**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Member Name – SUMIT KUMAR DHIBAR  Email ID - [kumarsumitprs@gmail.com](mailto:kumarsumitprs@gmail.com)  Ph No.- 8210367696 |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/Link/to/Repo> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  **Summary:-**  **Problem Statement**  The Play Store apps data has enormous potential to drive making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market.  Each app (row ) has values for category, rating, size and more. Another dataset contains customer reviews of the android apps.  Explore and analyze the data to discover key factors responsible for app engagement and success.  **Objective-**   * Installation of application by users according to the categories. * Mostly demanded applications in playstore. * Factors that affect the installation of application by the user.   **Steps involved to complete this project:-**   * loading the data into data frame * cleaning the data * extracting statistics from the dataset * exploratory analysis and visualizations * conclusion   I move to first step of data analysis by cleaning the data that will make the results more accurate. The steps I used are here :-  Data Preparation, Gathering data, Cleanse and validate data, etc. After completing the data cleaning operations I have done Exploratory Analysis and Visualization Exploratory data visualizations (EDVs) are the type of visualizations we assemble when we do not have a clue about what information lies within our dataset. I have done some visualizations using python libraries seaborn and matplotlib. Some of the graphs I visualised in project are here:-  Distribution of App Rating, Number of Installed applications for each category, Distribution of App Size, Distribution of Subjectivity, Apps installed according to its type, etc. |
| **Conclusion**  I started from scratch where the dataset we took was totally raw. I did a lot of cleaning on the data provided to bring it in a cleaner, representable form. Missing values were also removed in this process.   * The dataset contains possibilities to deliver insights to understand customer demands better and thus help developers to popularize the product. * Dataset can also be used to look whether the original ratings of the app matches the predicted rating to know whether the app is performing better or worse compared to other apps on the Play Store. * My motive in whole project was to analyse the data and find out main components that affects users decision to download app. * After completion of analysis I concluded that user prefer more of free apps. Most of the apps present in playstore are more or less of same size so size doesn’t affect their decision much. * I found most popular category of apps on two basis - Number of Installs and Number of reviews. Personalization wins in former criteria whereas Sports wins in later criteria |