Dataset + Project Proposal

Airline On-Time Statistics and Delay Causes Analysis in Major US Airports

Problem Definition:

This project analyzes the on-time performance and delays caused by airlines operating in the busiest airports in the US. By identifying patterns in flight delays, this project aims to provide insights into delays and make suggestions to reduce them

Data Source:

The data will be sourced from the Bureau of Transportation Statistics (BTS) within the U.S. Department of Transportation, specifically the "Airline On-Time Performance and Causes of Flight Delays" dataset.

Original Source: https://www.transtats.bts.gov/

Link to Dataset:

https://www.openml.org/search?type=data&sort=runs&status=active&qualities.NumberOfFeatures=lte_10&qualities.NumberOfInstances=gte_1000000&id=42728

The data is hosted on

https://www.openml.org/data/download/22044760/airlines_train_regression_10000000.arff.

Motivation:

Flight delays are a major inconvenience for passengers and lead to significant financial losses for airlines and airports. With the increasing demand for air travel, understanding the factors contributing to delays is crucial for improving operational efficiency and customer satisfaction. We are. interested in this problem because it involves analyzing large, complex datasets and has real-world applications that can benefit both travelers and the aviation industry.