

PLAY STORE APP REVIEW ANALYSIS

Almabetter, Bengaluru

<u>Team Members</u>	
Varsha Rani	Vivek Chandrakant Pawar
Rabista Parween	Tushar Gaikwad

1. Problem Statement:

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.

2. Introduction:

Here we are exploring and analyzing the data to discover key factors responsible for app engagement and success.

We have used the following Libraries :

- Pandas
- Numpy
- Matplotlib
- Seaborn

3. Objective:

The objective of this project is to deliver insights to understand customer demands better and thus help developers to popularize the product.

4. Steps Involved:

Exploratory Data Analysis

- **Univariate Analysis**

Univariate analysis is perhaps the simplest form of statistical analysis. Like other forms of statistics, it can be inferential or descriptive. The key fact is that only one variable is involved.

- **Skewness of Data**

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive, zero, negative, or undefined.

- **Bivariate Analysis**

Bivariate analysis is one of the simplest forms of quantitative analysis. It involves the analysis of two variables, for the purpose of determining the empirical relationship between them.

- **Correlation of Data**

Correlation is used to find the relationship between two variables which is important in real life because we can predict the value of one variable with the help of other variables, who is being correlated with it.

- **Categories vs Installs**

Maximum number of applications are being installed from the category game followed by communication category. So, you must bend your technologies, accordingly.

The most positively reviewed category have the highest installation rate.

- **Categories vs Reviews**

Reviews have the power to gain customer trust, and they encourage people to interact with the company. Customer interaction ultimately leads to improved profits for businesses. Gaming and Communication these two categories has highest percentage of reviews.

- **Content Ratings vs Installs**

The applications which have a content rating for everyone are being installed most than other. Play Store place a greater level of importance on ratings and reviews than ever before. Apps with higher ratings and reviews rank high in search. If an app ranks high then there's a better chance of it being found and downloaded.

Distribution

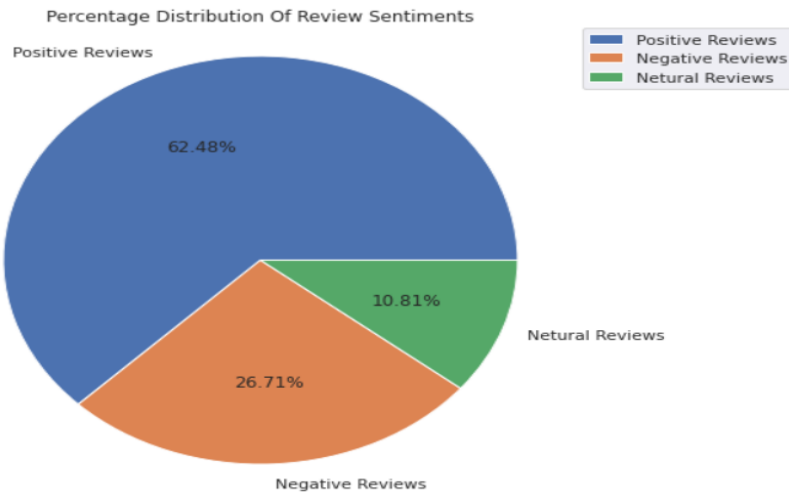
- **App Size Distribution**

This histogram of apps size distribution tell us about the optimum size range most liked by the user. The size of your application has an impact on how fast your app loads, how much memory it uses, and how much power it consumes.

- **Sentiment Subjectivity Distribution**

Sentiment is the emotion, feeling, opinion, or views held or expressed by users. Sentiment subjectivity is float number value whose range lies in between 0 to 1, where 0 is very objective and 1 is very subjective. Sentiment subjectivity determines the judgement of review writer's how happy, disappointed, frustrated they are with the service of the application.

- **Sentiment Distribution**



From the above pie chart, it can easily be understood that there is around 62 of user reviews sentiment is positive, around 27% of reviews sentiment is negative and the remaining around 11% of reviews sentiment is neutral. If some apps have a higher percentage of positive Reviews sentiments, then it is sure that the app is performing its intended work, and people are enjoying it, they may share the app with somebody thus increases the number of installations. So, need to keep an eye-tracking on the review sentiment it is what decides whether the app is going to feature on google play store. By featuring I mean visibility of apps when someone searches for a category. If the app is not visible in the top 10 or 12 apps range then there are fewer chances of the app being installed.

- **Sentiment Polarity Distribution**

Sentiment polarity is a float value ranging from negative one to positive one. i.e., range (-1, 1), (dtype = float) where -1 means negative statement 1 means positive statement.

- **Actual Number of Reviews for each kind of Sentiments**

<i>Sentiment</i>	<i>Sentiment Polarity</i>
Negative	10108
Neutral	4089
Positive	23642

- **Category vs Review vs Sentiment**

We always treat neutral reviews as insignificant, but it is not the case always, Neutral reviews are never neutral. A neutral review affects both negative and positive reviews and can also reinforce both. Statistically, neutral reviews are a good thing because without them, positive reviews will be overestimated, and negative ones will be underestimated. Neutrals give consumers a better understanding of a brand or a business.

- **Conclusion**

As per our EDA, an ideal application on the google play store should obey the following properties/characteristics:

1. **Reviews vs install:** We have experienced from the seaborn heatmap that reviews on the google play store are highly correlated with the rate of installation. Reviews are given by users as per their experience with the application. So reviews on the application should be examined properly to get to know the performance of the application, whether it is catering to the need of users, From review, we will get an idea on which aspect to work on.

2. Family Category has the most number of applications available on the google play store, and very few apps are available for the category beauty and parenting. We can see the people does not relay on apps for parenting and beauty. And for the comics the category is very saturated.
3. Game category has the highest percentage of installs that is 20.93% around 21% and second highest is communication have 19.48% percentage installs. So people's are more likely to connect with the people by using apps. And for entertainment purpose they are relaying on game as well.
4. Most space consuming category bar plot gives us the idea that which category has the most variety of applications available, and which has low. The category which consumes high space inside the google play store, it means it has more number of applications than other categories. So we have to take a significant decision to decide category for our future applications.