

Deliverable 1

1) Team

a) Members

Veda Samhitha Dyawanapally
Meghana Chikyala
Poornima Pulakandam
Amarnath Reddy Chinthapalli
Pavan Kalyan Reddy Madatala

b) Communication plan to include project artifact repository

A communication plan was created to outline how project artifacts were stored and shared among team members. In this case, We have used Google Spaces as the primary platform for discussions, while Google Meet was used for virtual meetings. Additionally, a Git repository will be utilized to share project files and other relevant information. The frequency of group meetings will be based on the project deliverables.

GitHub Repository Link: https://github.com/PavanKalyanReddyM/BDA_Project.git

2) Selection of data:

We have chosen this dataset from kaggle.

KickStarter Projects

DataSet Link: <https://www.kaggle.com/datasets/ulrikthygpedersen/kickstarter-projects>

Dataset Description :

The data set contains details about Kickstarter projects and whether they were able to achieve their funding goals or not.

3) Business Problem or Opportunity, Domain Knowledge

The dataset we have chosen is from Kaggle where it gives the data related to the Startup Project. We can use this data to derive some descriptive visualizations like how many projects were successful, which country has invested more in these projects and whether the Number of Projects being successful are increasing year by year etc.

Also, we use Machine Learning techniques and statistical analysis to give accuracy of the state of the project whether it is going to be successful or not and for developing accurate predictive models.

Domain Knowledge:

AWS S3, AWS Sage Maker for Analytics and ML Process.

AWS Quick Sight for Creation of Dashboard Visualizations.

4) Research Objectives and Question(s)

- 1) Which category of projects has the highest success rate?
- 2) What is the overall trend of crowdfunding projects over time, and how does this vary across different categories?
- 3) How many projects were successful and how many failed? What is the overall success rate of Kickstarter projects?
- 4) What is the Correlation between the Goal, Pledged and Backers attributes?
- 5) Are there any outliers in the dataset, in terms of extremely high or low amounts pledged, number of backers, or project goals?
- 6) How many projects were launched for each category with respective year?
- 7) What is the average funding goal and pledged amount for different categories and subcategories on Kickstarter?
- 8) Can we predict the success rate of a project (.e., the percentage of its funding goal that is met) based on its category, subcategory, country, launch date, and the number of backers?
- 9) How Important are Categories in predicting the success of Kickstarter Projects?
- 10) How Important are Sub Categories in predicting the success of Kickstarter Projects?