# Capstone Project Submission

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member's Name, Email and Contribution:** |
| --- |

***TEAM MEMBERS NAMES***


***1. Komal Sursawant***

komalsursawant029@gmail.com


- Upload dataset to Google colab.
- Analyze null values and filter them.
- Project summary
- Work on graphs.
- Correction of data types.
- Data visualization.


***2. Pratiksha Chimankar***

Pratikshachimankar5@gmail.com


- Upload dataset to Google colab.
- Analyze null values and filter them.
- Data cleaning.
- Data visualization.
- Correction of data types.
- Project Summary


***3. Dhruv Dubey***

ddhruv0407@gmail.com

- Upload  dataset to Google colab.
- Analyze null values and filter them.
- Data cleaning.
- PowerPoint presentation
- Correction of data types.
- Data  visualization.

***4. Sanket Kamble***

Kamblesanket1833@gmail.com

- Upload  dataset to Google Colab.
- Analyze null values and filter them.
- Data cleaning.
- Data  visualization.
- Correction of data types.
- Technical documentation

***5. Deepankar Singh***

Deepank1239singh@gmail.com

- Upload  dataset to Google colab.
- Analyze null values and filter them.
- Data cleaning.
- Correction of data types.

- ***CONTRIBUTION -  All MEMBERS EQUALLY CONTRIBUTE IN TEAM PROJECT.***

**Please paste the GitHub Repo link.**

https://github.com/Sanketkamble01/Playstore-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)**

Play Store is the biggest market for Android apps worldwide. In the play store there are so many different types of apps. On Google play store there are more than 5 million applications with more than 3 billion active users across more than 190 countries.

At the first step we import the libraries like numpy and pandas then we copy the file path in the notebook .This project provides us with two csv files of raw data. In this project we have given more importance to analyzing the data because we wanted to see what would be the output after analyzing the data. After that we have done the process of cleaning the data. In the process of analysis of data we observe some facts.

- ***THE GOAL OF THE PROJECT***

Our project goal is to predict the number of installs of apps by looking at app information
and its reviews.
We hope that this project will help app developers predict their number of installs or investors
who want to pick out the next big app.

- ***Introduction***

In this project, we investigate the different variables of Apps on Google Play Store that affect the application and the top most relevant user reviews. We attempt to use our analysis to answer the following questions:

- What are the top categories on playstore?
  - How important is the rating of application? *
  - Most percentage of content on Play Store?
  - Which category has the most number of installations?
  - check correlation
  - mostly used words on playstore reviews?
  - Relationship between content Rating in Free and paid apps?
  - What are the sentiments across databases?
  - How size impacts on installation of apps?

- ***Data cleaning***

After getting the dataset, our next step that must be done is preprocessing text. Preprocessing text aims to clean raw data into assorted data that is ready for use. The data preprocessing process can consist of many things, changing the data type in the column, changing or manipulating empty columns, deleting data with duplicate contents, and others.

- ***Data visualization***

1. Find the total numbers of apps in each category and number of Installs in each category.
2. Show the distribution of ratings.
3. Impact of ratings on Number of installations.
4. Compare between two interesting category free v/s paid apps.
5. Positive reviews for all apps are higher than negative and neutral.
6. Compare size with installation and see Size impact on no. of installation.

- ***Conclusion***

Here, we can see clearly in first graph that family category is on top position in terms of total number of apps, followed by game, personalization category and Beauty category is on bottom position in terms of total number of apps COMMUNICATION is on top then SOCIAL is on second in FAMILY category is not in the list of number of Installs in each category.
As a result, the app development companies could choose the type of programme that should be created and could also evaluate how it would perform in the future. They may also view the categorized reviews of every programme in one interface, which will enable them to determine which apps are well-liked by customers and which ones require further development