

## Data Collection and Preprocessing Phase

Date 15 March 2024

Date	10 July 2024
Team ID	740062
Project Title SmartLender -	Automotive kickstart
Maximum Marks	2 Marks

### Data Collection Plan & Raw Data Sources Identification Report:

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

#### Data Collection Plan:

Section	Description
Project Overview	The machine learning project aims to predict Successful or not based on applicant information. Using a dataset with features such as goal, Pledged, Backers, Category, Subcategory and Year, the objective is to build a model that accurately classifies project(Successful or Unsuccessful), facilitating efficient and informed decision-making in the growth process.
Data Collection Plan	<ul style="list-style-type: none"><li>● Search for datasets related to success or not, startup information, and applicant details.</li><li>● Prioritize datasets with diverse demographic information.</li></ul>
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle & UCI, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as Goal,

	Pledged, Backers, and State related details for machine learning analysis.
--	--

#### Raw Data Sources Report:

Source Name	Description	Location/URL Format Size	Access Permissions
Kaggle Dataset	The dataset comprises applicant details (ID, Name, Category, Subcategory, Goal, Pledged, Backers, Year State)	<a href="https://www.kaggle.com/datasets/ulrikthygepedersen/kickstarter-projects">https://www.kaggle.com/datasets/ulrikthygepedersen/kickstarter-projects</a> CSV 17 MB	Public
UCI	This data concerns credit card applications; a good mix of attributes	<a href="https://archive.ics.uci.edu/dataset/320/student+performance">https://archive.ics.uci.edu/dataset/320/student+performance</a> CSV 39.8 KB	Public