

Flight Price and Customer Satisfaction 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](#)

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.

Timeline:

Check your mail for the submission deadline of the project.

Project 2: Customer Satisfaction Prediction (Classification)

Project Title:	Customer Satisfaction Prediction
Skills Takeaway from This Project:	Python, Machine Learning, Data Analysis, Streamlit, Classification Models, MLflow
Domain:	Customer Experience

Problem Statement:

Build a classification model to predict customer satisfaction levels based on features such as customer feedback, demographics, and service ratings. Use the provided dataset to process, clean, and perform feature engineering. Deploy the model in a Streamlit application, allowing users to input customer data and receive a predicted satisfaction level.

Business Use Cases:

- Enhancing customer experience by predicting and addressing dissatisfaction.
- Providing actionable insights for businesses to improve services.
- Supporting marketing teams in identifying target customer groups.
- Assisting management in decision-making for customer retention strategies.

Approach:

Step 1: Data Preprocessing

- Load the provided dataset.
- Clean the data by handling missing values and duplicates.
- Encode categorical variables.
- Normalize or standardize numerical features as required.

Step 2: Customer Satisfaction Prediction

- Perform exploratory data analysis (EDA) to understand feature relationships.
- Use classification models like Logistic Regression, Random Forest, and Gradient Boosting to predict satisfaction levels.
- **Integrate MLflow:**
 - Log experiments and metrics for each classification model using MLflow.
 - Track parameters, metrics (e.g., accuracy, F1-score), and confusion matrices.
 - Save and organize all trained models in MLflow's model registry.

Step 3: Streamlit App Development

- Build an interactive Streamlit app that:
 - Visualizes customer satisfaction trends.
 - Allows users to input features such as demographics and service ratings.
 - Predicts customer satisfaction levels.

Results:

- Cleaned and processed dataset for analysis.
- Built classification models with predictions achieving high accuracy tracked with MLflow.

- Developed a user-friendly Streamlit app to analyze and predict customer satisfaction levels.

Project Evaluation Metrics:

- Completeness and accuracy of data preprocessing.
- Performance metrics for classification models (e.g., accuracy, F1-score, confusion matrix).
- 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, Classification, Streamlit, MLflow.

Dataset:

Dataset link: [Passenger_Satisfaction.csv](#)

The provided dataset includes:

- **Gender:** Gender of the passengers (Female, Male)
- **Customer Type:** The customer type (Loyal customer, disloyal customer)
- **Age:** The actual age of the passengers
- **Type of Travel:** Purpose of the flight of the passengers (Personal Travel, Business Travel)
- **Class:** Travel class in the plane of the passengers (Business, Eco, Eco Plus)
- **Flight distance:** The flight distance of this journey
- **Inflight wifi service:** Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)
- **Departure/Arrival time convenient:** Satisfaction level of Departure/Arrival time convenient
- **Ease of Online booking:** Satisfaction level of online booking
- **Gate location:** Satisfaction level of Gate location
- **Food and drink:** Satisfaction level of Food and drink
- **Online boarding:** Satisfaction level of online boarding
- **Seat comfort:** Satisfaction level of Seat comfort
- **Inflight entertainment:** Satisfaction level of inflight entertainment
- **On-board service:** Satisfaction level of On-board service
- **Leg room service:** Satisfaction level of Leg room service
- **Baggage handling:** Satisfaction level of baggage handling
- **Check-in service:** Satisfaction level of Check-in service
- **Inflight service:** Satisfaction level of inflight service
- **Cleanliness:** Satisfaction level of Cleanliness

- **Departure Delay in Minutes:** Minutes delayed when departure
- **Arrival Delay in Minutes:** Minutes delayed when Arrival
- **Satisfaction:** Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

Project Deliverables:

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

Timeline:

Check your mail for the submission deadline of the project.

Streamlit part:

need to create one streamlit app for both the projects in different pages.

REFERENCES:

1) PROJECT APPROACH GUIDE/RECORDING:

📖 Project Orientation Session : Flight Price Prediction (English)

2)STREAMLIT RECORDING:

📖 Special session for STREAMLIT(11/08/2024) (ENGLISH)

https://us06web.zoom.us/rec/share/JTr7DywhE1-SarjyIHBSCn4qnl7_uvJH6IGk06qAlkE0Ny1o_rqcq5FRFKuo93dm.iyM2o6I0h9aTUkNE (TAMIL)

3)STREAMLIT DOCUMENTATION: [Install Streamlit](#)

4)MLflow recordings:

Day1 : 📖 Time to Learn - 2024_12_23 10_28 GMT+05_30 - Recording.mp4

Day2 : 📖 Time to Learn - 2024_12_24 10_28 GMT+05_30 - Recording.mp4

5) MLflow DOCUMENTATION: [Getting Started with MLflow](#)

6) GitHub Reference : [How to Use GitHub.pptx](#)

7) Capstone Explanation Guideline : [Capstone Explanation Guideline](#)

8) Project Live Evaluation Metrics: [Project Live Evaluation](#)

9) EDA Guide: [Exploratory Data Analysis \(EDA\) Guide](#)

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.

Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday-Saturday (4:00PM to 5:00PM)

Booking link : <https://forms.gle/XC553oSbMJ2Gcfug9>

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.

Note: This form will Open on Saturday and Sunday Only on Every Week

Timing:

For DS and AIML

Monday-Saturday (05:30PM to 07:00PM)

Booking link : <https://forms.gle/1m2Gsro41fLtZurRA>

Evaluation Metrics : [Project Live Evaluation](#)