PYTHON PROJECT

FLIGHT TICKET PRICE PREDICTION

SHALINI ROY
MB23017

YASHASHVI TRIVEDI
MB23018

ANURAG SINGH
MB23039

PRANAV SHARMA
MB23040

0 24563 84926 54 2

PROBLEM STATEMENT

This project aims to predict flight ticket prices using regression models such as ridge, lasso, and decision tree. By analyzing historical ticket prices and various features, the goal is to provide accurate predictions for future prices.

Market Demand Analysis

The problem statement involves understanding the various factors affecting the pricing of airline tickets, such as seasonal demand, route popularity, and special events.

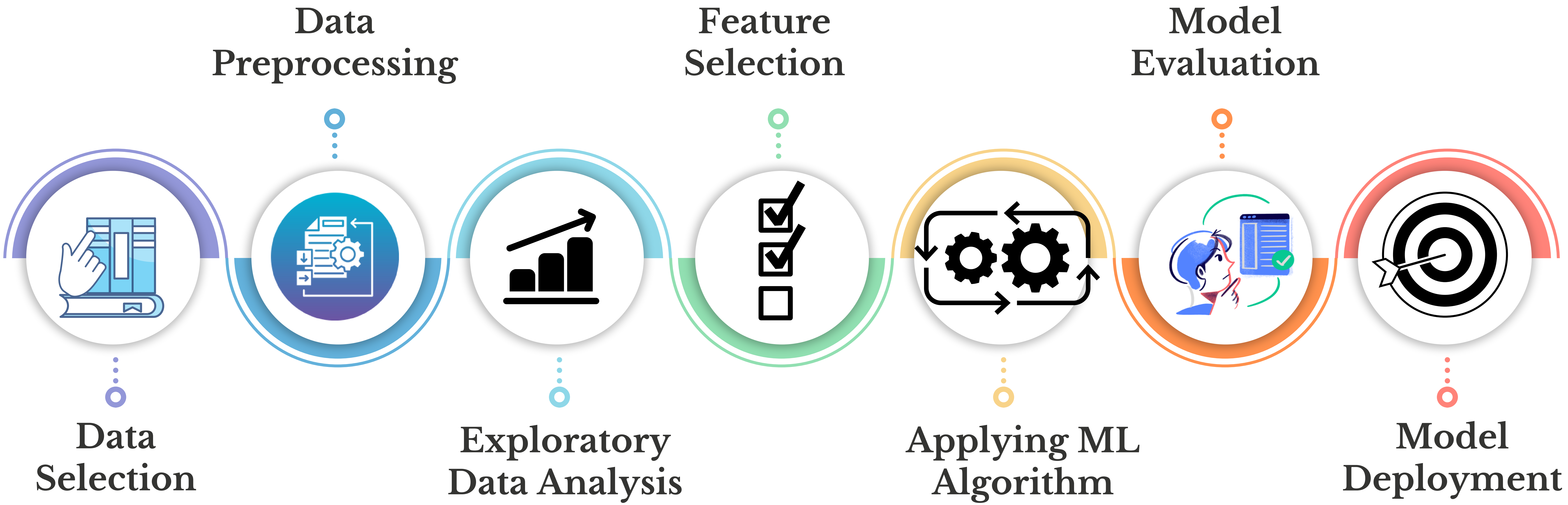
Data Collection Challenges

Obtaining comprehensive and reliable historical ticket prices and associated data is a daunting task due to the fragmented nature of the airline industry.

Performance Metrics

Here, we'll outline the evaluation metrics used to gauge the performance of the regression models and compare the accuracy of the predictions.

PROJECT STAGES



DATA COLLECTION AND PREPROCESSING

The datasets used for model building and predicting values in this project are collected from the internet, emphasing the real-world nature of the data and the need for thorough preprocessing to ensure its reliability and relevance to the problem at hand.

01 DATA CLEANING

By preventing biases and errors stemming from incomplete or redundant information, data cleaning becomes a fundamental step, shaping the reliability and effectiveness of machine learning models.

02 FEATURE-SPECIFIC CLEANING

By facilitating better model comprehension, feature-specific cleaning contributes to the overall effectiveness of machine learning models, ensuring they can capture and interpret key information for accurate predictions.

03 LABEL ENCODING

By translating categorical information into a format suitable for mathematical computations, label encoding facilitates the seamless integration of diverse data types, ultimately enhancing the model's ability to make accurate predictions.

