PES UNIVERSITY, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE18CS203

B.Tech, Sem III
Session : Aug-Dec, 2019

UE18CS203 – INTRODUCTION TO DATA SCIENCE

REPORT
ON
EXPLORATORY ANALYSIS ON
Google Play Store Apps

SECTION : A

#	SRN	Name	Contact No.	Email ID	Sign
1.)	PES1201800262	Navneeth	7019618817		
2.)	PES1201802127	Siddharth	8123333693		

ABOUT THE DATA SET

Google Play, formerly Android Market, is a digital distribution service operated and developed by Google. It serves as the official app store for the Android operating system, allowing users to browse and download applications. The Google Play Store Apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market. Applications are available through Google Play either free of charge or at a cost. They can be downloaded directly on an Android device through the Play Store mobile app or by deploying the application to a device from the Google Play website.

The dataset gives us data about all the apps that are available for the users to install from the play store based on different categories, ratings, reviews, etc.

The dataset has 10842 rows along with 13 columns.

In this section, a short description of the meaning of each column can be found.

1.) App: Application Name
2.) Category: Category the app belongs to
3.) Rating: Overall user rating of the app (as when scraped)
4.) Reviews: Number of user reviews for the app (as when scraped)
5.) Size: Size of the app (as when scraped)
6.) Installs: Number of user downloads/installs for the app (as when scraped)
7.) Type: Paid or Free
8.) Price: Price of the app (as when scraped)
9.) Content Rating: Age group the app is targeted at - Children / Mature 21+ / Adult
10.) Genres: An app can belong to multiple genres (apart from its main category).
11.) Last Updated: Date when the app was last updated on Play Store (as when scraped)
12.) Current Ver: Current version of the app available on Play Store (as when scraped)
13.) Android Ver: Min required Android version (as when scraped)

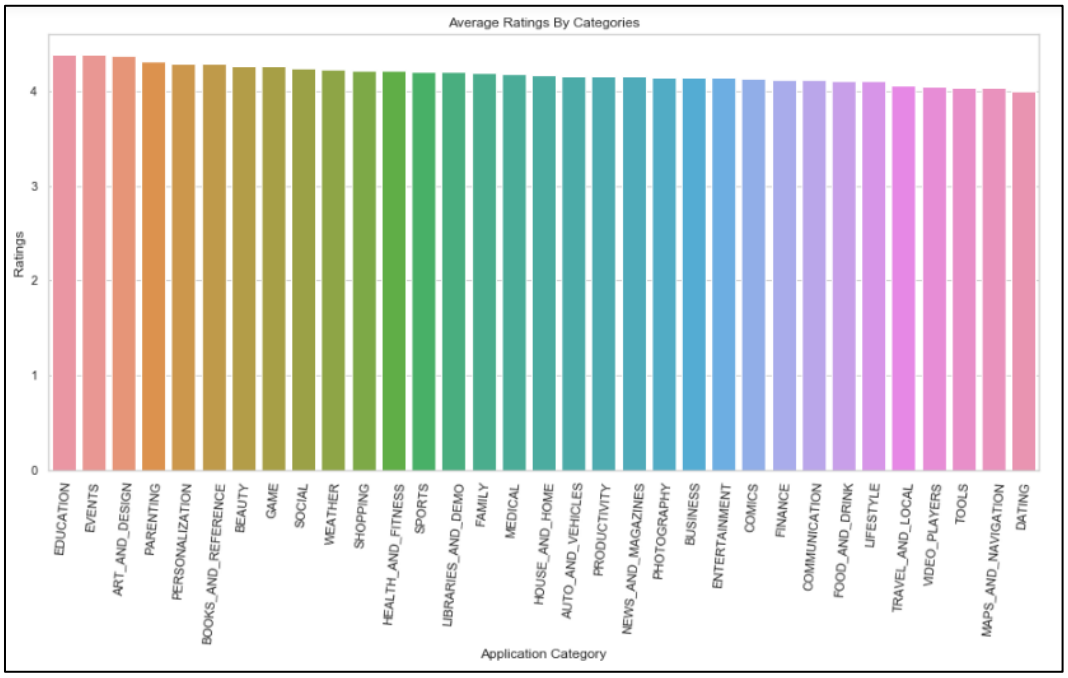
EXPLORATORY ANALYSIS

DATA CLEANING:

