PLAY STORE APP REVIEW ANALYSIS

By Ankita Hanamshet

Abstract:

Google Play Store or formerly Android Market, is a digital distribution service developed and operated by Google. It is an official apps store that provides variety content such as apps, books, magazines, music, movies and television programs. It serves an as platform to allow users with 'Google certified' Android operating system devices to donwload applications developed and published on the platform either with a charge or free of cost. With the rapidly growth of Android devices and apps, it would be interesting to perform data analysis on the data to obtain valuable insights.

1.Problem Statement

The 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.
Each app (row) has values for catergory, rating, size, and more. Another dataset contains customer reviews of the android apps.
Explore and analyze the data to discover key factors responsible for app engagement and success.

2. Introduction

Mobile apps are becoming popular every coming day as they are fast, practical and efficient. Their popularity is among users as well as in business owners also as the apps can be used for advertising. App is nothing but a computer program designed to run on devices like phone, tablet to provide users some services. Apps are offered in specific software repositories referred generally as App stores, where the largest shareholders are Google Play1, iPhone App Store 2 and Blackberry App World 3. Google play store is an official app store for Android operating system. These apps belong to various categories like games, communication, books, business, news, sports, and many others. In Google Play store, each app has its name and its rating. The app rating is the average of all the ratings given by the users. The users tend to download the apps with high ratings as it reflects the good experience of other users who already downloaded and used the apps. The study has showed that the app ratings affect the user’s intention to download and use an app . However, there is no application exists that provides relationship between the ratings and other factors of app so that the developer can ultimately decide how to develop good quality app. Also, many companies even use wrong ways to gain high ratings for their apps such as paid app ratings, so their apps can be more popular and visible in the play store. The proposed system provides an interface to visualize the

relationship between the ratings and the other factors of app for developers and the overall quality of app for users.

3. Description of App Dataset columns

1. App : The name of the app
2. Category : The category of the app
3. Rating : The rating of the app in the Play Store
4. Reviews : The number of reviews of the app
5. Size : The size of the app
6. Install : The number of installs of the app
7. Type : The type of the app (Free/Paid)
8. The price of the app (0 if it is Free)
9. Content Rating :The appropiate target audience of the app
10. Genres: The genre of the app
11. Last Updated : The date when the app was last updated
12. Current Ver : The current version of the app
13. Android Ver : The minimum Android version required to run the app

4. Data Prepration

Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. It is an important step prior to processing and often involves reformatting data, making corrections to data and the combining of data sets to enrich data.

Gathering data

This step is about getting to know the data and understanding what has to be done before the data becomes useful in a particular context. This can be done by reading the CSV file and doing initial statistical analysis.

Though the dataset may seem to have the correct datatypes for each column, we need to check it. Inconsistent datatypes will create issues while dealing with problems.

Cleanse and validate data

This step is crucial for removing faulty data and filling in gaps. Important tasks here includes:Removing extraneous data
Filling in missing values.
Conforming data to a standardized pattern.
Dataset may contain duplicate values for particular application:

Transform data

Transforming data is the process of updating the format or value entries in order to reach a well-defined outcome, or to make the data more easily understood by a wider audience.

5 . Data Cleaning

By diagnosing the data frame, we know that:

1. There are 13 columns of properties with 10841 rows of data.
2. Column 'Reviews', 'Size', 'Installs' and 'Price' are in the type of 'object'

3. Values of column 'Size' are strings representing size in 'M' as Megabytes, 'k' as kilobytes and also 'Varies with devices'.
4. Values of column 'Installs' are strings representing install amount with symbols such as ',' and '+'.
5. Values of column 'Price' are strings representing price with symbol '$'.

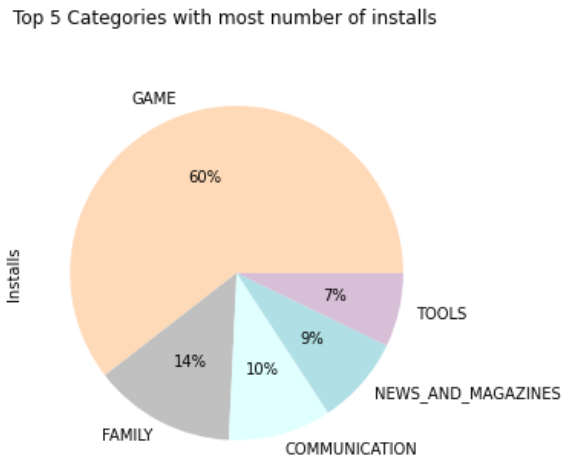
Hence, we will need to do some data cleaning.

