

Play Store App Review Analysis

Tushar Jain,
Shivansh Yadav,
Madhur Awasthi,
Rishabh,
Data science trainees,
AlmaBetter, Bangalore

Abstract:

The Google play store is one of the most popular and largest Android application stores. It is one of the most used digital distribution services around the world. Google is offering a large collection of android mobile applications in different categories like entertainment, gaming, health, education, e-books etc. Google play store is pre-installed on Android smartphones. it is a free platform for downloading mobile applications by users. users get inspired by the rating and review of the application and users can rate this app on the google play store with star ratings and reviews. users rate an app-only one but they can change their review and rating at any time. generally, the user tends to download highly rated apps because they follow the other user experience.

Google play store is millions of apps present on google play store of different categories and they have billions of users android mobile users around the world. if you are interested in digital goods from google. play store is the only place where you want to go.

Now we know about the data set of the google play store. they are two data sets in which one is play store data and the other is review data. in this data set have 10k+ and other we have 60k+ value with 13 and 5 features. in this data set they are many insufficient and missing values. So we aim to study and predict the data to discover key

factors responsible for app engagement and success. I have tried to discover the relationship among various attributes present in the dataset and also try to perform data analysis and prediction into the google play store application dataset. our project code can be found at

https://github.com/MadhurAwasthi/Play_Store_App_Review_Analysis-EDA

Keywords: *Google play store app, Data scraping, Analysis, Visualisation, Exploratory Data Analysis*

1. Problem Statement

The play store data has provided great opportunities for app-making businesses to succeed and engaged. The android market and mobile app industries expanding day by day and also raising more competition for those who create applications. due to competition, we analyze this data to help developers or investors to understand what type of app is likely to attract users and know the factors responsible for the success of applications for building an application which could reach to maximum no of users and is profitable, helpful and engaging.

So for developing or investing in an application it is required to know what is the motivation of the people to download

this an app . the description of this variable is given below :

- App: It specifies the name of the app.
- Category: It includes the category of the app.
- Ratings: The no of the rating given by the users for a particular app.
- Reviews: The reviews given by the users for the app
- Size: It includes the memory required for the app.
- Installs: It gives the count of installs associated with the app.
- Type: It indicates the type of the app (it is free or paid).
- Price: It indicates the price of the app (0 if it is Free).
- Content Rating: The appropriate target audience of the app.
- Genres: The genre of the app.
- Last Update: It gives the information of the date when the app was last updated.
- Current version: It specifies the current version of the app.
- Android version: It gives information about the android version required to install the app

2. Introduction

Mobile apps are everywhere. They are easy to creat and can be luctative because of these two factors, more and more apps are being developed. Android market has capured around 74% of the total market which is says that huge amount of population using android.

Mobile application are one of the fastest growing segments of downloadable software

application markets. So we choose Google play store dataset due to as increasing popularity. In today's era the Google play store is the largest and most popular android app store 94% of apps in the store were freely available. So analyzing play store data set provide wide collection of features like rating, prices ,no of downloads and apps discription which help developer to making a app engagement and success.

3. Study

Before going ahead we have to study what we are looking for and why this is needed so let's discuss this.

We have studies from various resources like Google, Youtube and many more and got to know that in the current fast-paced environment every business wants to go online and also investors want to invest in the online businesses, Tech-savvy investors can do some research and they might make a decision, however, what about those who do not have that much tech-savvy and unable to make decisions.

We have studied investors or business that needs help on this and they want the correct and meaningful insights which can help them to build strong decision based on the data.

4. Software required specification

Now, when we have the clarity of the Introduction and we have studied the

problem statement, let's see what we needed to get the Insights into this.

There are a few things we needed before getting the insights discusses below:

- **Data Set**

Data Set is most important because we are going to provide the Insights based on the Data set only, We have the google play store applications Data set downloaded from google.

- **Google Colab**

We are going to use Google Colab, it allow us to use the Python Arbitrary codes in single Notebook files, also we can share it among our team so anyone can access it and make changed.

- **Python, Pandas, Numpy, Matplotlib, Seaborn**

Now, when we do have the dataset so we can start the Analyzing Part and for the same, we are going to use Python, we are using python here because Python has vast libraries which can help us to Analyze the Data without anomalies can give us a meaningful Insights.

