**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| Name - KULDEEP SINGH  Email - [kldp0176@gmail.com](mailto:kldp0176@gmail.com)  Project - Exploratory Data Analysis  Topic - Play Store App Review Analysis  Contribution :-   1. Data Wrangling:  * Play store App Dataset * Reviews Dataset  1. Cleaning Data:  * Fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data. * Dealing With NaN values.  1. Exploratory Data Analysis:  * Correlation * Top Categories * Most Installed Category * Distribution of the ratings * Free & Paid App Comparison * Top Expensive Apps * Content Rating * Genres * Positive - Negative sentiment * Sentiment\_polarity and Sentiment\_Subjectivity |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/kd0176/Play-Store-App-Review-Analysis> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Project - Exploratory Data Analysis**  **Topic - Play Store App Review Analysis** Google Play Store, formerly known as Android Market, is the official distribution storefront for Android applications and other digital media, such as music, movies and books, from Google. It is available on mobile devices and tablets that run the Android operating system (OS), supported Chrome OS devices and on the web.The Play Store apps data has enormous potential to drive app-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.What we need to do in this project is that we need to explore and analyze the data to discover key factors responsible for app engagement and success.As the first step, we need to check the structure and the manner in which the data is organized, also we need to check and see that the dataset is ready for performing several exploration operations or not.We can check the structure and the manner in which the data is organized **using pandas library's attributes and methods:**   * **shape**: It returns a number of rows and columns. * **columns**: It returns column labels. * **nfo()**: It gives a concise summary of a DataFrame. * **head()**: Returns the first n rows. * **tail()**: Returns the last n rows   Second step is cleaning data, Dataset can contain missing data, numerical string value, and various cues. If we can clean them, It will make our analyzing process easy.  We need to fix or remove incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset, and also handle Null or missing values.  Final step is perfoming Exploratory Data Analysis,  Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns,to spot anomalies,to test hypotheses and to check assumptions with the help of summary statistics and graphical representations. |