**Play Store App Review Analysis**

**ASIT KUMAR PANDA**

**DATA SCIENCE TRAINEE,**

**ALMABETTER**

**Abstract:**

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 category, 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.

**1.Introduction**

**Google Play**, also branded as the **Google Play Store** and formerly **Android Market**, is a [digital distribution](https://en.wikipedia.org/wiki/Digital_distribution) service operated and developed by [Google](https://en.wikipedia.org/wiki/Google). It serves as the official app store for certified devices running on the [Android operating system](https://en.wikipedia.org/wiki/Android_(operating_system)) and [its derivatives](https://en.wikipedia.org/wiki/Google_Operating_System) as well as [ChromeOS](https://en.wikipedia.org/wiki/ChromeOS), allowing users to browse and download applications developed with the [Android software development kit](https://en.wikipedia.org/wiki/Android_software_development) (SDK) and published through Google. Google Play has also served as a [digital media](https://en.wikipedia.org/wiki/Digital_media) store, offering music, books, movies, and television programs. Content that has been purchased on [Google Play Movies & TV](https://en.wikipedia.org/wiki/Google_TV_(service)) and [Google Play Books](https://en.wikipedia.org/wiki/Google_Play_Books) can be accessed on a [web browser](https://en.wikipedia.org/wiki/Web_browser), and through the [Android](https://en.wikipedia.org/wiki/Android_(operating_system)) and [iOS](https://en.wikipedia.org/wiki/IOS) apps.

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 [exclusive](https://en.wikipedia.org/wiki/Proprietary_software) Google Play Store [mobile app](https://en.wikipedia.org/wiki/Mobile_app) or by [deploying](https://en.wikipedia.org/wiki/Software_deployment) the application to a device from the Google Play website. Applications operating hardware capabilities of a device can be targeted to users of devices with specific hardware components, such as a motion sensor (for motion-dependent games) or a front-facing camera (for online video calling). The Google Play Store had over 82 billion app downloads in 2016 and reached over 3.5 million apps published in 2017, while after a purge of apps is back to over 3 million. It has been the subject of multiple issues concerning safety, in which [malicious software](https://en.wikipedia.org/wiki/Malware) has been approved and uploaded to the store and downloaded by users, with varying degrees of severity.

Google Play was launched on March 6, 2012, bringing together Android Market, Google Music, Google Movies and the Google eBookstore under one brand, marking a shift in Google's digital distribution strategy. Following their re-branding, Google has expanded the geographical support for each of the services. Since 2018, Google has gradually sunsetted the Play brand: [Play Newsstand](https://en.wikipedia.org/wiki/Google_Play_Newsstand) was rebranded as [Google News](https://en.wikipedia.org/wiki/Google_News) in 2018; [Play Music](https://en.wikipedia.org/wiki/Google_Play_Music) was withdrawn in favor of [YouTube Music](https://en.wikipedia.org/wiki/YouTube_Music) in 2020; and Play Movies & TV was rebranded as Google TV in 2021. In 2022, [Play Games](https://en.wikipedia.org/wiki/Google_Play_Games) is expected to shut down its mobile app in favor of an Android emulator for Windows with the same name. The remaining standalone mobile app will be Play Books.

1. **Description of Dataset**

### Dataset consist of two csv file.

1. Playstrore.csv: In this file total number of columns is 13 and number of rows is 10839.
2. User\_reviews.csv: In this file total number of columns is 5 and total number of rows is 64294

Playstrore.csv:

|  |  |  |
| --- | --- | --- |
| Sr. No | Column | Description |
| 1 | App | Name of the App and Game. |
| 2 | Category | Category of App |
| 3 | Rating | Rating of App |
| 4 | Reviews | Total Number of reviews |
| 5 | Size | Size of app in MB |
| 6 | Install | Number of installs |
| 7 | Type | Free type or paid type |
| 8 | Price | App price |
| 9 | Content Rating | Audience of app |
| 10 | Genres | Different genres under app falls |
| 11 | Last Updated | Last update date |
| 12 | Current Version | Current version of app |
| 13 | Android Version | Android version required |

User\_reviews.csv

|  |  |  |
| --- | --- | --- |
| Sr. No | Column | Description |
| 1 | App | Name of the App |
| 2 | Translated reviews | User comments on app |
| 3 | Sentiment | User sentiment |
| 4 | Sentiment Polarity | Polarity refers to the strength of an opinion. |
| 5 | Sentiment Subjectivity | Subjectivity refers to the degree to which a person is personally involved in an object. |

## **3. Filling Null Values**

In this we observe all the values of dataset, if there is any missing value, null value, empty cell or blank cell.