04 FEATURE CONCATENATION

The objective of this step was to integrate processed categorical and numerical features into a consolidated dataset, essential for comprehensive model training.

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is essential for gaining a comprehensive understanding of a dataset before applying machine learning algorithms. It aids in uncovering insights, identifying data quality issues, and guiding preprocessing and modeling decisions.

01

TREND ANALYSIS

Exploring the trends in ticket prices based on factors such as time of year, day of the week, and destination.

02

CORRELATION EXAMINATION

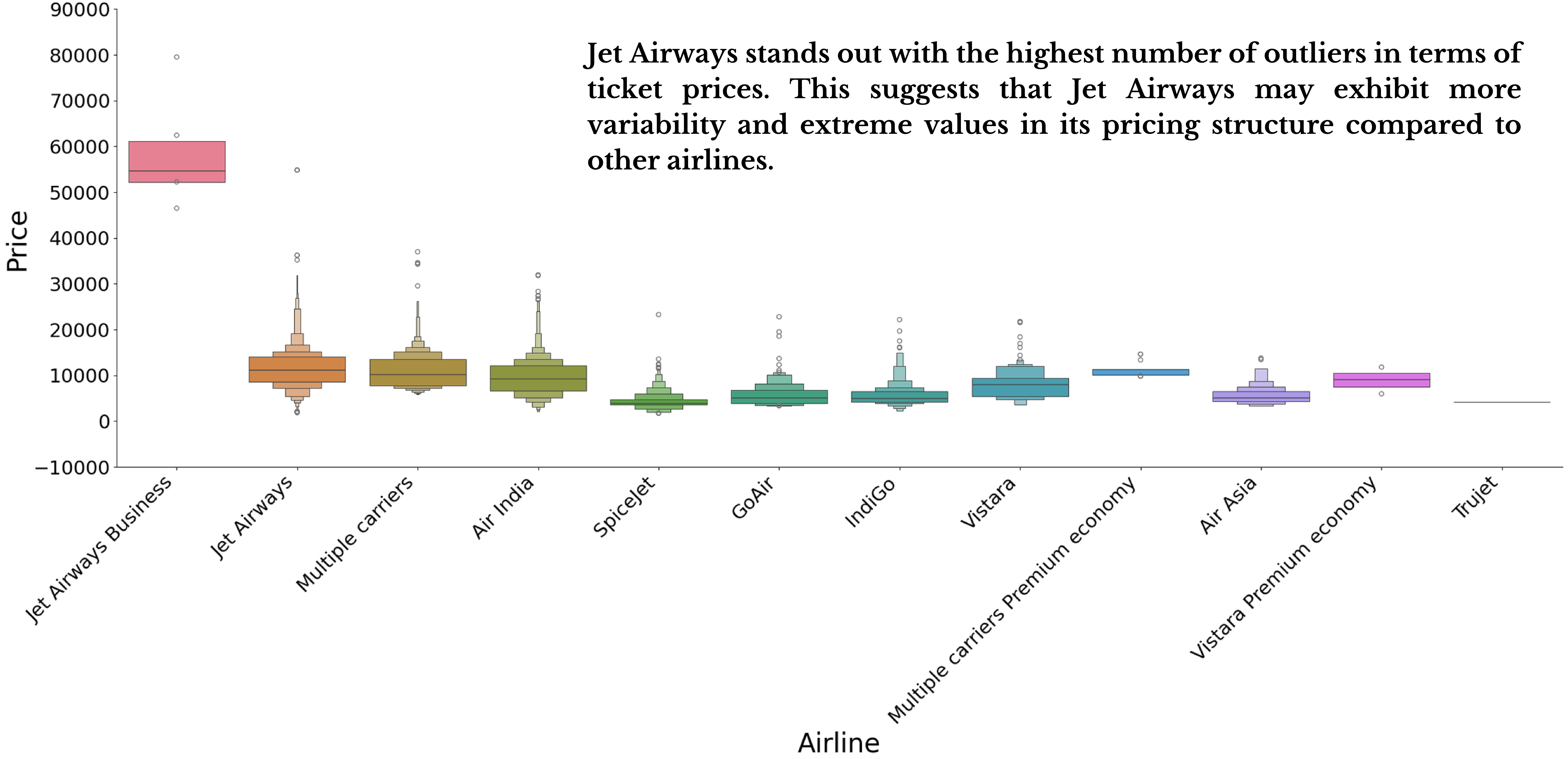
Investigating the relationships between ticket prices and variables like departure time, airline, and route.