In our dataset, we are dropping row **10472** due to missing value in the 'Category' column. We had to clean the 'Size' column by stripping the 'M' (Megabytes) and 'k' (kilobytes) and convert the kilobytes to megabytes to maintain the uniformity and the places having 'Varies with Device' are replaced with NaN values. The NaN values are replaced with the mean. We had to clean the 'Installs' column by stripping all the '+' and removing all the ',' present in that column. The 'Price' column was cleaned by stripping the '$' symbol. We have cleaned the 'Android Ver' column and converted them to float. The 'Reviews', 'Installs' and the 'Price' column have been converted to float/int.

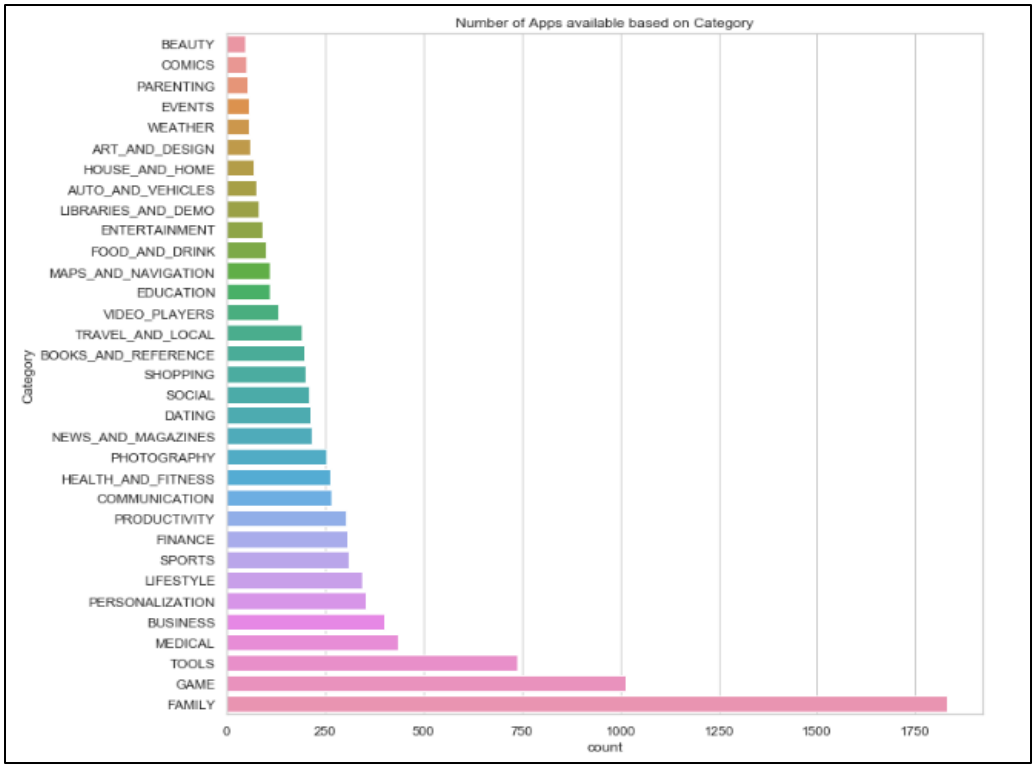
INSIGHTS DRAWN:

