Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email, and Contribution:
1. Name: - ANIL YADAV Email ID:- anilyadavnoorgarh@gmail.com • Contributed To notebook and helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA Visualization • Contributed to the contents of PPT. • Contributed to Technical Documentation in the content of the problem statement goal of the project and steps involved.
Please paste the GitHub and drive link Repo link.
ANIL GitHub Link:- https://github.com/anilyadav67/playstoreappreview.git Drive link's:- https://colab.research.google.com/drive/1wRVhetHydfJ_4T0VQUX3tU9vOX9giU3w?usp=sharing
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words) Data science can be summarized into five steps: capture, maintain process, analyze, and communicate. The analysis of Google Play Store application aided to build most reliable and more interactive applications. This would be very useful for app developers to build an application focused on certain

discussed category in this analysis. This analysis will help in building the application with precise and accurate objectives.

- Android is the most popular operating system in the world, with over 2.5 billion active users spanning over 190 countries.
- Google Play was launched on March 6, 2012, bringing together Android Market marking a shift in Google's digital distribution strategy .
- Android is the dominant mobile operating system today more than 85% of all mobile devices running Google's OS. The Google Play Store is the largest and most popular Android app store.
- There are more than 3.04 million apps found on Google Play Store.
- 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. The main goal of our project is-

1) The purpose of our project is to gather and analyze detailed information on apps in the Google Play Store in order to provide insights on app features and the current state of the Android app market.
2) The Objective of the project to Explore and analyze the data to discover key factors responsible for app engagement and success.
3) The paly 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 has value for category , rating ,size ,and more .Another data set contains customers reviews of the android apps .Explore and analyse the data to discover key factors responsible for app

❖ Percentage Of Paid Apps v/s Free apps: - We Observed that 92.6% of Apps are free and only 7.4% of Apps are paid in Play store.

❖ Correlation Heatmap: -

	Sentiment Polarity	Sentiment Subjectivity
Sentiment Polarity	1.000000	0.261587
Sentiment Subjectivity	0.261587	1.000000

Note:- After Analyzing Above Heatmaps, We can infer that we don't have any sort of good Correlations between the different Columns of both the data sets.

Basic Pie chart to View Distribution of Apps Across various Categories:

- Rating is Highest for family and Game category apps.
- Rating is low for Events and Beauty category Apps.

Mean Install Of Each Category:-

- The Average installs is Higher For Communication category apps followed by social category apps.
- The Average Installation is Low for Category such as Beauty, Comics, dating, events, Medical and parenting.
- Even through the Average Rating is quite high for event category but the mean install is quite low.

Top 10 Genres of Apps: -

- Tools Genre have the Highest count followed by Entertainment.
- Lifestyle genre have the Lowest count followed by Communication

The Top 10 Expensive Apps in Play store?

- From the above graph we can interpreted that I AM RICH App is the most Expensive App in the play store. But this seems to be like a uncap. We need to further analyse If it is a junk app or not by deploying machine learning models in it

Top 10n Installed Apps: -

- From the above the graph, we can see that Communication Category Messenger – Text and Video chat for free, WhatsApp Messenger, Gmail has the highest installs. In the same way we by passing different category names to the function, we can get the top 10 installs apps.

Does the last update have an effect on rating?

- From above graph, we can conclude, the app gets more recent updates chance of getting a higher rating increase

The Distribution of rating per number of installs and type (paid or free)

Findings

- Rating per Install Category and Type

- Looks like rating is Distributed around 4.5. when it’s categorized per install category.

- Google Play store have very few paid apps.

IS there any co- relation between range reviews and price columns together

Finding:

- We can see that most of the Rating are between 4 around 4.5 and 5.
- As far as Reviews are concerned for most of the Apps Reviews are not given
- Also for Price ,most of the Apps are Free .

Histogram Of Subjectivity

Findings:

0- Objective (fact)
1- Subjective (opinion)

- It can be seen that maximum number of Sentiment Subjectivity lies between 0.4 to 0.7 from this we can conclude that maximum number of the users give reviews to the applications, according to their experience.

IS Sentiment subjectivity Proportional to Sentiment Polarity

Findings:

- From the above scatter plot it can be concluded that sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of cases. It shows a proportional behavior, when variance is too high or low

Percentage of review Sentiments

Findings:

- Positive reviews are 64.30%
- Negative reviews are 22.80%
- Neutral reviews are 12.90%

Co-Relation in Merge data frame

Finding:

- In this correlation matrix, there is not a significant relationship between Rating, Reviews, Size and installs with respect to the sentiment polarity and sentiment subjectivity.

Challenges Faced:

- Reading the data set and comprehending the problem statement.
- Examining the business KPIs for app Development and devising a solution to the problem
- Handling the error, Duplicate and Nan values in the dataset.
- Designing multiple visualizations to summarize the information in the data set and successfully communicate the results and trends to the reader.

Conclusion's:

- Rating is very important factor for Installation of apps as user mostly like to watch rating before using app. So developer should also work on updating their content as per the ratings

- From all above we analyze rating and installation are related, so owners should encourage to write review of their apps.
- Number of installations of free app is more compared to Paid apps, so Developer can also consider this point for high reach.
- They need to focus on updating their apps regularly as it attract more people .
- More Installation of App in game app, followed by a communication, Productivity so if one can thinking to developer an app will go for these.
- We can also see social media app is almost free, so keeping social media app free may be a great advantages over others.
- App category like events and beauty have not much reach ,so one can also kept this consideration.
- Most of the apps are downloaded by teens, so users of the other age category, must also be encourage to install the apps.
- From our above analysis we conclude that sentiments of the user keep varying, so owner need to keep updating their app regularly basis on user feedback.