Play Store App Review

V Rupesh Kumar Patro
Shashank Maindola

Abstract:

In today's world mobile applications find their utter usefulness because specific software is required for almost every purpose be it personal, social, business & for any such functions. The android OS platform has a market share of 71% (Sept 2022) which means there's huge potential for app developer in this market. The Google Play Store is official digital distribution service operated and developed by Google. It serves as the official app store for certified devices running on the Android operating system, allowing the users to browse and download applications developed with the Android software development kit (SDK) and published through Google. Google Play store allows access to download a wide range of apps, books, movies, and television Content either free of charge or at a cost.

Introduction

Currently present on the devices around 190 nations worldwide, the Google Play Store hosts more than 3.5 million Android Apps. Maximum applications in the play store are free which is the main reason why it is so popular worldwide. Globally there are more than a billion users worldwide. On an average, around 3739 apps are added to the play store daily. In terms of revenue, the game apps generate the most of the revenue. It is projected that this category will responsible for more than 50% revenue generated by end of the year 2022. The android app market is growing steadily and as per **a report by Sensor Tower,** user spending on Google Play is projected to reach **$60 billion by 2023 from $53 billion in 2022, at a CAGR of 19%**. But still maximum app developers fail in grabbing that market.

Our goal here is to draw actionable insights so that it can be useful to the developer to work and capture the android market.

Methodology

- **Data Understanding**
- **Data Cleaning**
- **Missing/Null Value Treatment**
- **Correcting Data Type**
- **Visualization/ Insights**
- **Conclusion**

Data Understanding

We are provided with two data sets namely **playstore.csv**: contains all the details of the applications on Google Play. There are 13 features that describe a given app.

Playstore.csv

The App Data set is of dimension 10841x13

Data	Description
1. **App**	The name of the application
2. **Category**	The category to which the app belongs

3. Rating	The rating of the application out of 5
4. Reviews	The user reviews given to the application
5. Size	Size of the application.
6. Installs:	Number of installations across the devices.
7. Type:	Free or paid
8. Prices:	Price of application in dollars.
9. Content Rating:	Age based application use.
10. Genres:	Types of genre an application belongs.
11. Last Updated	Most recent update
12. Current Ver:	The most recent version of the application.
13. Android Ver:	The minimum device android version needed to install

user_reviews.csv:
contains reviews for each app, most helpful first. The text in each review has been preprocessed using NLP and attributed with three new features: Sentiment (Positive, Negative or Neutral), Sentiment Polarity and Sentiment Subjectivity.

Data Understanding

To begin with we need to work as per the following steps:

- Import the respective libraries
- Mount the data from the drive and import the files in the csv format.
- Viewing all data information.
- Check for unique values

Data Cleaning

Data cleaning is one of the most essential subtasks of any data science project. We observe that some entries in the columns like Installs, Price and Size have a few special characters (+, $, M, k). We want to convert it to numeric but special characters are preventing the columns from being purely numeric, making it difficult to use them in subsequent future mathematical calculations. Hence these special characters are filtered out of our dataset

Null/Missing Value Treatment

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them or treated them at the beginning of our project in order to get a better result.

Correcting Data type

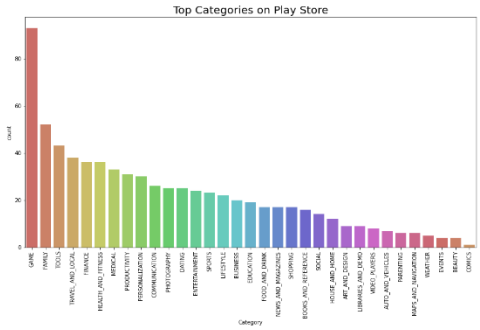
To have more and more numerical data for better analysis, we need to convert the data type of the columns where the datatype is object but we know it's numerical data. We found Installs, Size and Price are the ones which contain numerical data; hence we convert their data type for further use.