1.) Apps of different categories and their ratings



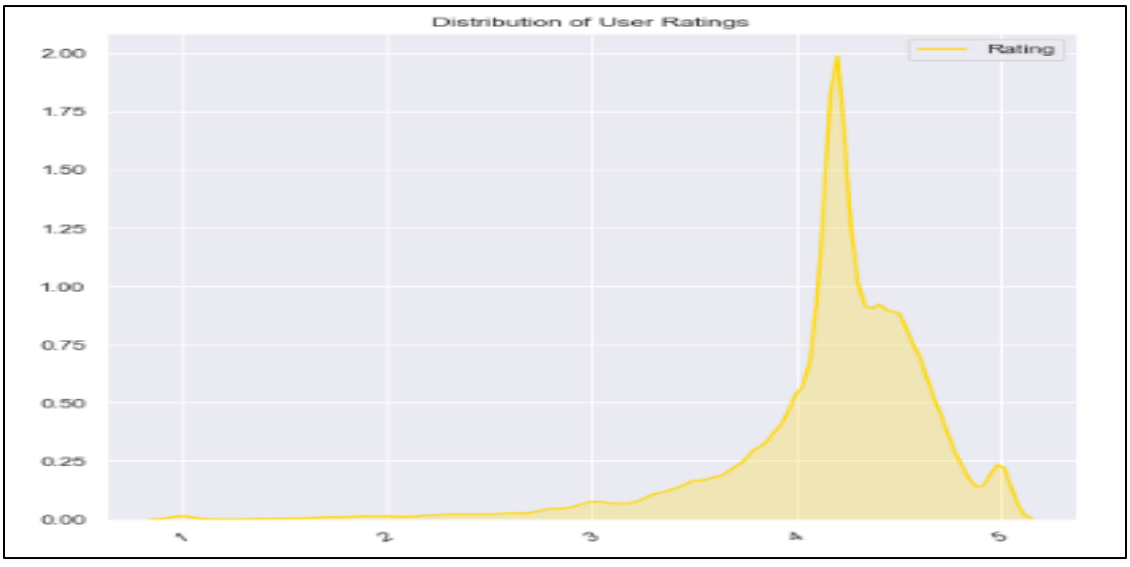
This shows that the Apps belonging to the '**Education**' category have a higher rating while the Apps belonging to the '**Dating**' category has lower ratings when compared to the other categories.

2.) Number of Apps available based on 'Category'



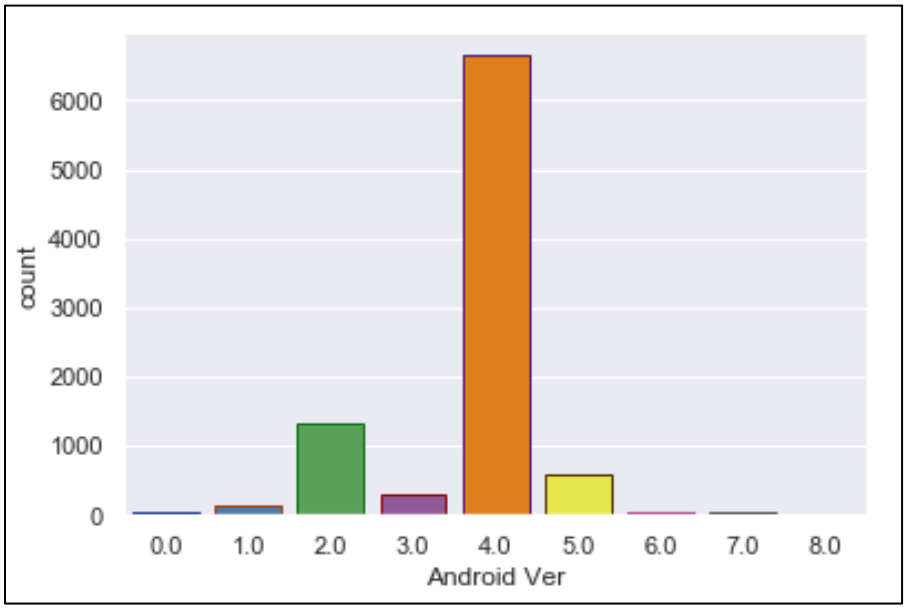
From this we are able to find out that there are less than 250 Apps in the '**Beauty**' category, while the '**Family**' category has more than 1750 Apps.

3.) Distribution of User's Ratings



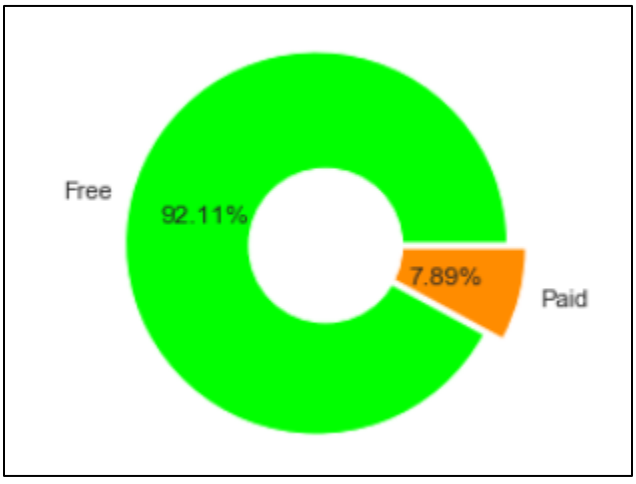
Most of the apps, have a rating that lie in the range of 4.1 to 4.7. Some apps also have a rating of 5.0. Very few apps have a rating below 3.5.

