

# Capstone Project Submission

## Instructions:

- iii) Please fill in all the required information.
- iv) Avoid grammatical errors.

## Team Member's Name, Email, and Contribution:

### Team Member's Role:-

- **RAM MANOHAR THAKUR**  
eMail- [rayanka148@gmail.com](mailto:rayanka148@gmail.com)
  - **ROLE-** Data Understanding
  - Feature Engineering
    - Null value check
  - Data Visualization
    - Donut plot, box plot
  - Multivariate Analysis
    - Heat map
  - Research Analytics
    - Technical documentation
- **SAYAN BISWAS**  
eMail- [sb5798512@gmail.com](mailto:sb5798512@gmail.com)
  - **ROLE-** Data Understanding
  - Feature Engineering
    - Missing value
  - Data Visualization
    - Count plot
  - Multivariate Analysis
    - Correlation matrix
  - Research Analytics
    - Technical documentation
- **SNEHAL CHANDU RAGIT**  
eMail- [sragit12@gmail.com](mailto:sragit12@gmail.com)  
**ROLE-**
  - Data Understanding
  - Feature Engineering
    - Duplicate Value
  - Data Visualization
    - Scatter Plot
  - Multivariate Analysis
  - Research Analytics
    - Technical documentation
- **SHRADHA ANILRAO LAKHADIVE**  
eMail- [shraddhalakhadive14@gmail.com](mailto:shraddhalakhadive14@gmail.com)
  - **ROLE-** Data Understanding
  - Feature Engineering
    - Replacing null value

- Data Visualization
  - Pie Chart & helping plot making
- Multivariate Analysis
- Research Analytics
  - Technical documentation

**Please paste the GitHub Repo link.**

Github Link:- <https://github.com/sayan3600/SAYAN.git>

**Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)**

#### **PROBLEM STATEMENT:**

This data contains a description of different types of apps in each row. This method is considered to be one of the most important methods of data science. Through this method we can perform other tasks besides making basic business decisions. We have a lot of problem statement:

1. How many apps are free or paid?
2. Which apps are at the top of their popularity?
3. Which apps category have been downloaded the most?
4. The most popular category among the categories that appear on the Play Store dataset?
5. What percentage of apps are free?
6. The name of the most downloaded apps among the paid apps?
7. How many free apps have been downloaded over a million?
8. Which apps received the most positive and the least negative reviews?

#### **APPROACH:**

Exploratory Data Analysis (EDA) is an approach to analyze data. The first and foremost task is that the data analysis to view the data and tries to make some sense out of it. Later we figure out what question we want to ask and how to use the available data to get the insights and all the answers that we need from the data set. EDA helps us to

- Look into the dataset
- Examine the relationships among the variables
- Identify the interesting observation
- Develop an initial idea of possible associations among the predictors and the target variable.

#### **CONCLUSION:**

After analysis we have got answers to some of questions. We visualized trends and relations between features of apps using different types of plots such as bar graph, distribution plot, scatter plot, box plot, pie chart and more. Our analysis can be useful for future app developers and marketers in making their app a big success.

- a) First of all we can say that 91.9% apps are free & 8.1% apps are paid.
- b) Most apps are available in Family category.
- c) Most favorite category of users is Game.

- d) The most disliked category by users is medical.
- e) Personalization is the most popular category among paid apps.
- f) Free apps & Paid apps have an average rating of 4.18 & 4.26
- g) People prefer small size apps more.
- h) Most of the paid apps cost less than \$100
- i) Paid apps have lower negative sentiment polarity than free apps.

In this way, our analysis shows how these factors can contribute to app engagement and success.