Data Visualization

DSCI 5360

Analyzing Google Play store Apps data to understand the Android Market

Sritha Darbha

Sravanthi Peri

Executive Summary:

The aim of our project is to analyze the dataset of 10K Play store apps for analyzing the Android Market. The data from Play Store apps has a lot of potential for helping app developers succeed. Developers may draw actionable insights to work on and capture the Android market.

While many public datasets (on Kaggle and elsewhere) give Apple App Store data, there are few Google Play Store app datasets available anywhere on the web. Further investigation revealed that the iTunes App Store website has a beautifully indexed appendix-like structure to facilitate web scraping. Google Play Store, on the other hand, employs advanced modern-day techniques (such as dynamic page loading) that make scraping more difficult. This dataset from Play Store apps has a lot of potential for helping app developers to draw actionable insights to work on and capture the Android market.

Analysing the Dataset:

The main aim of this dataset is to help developers draw insights to work on and capture the Android market.

The dataset we chose is a second-hand dataset from Kaggle.com. It is taken from the following weblink:

<https://www.kaggle.com/lava18/google-play-store-apps>.

We have edited and renamed some of the variables of the dataset to best suit for the Project. The dataset file has been attached separately with the Project submission.

We have used Tableau 2021.2 to visualize and analyze the dataset.

This dataset has values for category, rating, size, and more. The dataset has 10805 entities and 13 variables. The variables are:

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Variable** | **Description** |
| 1 | App Name | Name of the Application |
| 2 | Category | Category of the Application |
| 3 | App Rating | Rating of the Application |
| 4 | Reviews | Number of Reviews for the Application |
| 5 | Size | Memory size of the Application |
| 6 | Installs | Number of times the Application has been installed |
| 7 | Type | Is the Application Paid or Free |
| 8 | Price | Price of the Application if Paid |
| 9 | Content Rating | Age Appropriate for this content |
| 10 | Genres | Genre of the Application |
| 11 | Last Updated | Date the Application was last updated |
| 12 | Current Version | Current version of the Application |
| 13 | Android Version | Android version required to download the Application |

We have created a total of 11 sheets, 3 dashboards and 1 story in Tableau to explore our dataset.

Issue, Problems and Solutions:

The aim of our project is to analyze the dataset of Play store apps for helping developers to draw actionable insights to work on and capture the Android market.

Firstly, we want to see the different Category and Genres of the Apps will affect the number of Apps and App Ratings for each one of them. We have created four sheets for the same.

We have created first sheet with the number of Apps per each Genre. We found the highest number of Apps for Genres Tools, Entertainment, Education and Business.

Chart

Description automatically generated

We then created a second sheet with the number of Apps per each Category. We found the apps under Categories Family and Game have the highest count. This was followed by category Tools.

Chart, bar chart

Description automatically generated

We have created a third sheet comparing App ratings for each Genre. We found the Genre Tools, Education and Entertainment have the highest App Ratings.

Chart, bar chart

Description automatically generated

We have created a fourth sheet for App ratings for each Category. We found categories Family, Game and Tools have the highest App ratings.

Chart, bar chart

Description automatically generated

We have now created a Dashboard with these four sheets.

Chart

Description automatically generated with low confidence

From this, we took insights that the Genre and Category of Apps with high App count have high ratings.

Next, we want to explore how the number of reviews are affected on different factors. We have created three sheets for the same.

We created the first sheet comparing the Android Version based on the Count of Reviews. We found with the Android Version 4.1 and up have the highest count of Reviews.Chart, histogram

Description automatically generated

Next, we compared between the Last Version Updated year for an App with the Reviews. We found the year 2018 have the highest count of Reviews.

Graphical user interface, application

Description automatically generated

We now compare the number of Reviews with the number of Installs. Here, we see Apps with over million installs have highest count of reviews.

Chart

Description automatically generated

We now created a Dashboard with these three sheets.

Graphical user interface

Description automatically generated with low confidence

From this, we observed Apps which are on phones 4.1 and higher and with at least a million installs and latest updates have high count of reviews, which means have high engagement.

Next, we want to compare over App rating for different factors. We have created three sheets for the same.

Firstly, we compared App ratings over with Content ratings and found the Content for Everyone with highest sum of App ratings.

Graphical user interface

Description automatically generated

We then compared for App ratings with the App Type. We found the Free Apps with highest App ratings compared with Paid Apps.

Chart, pie chart

Description automatically generated

We now compared App ratings with the number of Installs. We found Apps with at least a million downloads to have high App ratings.

Chart, treemap chart

Description automatically generated

We have now created a Dashboard with these three sheets.

Chart

Description automatically generated

We took insights from this that Apps with Content type suitable for Everyone and are Free with at least a million downloads have high ratings.

We now wanted to create a sheet to check what Category of Apps have highest Paid Apps. We found Finance, Lifestyle and Family Apps to be the highest ones.

Chart, treemap chart

Description automatically generated

We have created a Story with all our Analysis together at the same place.

Chart, waterfall chart

Description automatically generated

Conclusion:

Thus, we visualized our dataset using Tableau to better understand the different factors affecting in Installing, Rating and Reviews of the Google Playstore Apps.

Roles:

We both have done the project together : From searching for the dataset to writing the reports and analyzing on tableau.

References:

1. <https://www.kaggle.com/lava18/google-play-store-apps>.