Pandas is the library of python which we are going to use for creating the Dataframes and for removing the anomalies from the Data.

Matplotlib we are going to use for the Visualization part so that stakeholders can easily understand the data.

4. Steps involved:

- **Data Preparation and cleaning**

Now when we do have the clarity of the problem and we know the tools which can be used so now the most important part is Data preparation and cleaning.

We can understand when we download the raw data from any source so they contain many duplicate values, Null values etc.

- **Null values Treatment**

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

- **Duplicate Values Treatment**

In our Dataset, we also have some duplicate values, like we have some apps with the same categories which are presented more than once, as they have the same ratings, the same Category so we can conclude that it is a redundant data which we can remove.

- **Column Selection**

We know that in our data set we do have many rows like we do have approx 13 rows which all can not be used for getting the Insights, so we have selected the columns which can help us to get meaningful insights like:

- **Rating:** Rating is most important to check the app engagement because using

the Rating column we can predict how famous the application is, however, we also have to check the Installs because it might happen an App do have the 5 rating and 100 Installs, It is not a successful application at all.

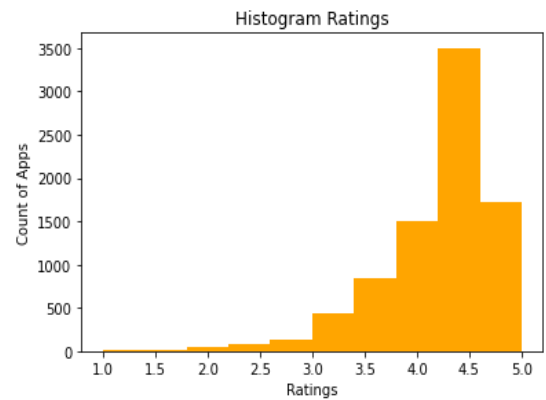
- **Installs:** Installs are most important to get the insights on the play store Data set because based on the Installs only we can say which app is being more used among the people.
- **Size:** Size is also important for the App because generally, As in my opinion users want to use the lower size application, we will see in the conclusion section what our Data says.
- **Price:** We can use price as well because generally everyone wants to use free applications.

- **Visualization:**

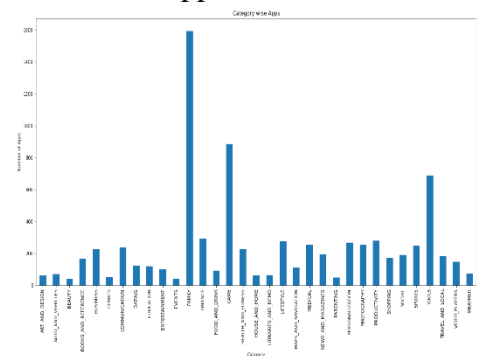
Visualization is the most important part here because stakeholders can understand the Data based on Graphs easily so we have to build Graphs for the same.

- **Rating Histograms:** We have built histograms for the ratings so that we can see the number of apps which contains high ratings, therefore we can analyze the

apps which have good ratings, In the below histograms we can see that it is left-skewed which means that most of the applications we have with good ratings with a rating of 4.5.



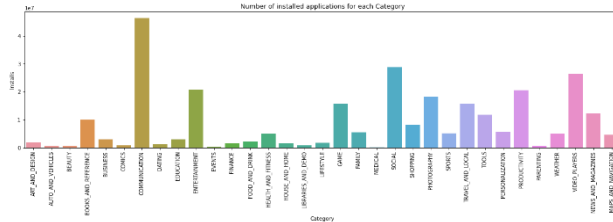
- **Category wise apps:** Let's see the apps according to categories so that we can see which category contains the most of the applications.



From this plotting, we know that most of the apps in the play store by categories are from 'Family', 'Game' and 'Tools'.

- **Number of apps Installed in each category:** Now let's see

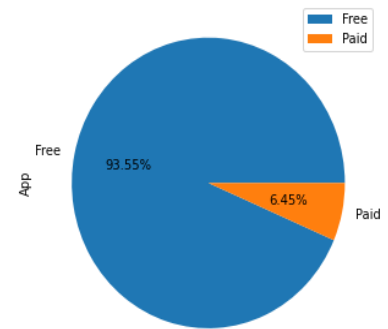
which category has the more number of Installed applications so that we can conclude the most successful category.



