Analysing Google Apps Store dataset in terms of App downloads and Rating

Soumita Paul

Batch: Full Stack Data Analytics

Document Version Control

Date Issue	Version	Description	Author
14-10-2022	1.0	Internship Report on Analysing Google Apps Store Dataset	Soumita Paul

Approval Status

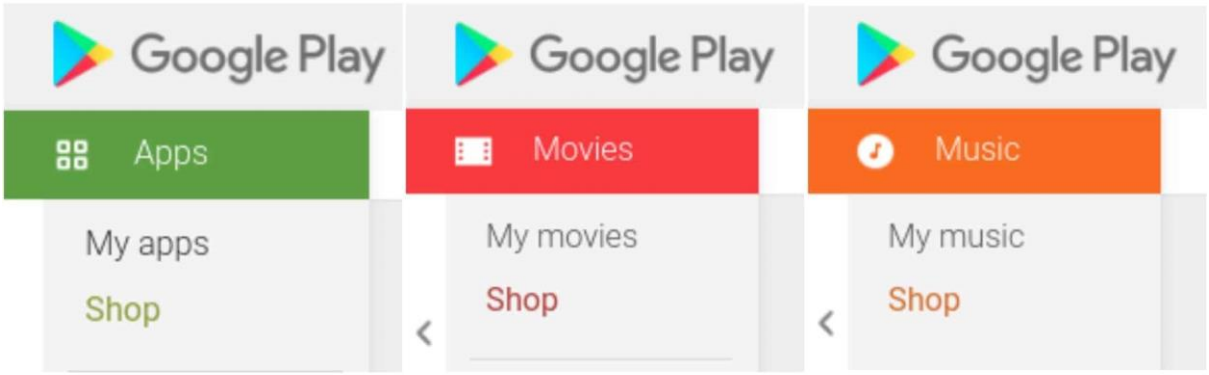
Version	Review Date	Reviewed By	Approved By	Comments

Contents

ABSTRACT

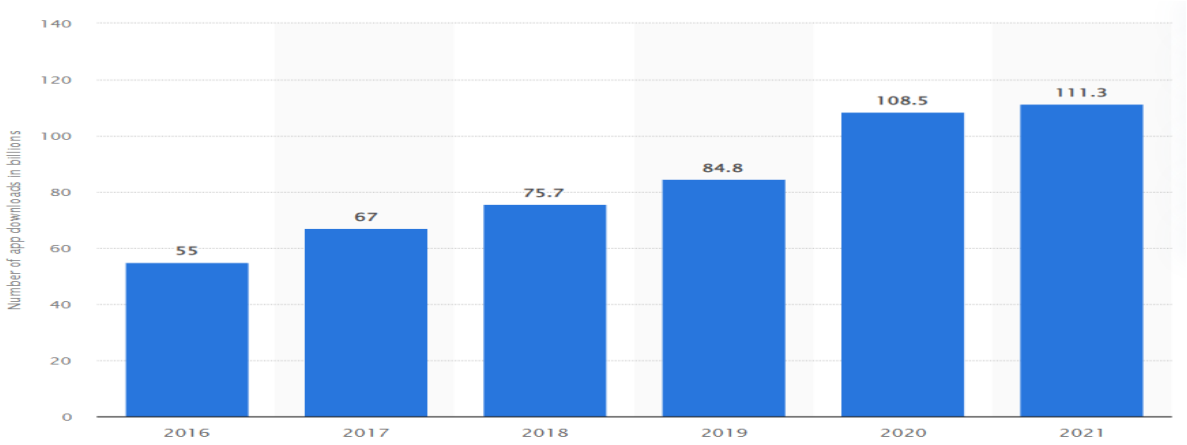
What is Google Play Store? Well, from the mobile standpoint, it's an app, though it's actually a platform. So, it's essentially far more than that, which is something we'll get down to in a minute.

It's Google's platform for offering various digital content to its consumers. Contrary to what some people may think, the Google Play Store is not just an app store, not at all. You will find all sorts of content available here. The Google Play Store is home to music, movies, books, and games… in addition to your regular Android applications.



Many people are using the Play Store to get access to the newest music in the market, or get a book to read. Movies are still quite pricey, at least in the vast majority of regions, but that is an option as well. From this point on, the Google Play Store library can only grow further, and who knows, Google may even add TV shows to the mix at some point.

One of the main reasons for its popularity is that The Google Play Store had over 82 billion app downloads in 2016 and reached over 3.5 million apps published in 2017.



In 2021, Google Play users worldwide downloaded 111.3 billion mobile apps, up from 76 billion apps in 2018. Generating significantly more downloads than it's iOS counterpart, the Apple App Store.