4.) Count Plot of the various Android Versions



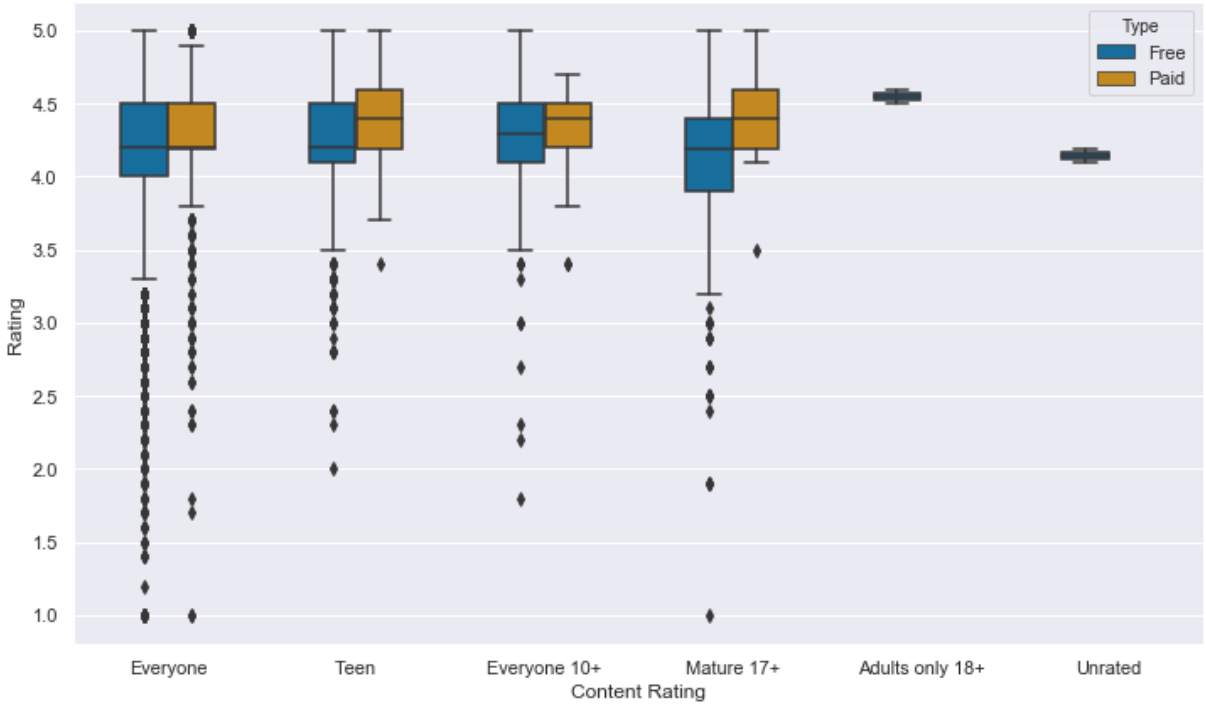
From the count plot we are able to infer that most of the apps support **Android Version 4.0**. More than 6500 apps support Android Version 4.0. This is because **Google** supported Android Version 4.0 for a span of **seven** years and lot of apps have been developed in those years. While around 1200 apps support Android Version 2.0 and less than 1000 apps support version 5.0 and above.

