**Capstone Project -1 Submission**

**Play Store App Review Analysis**

****

**Abhinov Anand**

**Saurav Kumar**

**Vivek Kumar**

**Data Science Enthusiast**

**AlmaBetter, Bangalore**

---------------------------------------\*\*\*\*------------------------------------

**Abhinov Anand-** <abhinov.1rn14me001@gmail.com>

**Saurav Kumar-** <kumarsaurav1894@gmail.com>

**Vivek Kumar –** <vivekkumarangad@gmail.com>

**GitHub Link :-**

**Abhinov Anand-**<https://github.com/abhinov311297/Play_Store_Review_Analysis>

**Saurav Kumar-** <https://github.com/Saurav1894/Play-Store-App-Review-Analysis-Capstone-Project>

**Vivek Kumar-**<https://github.com/vivekkumar2100/Play-Store-App-Review-Analysis>

**Content**

**Title Page no**

**1.Abstract 01**

**2.Problem Statement 02**

**3.Introduction 03**

**4. EDA on given Data set 04**

**5. Distribution of Category 05**

**6. Category wise mean rating 06**

**7. Paid vs Free apps 07**

**8. How the Play Store App work 08**

**9. What Review Saying? 09**

**10.Steps Involved 09**

**11.Conclusion 10**

**1.Abstract**

Google play store is engulfed with a few thousands of new applications regularly with a progressively huge number of designers working freely or on the other hand in a group to make them successful, with the enormous challenge from everywhere throughout the globe.

The play store app review analysis seeks to overview of the developments of the apps. In which it contains the size of apps, number of installs, rating, content rating and this overall features of app from the play store.

Our experiments can help to understand what is the criteria to installs any type of apps. So the business problems can judge quickly and solve them by using sorting, cleaning and data analysis.

**Keywords:- Data cleaning, Data analysis.**

**2.Problem Statement**

* Data provides an (row) has values for category, rating, size, and more. Another dataset contains customer reviews of android.
* Each app has been classified in different types of category which has given restriction for some users or we can say that its divide into mature, teen, 18+, etc.
* The main objective is to build a EDA which helps in the overview of the development of apps from the play store very quickly and efficiently.

**Apps:** Different apps from play store.

**Category:** Each app has been classified according to their types of works. Ex. Photo editor is belonging to Art and design.

**Rating:** The rating of apps varies from 1.0 to 4.0.

**Size:** The total size of the apps in play store.

**Installs:** The number of installs for apps.

**Type:** There are two types of apps, paid and free.

**Content:** This classified as according to the age of customer.

**Last Update:** The date in which the app is updated.

**Current Version:** The current version of the apps.

**Android version:** The required version of apps.

**3.Introduction**

The Google Play Store started life as the “Android Market” in 2008. It launched alongside the very first Android devices, and its purpose was to distribute apps and games. The Android Market was extremely basic at the beginning. It didn’t support paid apps and games until 2009. However, as the Android platform grew, so did the Android Market. By 2012, it featured over 450,000 Android apps and games.

By this time, Google’s ecosystem had expanded greatly compared to the humble beginnings of the Android Market. In fact, the Android Market was just one of the company’s online markets. At the time, this was the only place that Google had to sell goods. As the company’s hardware efforts grew, it was time for a new store. Posting reviews online has become an increasingly popular way for people to express opinions and sentiments towards the products bought or services received. Analysing the large value of online reviews would produce useful actionable knowledge that could be of economic values to vendors and other interested parties.

**4.EDA on given Data set**

There are two dataset:

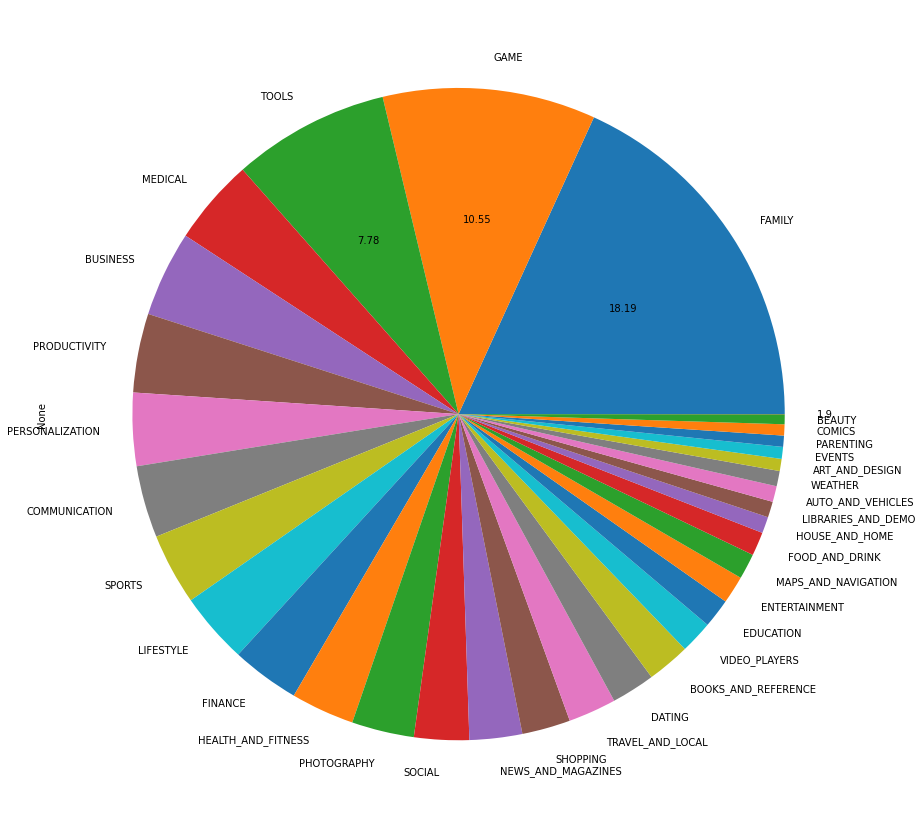
1. **Play Store Data** (App, Category, Rating, Review, Size, Install, Type, current rating, genres, Last update, Current Var, Android Var)
2. **User Review Data** (App, Sentiment, Sentiment Polarity, Sentiment Subjectivity)

By diagnosing the data frame, we know that:

* There are 13 columns of properties with 10841 rows of data.
* Column 'Reviews', 'Size', 'Installs' and 'Price' are in the type of 'object'
* Values of column 'Size' are strings representing size in 'M' as Megabytes, 'k' as kilobytes and also 'Varies with devices'.
* Values of column 'Installs' are strings representing install amount with symbols such as ',' and '+'.
* Values of column 'Price' are strings representing price with symbol '$'.

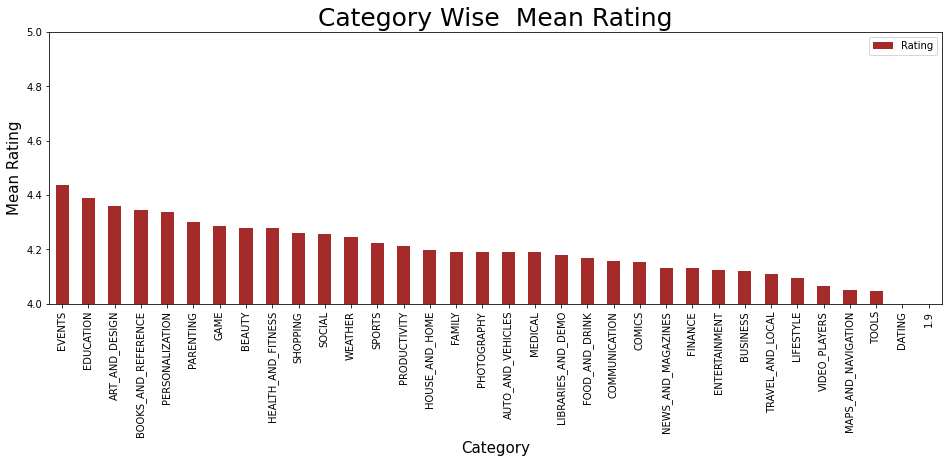
**5.Distribution of Category**

This pie chart shows us the Family related apps are the most number in availability, followed by Games which is generally preferred by Children. This two constitute nearby 30% of the available Apps.



