**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Nada Nasser – [nadanasser2468@gmail.com](mailto:nadanasser2468@gmail.com)   -Preparation of Individual and become a part of team Colaboratory Notebook preparation. Initial investigation of the given data for analysis. Data Processing, Data Cleaning, Data Manipulation, and Visualization, Plot Preparation.  -Power point Presentation preparation.  -Frame work of project.  -Debug errors.  -Technical Documentation preparation.  -Summary preparation.   1. Mohammad Ammaar– [mohammad.ammaar1@gmail.com](mailto:mohammad.ammaar1@gmail.com)   -Preparation of Individual and team Colaboratory Notebook preparation.  -Power point Presentation preparation.  -Data Analysis, Data Visualization, Plot preparation  -Frame work of project.  -Debug errors.  -Technical Documentation preparation.  -Summary preparation.   1. Mohammed Shahzeb Khan – [szebk1997@gmail.com](mailto:szebk1997@gmail.com)   -Organize team meetings and knowledge sharing.  -Preparation of Individual and team Colaboratory Notebook preparation.  - Power point Presentation preparation.  -Data Analysis, Data Visualization, Plot preparation  -Frame work of project.  -Debug errors.  -Technical Documentation preparation.  -Summary preparation.   1. Gayathri Singh -   -Organize team meetings and knowledge sharing.  -Preparation of Individual and team Colaboratory Notebook preparation.  - Power point Presentation preparation.  -Data Analysis, Data Visualization, Plot preparation  -Frame work of project.  -Debug errors.  -Technical Documentation preparation.  -Summary preparation.   1. Mansur Shikalgar -   -Organize team meetings and knowledge sharing.  -Preparation of Individual and team Colaboratory Notebook preparation.  - Power point Presentation preparation.  -Data Analysis, Data Visualization, Plot preparation  -Frame work of project.  -Debug errors.  -Technical Documentation preparation.  -Summary preparation. |
|  |
| Github Link:- <https://github.com/nadanassern?tab=repositories> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| The **Google Play Store** is a [digital distribution](https://en.wikipedia.org/wiki/Digital_distribution) service operated and developed by [Google](https://en.wikipedia.org/wiki/Google). It serves as the official app store for certified devices running on the [Android operating system](https://en.wikipedia.org/wiki/Android_(operating_system)) and [its derivatives](https://en.wikipedia.org/wiki/Google_Operating_System) as well as [Chrome OS](https://en.wikipedia.org/wiki/ChromeOS), allowing users to browse and download applications developed with the [Android software development kit](https://en.wikipedia.org/wiki/Android_software_development) (SDK) and published through Google. Google Play has also served as a [digital media](https://en.wikipedia.org/wiki/Digital_media) store, offering games, music, books, movies, and television programs be. Content that has been purchased on [Google Play Movies & TV](https://en.wikipedia.org/wiki/Google_TV_(service)) and [Google Play Books](https://en.wikipedia.org/wiki/Google_Play_Books) can be accessed on a [web browser](https://en.wikipedia.org/wiki/Web_browser), and through the [Android](https://en.wikipedia.org/wiki/Android_(operating_system)) and [iOS](https://en.wikipedia.org/wiki/IOS) apps.  Problem definition:  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. Objective of the project to Explore and analyze the data to discover key factors responsible for app engagement and success.  Exploratory Data Analysis on given Data set:  There are two dataset:   * Play Store Data   (App, Category, Rating, Review, Size, Install, Type, current rating ,genres , Last update, Current Version ,Android Version)   * User Review Data   (App, Sentiment ,Sentiment Polarity, Sentiment Subjectivity) Digging into data we understand.  • There are 13 columns of properties with 10841 rows of data.  • Column 'Reviews', 'Size', 'Installs' and 'Price' are in the type of 'object'  • Values of column 'Size' are strings representing size in 'M' as Megabytes, 'k' as kilobytes and also 'Varies with devices'.  • Values of column 'Installs' are strings representing install amount with symbols such as ',' and '+'.  • Values of column 'Price' are strings representing price with symbol '$'.  From the given data, we have to analyze the required data and clean the whole information for further analysis. Analysis of information is done and conclusions are being made which is useful for the business aspect.  Conclusion:  The Google Play Store Apps report provides some useful details regarding the trending of the apps in the play store. As per the graphs visualizations shown above, most of the trending apps (in terms of users' installs) are from the categories like GAME, COMMUNICATION, and TOOL even though the amount of available apps from these categories are twice as much lesser than the category FAMILY but still used most. The trending of these apps are most probably due to their nature of being able to entertain or assist the user. Besides, it also shows a good trend where we can see that developers from these categories are focusing on the quality instead of the quantity of the apps.  Some insights on which we worked are as follows:   * Top categories of application installations, most of the apps are under the categories Family and Game. * 4.1 rating has the highest count in the given list of applications. * Distribution of ratings, large number of apps were in between 3.8 and4.8 rating. * Average rating of (active) apps on Google Play Store is 4.2. * Most of the apps in the play store are in the size range between 10MB and 100MB. * 100,00,000 is the maximum installation rate. * Percentage of paid and free apps are plotted using a pie chart. * Top 10 installed apps are plotted. * Correlations check among ‘Install’ label and other labels of the datasets. * Visualization using Histograms and Bar plots. * Insights over the content rating with the number of apps. * Word cloud for better understanding of repetitive words. * Category Line plot for positive as well as negative sentiments opinion. * If we see individually app wise the communication app like Facebook and Whatsapp get highly reviewed app. It shows that people are regularly active on that and give the feedbacks * Medical and Family apps are the most expensive and even extend up-to 80$. * Users tend to download a given app more if it has been reviewed by a large number of people. * More than half users rate Family, Sports and Health & Fitness apps positively. Apps for games and social media get mixed reviews, with 50 percent positive and 50 percent negative responses. |