03

OUTLIER DETECTION

Identifying and understanding anomalies in ticket prices or features that could impact the modeling process.

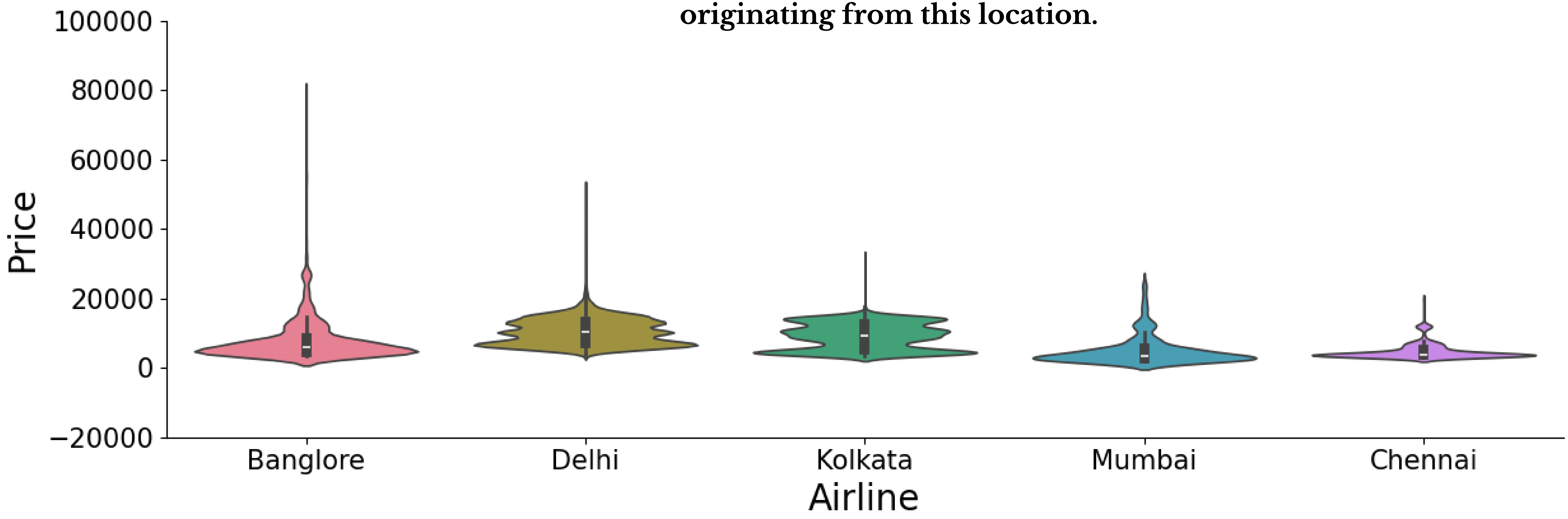
PRICE VS AIRLINE PLOT



Jet Airways stands out with the highest number of outliers in terms of ticket prices. This suggests that Jet Airways may exhibit more variability and extreme values in its pricing structure compared to other airlines.