PROBLEM STATEMENT

Technology is the increasing need nowadays and used everywhere. One of the features of Technology is android. Which we all use in our daily life. Android is a mobile operating system based on a modified version of the Linux kernel and other opensource software, designed primarily for touchscreen mobile devices such as smartphones and tablets.

OBJECTIVES

The objective of the project is to perform data visualization techniques to understand the insight of the data. This project aims apply various Business Intelligence tools such as Tableau or Power BI to get a visual understanding of the data.

- To collect data and details of applications like apps rating from Google Play Store.
- To develop a data analytics system to gain insight and trends of android applications on Google Play.
- To extract user knowledge hidden behind the data.

TOOLS USED

The different types of tools used to create the data visualization are

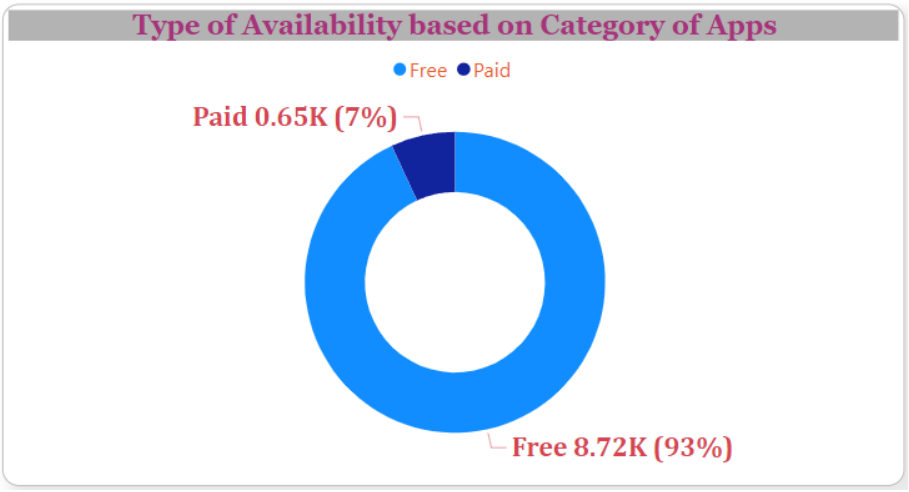
- Microsoft Excel
- Microsoft Power BI

DESCRIPTION OF THE DASHBOARD CONTENT

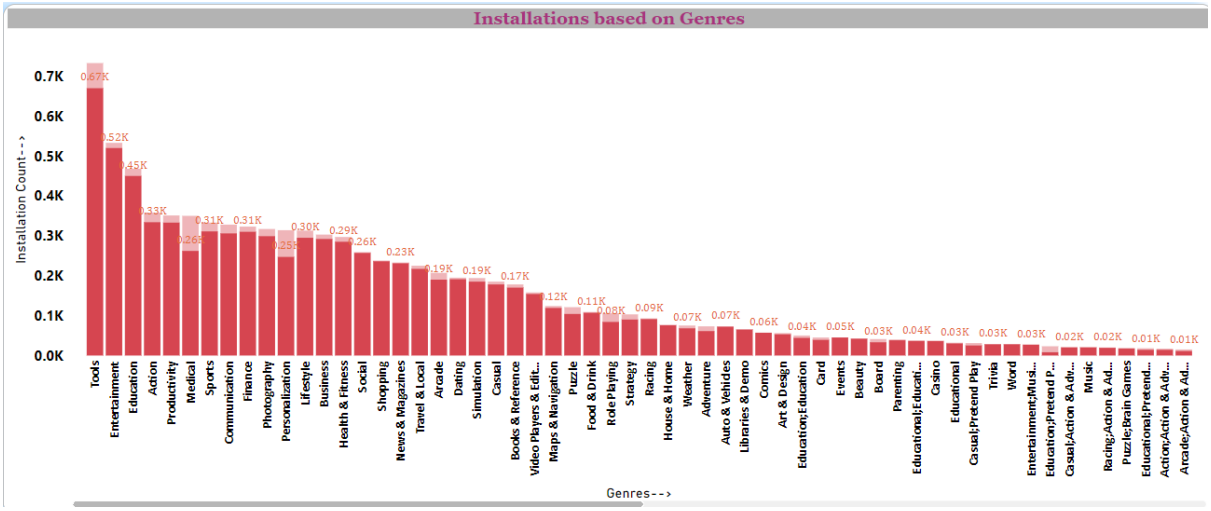
- **Which type of availability (Free or Paid) has more downloads?**

The larger the section in the round of the donut chart will determine which availability type based on category apps is more dominant, and vice versa. As shown in the donut chart, there are 8.72K from overall analysed apps are free type and only 0.65K are paid-types.