5.) Percent of Free and Paid Apps



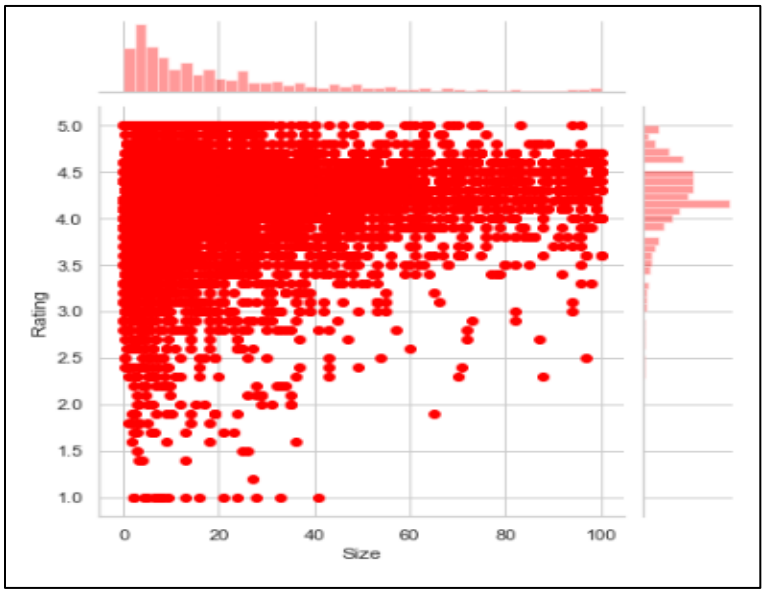
We were able to find out that about **92.11%** of the apps belonged to the **Free** category, while **7.89%** of the apps belonged to the **Paid** category.

6.) Comparing the Free and Paid Apps with respect to their Ratings grouped by their Content Rating



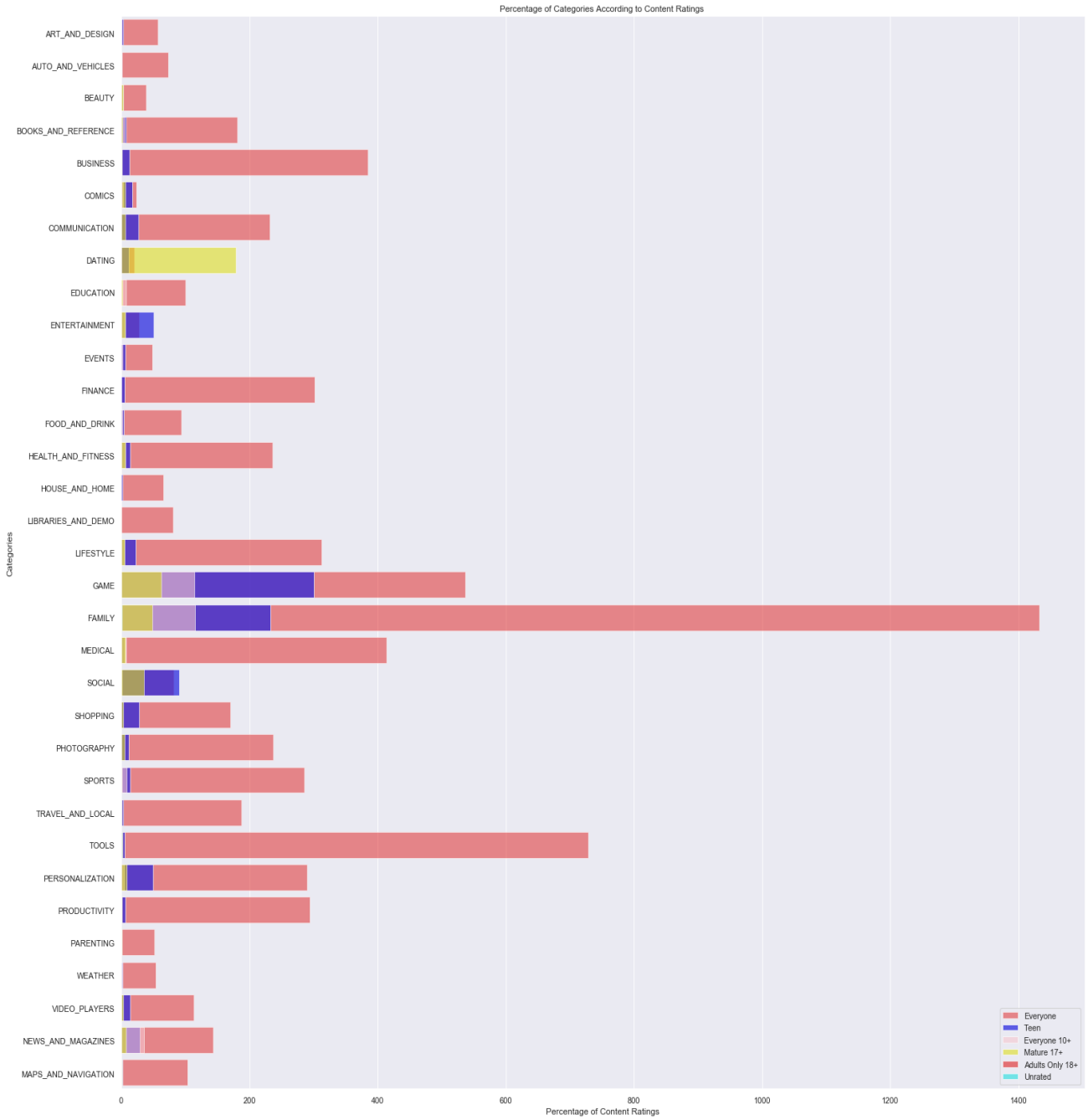
Clearly, from this graph we can find out that the **Paid Apps** tend to have higher rating than the **Free Apps.** We can come with something like **"People are tend to vote higher when they pay for it"**.