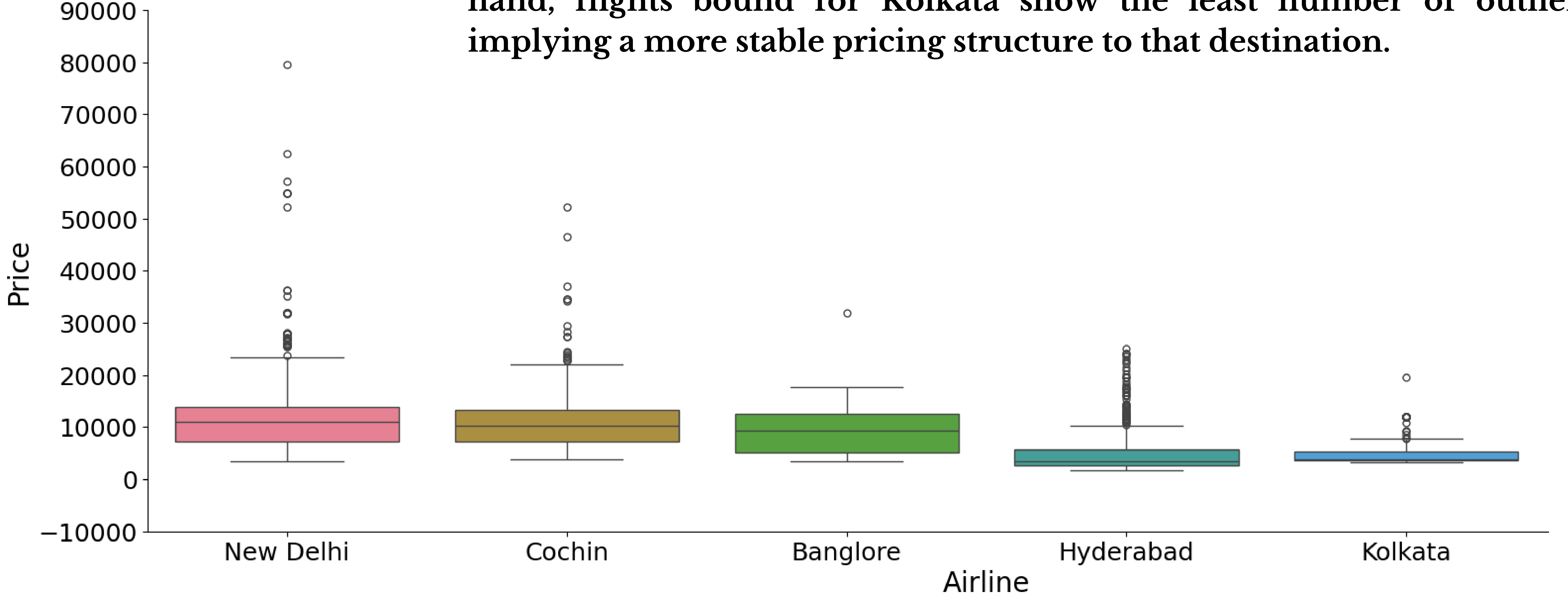
PRICE VS SOURCE PLOT

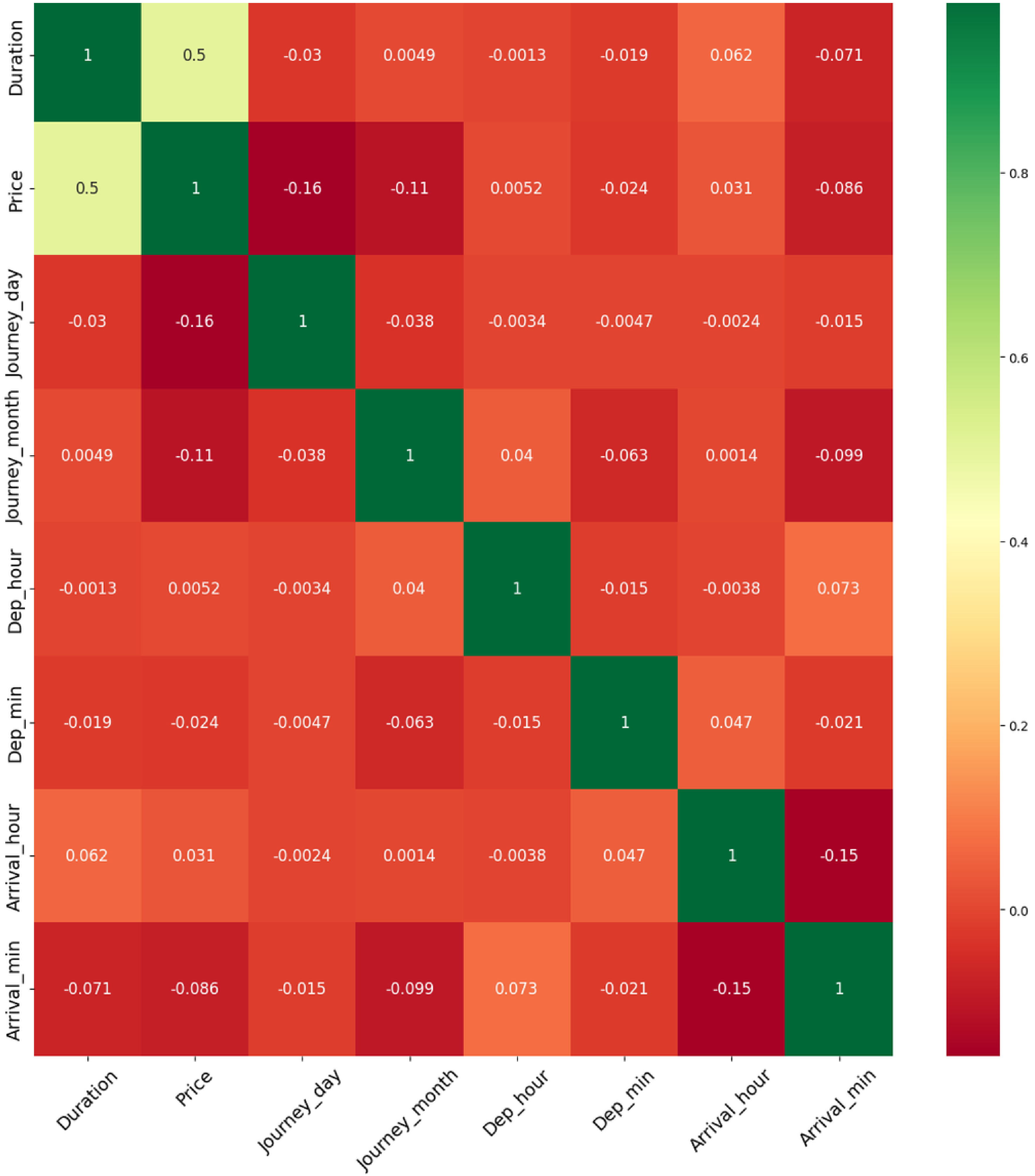
Flights departing from Bangalore exhibit the most outliers, indicating a wider range of ticket prices. Conversely, Chennai stands out with the least number of outliers, suggesting a more consistent pricing pattern for flights originating from this location.



