Play Store App Review Analysis

By-Dipanshu Kumar,

Data Science Trainee,

AlmaBetter, Bangalore

**Abstract:**

Apps have reshaped nearly every aspect of our lives, from ordering food, making a grocery list, checking account balances, and communicating with others. For everything, there's literally an app. Our everyday activities have been improved and simplified by technology, thereby making things easier.

But with the growing Industrial revolution in the form of Digitalization has boosted the development of more and more application

So, we’ll use some Data sets (Specifically Play store data and user review csv files) to analyze the key factors behind the growth success or failure of an App.

**Problem Statement:**

While innovating, many companies fall into errors that can lead to failure. We’ll try to highlight which are the most common and how they can be solved.

Mobile applications have been one of the main drivers of the technological development of companies, both in the B2B sector and B2C. The reception of mobile technology is the result of the recognition of the benefits that mobile applications provide to companies, which can certainly be used by companies of any size.

There are many factors that play their role in the success of a mobile application. An app can be downloaded many times, but, in turn, register a very large number of uninstallations. One of the most common factors that cause this to happen lies in the application’s interface. Among the errors in the development of apps, is in the design, in the interface. It will be the first thing we have to take care of. But there are many others.

However, not everything is rosy. Among so many apps only a third of mobile applications are established in the market. The hard truth is one out of every three developments is successful, this implies that companies must be aware of the risks, but more importantly of the way in which these risks can be identified and diminished. This doc. will allow you to know the most common problems in the application of mobile technologies and presents solutions that help reduce the risk of companies seeking to develop mobile applications.

**The most common problems in the development of mobile applications are as follows:**

**1. Lack of vision about the use of mobile technologies**

**2. Ignorance of mobile platforms in the market**

**3. Lack of knowledge or training of the members of the company on the use of mobile technology**

**4. The notifications**

**5. Complicated user interface**

**6. Extensive use of tutorials**

**7. Endless forms**

**And the count goes on.**

**Introduction:**

Ratings and reviews are important. They provide valuable quantitative and qualitative feedback on your users’ reported experience of your app or game, and the broader service that you offer. That’s why they’re one of the signals people use when deciding what to download.

We’ve heard from both Play Store users and developers that ratings and reviews could be more helpful. This is especially true when ratings from one area unfairly impact another — like when a bug that only impacted a single country negatively affects the app’s rating everywhere; or when positive improvements in a tablet experience are overlooked because of the number of users on phones.

Both play an important role in marketplaces. Not only do they inform potential users, but developers can also use that feedback to improve the app. However, it’s not uncommon for suspicious reviews to flood the reviews section, which doesn’t help anyone.

They help people decide which apps to download and they are taken into consideration for featuring and placement on Play Store. But because the app experience can vary depending on the user’s region and device type, aggregate ratings don’t always tell the whole story.

The market has increased to over 3.5 million Apps and around 3000+ apps are being added per day as per a Google survey report. Thus, the market, in turn, led to around 5 billion users downloading all over the world. It is evident that 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 dynamics of popular app markets impacts both the developers and users. We try to analyzes the actionable insights for the developers to work on and capture the Android market and also analyzes factors for app engagement and success with classifier models used for finding the user engagement, success parameters, and the complete Data visualization.

While working on the project we followed the following steps:

* Getting the data
* Exploratory Data Analysis
* Breakdown of dataset
* Examining the values
* Data cleaning
* Modelling and Data visualization using Charts and Plots
* Getting the results.

**Important steps involved:**

* **Exploratory Data Analysis:**

After loading the dataset, we critically performed initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions to represent the summary and graphical representations.

It gave us a briefness regarding how to move forward and what is to be done.

* **Datasets Description:**

Play Store Data.csv: It consists of name of applications and a total of 13 features describing the app.

User Reviews.csv: This file consists of 3 important features on how the use feel regarding the app

On a general tone both the collection contains related, discrete items of related data that may be accessed individually or in combination or managed as a whole entity.

