Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	739930
Project Title	Airline Review Classification
Maximum Marks	2 Marks

Data Quality Report:

This data quality report serves as a foundational document, outlining the steps taken to ensure the reliability and usability of Airline data in the project. It underscores the commitment to maintaining high data standards and reliability in achieving the project's objectives effectively.

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing values in 'Gender', 'Age', 'Review', 'Airline', 'Rating' columns	Moderate	Use mean/median imputation to fill missing values in numerical columns (Age, Rating). For categorical columns (Gender, Review, Airline), use mode imputation to fill missing values.
Kaggle Dataset	Categorical data in the dataset	Moderate	Perform encoding on categorical variables to convert them into numerical format suitable for modeling. Use one-

			hot encoding for columns with multiple categories (Airline) and label encoding for binary columns (Gender).