1. We realized that there are 1474 rows having null values under column 'Rating'. Hence, we decided to replace the null values with Mode of overall 'Rating' values.
2. Features like Type,Content Rating, App Ver and Android version had 1,1,8,3 null values so dropping those columns
3. Also deleting the duplicate recordsColumn 'Reviews', 'Size', 'Installs' and 'Price' are in the type of 'object' so converting them to numeric type
4. Values of column 'Size' are strings representing size in 'M' as Megabytes, 'k' as kilobytes and also 'Varies with devices'. Replacing these values with integer and also dropping ‘Varies with Device’ because
   it will be not a better idea to replace t hose value with mean or mode since, the size of some apps would be too l arge and some of them too small.
5. Values of column 'Installs' are strings representing install amount with symbols such as ',' and '+'. Remiving ‘,’ and ‘+’
6. Values of column 'Price' are strings representing price with symbol '$'. Removing ‘$’

6 . Exploratory Analysis and Visualization

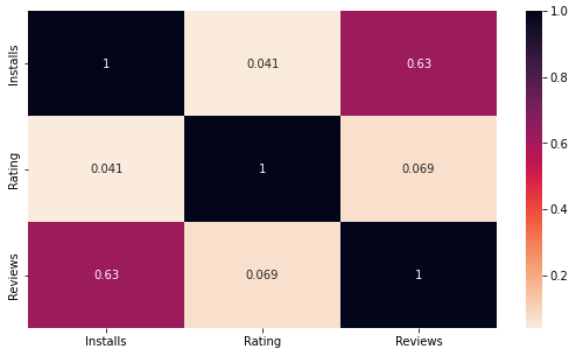
Exploratory data visualizations (EDVs) are the type of visualizations we assemble when we do not have a clue about what information lies within our dataset.

Analysis 1: Categories which have most number of installs?



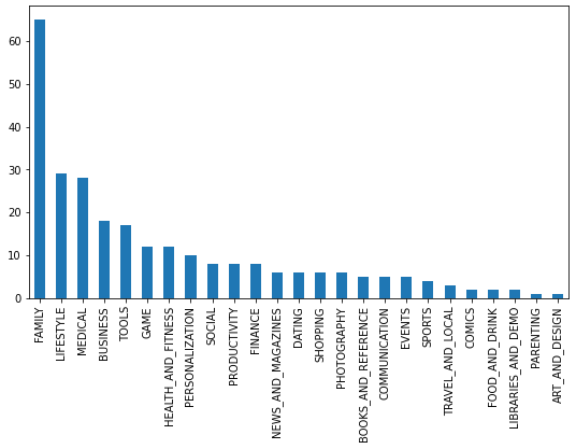
- When we consider top 5 app install categories. The category is dominated by Gaming by having 60% presence

Analysis 2: Correlation between Installs, Rating and Review



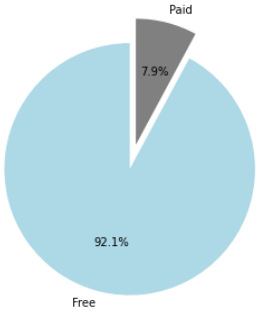
From the graph it is clear that Installs and R eviews are correlated which means that thes e both are directly proportional

Analysis 3: Categories to which highest rated(Rating=5) apps belong



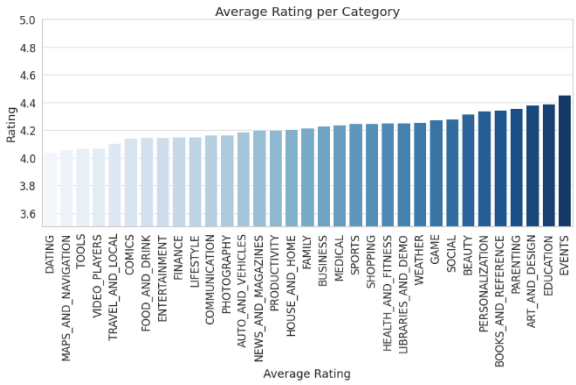
- The maximum rating in Google app store is 5
- From the analysis, Family has most no of max rated apps followed by Lifestyle and Medical

Analysis 4: Finding the percentage of paid and free apps in Play Store.