Also gave us an approach on which dataset to start first.

* **Data Cleaning:**

Incorrect, inconsistent data can lead to false conclusion and misdirects the scales. So, detecting and correcting of inaccurate records is necessary to maintain the flow.

When we were using the data sources, there came many opportunities for data to be duplicated or mis leaded. 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 so to know what we are doing in the right way every time.

This process also involved conversion of datatypes, using the null values, filling those null values to get a manageable dataset.

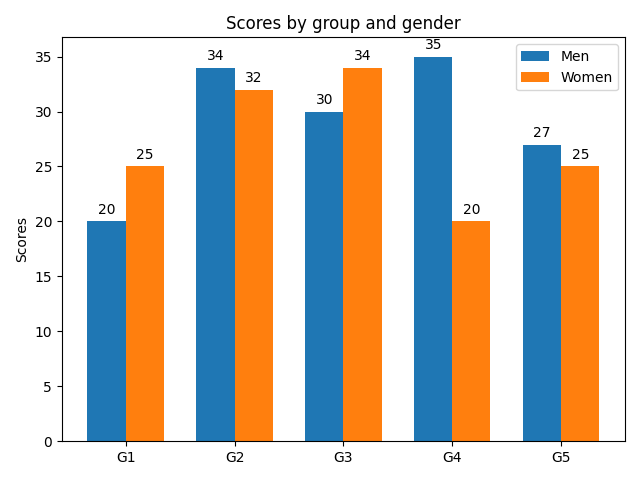
* **Modelling and Data Visualization using charts and plots:**

To understand the complex datasets, we have made the utilization of the following charts or plots (here we are highlighting the technical aspects of Plots used).

**Bar chart:**

Bar graphs are the pictorial representation of data (generally grouped), in the form of vertical or horizontal rectangular bars, where the length of bars is proportional to the measure of data.

A bar graph shows comparisons among discrete categories. One axis of the chart shows the specific categories being compared, and the other axis represents a measured value. Some bar graphs present bars clustered in groups of more than one, showing the values of more than one measured variable.

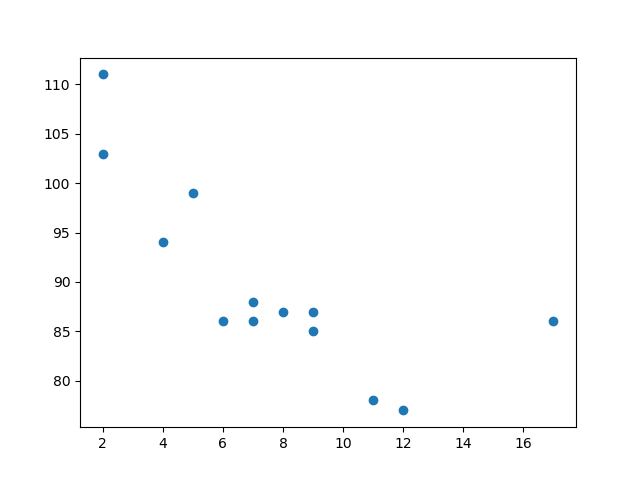


Advantages Bar graph: summarises the large set of data in simple visual form. It displays each category of data in the frequency distribution. It clarifies the trend of data better than the table. It helps in estimating the key values at a glance

**Scatter Diagram:**

A two-dimensional graph in rectangular coordinates consisting of points whose coordinates represent values of two variables under study.

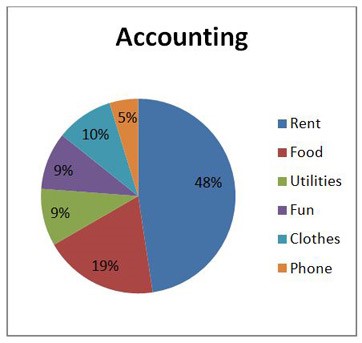
A scatter plot is a diagram where each value in the data set is represented by a dot. The Matplotlib module has a method for drawing scatter plots, it needs two arrays of the same length, one for the values of the x-axis, and one for the values of the y-axis



Advantages of scatter diagram is a show the relationship between two variables. It is the best method to show you a non-linear pattern. The range of data flow, like the maximum and minimum value, can be determined. Patterns are easy to observe.