7.) How the App Sizes impact the App rating



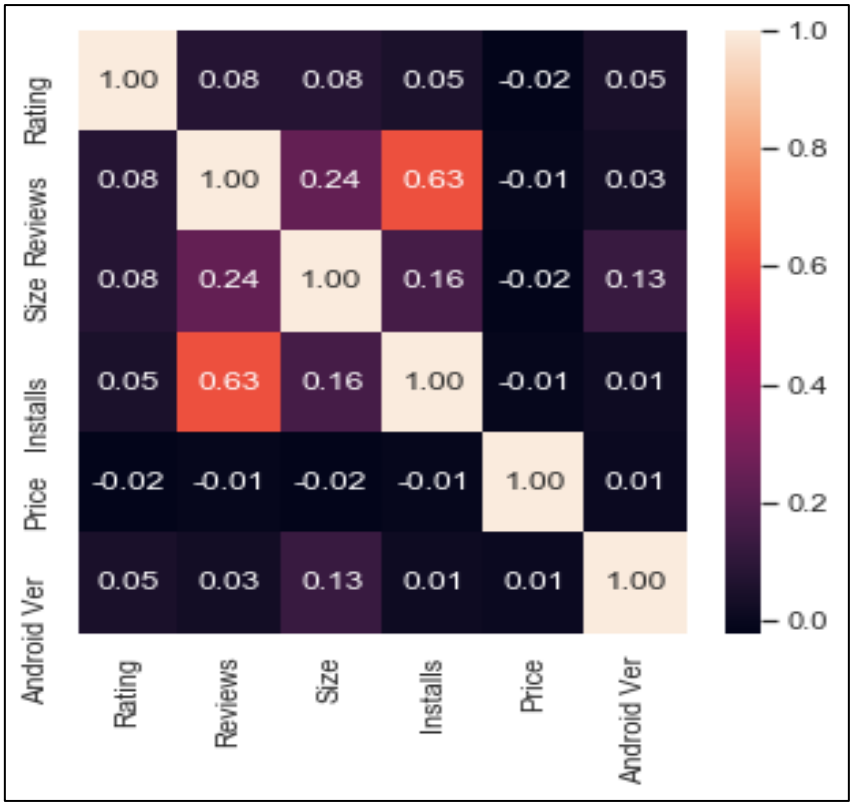
Most top-rated apps are optimally sized between **~2MB to ~40MB** neither too light nor too heavy.

