**Capstone Project Submission**

Play Store App Review Analysis

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. **Prashant Singh**([prashantsingh87884@gmail.com](mailto:prashantsingh87884@gmail.com))  * Upload dataset to Google colab and explain dataset to team members. * Analyze null values and filter them. * Data cleaning. * Correction of data types * Data wrangling * Data Visualizations * Technical Write up * PowerPoint presentation * Project summary  1. **Sachin**([sshokeen1996@gmail.com](mailto:sshokeen1996@gmail.com))  * Upload dataset to Google colab and explain dataset to team members. * Analyze null values and filter them. * Data cleaning. * Correction of data types * PowerPoint presentation * Project summary  1. **Jaanhvi Khandelwal(**[jvkhandelwal2709@gmail.com](mailto:jvkhandelwal2709@gmail.com)**)**  * Data wrangling * Data Visualizations * Technical Write up * PowerPoint presentation * Project summary  1. **Pankaj Pandit(**[**pankajpandit527@gmail.com**](mailto:pankajpandit527@gmail.com)**)**  * Upload dataset to Google colab and explain dataset to team members. * Analyze null values and filter them. * Data cleaning. * Correction of data types * Data wrangling |
| **Please paste the GitHub Repo link.** |
| Sachin Github Link:- [Sachin8700/Sachin-EDA-Capstone-Project-Play-Store- (github.com)](https://github.com/Sachin8700/Sachin-EDA-Capstone-Project-Play-Store-)  Prashant Github Link:-<https://github.com/Prashant87884/EDA-CAPSTONE-PROJECT.git> |
| Mobile apps are everywhere. They are easy to create and can be lucrative, because of these two factors, more and more apps are being developed. In this notebook, we will do a comprehensive analysis of the Android app market by comparing over ten thousand apps in Google play store across different categories  The Google play store app dataset consists of enormous data that can be used to create effective insights. There are various key factors that play a major role in the success & engagement from the user’s end. Our problem statement is quite inevitable in comparison with the present Google Play store App market. Upon doing several pieces of research it can be seen that every day around 3000+ apps is being added to the play store library. Therefore enormous datasets & variety of insights can be concluded for business improvements.  In this EDA project we were provided with two datasets   * Playstore.csv -> contains all the details of the applications of Google Play. There are 13 features that describe a given app. * User\_reviews.csv -> contains 100 reviews for each app, most helpful first. The text in each review has been pre-processed and attributed with three new features * Sentiments (Positive, Negative, Neutral) * Sentiment Polarity * Sentiment Subjectivity   **Approach:-**   * At first, we break down the datasets by importing necessary library classes, followed by checking unique values, converting the data types to similar objects, removing special characters as the analysis demands & making the entire dataset ready for analyzing & plotting actionable insights. * After importing libraries then we import our csv file using read.csv function in pandas library. * Then we explore our dataset using different function of pandas libraries like info, head and describe. * Upon exploring we found that 2 columns price and installs have unnecessary symbols intact with them so in first task of data cleaning we removed $ sign from price column and then remove + sign from the installs column. * Upon further exploring we found out that there is a wrong entry in the dataset so then we have removed that value from the dataset. * After examining null & missing values from the dataset we directly went deep into the visualization steps.   Some insights on which we worked are as follows:   * **Count of all the applications in each category of the Dataset.** * **Number of installs in all the apps genres of the Dataset.** * Heatmap visualization for finding out the correlation of different attributes of Dataset. * Visualization of Ratings using violin graph. * Visualization by bar graph of number of installs type wise according to category. * Distribution of subjectivity using a histogram. * Visualization of sentiments of each category using bar graph. * Does sentiment subjectivity proportional to sentiment polarity? * A pie chart representing percentage of review sentiments. |