**Pie Chart**:

A circular chart cut by radii into segments illustrating relative magnitudes or frequencies

A pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice (and consequently its central angle and area) is proportional to the quantity it represents. While it is named for its resemblance to a pie which has been sliced, there are variations on the way it can represented. 

Advantages of pie chart is representing data visually as a fractional part of a whole, which can be an effective communication tool for the even uninformed audience. It enables the audience to see a data comparison at a glance to make an immediate analysis or to understand information quickly.

**Correlation Heat Maps:**

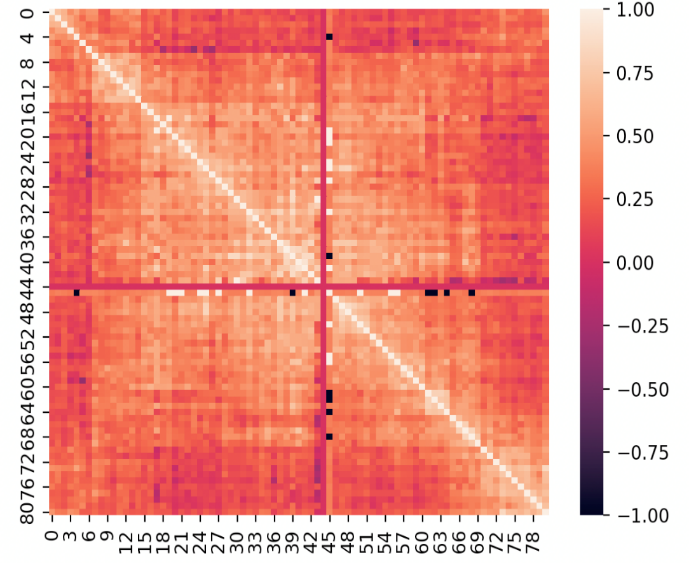
Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate in relation to each other.

A Positive correlation indicates the extent to which those variables increase or decrease in parallel A Negative correlation indicates the extent to which one variable increases as the other decreases. A No correlation indicate there is no relationship between the two variables. In other words, as one variable moves one way, the other moved in another unrelated direction.

Advantages of correlation is neither variable goes through a manipulative process.Two different data collection methods are available with correlational research.The results from correlational research are more applicable.It offers a beneficial starting position for research.Researchers can determine the direction and strength of each relationship.

A heat map is a data visualization technique that’s a graphical representation of data in two dimensions, using colours to demonstrate different factors.

Practical data science using python in heatmap. Where heatmap contains values representing various shades of the same colour for each value to be plotted. Usually, the darker shades of the chart represent higher values than the lighter shade. For a very different value a completely different colour can also be used.

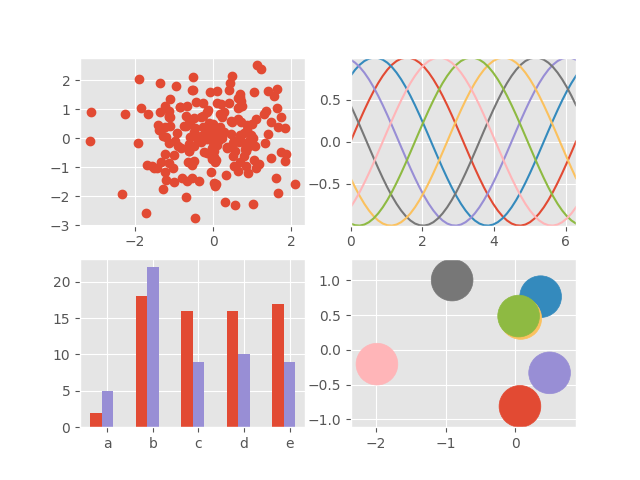


Advantages of heat map help you to understand how users and customers interact with your website, giving insight into things like where they're looking and how far down, they're scrolling. This information can help you set up your site in a more user-friendly way.