8.) Percentage of Categories According to Content Rating



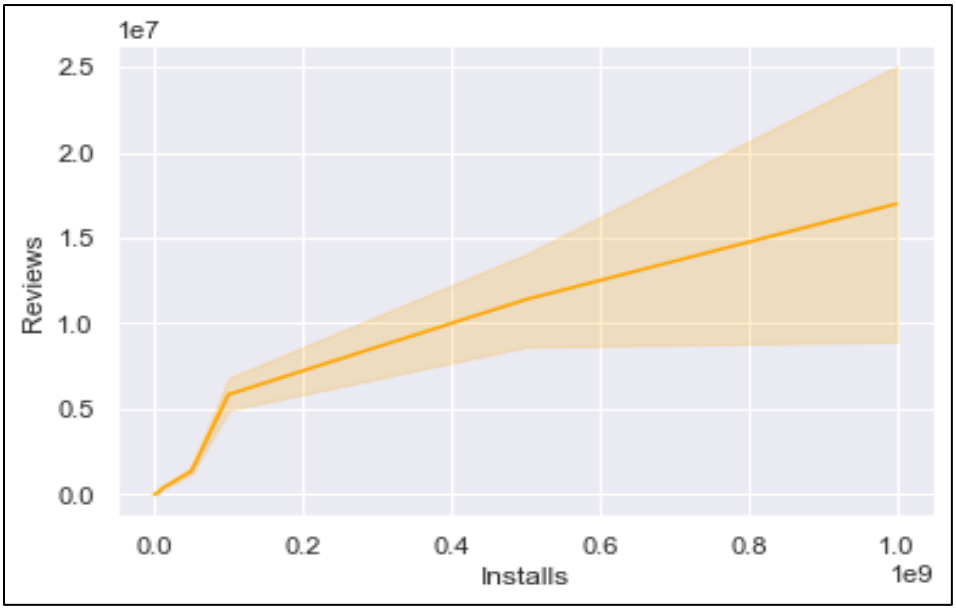
From this graph, one can easily observe that nearly all the applications are targeting 'Everyone', except 'Dating', 'Entertainment' and 'Social' applications.

9.) Heatmap

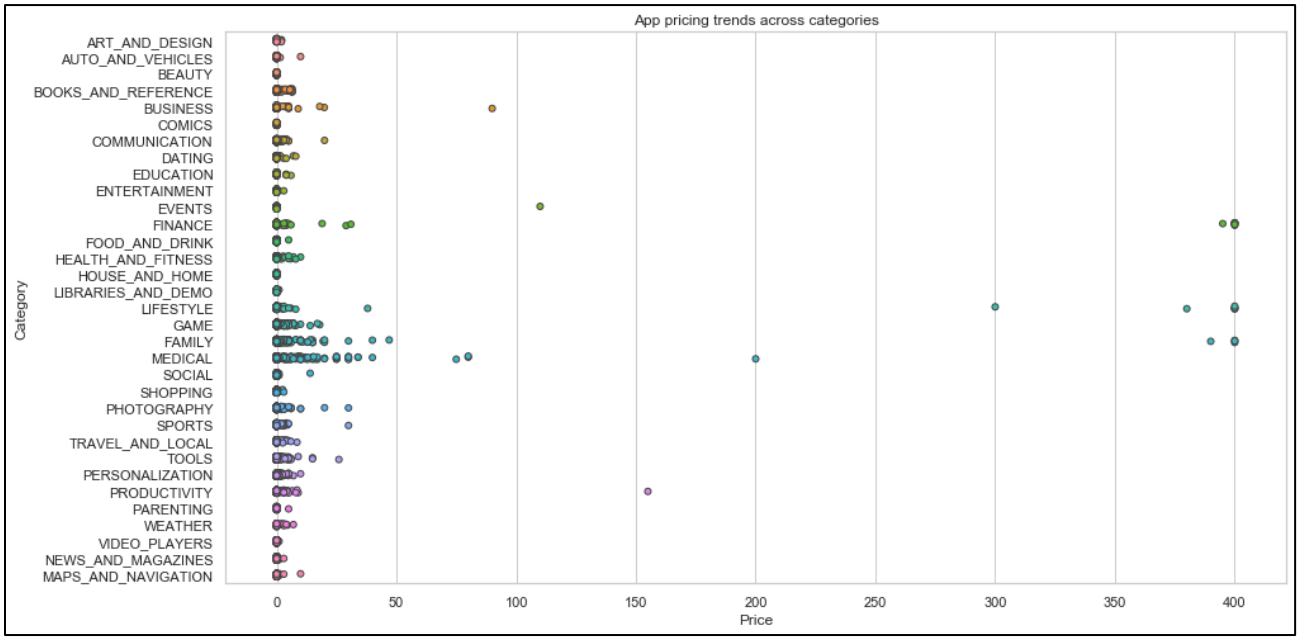


From the Heatmap, we are able to determine that the **'Installs'** and **'Reviews'** are highly co-related.

