**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| 1. **Vashu Garg (Team Lead) [Email:-** [**vashugarg466@gmail.com**](mailto:vashugarg466@gmail.com)**] Contribution:-**  * Initial Libraries and Dataset Import setup. * Data Exploration and Cleaning  1. Remove Corrupted Data (Nan or Null) 2. Make Columns Datatype Compatible 3. Replace with mean and mode values 4. Drop duplicate rows and columns  * Category-wise Playstore app data analysis * Pie-Chart for apps availability under different categories * Correlation graph among factors affected on market * All ‘Google’ Listed products data operation * Positive and Negative Sentiments analysis. * Bar graph and Pie-Chart on Different App Categories. * Conclusion for the data analysis.  1. **Somya Jain [Email:-** [**Somya.jain211999@gmail.com**](mailto:Somya.jain211999@gmail.com)**] Contribution:-**  * Identifying Best Gaming and Art and Design Apps * Updated date of apps analysis using line plot * Max Installed Applications * Bar for Paid and Free Apps * Application having least rating analysis  1. **Jyoti Singh [Email:- js3617095@gmail.com ] Contribution:-**  * Best Rating Application Analysis * Applications have more than 50M installs. * Study on marketplace value from Rating. * Mature 17+ Content Rating Applications * Indentifying the best Category as per the review  1. **Deepika Gupta [Email:-** [**deepika2.gupta1111@gmail.com**](mailto:deepika2.gupta1111@gmail.com)**] Contribution:-**  * Analysis of top 5 food and drink category apps * Bar to Understand content rating provided by ‘Teens’ * Estimation of revenue of the apps * User Sentiment Analysis by Bar Graph * Average number of words used in translated review. * Pie-Chart of sum of Sentiment polarity rate per Category * Applications with sentiment polarity and subjectivity > 0.5  1. **Palak Bindal [Email:-** [**palakbindal1993@gmail.com**](mailto:palakbindal1993@gmail.com)**] Contribution :-**  * Import the data set and reading data sets * Cleaning the data :  1. Cleaning the data of columns one by one 2. Replacing nan with mode values 3. Converting str values into float for the ease of calculations 4. Dropping the row which has absurd values  * Recognizing best applications with maximum Installs * Picking out the applications with 4+ Ratings * Highest paid application Analysis * List of the applications famous among ‘Teens’ * Finding applications generated the highest revenue. * Correlation between installs and rating * Bar to analyze installs per category * Analysis that most of the apps got the rating 4.3 * Survey of ratio of free and paid apps through pie-chart. |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| In today’s world everything has become very easy and quick only because of the digital environment and technology, so mobile and desktop applications have played a big role in it.  Google playstore is the environment provided by “Google” for different types of applications operated on android devices. This is the official Google app store for the Android operating system. It has millions of apps with their reviews so that we can use such an amount of data for any data science task and analysis.    **Problem Statement & Approach:**    To manage and analyze so much data of applications is a challenging task. So we have provided some raw datasets on which we have to analyze and do some operations to find out the meaningful results, so that developers and the market can easily understand the data for future predictions and provide insights for meaningful business decisions.  Here we were provided two data sets : 1. Play store data which contains mainly all the applications name, their ratings, reviews, type if they are free or paid, category, content rating, installs, size, Genres, last updated dates, Current versions, Android versions. 2. User reviews data which have translated reviews with user sentiments.  By these raw datasets, we can not find the insightful data required for taking decisions, so we have done data wrangling and data cleaning of the data column wise. So that we can work on data smoothly without facing any errors. It includes removing any rows with absurd data if any. Also to replace nan values with the meaningful values like mode and mean.  **Conclusion**:   1. Best gaming app based on rating is Monster Ride Pro 2. Best Art and Design app based on rating is Spring flowers theme couleurs d t space 3. Apps which have maximum installs are mostly free. 4. Total No. of Free apps are 8719 & Paid apps are 647 5. App with least rating is DS Creator 2.0 is of Category ‘Tools’ 6. Apps of Category ‘Weather’ has maximum reviews 7. Best Food and drink app based on rating is Bar B-Q Rib House 8. App which generated the highest revenue is Minecraft 9. Average no. of words used in Translated reviews is 18 10. Most of user sentiments are Positive 11. Game Category has the highest sentiment polarity rate. 12. Maximum installed apps are Subway surfers,Facebook,Messenger & Google Drive 13. Category with has max no. of apps with 4+ rating is ‘Family’ 14. Most expensive app on Play store is “I’m Rich” priced at 400$ 15. 1208 apps are famous among teens out of 10841 apps 16. Highest revenue generated app belongs to Category ‘Family’ 17. Most of the apps has got the rating between 3.5 to 5 18. Maximum no. of installations belongs to Category ‘Communication’ and ‘Social Media’ 19. Most of the apps available on play store belongs to Category ‘Family’ i.e, 20.3% 20. Application reviews and installs are highly correlated i.e, 0.64 21. 48 apps are Google listed and their average rating is 4.3 22. Genre ‘Action’ has highest - 358 Gaming applications 23. Maximum no. of applications got updated in year 2018 24. Out of all Categories, ‘Game’ has got the most positive sentiments |
| **Please paste the GitHub Repo link:-**  **https://github.com/Somya2119/CapstonProject-play-store-app-data-analysis-new.git** |
|  |