Visualization/Insights

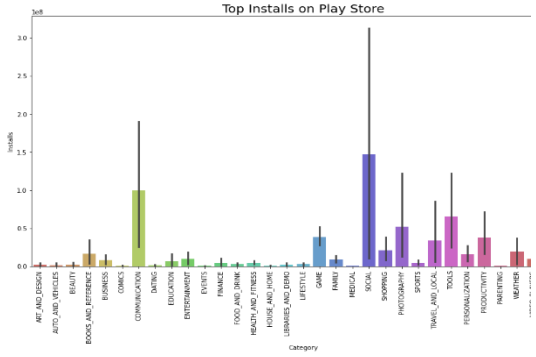
In this step, we will make use of visualizations to get some insights out of our data. Many insights can be made out of our data using different visualization techniques, but we have grouped our visualizations and based on those we have summarized our insights into 5 different observations.

1. Observation 1: The Top Apps

We plotted a count plot of top categories present in play store

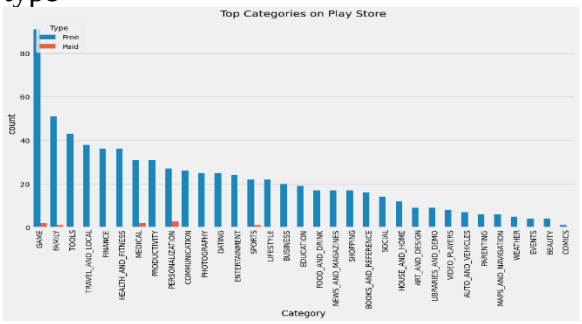


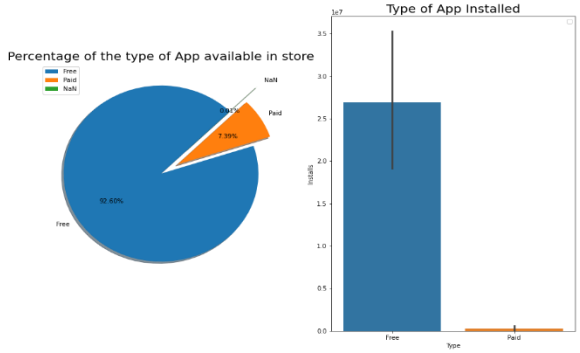
Game was the most dominating category in it. Then we plotted a count plot of the top installs in the play store.



It was to our shock that social and communications were the one which were actually downloaded the most

Then we plotted graphs based on the type

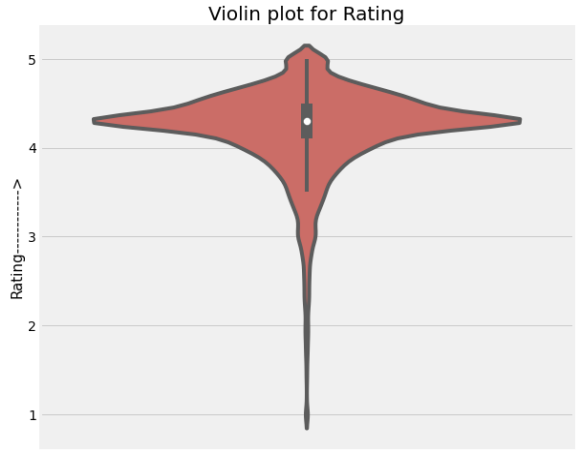


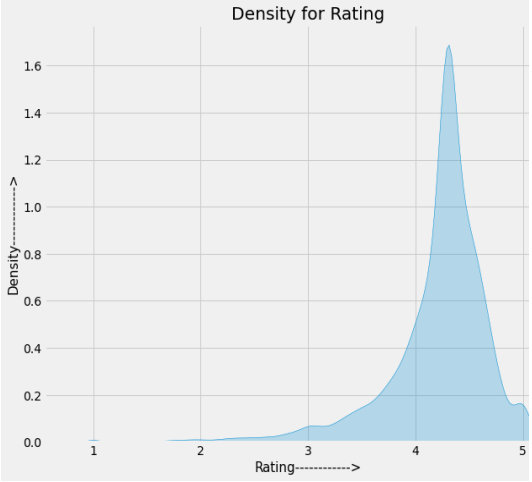


We found out, that even though the paid apps contributed to almost 8 percent of the total apps. The downloads are very marginal when compared with the free apps.