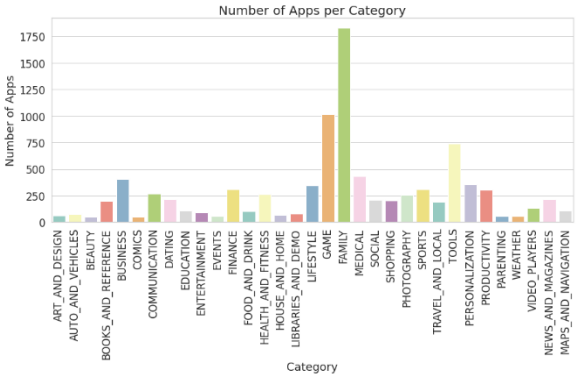
From the plot we can imply that majority of the apps in the Play Store are Free apps.

Analysis 5: What is the average rating per category?



Average ratings per category are distributed between 4.0 and 4.5

Analysis 6: How many apps are there in each category?

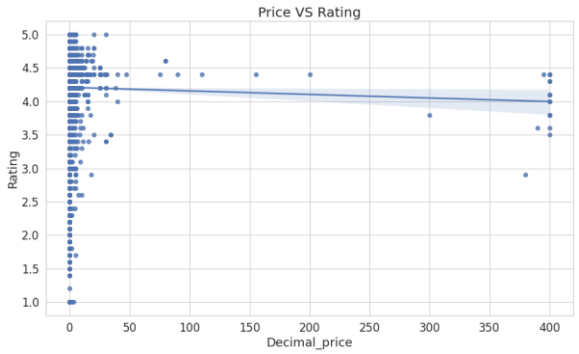


- Art And Design and Events category is having the least number of apps

but they have the highest average rating

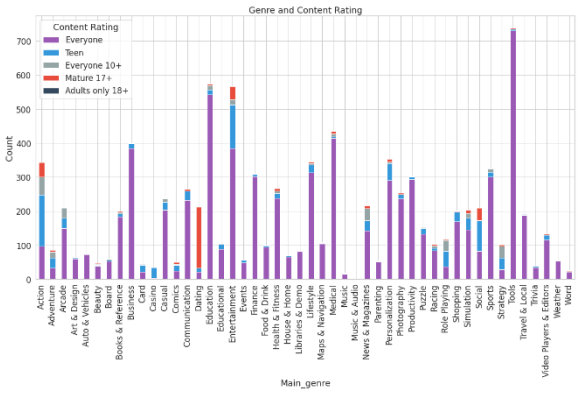
- Game and Family category apps have outnumbered the other apps

Analysis 7: Do we have a correlation between price of the app and rating?



- From the graph price and rating seems to be inversely proportional
- Apps either have a price lower than 100, or price greater than 300

Analysis 8: Which genres addresses which audience?



- Almost every app genre is for everyone
- But dating apps are for mostly for mature individuals rather than everyone :)
- Actions genre is mostly for Teens

7. Conclusion

The Google Play Store Apps report provides some useful insights regarding the trending of the apps in the play store. As per the graphs visualizations shown above, most of the trending apps (in terms of users' installs) are from the categories like GAME, COMMUNICATION, and TOOL even though the amount of available apps from these categories are twice as much lesser than the category FAMILY. The trending of these apps are most probably due to their nature of being able to entertain or assist the user. Besides, it also shows a good trend where we can see that developers from these categories are focusing on the quality instead of the quantity of the apps.

Other than that, the charts shown above actually implies that most of the apps having good ratings of above 4.0 are mostly confirmed to have high amount of reviews and user installs. There are some spikes in term of size and price but it shouldn't reflect that apps with high rating are mostly big in size and pricy as by looking at the graphs they are most probably are due to some minority. Futhermore, most of the apps that are having high amount of reviews are from the categories of SOCIAL, COMMUNICATION and GAME like Facebook, WhatsApp Messenger, Instagram, Messenger – Text and Video Chat for Free, Clash of Clans etc.

Eventhough apps from the categories like GAME, SOCIAL, COMMUNICATION and TOOL of having the highest amount of

installs, rating and reviews are reflecting the current trend of Android users, they are not even appearing as category in the top 5 most expensive apps in the store (which are mostly from FINANCE and LIFESTYLE). As a conclsuion, we learnt that the current trend in the Android market are mostly from these categories which either assisting, communicating or entertaining apps.

References-

1. GitHub
2. GeeksforGeeks
3. Analytics Vidhya
4. Jovian AI