plot, 0.03% and 0.02% respectively.
    - So it can be concluded that most of the contents are generic.

# Comparison of free vs paid apps:

Here we compare the number of installs as compared to free and paid app



# 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 bushiness to get the data flow and trends of current time!!

### References-

1. Stack overflow
2. Geeks for Geeks
3. matplotlib.org
4. W3School.com