



FLIGHT PRICE PREDICTION



Author: Marife Ramoran



SUMMARY

- This project endeavors to forecast domestic flight prices within India for the period spanning March 2019 to July 2019, employing linear regression and data analysis methodologies. The objective is to deliver precise predictions, aiding in strategic planning and budgeting for airlines regarding their flights.



OUTLINE

- Business Problem
- Data & Methods
- Results
- Conclusion





BUSINESS PROBLEM

- For business owners in the travel industry, the challenge lies in the absence of a robust predictive model that utilizes linear regression and data analysis techniques. This deficiency not only impedes the optimal allocation of travel budgets but also erodes the satisfaction and confidence of travelers in the planning process. Thus, our business problem centers on the critical necessity to address this gap by developing an advanced predictive tool. This solution aims to empower business owners with precise flight price forecasts, thereby revolutionizing the landscape of travel planning and enhancing overall customer satisfaction.

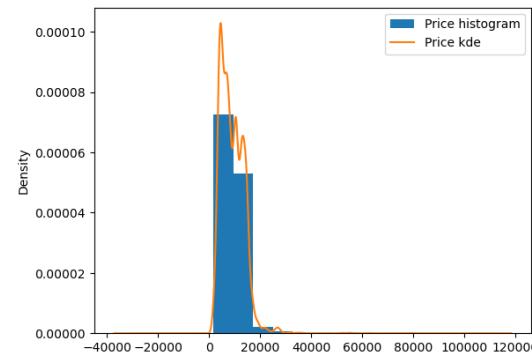
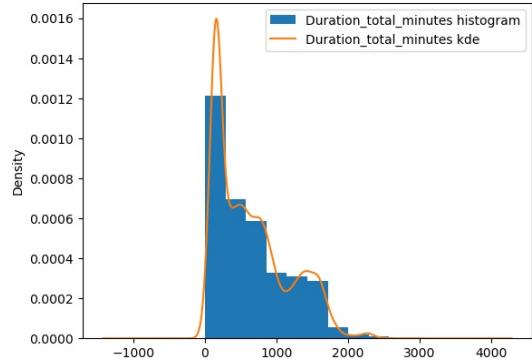
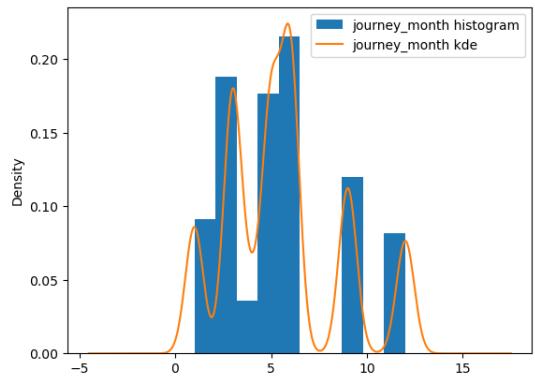




DATA & METHODS

Utilizing visualization techniques, we will analyze the distribution of flight prices, discern trends, and investigate potential correlations among various features. This phase is instrumental in guiding informed decisions related to feature selection and gaining insights into the inherent patterns within the dataset. The dataset comprises 10,682 entries encompassing comprehensive details about flights, including information about the airline, number of stops, month of journey, and duration.