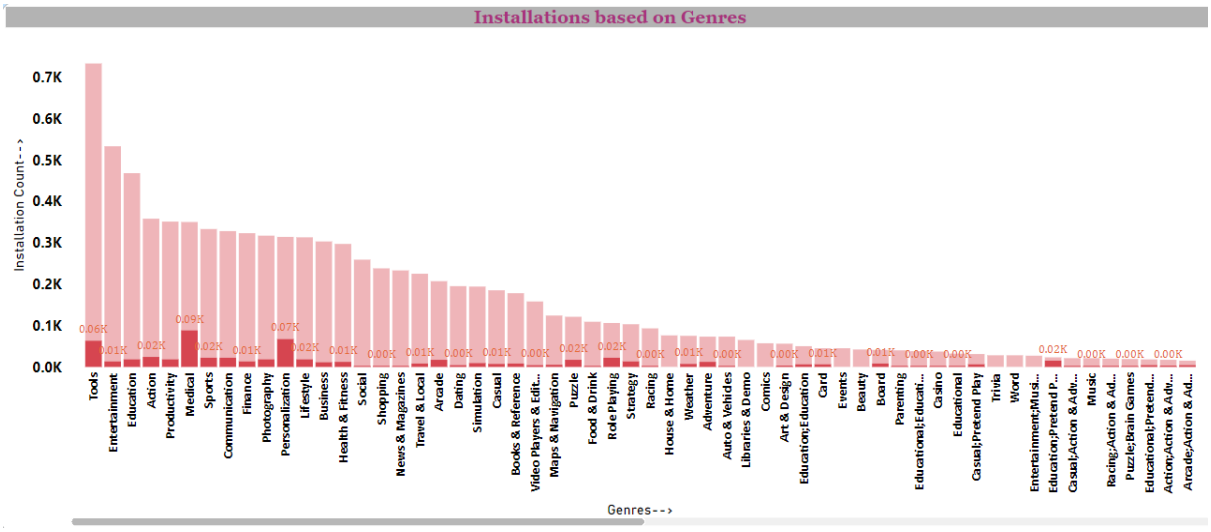
The bar chart shows that when the count of installations when the apps are free and when they are paid. Free type apps generally receive more downloads than paid types because they are expected to cost nothing to the user. Leading on from this, free-type apps make it easier to get and access from tons of more users. It is proven in the bar chart as we can see the installation number of free apps are much higher than those in paid- type.



