## **Flight Price Prediction**

## **Project 1: Flight Price Prediction (Regression)**

|  |  |
| --- | --- |
| **Project Title:** | Flight Price Prediction |
| **Skills Takeaway from This Project:** | Python, Streamlit, Machine Learning, Data Analysis,MLflow |
| **Domain:** | Travel and Tourism |

### **Problem Statement:**

Build an end-to-end project to predict flight ticket prices based on multiple factors such as departure time, source, destination, and airline type. Use the provided dataset to process, clean, and perform feature engineering. Train a regression model to predict flight prices and deploy the model in a Streamlit application. The app should allow users to input filters (route, time, and date) and get a predicted price for their flight.

### **Business Use Cases:**

* Helping travelers plan trips by predicting flight prices based on their preferences.
* Assisting travel agencies in price optimization and marketing strategies.
* Enabling businesses to budget for employee travel by forecasting ticket prices.
* Supporting airline companies in identifying trends and optimizing pricing strategies.

### **Approach:**

#### **Step 1: Data Preprocessing**

* Load the provided dataset.
* Clean the data by removing missing or duplicate entries.
* Convert date and time columns into standard formats.
* Perform feature engineering to calculate new features (e.g., price per minute).

#### **Step 2: Flight Price Prediction**

* Perform exploratory data analysis (EDA) to identify trends and correlations.
* Use regression models like Linear Regression, Random Forest, and XGBoost to predict flight prices.
* **Integrate MLflow:**
* Log experiments and metrics for each regression model using MLflow.
* Track parameters, metrics (e.g., RMSE, R-squared), and artifacts (e.g., model files, visualizations).
* Save and organize all trained models in MLflow’s model registry.

#### **Step 3: Streamlit App Development**

* Build an interactive Streamlit app that:
  + Displays visualizations of flight price trends.
  + Allows users to filter by route, airline, and time.
  + Predicts flight prices based on user inputs.

### **Results:**

* Cleaned and processed dataset for analysis.
* Built regression models with predictions achieving high accuracy.
* Developed a user-friendly Streamlit app to analyze and predict flight prices.

### **Project Evaluation Metrics:**

* Completeness and accuracy of data preprocessing.
* Performance metrics for regression models (e.g., RMSE, R-squared).
* Functionality and usability of the Streamlit app.
* Visual appeal and informativeness of the app’s charts and tables.
* Effectiveness of MLflow integration for model tracking and management.

### **Technical Tags:**

Python, Data Cleaning, Feature Engineering, Machine Learning, Regression, Streamlit, MLflow

### **Dataset:**

Dataset Link : [Flight\_Price.csv](https://drive.google.com/file/d/1RrSe3M0Ia-ekihZzWZXLTYFa-_bsba-C/view?usp=sharing)

The provided dataset includes:

* **Airline:** Name of the airline.
* **Date\_of\_Journey:** Date of takeoff.
* **Source:** Starting airport location.
* **Destination:** Final landing airport location.
* **Route:** The route from where the plane will go and stops.
* **Dep\_Time:** Departure time.
* **Arrival\_Time:** Arrival time of the plane landing.
* **Duration:** How long the flight lasted.
* **Total\_Stops:** Number of stops between flights for fuel, etc.
* **Additional\_Info:** Additional notes from the airline (e.g., meal not included).

### **Project Deliverables:**

* Python scripts for data preprocessing, model training, and MLflow integration.
* A clean CSV file containing processed flight data.
* Regression models for price prediction logged and managed using MLflow.
* A Streamlit app for data visualization and prediction with MLflow metadata integration.
* Documentation covering methodology, analysis, and insights.