

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 Collection Plan & Raw Data Sources Identification Template

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 Template

Section	Description
Project Overview	The aim of the Flight Delays Prediction (FDP) project is to accurately forecast flight delays using historical and real-time data, enhancing operational efficiency for airlines and airports while improving the travel experience for passengers through proactive delay management using Machine Learning techniques.
Data Collection Plan	Data is collected from public datasets available on Kaggle, OpenSky Network, and the other Airlines Reporting Corporation sources.
Raw Data Sources Identified	Raw data includes flight schedules, departure and arrival times, delay causes, weather conditions, and air traffic information from airlines, airports, and tracking systems.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	Kaggle flight datasets typically include flight number, airline, origin/destination airports, scheduled/actual times, delays, cancellation status, flight date, distance, and sometimes weather data.	https://www.kaggle.com/datasets/abderrahimlakouche/flight-delay-prediction	CSV	100MB	Public
OpenSky Network Dataset	The OpenSky Network dataset provides real-time and historical flight trajectory data, including aircraft positions, velocities, and timestamps, useful for tracking and predicting flight delays.	https://opensky-network.org/data/datasets	CSV	100MB	Public (with OpenSky account)