

FLIGHT TICKET PRICE PREDICTION PROJECT

ENHANCING PRICING INSIGHTS FOR AIRLINES AND PASSENGERS

**BY
GOWCIGAN M**





Introduction

- Flight ticket pricing is a complex and dynamic challenge in the airline industry.
- Accurate prediction of ticket prices is crucial for optimal decision-making.

Benefits for Airlines:

- Enhanced Revenue Management: Accurate pricing predictions enable airlines to optimize seat pricing and maximize revenue.
- Competitive Edge: Improved pricing strategies provide a competitive advantage in the market

Benefits for Passengers:

- Cost Savings: Predictable pricing allows passengers to plan and book flights at the most cost-effective times.
- Planning Convenience: Insight into pricing trends helps passengers make informed decisions on travel dates.

Data Science Approach

Data Collection:

- Gather historical flight data with labeled ticket prices from reliable sources.
- Ensure data completeness and accuracy for effective model training.

Data Cleaning and Preprocessing:

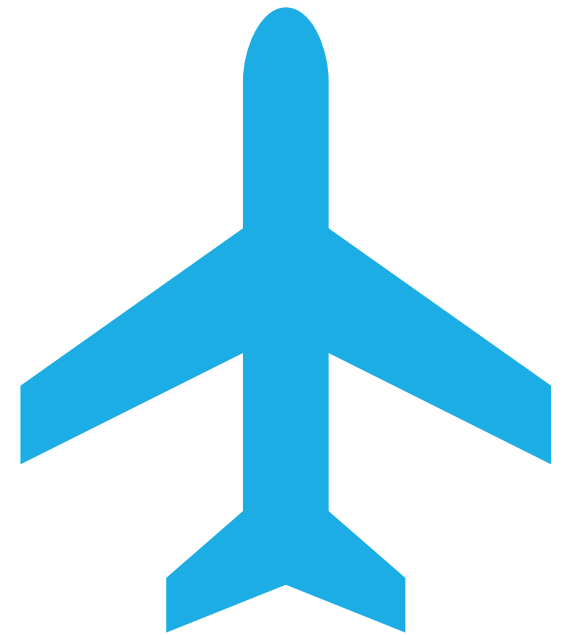
- Identify and handle missing values, outliers, and inconsistencies in the dataset.
- Standardize and normalize features for uniformity in model training.

Feature Engineering:

- Explore relevant features and create new variables that may enhance model performance.
- Transform and encode categorical variables for compatibility with machine learning models.

Model Selection:

- Choose a suitable regression model for predicting flight ticket prices.
- Justify the selection based on model performance metrics and problem requirements.



Model Training:

- Split the dataset into training and validation sets.
- Train the selected model using the training data, adjusting hyperparameters for optimal performance.

Model Evaluation:

- Assess the model's performance on the validation set using appropriate metrics (e.g., Mean Squared Error, R-squared).
- Iteratively refine the model based on evaluation results.

Results Analysis:

- Interpret the model's predictions and understand its effectiveness.
- Extract insights into pricing trends and factors influencing ticket prices

Visualization:

- Utilize Plotly for creating interactive visualizations to enhance the understanding of pricing trends.
- Present key findings through graphs and charts.

Deployment:

- Prepare the model for deployment in a production environment.
- Ensure seamless integration with existing systems for real-time pricing predictions.



Supervised Learning - Regression

Definition: Supervised learning is a type of machine learning where the model is trained on a labeled dataset, meaning the input data is paired with corresponding output labels.

Regression as the Problem Type:

Definition: Regression is a type of supervised learning used for predicting continuous outcomes, such as prices, temperatures, or stock values.

Reasons for Choosing Supervised Learning - Regression:

- 1. Nature of the Target Variable:** Ticket prices are continuous and can take any numeric value, making regression the appropriate approach.
- 2. Availability of Labeled Data:** Historical data includes instances where ticket prices are known, enabling the model to learn the relationship between input features and price variations.
- 3. Quantitative Prediction Requirement:** The goal is to provide a quantitative prediction of ticket prices rather than assigning them to predefined categories



Tools and Libraries

Machine Learning (Sklearn):

Role in Building Models:

- Scikit-Learn (Sklearn) is a powerful machine learning library in Python.
- It provides a wide range of tools for building and implementing machine learning models, including regression models for our flight ticket pricing prediction