Free vs Paid Apps Availability



Installation of Apps through Bar chart when the apps are free

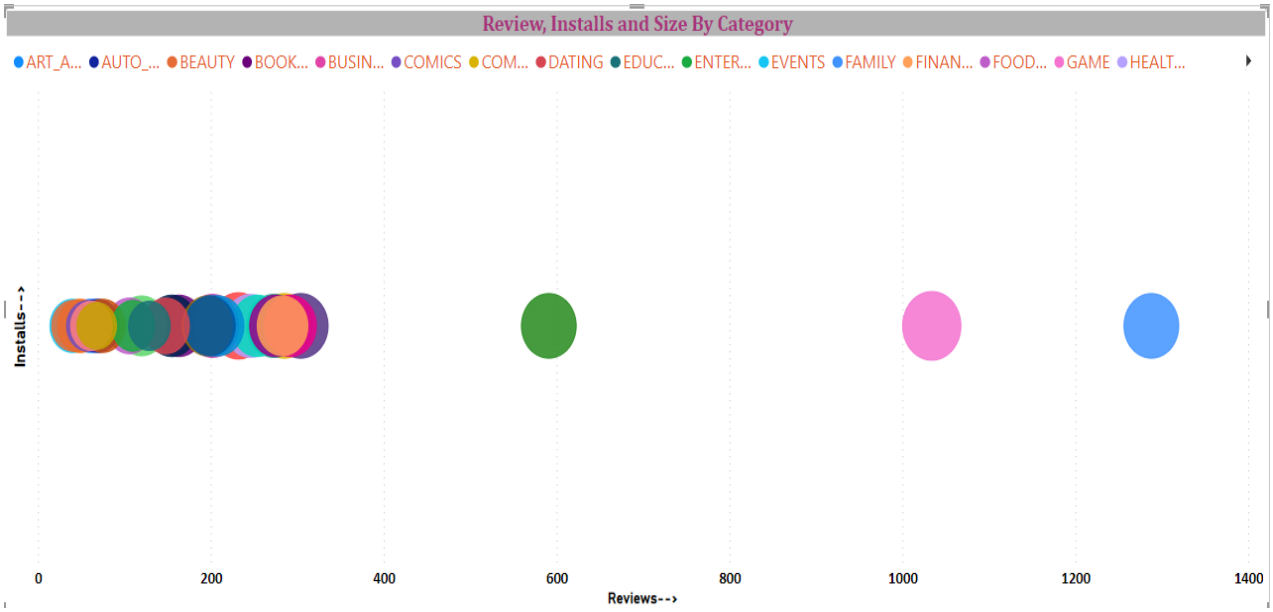


Installation of apps through Bar chart when the apps are paid

- **What is the popular category based on the Number of installs and reviews?**

The bigger the number of installations, the more reviews were left on that category of app. The size of the dot inside the scatter plot also plays a vital role as, the bigger the dot the bigger the size of apps.

From the scatter plot chart, it can be observed that the blue circle depicts the category “Family” with a greater number of reviews and installs, whereas the “Game” category comes in second. This concludes that size doesn’t have a direct relationship with number of installations and reviews as size does not matter but installations and reviews are parallel as one increases, the other one will increase as well.



Scatter Plot: Reviews vs installation count

- **Which Android version has more users?**

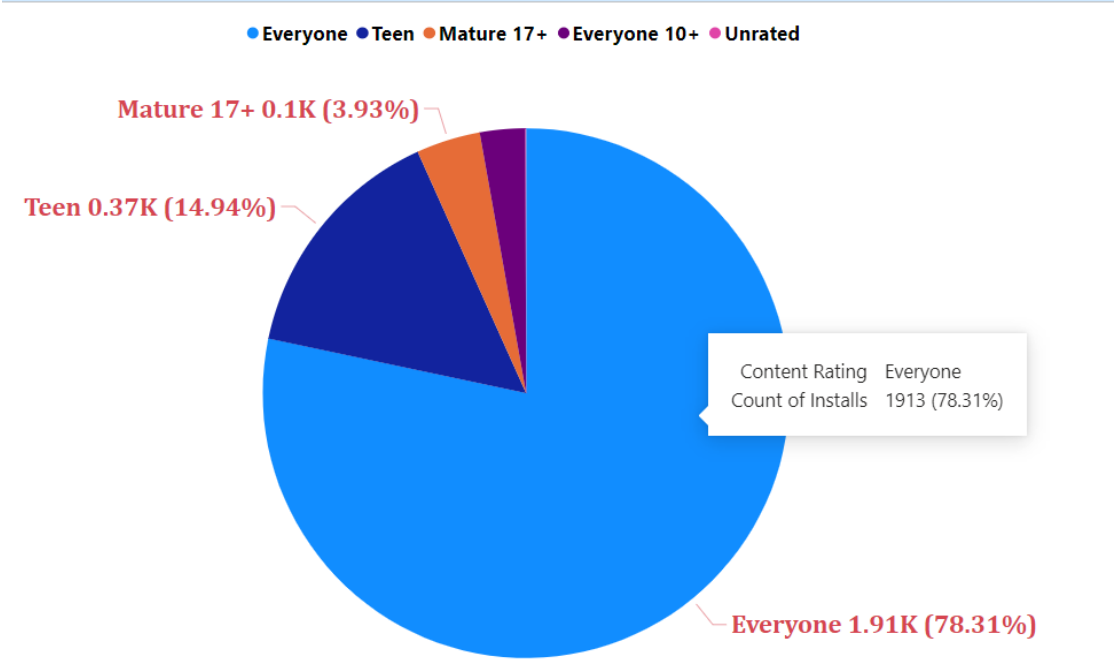
The size of the rectangle within the tree map is key to determine which android version has more users. The bigger the size of the rectangle shows that the numbers of users for that android version that install applications is many and vice versa for the smaller size of the rectangle.

The most installed android version of application is that of 4.1 and up and count of installs are 2059. The second in number is “Varies with device” and so on.

Tree Map

- **What is the most popular age group on content rating?**

The pie chart shows the most popular age group which has highest number of installations and reviews is “Everyone” where the installs are 1.91K Approx.



THE FINAL DASHBOARD

The power bi dashboard depicts the Google play dataset where we can see and visualize with the help of the card filters the total number of Genres, Reviews, Installs and total number of Apps available in Google play store.

The Gauge Indicates the average rating with respect to the category and content rating of the apps. Therefore, with the help of this project, one can understand the importance of these apps and how much the total population is using it based on the age category. However, the apps which are free are widely used all over than the paid apps.