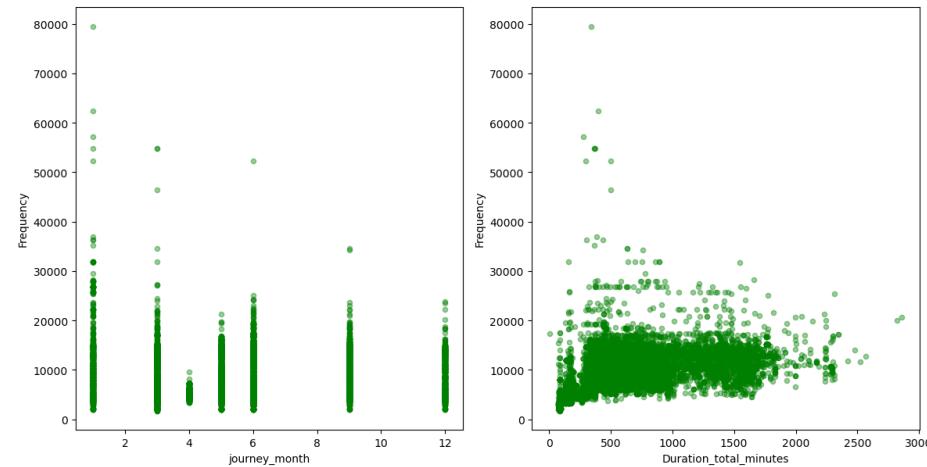




OLS Regression Results

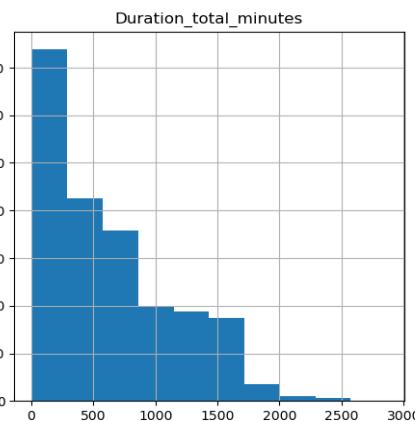
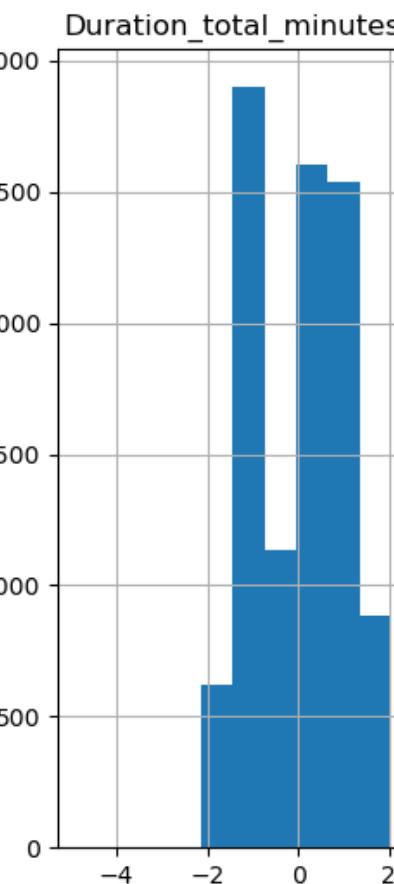
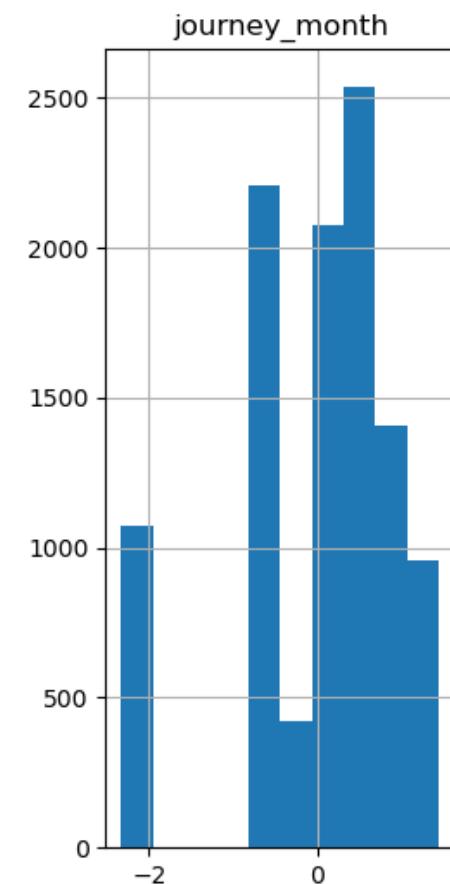
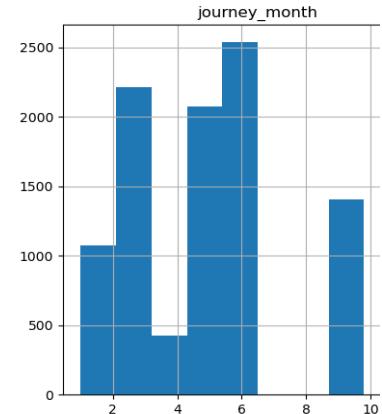
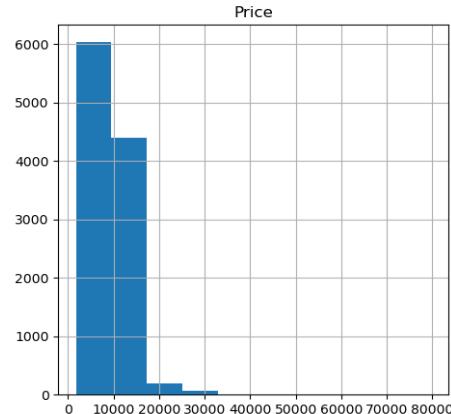
Dep. Variable:	Price	R-squared:	0.601
Model:	OLS	Adj. R-squared:	0.601
Method:	Least Squares	F-statistic:	730.9
Date:	Mon, 18 Dec 2023	Prob (F-statistic):	0.00
Time:	10:37:49	Log-Likelihood:	-1.0036e+05
No. Observations:	10682	AIC:	2.008e+05
Df Residuals:	10659	BIC:	2.009e+05
Df Model:	22		
Covariance Type:	nonrobust		