2. Observation 2: The User Rating

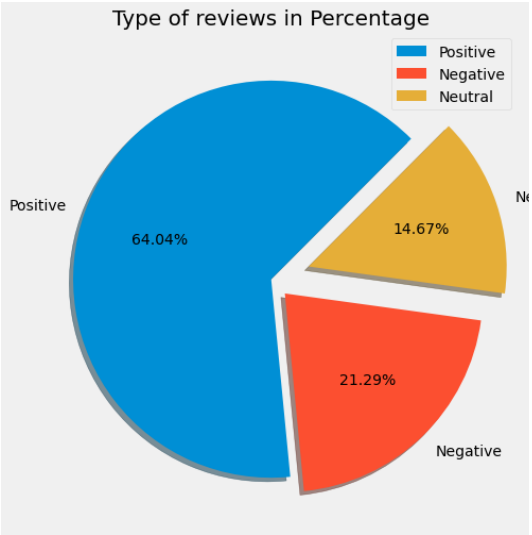
We did plot violin plot and density plot to understand the density of the rating





From our observation, we found that the average volume of ratings across all app categories is between 4-4.5. The histogram plot is skewed to the left indicating that the majority of the apps are highly rated with only a few exceptions in the low-rated apps.

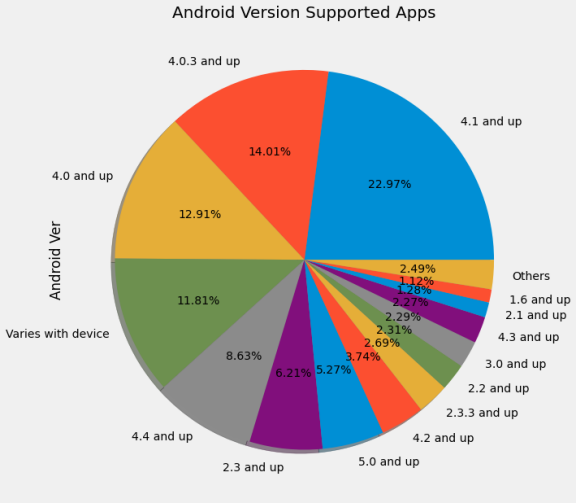
Now we plotted a pie chart to understand the percentage of the sentiments



Even though, almost 40% people have negative and neutral outlook in reviews, people generally rate the apps as good and very good.

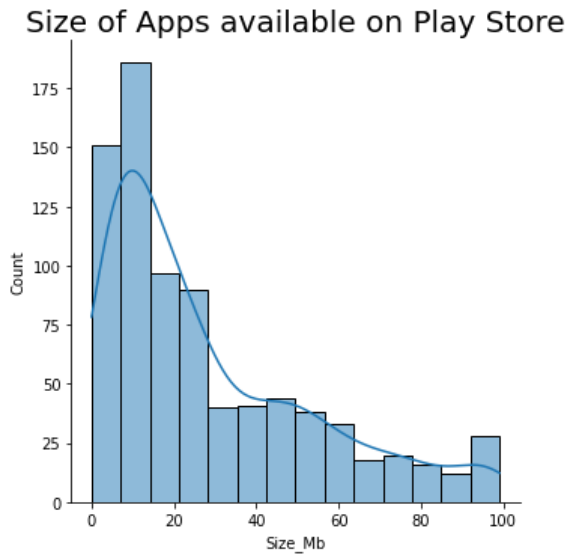
3. Observation 3: Type of Apps

We plot a pie chart of the android version supported Apps.

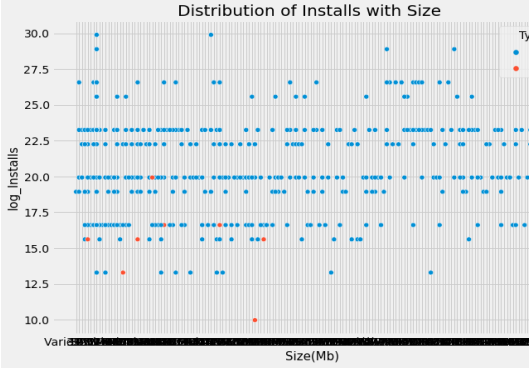


We found that maximum apps still support the older versions of android which means older version of android phone are still being used by the people.

