Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	739930
Project Title	Airline Review Classification
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	Airline review classification aims to analyze customer reviews and classify them into categories, such as positive, negative, or neutral. This project leverages natural language processing (NLP) and machine learning techniques to gain insights from customer feedback, helping airlines to improve their services.
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 .

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	size	Access permissions
Kaggle Dataset	The dataset comprises customer details(airline details)	https://www.kaggle.com/datasets/khushipitroda/airline-reviews	CSV	19.5MB	Public