Data Collection and Preprocessing Phase

Date	23 September 2024
Team ID	LTVIP2024TMID24998
Project Title	Flight Delays Prediction using Machine Learning.
Maximum Marks	2 Marks

Data Quality Report Template

Data Quality Report template summarize data quality issues from the selected dataset, including their severity levels and the proposed resolution plans. This report helps systematically identify and address any discrepancies or issues with the dataset to ensure high-quality input for the machine learning model.

Data Source	Data Quality Issue	Severity	Resolution Plan
Flight Dataset	Issues arise from Missing or Incomplete Data and Inaccurate Timestamps	High	Implement data validation checks during data entry to ensure all required fields are completed.
Flight Dataset	Inconsistent Flight Status Reporting	Moderate	Develop a standardized set of definitions and categories for flight status (e.g., delayed, canceled).
Flight Dataset	Outdated Information	High	Set up automatic updates to ensure the dataset reflects the latest information and regularly review historical data for relevance and remove or archive outdated entries.

Flight Dataset	Duplicate Entries	Moderate	Implement deduplication algorithms to identify and merge duplicate records automatically.
Flight Dataset	Lack of Standardization	Moderate	Create and enforce a data dictionary that specifies formats, units, and naming conventions for all data fields.
Flight Dataset	External Data Dependencies	Moderate	Evaluate the reliability of third-party data sources and implement fallback mechanisms to handle instances where external data is unavailable or inconsistent.
Flight Dataset	Minor Formatting Inconsistencies	Low	Correcting the data in capitalization or spacing in text fields.