Google Playstore Apps Review Analysis

Team members

Utsav Ponda, Sivaramaguhan S, Vikramaditya Sah, Suhas Jagadish
Data science trainees,
AlmaBetter, Bangalore

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. Since most Play Store applications are free, the income model is very obscure and inaccessible regarding how the in-application buys, adverts and memberships add to the achievement of an application

Our experiment can help to discover the relationships among various attributes present in my dataset such as which application is free or paid, about the user reviews, rating of the application.

Keywords: exploratory data analysis, python, visualisation, exploration

1. Problem Statement

In today's scenario we can see that mobile apps playing an important role in any individual's life. It has been seen that the development of the mobile application advertise has an incredible effect on advanced innovation. Having said that, with the consistently developing versatile application showcase there is additionally an eminent ascent of portable application designers inevitably bringing about high as can be income by the worldwide portable application industry.

1.1 Google Play store Dataset

This dataset contains each App play store data in the following order:

1. Apps Name.
2. Category.
3. Rating.
4. Review.
5. Size.
6. Installs.
7. Type.
8. Price.
9. Content Rating.
10 Genres
11 Last Updated
12 Current Ver
13 Android Ver

The dataset consists of Google play store application, which is the world's largest community for data scientists to explore, analyze and share data.
This data set is for Webscratched information of 10k Play Store applications to analyze the market of android. Here it is
a downloaded dataset which a user can use to examine the Android market of different use of classifications music, camera etc. With the

assistance of this, client can predict see whether any given application will get lower or higher rating level. This dataset can be more over used for future references
for the proposal of any application. Additionally, the disconnected dataset is picked so as to choose the estimate
exactly as online data gets revived all around a great part of the time. With the assistance of this dataset I will examine
various qualities like rating, free or paid and so forth utilizing Hive and after that I will likewise do forecast of various traits
like client surveys, rating etc.

1.2 User Reviews Dataset

This dataset contains each user review per app in the following order:

1. Apps Name.
2. Translated Review.
3. Sentiment
4. Sentiment_Polarity
5. Sentiment_Subjectivity

2. Introduction

Mobile applications are one of the fastest growing segments of downloadable software application markets. Out of all of the markets we choose Google Play store due to its increasing popularity and recent fast growth One of the main reasons for this popularity is the fact of the apps are free of cost.

Developers and users play key roles in determining the impact that market interactions have on future technology.
However, the lack of a clear understanding of the inner working and dynamic of popular app markets impacts both the developers and users, here we seek to shed light on the dynamics of the Google Play Store and how we can use different features from this data set for prediction purposes. In this capstone project, we will provide a longitudinal study of Google Play app metadata which will give u unique information that is not available through the standard approach of capturing a single app snapshot.

we try to do data cleaning on the data set to reduce the error percentage. And also analyse the data set using different plots and re-move the stuff not needed from the data set. Our Analysis is divided into three phases: data extraction, data cleaning, data visualization.

3. Steps involved:

- **Exploratory Data Analysis**
- **Null values Treatment**
- **Data Cleaning**
- **Key Factor for Engaging and Successful apps.**
  To know the practical insight, apps must qualify above threshold benchmark based on below factors:

Installs: No. of Installs of individual app should be greater than mean of installs of entire apps in data set.

Reviews: No. of Reviews of individual app should be greater than mean of installs of entire apps in data set.

Most Engaging Apps by Categories : The apps which have more reviews, have the higher user interaction and can be termed as most engaging so we have taken percentage ratio of review to installs will give most engaging apps by categories.
We observed that categories like 'GAME', 'Family', 'Health and Fitness', 'Tools' are some the prominent categories in the data set.
.

8. Conclusion:

Finally concluding whole project Starting with loading the data so far we have done EDA , null values treatment, data cleaning, data visualisation using various plots and determining key factors which led to most engaging and successful apps categories.

References-

1. MachineLearningMastery
2. GeeksforGeeks
3. Analytics Vidhya