After observing the first file which is Playstore.csv we found the column rating, current version and android version have null values.

## **4. Data Cleaning**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time.

Data cleaning means fixing bad data in your data set.

Bad data could be:

1. Empty cells
2. Data in wrong format
3. Wrong data
4. Duplicates

Now we discuss the cleaning of our Dataset.

We start cleaning data from Reviews column. The datatype of this column is object and we want it to convert it to integer. We do some coding and changed it to integer successfully. Now we take size column from dataset. In this column there is outlier(‘M’) is present and datatype is object. We want to remove the outlier(‘M’) from it and change the datatype as float. We do some coding and store the data of size in new column called as size\_in\_mb. Now we take Installs column. In this column, outliers (‘+’) are present and datatype is object. We want datatype as integer. So, we did some coding and remove the outlier (‘+’) and change datatype as integer. Now we take price column. ‘$’ outlier is present in price column and datatype is object. We want to remove the outlier and change the datatype as float. We did some coding and remove the outlier (‘$’) and change datatype as float. After we take Android ver column. The datatype is object and we convert it as float and store in new column named it as min\_andriod\_ver. After that we take last updated column and using this column, we make new column and named it as year. In this we store year of each application. Now we done with dataset and clear all duplicates, empty cell and wrong data. Now our data is clean and using this data we create new Dataframe named it as df\_new.

1. **Data Visualization**

Data visualization is the discipline of trying to understand data by placing it in a visual context so that patterns, trends, and correlations that might not otherwise be detected can be exposed.

Python offers multiple great graphing libraries packed with lots of different features. Whether you want to create interactive or highly customized plots, Python has an excellent library for you.

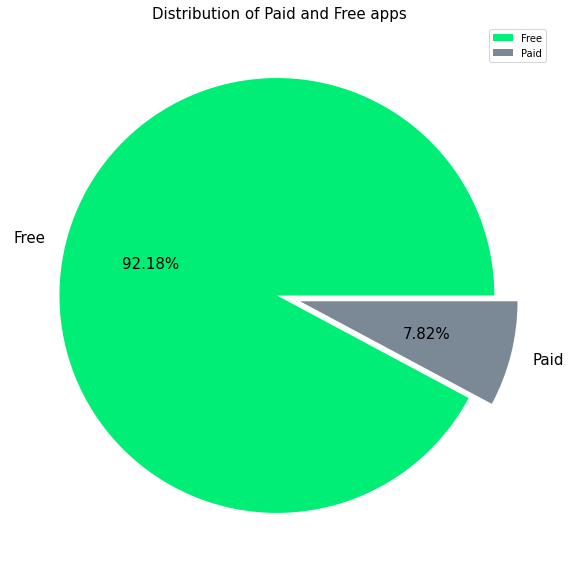
To get a little overview, here are a few popular plotting libraries:

1. Matplotlib: low level, provides lots of freedom
2. Pandas Visualization: easy to use interface, built on Matplotlib
3. Seaborn: high-level interface, great default styles
4. Plotly: can create interactive plots

Using these libraries, we do some visualization.

1. **Observation-1: (Number of Paid Apps & Free Apps):**

In this observation we find out the number of Free apps & Paid Apps in google playstore with visualisation graph.