Now we plotted the graph to find distribution of the size of apps available



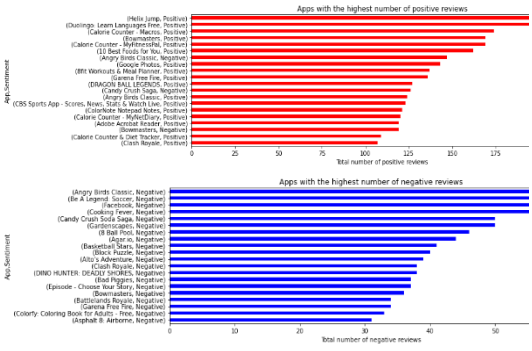
And yes we also plotted the distribution of the size of the installed apps.



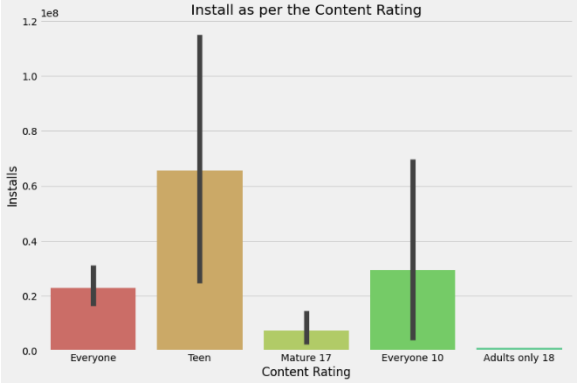
We found that size may impact the number of installation, the bulkier the application the less likely it is going to be installed by user.

4. Observation 4: The User Experience-I

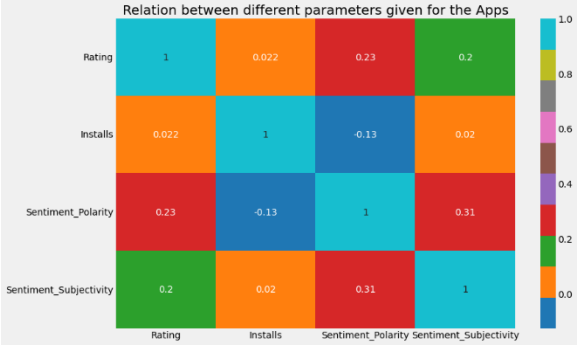
We plot a graph to find out the most positive and negative review and we found negative doesn't means it wasn't downloaded at all



Then we plotted a graph install based on the content rating



It was found that the category 12 and above had the maximum downloads. The adult category attracted very minuscule amount of install by the users.
We also plotted a correlation plot to find out the relation between different parameters.

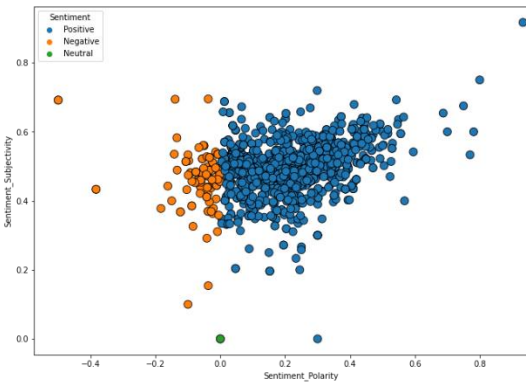


We observed that there is a direct relation between the sentiment polarity, rating, sentiment subjectivity and installs. If the sentiments are negative then there can be decrease in the number of installs by the users.

5. Observation 4: The User Experience-II (Sentiments)

User sentiments are an important part of app development as they provide feedback to the developer

and in return developer can improve the user experience for the users.
We started with a scatterplot of the sentiment polarity and sentiment subjectivity



Then we used word cloud to find the positive and the negative stop words.

We observed that the sentiment subjectivity is not always proportional to the sentiment polarity but in maximum number of cases it shows a proportional behavior

Using the sentiment data, we observed that reviews play a major role in app development process. It not only helps the developer to know the shortcoming in the app so that the developer can provide a great user experience.

Conclusion:

- The smaller apps have more audience and hence app development companies must focus on small but feature rich apps.
- It is better to develop apps that have support of older versions of devices as well.
- Developer should focus on developing apps which are rated PG 12+, as that category has the maximum installs. The social,

communications and tools have the maximum installs.

- The negative feedbacks directly impact the number of installs by the users, hence developer need to keep a eye to on the user reviews and feedbacks.

References-

1. Key Google Play Store Statistics in 2022 You Must Know (appinventiv.com)
2. GeeksforGeeks
3. Stackoverflow