**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **1. Trupti Patil**  Gmail – trupti3033@gmail.com   1. data 2. Data Wrangling   1. Play Store App Review csv  2. User Reviews csv  2. Data Cleaning and manipulation  1. Data exploration  2. Dealing with missing values  3. Duplicates values  4. Reshaping data  3. filtering data  4. Dealing with outliers  3. EDA – Exploratory Data Analysis  1. Highest price apps  2. Highest Content Rating  3. Category vs Price  4. Category vs Installs  5. Earnings  6. Sentiment Analysis  7. Most liked Genre  8. Expensive Apps  9. Free vs Paid Apps |
| **Please paste the GitHub Repo link.** |
| GitHub:-  Drive Link: - |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| Play store app is most popular and largely used Android app. which stores the millions of applications with different – different categories. For EDA we have provided two datasets play store analysis and user review data. In this project with the help of different python libraries like Pandas, numpy, Matplotlib and seaborn etc. I will analyze various apps present in play store.    As the first step we perform data wrangling over the raw data. Which deals with data exploration, dealing with missing values, reshaping data, filtering data, deal with categorical values.    In Category wise analysis, we include that what are the category of app which installed most of the times and what are the category of app which have highest price by using sum aggregate function and done visualization like category verse installs graph and category verses price graph.    Then we include the distribution of rating of apps. With the help of count plot, we give the info that maximum apps having how much rating.    Then we compare the app category wise with rating. which was gives us detail summary which app is having high rating and which is having low rating. Apps with a rating above indicator is game, auto and vehicles which indicates positive feedback for these categories    Then we also done analysis on what percentage of Apps are paid and free. By doing this we analyzed that most of the apps are free which helps most of the population to access the apps.    We focused on which Genre is most preferred by the population on play and visualized by bar plot and found that action and causal are the category which installed by most of the population.    Also, from user review dataset we included the sentiments analysis of the users regarding play store application that weather it is positive, neutral and negative. Then with the help of histogram we understand the Sentiment Subjectivity of user’s review.    At last, after doing this analysis we find correlation between the numerical data to know that what is the relationship between data and visualize in heatmaps. |