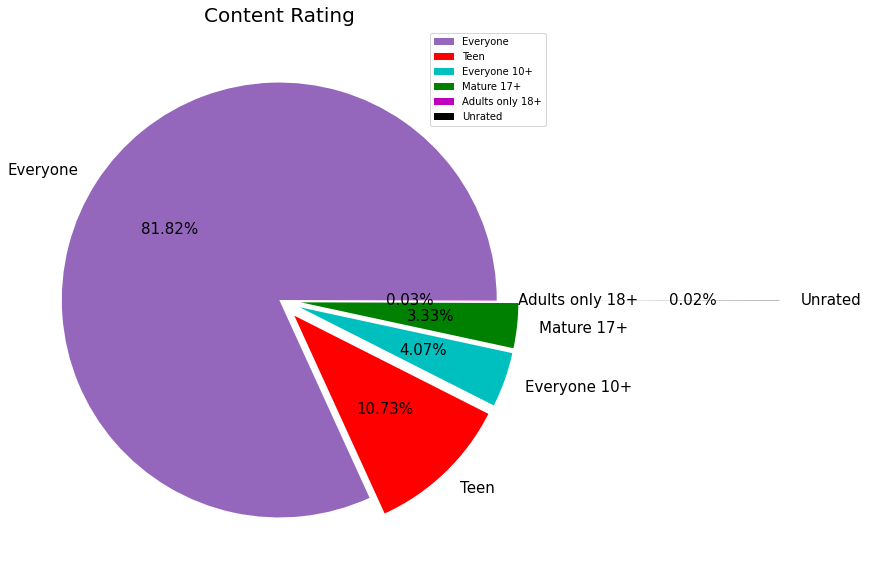


In this graph we visualize the percentage of free and paid apps. After distribution of all apps from google play store we found 92.18 % of apps available free and 7.82 % of apps found paid version.Means most of the apps are available free in Play store.

1. **Observation-2: (Content Rating column which category apps are found more on play store):**

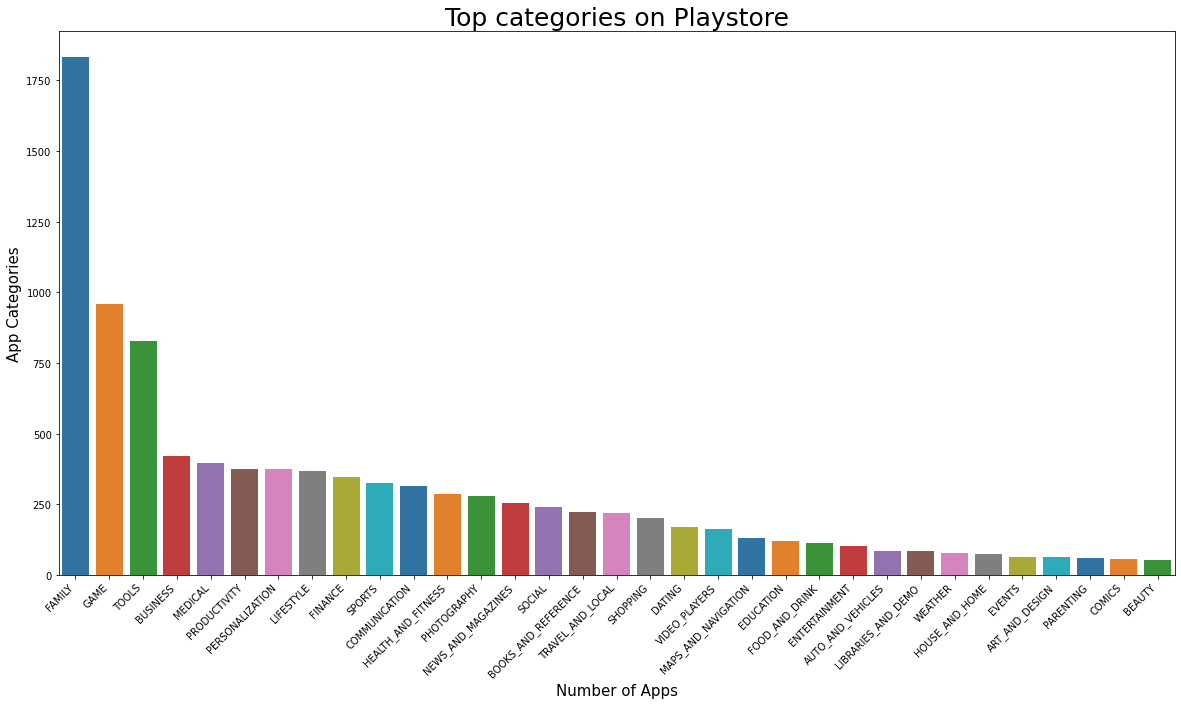
In this observation we visualize which app is more found on play store.

We found that Everyone category app is found more in play store with 81.82 %.



