**Capstone Project Submission**

| **Team Member’s Name, Email and Contribution:** | | |
| --- | --- | --- |
| **Name**  Shreya Agrawal  Shubhangani  Tiwari  Priyanshu Tiwari  Husain Bharmal  Vaibhav Lawate | **Email**  [agrshreya11@gmail.com](mailto:agrshreya11@gmail.com)  [tiwarishubhangani@gmail.com](mailto:tiwarishubhangani@gmail.com)  priyanshutiwari384@gmail.com  [hkbharmal72@gmail.com](mailto:%20hkbharmal72@gmail.com)  lawate.v17[@gmail.com](mailto:pranjalitete@gmail.com) | **Contribution**  Data Analysis, Data cleaning on play store app data, PowerPoint Presentation.  Data Visualization, Data Cleaning on user reviews data, PowerPoint Presentation.  Data Visualization on play store app, EDA on merged dataset.  Data Visualisation on user reviews data, EDA on merged dataset.  Data Analysis, EDA, Final touch to the PowerPoint Presentation. |
| **Github Repository:** | | |
| <https://github.com/Shreya-data/Play-Store-App-Reviews-Analysis> | | |
| **Project Summary:**  Playstore is an application for android users which allows the users to download millions of applications  for various purposes.  In this capstone project we have compared thousands of applications across various features.  We have been provided with 2 Dataset files – ‘Play\_store\_csv’ and ‘User\_Reviews”. One containing 13 features namely ‘App’, ‘Category’, ‘Ratings’, ‘Reviews’, ‘Types’, ‘Size’ , ‘Installs’ , ‘Genres’ , ‘Price’ , ‘Content\_Rating’ , ‘Last\_Updated’ , ‘Current\_Version’ and ‘Android\_Version’ and another file containing features namely ‘App’ , ‘Translated\_Review’ , ‘Sentiment’ , ‘Sentiment\_Polarity’ and ‘Sentiment\_Subjectivity’.  First we have performed Data Wrangling over the raw data in which we removed or replaced all the duplicates, NaN values and errors that are present in our data. Then we filtered it one by one. After this we changed the datatype of categorical columns which is countable into numerical.  After this we performed EDA, we plotted piechart, countplot, distplot, barplot for univariate and bivariate data, to get the useful insights:   * For categories, we observed that ‘Family’ , ‘Games’ and ‘Tools’ are the top three ones. * For genres, we observed that ‘Tools’ is on the top which followed by ‘Entertainment’. * For rating, we observed that the most common rating is around 4.3. * For type, we observed that there are approx. 92% free apps and 8% are paid. * For Content\_Rating, we observed that there are 81% apps are for everyone. * For Android\_Version, we observed that most of the apps required android version 4.0 and above. * For Sentiment, we observed that Positive, Negative and Neutral Sentiment are 64%, 23% and 13% respectively. * Further, we plot Pairplot and Heatmap for correlation of the merged data.   We have added more insights in our colab notebook and ppt.  These observations can clearly help the developer to capture the android market. | | |
|  | | |