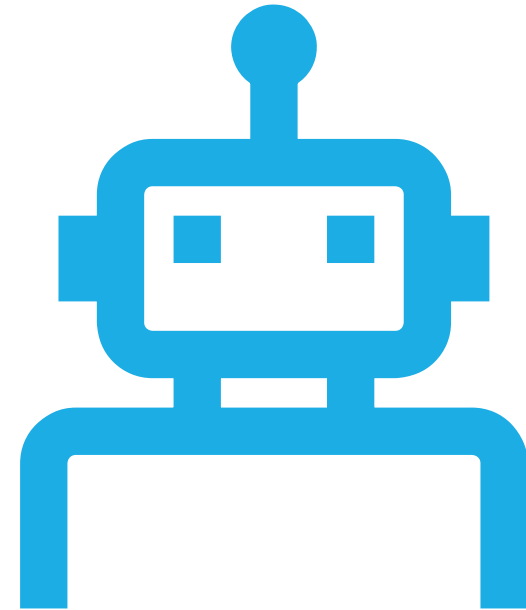
Role in Evaluating Models:

- Sklearn offers various metrics for evaluating model performance, such as Mean Squared Error (MSE), R-squared, and others.
- It simplifies the process of assessing how well the model is capturing the patterns in the data, aiding in the iterative refinement of the regression model.

Mathematics (Numpy):

Use for Mathematical Operations:

- Numpy is a fundamental library for numerical operations in Python.
- It provides efficient and fast array operations, enabling complex mathematical computations required for data preprocessing, feature engineering, and model training.



Visualization (Plotly):

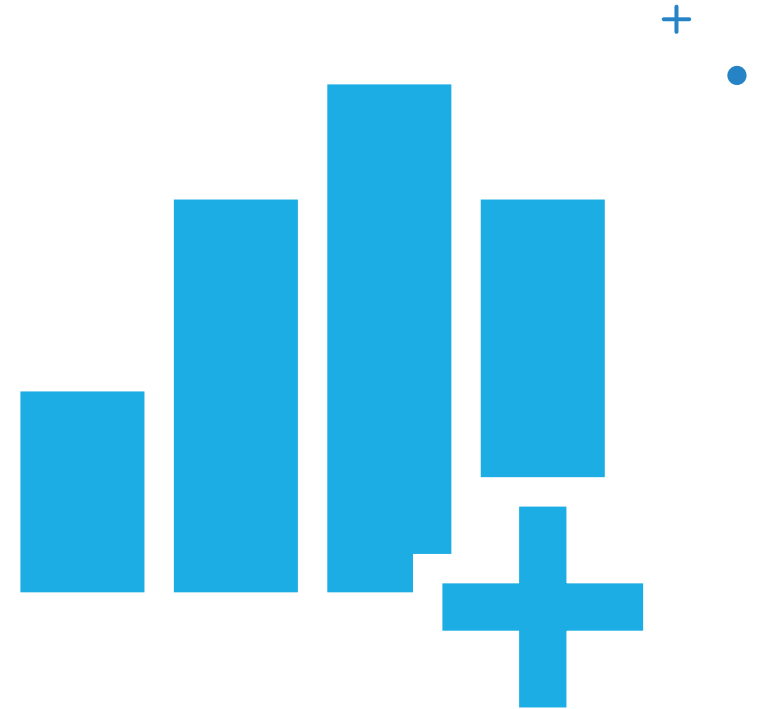
Use for Creating Interactive Visualizations:

- Plotly is a versatile and interactive visualization library.
- It facilitates the creation of visually appealing and interactive charts and graphs, aiding in the exploration and communication of data patterns.

Statistics:

Role in Analysis and Hypothesis Testing:

- Statistical analysis is integral to understanding data patterns, making informed decisions, and validating hypotheses.
- Key statistical techniques, such as hypothesis testing, may be employed to assess the significance of factors influencing ticket prices.
- Statistics play a vital role in ensuring the robustness and reliability of insights derived from the data, contributing to the overall credibility of the project.



Result



+

•

○

	Flight_ID	Flight_Price
0	F45001	877.90
1	F45002	810.43
2	F45003	730.97
3	F45004	703.02
4	F45005	758.37
...
4995	F49996	721.69
4996	F49997	756.99
4997	F49998	754.57
4998	F49999	779.34
4999	F50000	750.37



"DATA ARE JUST SUMMARIES OF THOUSANDS OF STORIES – TELL A FEW OF THOSE STORIES TO HELP MAKE THE DATA MEANINGFUL." - CHIP HEATH