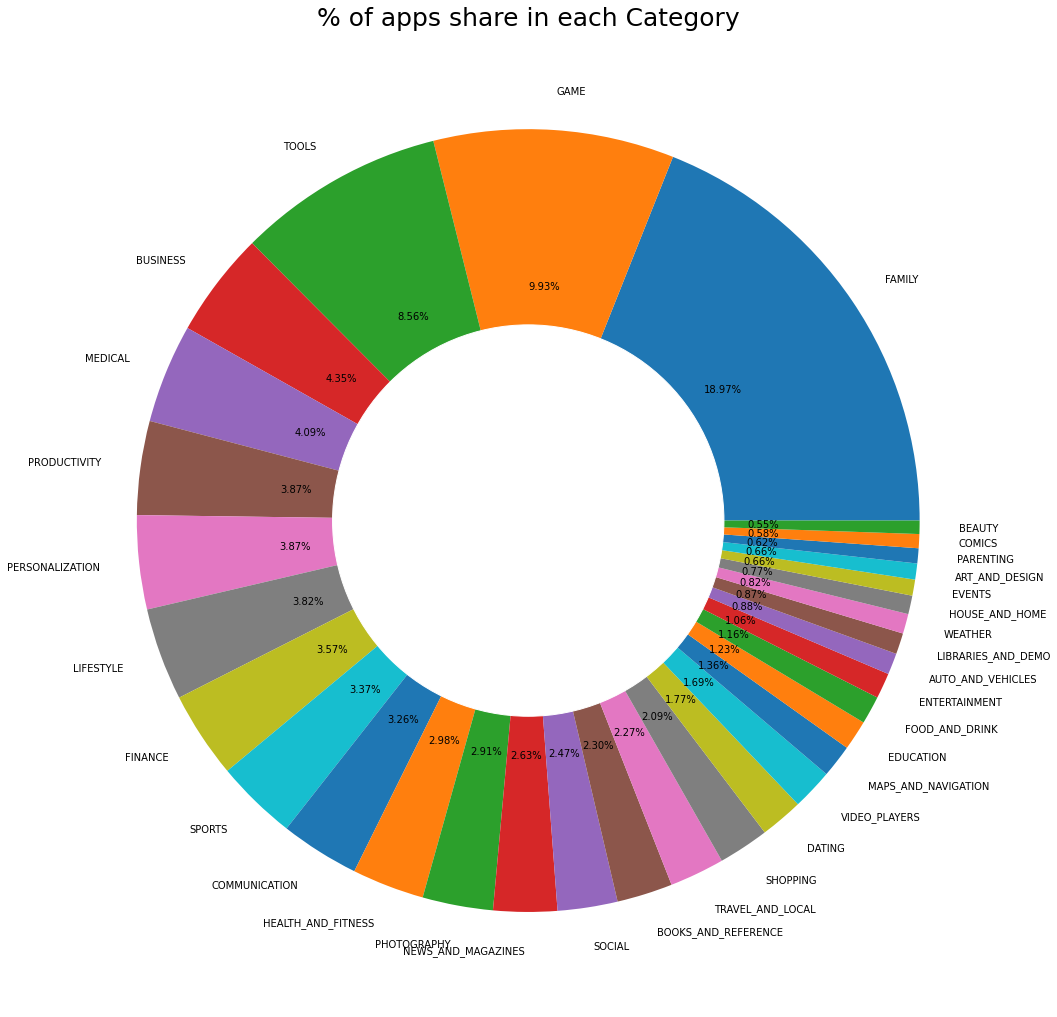
1. **Observation-3: (Top categories Apps on Google Play store):**

In this observation we present a bar graph to visualize the top category apps on google play store. We found Family , game and tools apps are top 03 apps on play store.



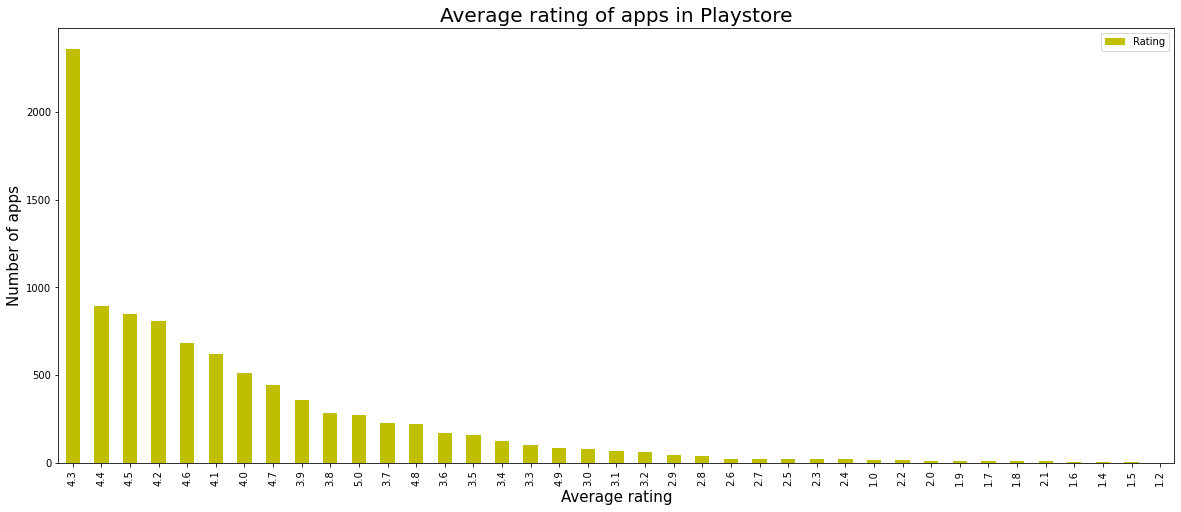
1. **Observation-4: (Percentage of apps in each category in the play store):**