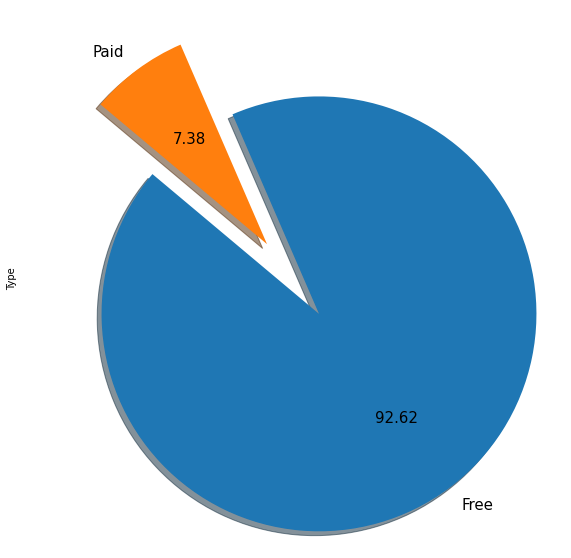
**6. Category wise mean rating**

This shows the Education related Apps also more desirable for the users, their high rating suggest why the current generation knows going for the online education. Online education provide opportunities to the student especially girls in remote areas to access the learning material.

****

**7.Paid vs Free apps**

Results are quite astonishing! Only 7-8% are paid Apps thus makes Google Play Store most popular among the users. Even middle class or poor can use this store optimally.

****

**8.How the Play Store App work**

At its basics, the ranking algorithm of both Google Play Store and App Store work the same. However, I also noticed some concrete differences regarding the importance of some factors during my work with various apps.

So this are the key points in which App Store differs from Google Play App Store is stronger focused on Keywords (Optimize your app meta data - Title, Description, Keywords!)- Only ratings and reviews for the latest app version are displayed on App Store (encourage your users to leave some reviews).

There’s a higher focus on App Quality on App Store (updating your app is not as easy as on Google Play) Hope it was helpful! **Disclaimer**: I am the CEO of **App Radar**. We create solutions to optimize your app publishing process and make it easier to optimize your apps for a higher visibility within the app stores.

**9.What Review Saying?**

The reviews are of three types:

* Positive Sentiment- The customer which allow to give their positive feedback regarding their apps.
* Negative Sentiment- The apps not perform well accordingly customer. So the customer allowed to give their negative feedback for the apps.
* Neutral sentiment- This type of sentiment are classified as neutral reviews.

**10.Step Involved**

* **Loading the data sets:** Two data sets, first Play Store app dataset and user reviews datasets.
* **Importing libraries:** NumPy, Pandas, Seaborn and Matplotlib.
* **Data cleaning:** In this process we convert all string contained data to numeric data. As instead of this we will unable to do the EDA on the data of play store.
* **Exploratory Data Analysis:** In this procedure we simultaneously work with each feature from our dataset and clearly visualize each and every point of aspects. This gives a graphically representation of entire dataset.
* **Extraction of un-useful Data:** In entire data some features has un-useful for the data analysis. So it’s better to extract this kind of features from our dataset.

**11.Conclusion**

* We must to know what is data cleaning? So we just giving edge to our concept that everyone should know data cleaning. Then we move with EDA which was contained of different type of analysis, we look individually on category of apps.
* After several operation we come to know about the various aspect of the Dataset of Play Store. like Which Category of Apps are most available, their Installs Quantity, Reviews numbers as well as Size and Ratings they got and found the Games related Apps are the dominant one.
* We also gone through the information related to the Price where we have seen the 92% Apps are free and got the list of most expensive Apps. Age wise Content Rating also touched to analyze the pattern.
* We also tried to figure out the number of Apps shared on different version of Android to see the technological penetration in App market.
* At last we tried to map a correlation of different data and got the stronger positive relation between Review and Installs.
* Quiet long!but this analysis gave information not only about Apps but also the interest of the users as well as population which ll later help developer to make more user friendly Apps and can be utilized by everyone.