We have also plotted a line plot which goes according to the above statement.



10.) App pricing trends across Categories

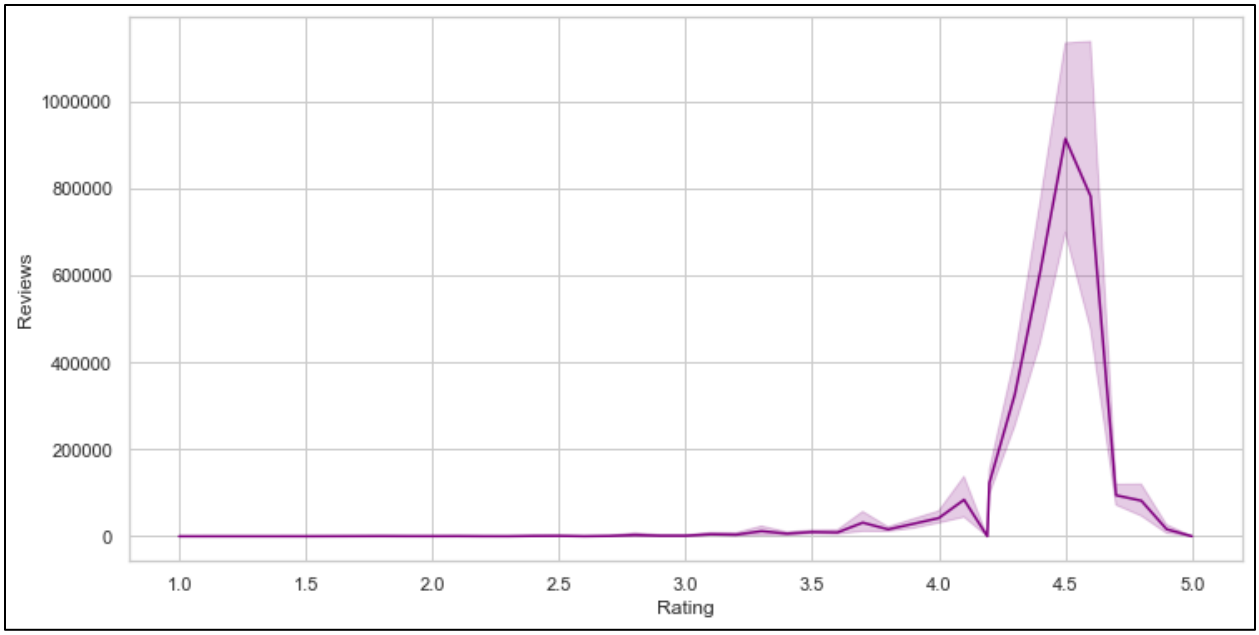


This graph depicts the relation between the category and pricing. We can see that there are apps that are **greater than $200**. The table below shows the categories that have apps with pricing greater than $200.

	Category	App
0	FAMILY	4
1	FINANCE	7
2	LIFESTYLE	6

The categories **'Family', 'Finance', 'Lifestyle'** have apps that are priced greater than $200.

11.) Relation between Reviews and Ratings



From the graph, we can conclude that the apps rated from **4 to 5** have highest number of **Reviews.**

CONCLUSION

- Most of the top-rated apps are optimally sized between ~2MB to ~40MB - neither too light nor too heavy.
- ‘Education’ Category has the highest average rating while the ‘Dating’ Category has the lowest average rating.
- There are less than 250 apps in the ‘Beauty’ category while the ‘Family’ category has more than 1750 apps.
- Most of the apps are rated from 3.5 to 4.5.
- The Android Version 4.0 has the most supported apps.
- People vote higher when the apps belongs to the Paid category.
- Most applications are targeting everyone except ‘Dating’, ‘Entertainment’ and ‘Social’ applications.
- ‘Reviews’ and ‘Installs’ are highly co-related,