In this visualization we observe that percentage of apps present on google play store in each category.Family, Game and tools apps are top 03 apps in play store which having 18.97%, 9.93% and 8.56% respectively.



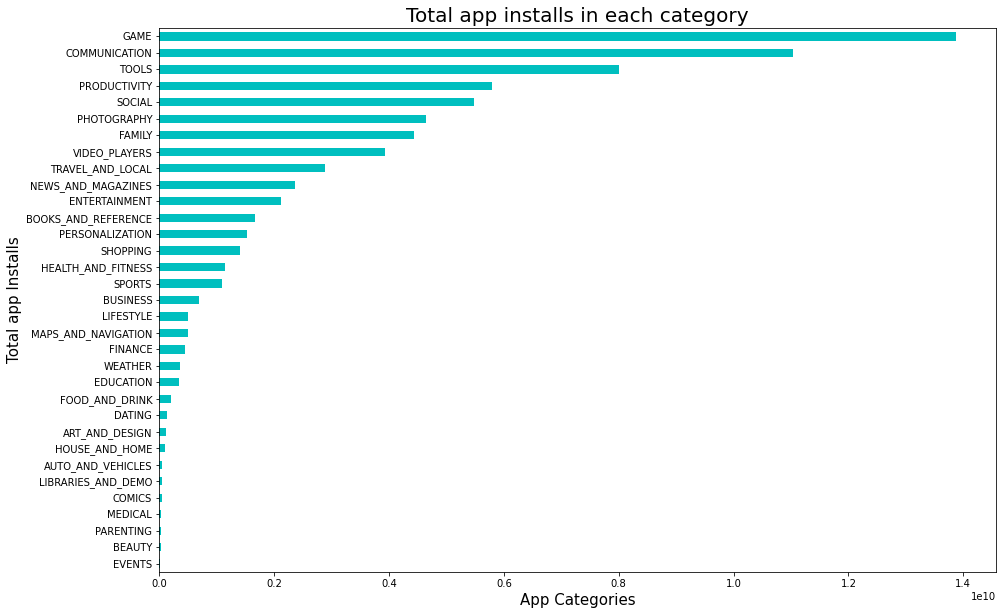
1. **Observation-5: (Average rating of the apps):**

In this visualize we present the average rating of apps in google play store with the bar graph.in average rating most of the apps getting 4.3 rating from the customer.