**Plt.Style.Use** (ggplot):

A plot style is an object property, similar to line type and colour. A plot style can be assigned to an object or a layer. A plot style controls an object's plotted properties, including Colour, Size etc.

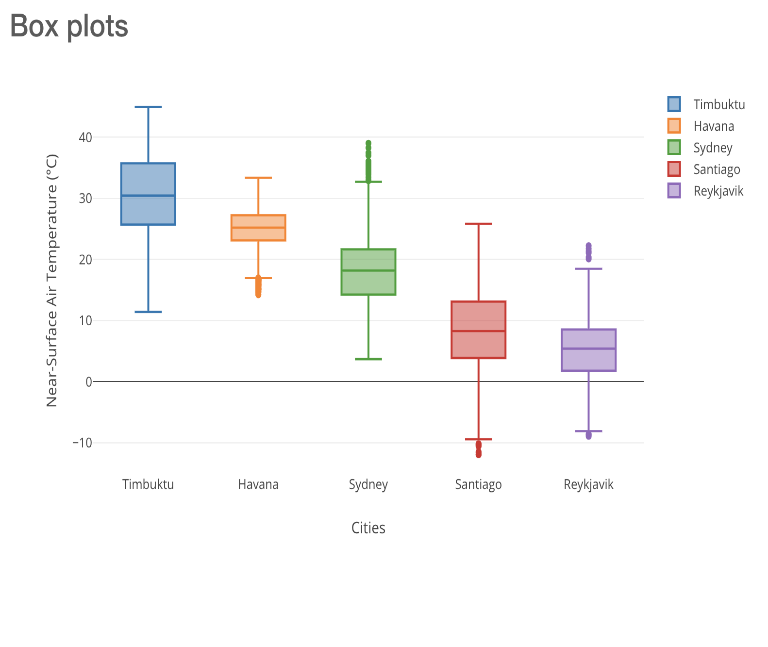
GG Plot is a plotting package that provides helpful commands to create complex plots from data in a data frame. It provides a more programmatic interface for specifying what variables to plot, how they are displayed, and general visual properties.



Advantages of style plot is consistent underlying grammar of graphics plot specification at a high level of abstraction. very flexible. theme system for polishing plot appearance (more on this later) mature and complete graphics system. many users, active mailing list.

**Box Plot:**

A box plot is a chart that shows data from a five-number summary including one of the measures of central tendency. It does not show the distribution in particular as much as a stem and leaf plot or histogram does. But it is primarily used to indicate a distribution is skewed or not and if there are potential unusual observations (also called outliers) present in the data set. Boxplots are also very beneficial when large numbers of data sets are involved or compared.



A box plot is a highly visually effective way of viewing a clear summary of one or more sets of data. It is particularly useful for quickly summarizing and comparing different sets of results from different experiments.

**Conclusion:**

In this study the models were proposed and tested intended to explain the relationship between the apps, which provides knowledge management practices to determine apps performances.

The major purpose of this study was to evaluate the impact of the apps on organizational performance.

To test the mediating relationship, a number of steps were included primarily speaking Exploratory Data Analysis, Data Cleaning, Data Visualization.

Significant and positive results were found which may be a great asset on transforming how organization works and makes their business choices.

**Acknowledgement:**

This project was completed by Dipanshu Kumar. I am extremely grateful to the celebrated authors whose precious works have been consulted and referred in this project work. I also wish to convey my appreciation to our peers who provided encouragement and timely support in the hour of need.

**References:**

1. medium.com
2. Geeksforgeeks
3. Jovian
4. Matplotlib.org
5. Sciencing.com