PRICE VS DESTINATION PLOT

Flights heading to New Delhi experience the highest number of outliers in ticket prices, indicating potential pricing variations. On the other hand, flights bound for Kolkata show the least number of outliers, implying a more stable pricing structure to that destination.





CORRELATION HEATMAP

The heatmap suggests that price is primarily influenced by factors within the flight itself, like duration and possibly departure time and airline, rather than external factors like day or month of travel.

REGRESSION MODELS

1. RIDGE REGRESSION

- Ridge Regression, a regularized linear regression technique, was chosen for its effectiveness in addressing multicollinearity within the dataset.
- By mitigating multicollinearity issues, Ridge Regression provides more stable coefficient estimates and better handles scenarios where predictor variables exhibit high correlation.

2. LASSO REGRESSION

- In scenarios where multicollinearity is present, Lasso Regression not only mitigates the issues associated with correlated predictors but also possesses the unique ability to encourage sparsity in the model.
- Lasso Regression acts as an effective feature selection mechanism, identifying and emphasizing the most influential variables while disregarding less impactful ones.

3. DECISION TREE REGRESSION

- In contrast to linear models, Decision Trees can represent and model non-linear patterns and interactions among features.
- This non-linear flexibility allows Decision Tree Regression to adapt well to datasets with non-linear relationships, making it particularly suitable for scenarios where the target variable's dependence on predictor variables involves intricate and nonlinear patterns.

MODEL COMPARISON

MODEL	RIDGE REGRESSOR	LASSO REGRESSOR	DECISION TREE
TRAIN RMSE	3558.67	3560.85	370.82
TRAIN MAPE(%)	32	32	1
TRAIN R-SQUARED	0.4151	0.4143	0.9936
TEST RMSE	3457.60	3459.38	1949.12
TEST MAPE(%)	32	32	9
TEST R-SQUARED	0.4244	0.4238	0.8171

RESULTS

SNO.	ACTUAL VALUES	PREDICTED VALUES
1.	17996	16856.79
2.	3873	3971.46
...	......	......
3139.	4823	4244.76
3140.	15129	15646.54

CONCLUSION

Regression Models Evaluation

Successfully implemented and evaluated various regression models, including Ridge Regression, Lasso Regression, and Decision Tree Regression.

Insights into Flight Prices

Provided valuable insights into the complex factors influencing flight prices, contributing to a deeper understanding of pricing dynamics.

Predictive Accuracy Metrics

Demonstrated the predictive accuracy of the models through metrics such as Root Mean Squared Error and R-Squared.

FUTURE ENHANCEMENT

Integration of Real-Time Data Feeds
Enhance model accuracy by integrating real-time data for dynamic market changes.

Advanced ML Techniques
Explore ensemble methods and deep learning for intricate pattern recognition and improved predictions..

Refined Feature Engineering
Include seasonal trends, airline-specific factors, and geopolitical events to broaden model considerations.

Web App or API Deployment
Offer a user-friendly platform for accessing real-time flight price predictions and insights.

Continuous Monitoring and Retraining
Establish regular protocols for monitoring and retraining to ensure sustained accuracy in the evolving aviation industry.

THANK YOU