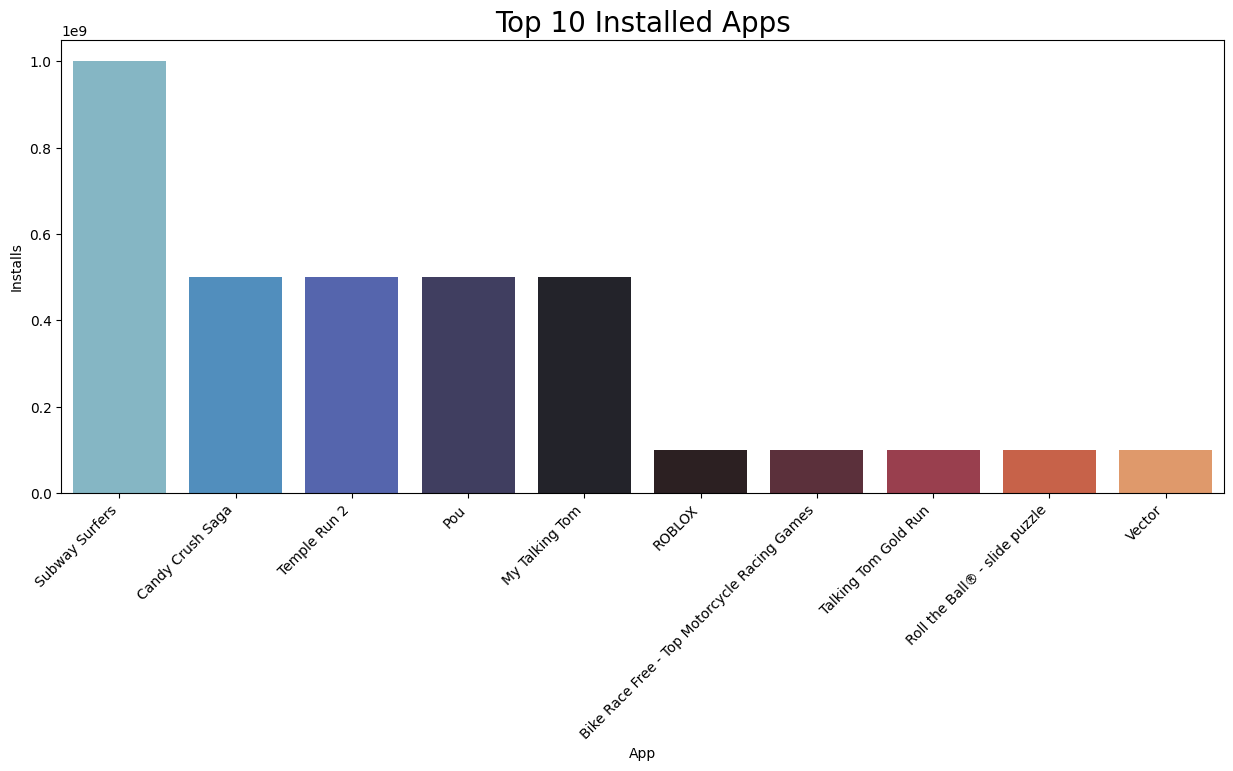
1. **Observation-6: (Most number of installs App's in Google play store):**

In this presentation we visualize the most number of install apps in play store . We found Game, communication and tools app installed most in the google play store by the customer.



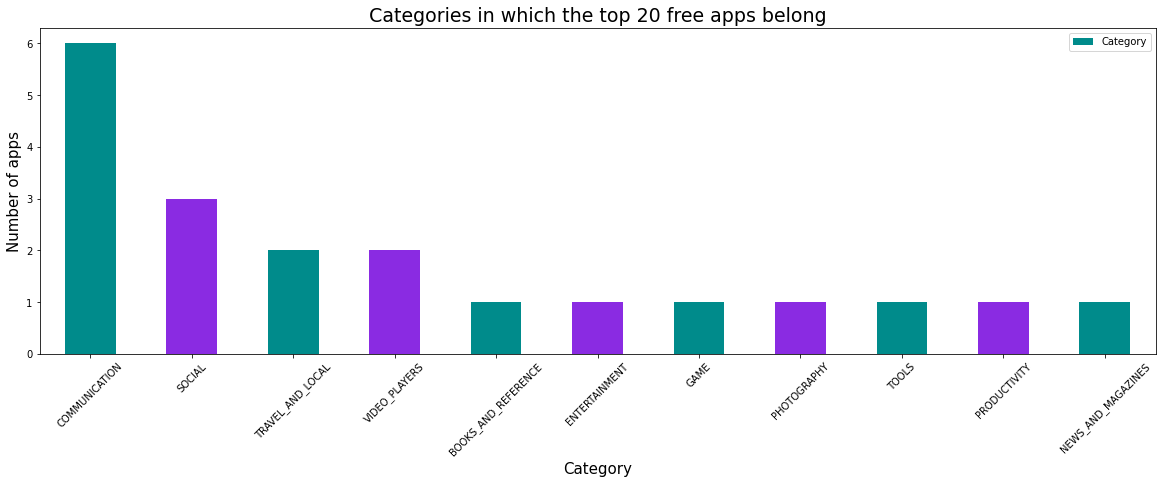
1. **Observation-7: ( What are the Top 10 installed apps):**

In this visualization we found the top 10 installed apps from google play store. Subway surfers is the number one installed apps in the play store. After that candy crush saga & Temple run-2 is the second and third installed apps from the google play store.



1. **Observation-8: ( Top Free apps):**

In this presentation we found that categories wise top 20 free apps in the play store. Communication category apps are mostly free in the google play store.



**6. Conclusion:**

After the completion of project. I have learned and got exposure to different tools and techniques in data analysis. I was able to complete the project successfully with the help of tools like Python, Pandas, Matplotlib, NumPy, Seaborn, and Plotly. Also, I learned different techniques like Data Cleaning, Data Preparation, Data Exploration and visualization, and Data Interpretation.