**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: - Vinayaka R Kulkarni  * Email ID:- [vinayakark94@gmail.com](mailto:vinayakark94@gmail.com) * Contributed In notebook helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA Visualization * Contributed for the contents of PPT. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved.  1. Name: - Prajwal Singh  * Email ID: - [singhprajwal048@gmail.com](mailto:singhprajwal048@gmail.com) * Contributed in notebook for data cleaning, data manipulation, and in EDA Visualization and finalizing the conclusion. * Contributed in PPT by making sure all the points to be covered. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved.  1. Name: - Jayanthi Challagunda  * Email ID:- [jayanthi0c@gmail.com](mailto:jayanthi0c@gmail.com) * Contributed In notebook helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA Visualization * Contributed for the contents of PPT. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved.  1. Name: - Samarth Kushwaha  * Email ID: - [kushwahasa@rknec.edu](mailto:kushwahasa@rknec.edu) * Contributed in notebook for data cleaning, data manipulation, and in EDA Visualization and finalizing the conclusion. * Contributed in PPT by making sure all the points to be covered. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved. |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/PrajwalSingh048/Playstore-App-Review-Analysis> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| 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.  The data comes from the Google Play store dataset. There are various entries regarding each app in each row. This data set will be used for exploratory data analysis, which is an important step in the data science process because it prepares data for further modelling used in machine learning algorithms as well as securing very initial business decisions. To help us draw some preliminary conclusions about the chances of a newly launched app's success, we will structure the data, clean it, and present certain trends we observe.  Dataset :--   * Google play store dataset * User Review Dataset   Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset.  Google play store dataset has following columns -App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Genres, Last Updated, Current Ver and Android Ver.  User Review Dataset has columns -**Translated\_Review,Sentiment, Sentiment\_Polarity and Sentiment\_subjectivity.**  Our motive in whole project was to analyze the data and find out main components that affect users’ decision to download app.  EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. In this article, we will understand EDA with the help of an example dataset. We will use **Python** language (**Pandas** library) for this purpose . Free vs Paid We write a function play store info (), that will display 5 attributes about  the columns: Data type, Count of non-null values,  We start of by Finding the percentage of null values in each column by  using isnull().sum()function  We drop the null values from columns ‘Current Ver’, ‘Android Ver’ and  ‘Type’ from our dataset using the .notna() function of the pandas library.  We head on to find mean and median in the Rating column excluding the  NaN values using the median() and mean()  Moving ahead we plot an sns.distplot of the “Rating” Column. We  replaced all the “NAN” values with the median value of Rating column  using .fillna() function of the pandas library.  we can see that 92.2% apps are free, and 7.80% apps are paid on Google Play Store, so we can say that Most of the apps are free on Google Play Store.  The mean of the average ratings (excluding the NaN values) comes to be 4.2. The median of the entries (excluding the NaN values) in the 'Rating'  column comes to be 4.3. From this we can say that 50% of the apps.  an average rating of above 4.3, and the rest below 4.3. From the distplot visualizations, it is clear that the ratings are left  skewed.  We know that if the variable is skewed, the mean is biased by the values at the far end of the distribution. Therefore, the median is a better representation of the majority of the values in the variable  There are a total of 33 categories in the dataset from the above output we can come to a conclusion that in play store most of the apps are under FAMILY & GAME category and least are of EVENTS & BEAUTY Category  Sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of case, shows a proportional behavior, when variance is too high or low.  Sentiment Polarity is not highly correlated with Sentiment Subjectivity.  The dataset contains immense possibilities to improve business values and have a positive impact. It is not limited to the problem taken into consideration for this project. Many other interesting possibilities can be explored using this dataset.  From the results and process we have implemented, we can conclude that we have achieved this group project objective which is analyzing the Google Play Store apps and determine trends of the Google Play Store and both of our research questions.   |  | | --- |  |  | | --- |  |  |  |  |  | | --- | --- | --- | --- | |