From here it is seen that applications of the category 'BOOKS_AND_REFERENCE', 'COMMUNICATIONS', 'ENTERTAINMENT', 'GAME', 'FAMILY', 'SOCIAL', 'PHOTOGRAPHY', 'TRAVEL_AND_LOCAL', 'TOOL', 'PRODUCTIVITY', 'VIDEO_PLAYERS' and 'NEWS_AND_MAGAZINES' are installed the most (1 billion+ times). Along with these, 'HEALTH_AND_FITNESS' is also installed approximately 4.5 million+ times.

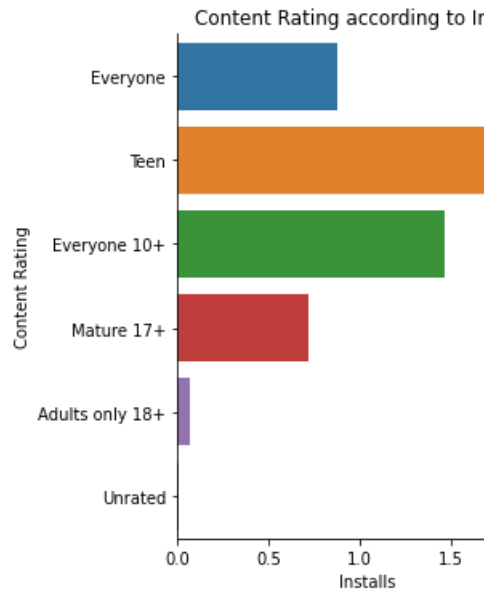
- **Type price Distribution:** Let's see the distributions of the price using a pie chart.

A Pie Chart Representing Percentage of Type Distribution

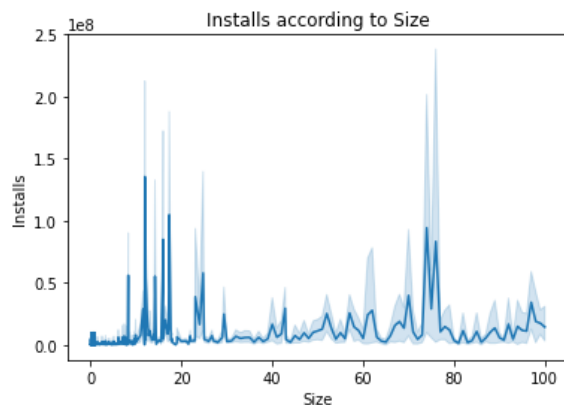


In the pie chart, we can see, Free apps are installed more than paid apps.

- **Content Rating according to Installs:** Now, it's time to see which generations are using the applications most based on Installation. Apps targeted at any age group other than those labelled 'Adults only 18+' and 'Unrated' are installed. It can however be seen that apps labelled 'Teen' and 'Everyone 10+' is installed 30 million+ times.



- **Installs according to Size:**
From the chart, we can say that the applications whose sizes are from 15-to 20 have the maximum number of installs.



5. Conclusion: This is our last part of the project, here we will conclude with the things we found.

The application's success is majorly perceived by its number of 'Installs' and its

'Rating' and there are a few key factors responsible for both.

So we found that apps with a maximum no of 'Installs' belong to Category:

- Communication
- Social
- Video_Players
- Productivity
- Entertainment

The Content Rating should be:

- Teen, Everyone 10+
- Everyone

So if we want to build an app we should keep in mind the category as well as the content rating.

In our analysis of the top 10 apps, we found the most successful app would be of Category:

- Communication
- Social

Content rating should be:

- Teen
- Everyone

If we want to create an App, the most favourite categories are:

- Communication
- Social
- Video_Players
- Productivity
- Entertainment

Most favourable content:

- Teen
- Everyone 10+
- Everyone

- All most all of the top 10 apps are in the size range of 10-25 MB
- Almost all top apps are Free.

we can deduce an application of:

- Category (Communication)
- Content (Teen)
- Size (10-25) MB
- Free

would most probably Succeed.