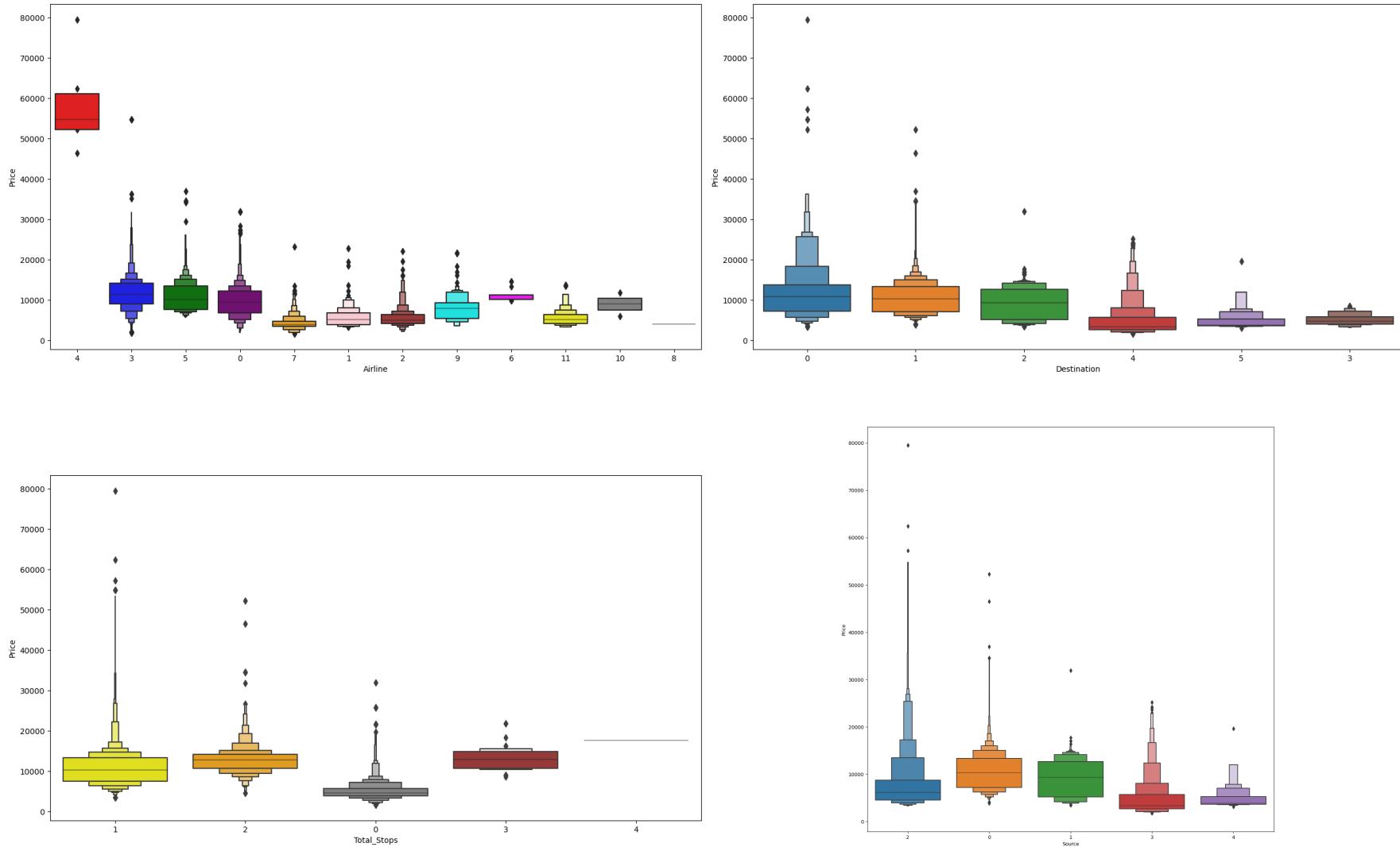
OLS Regression Results

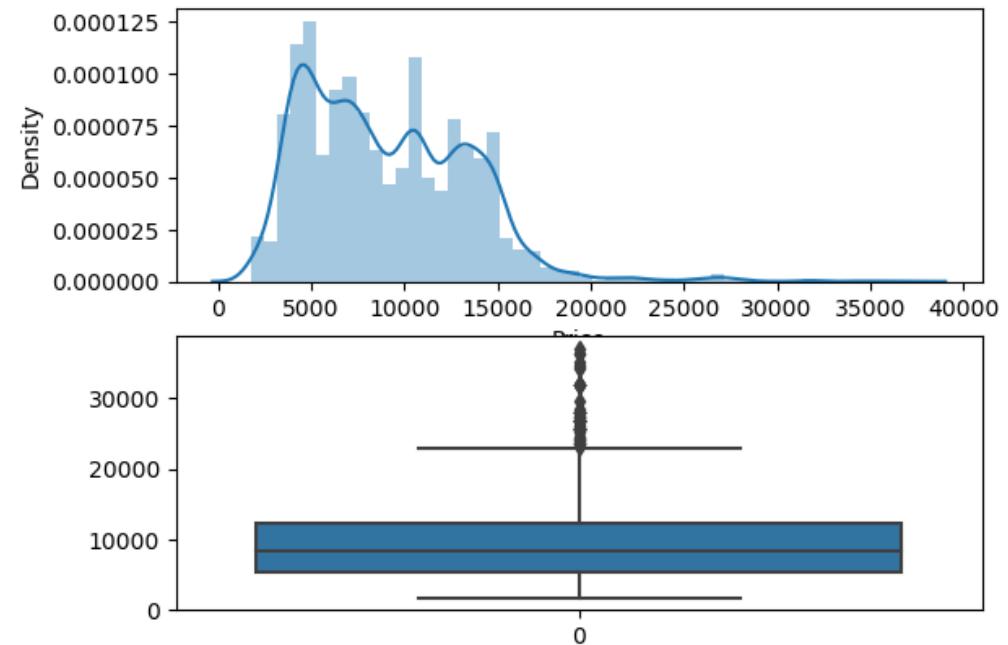
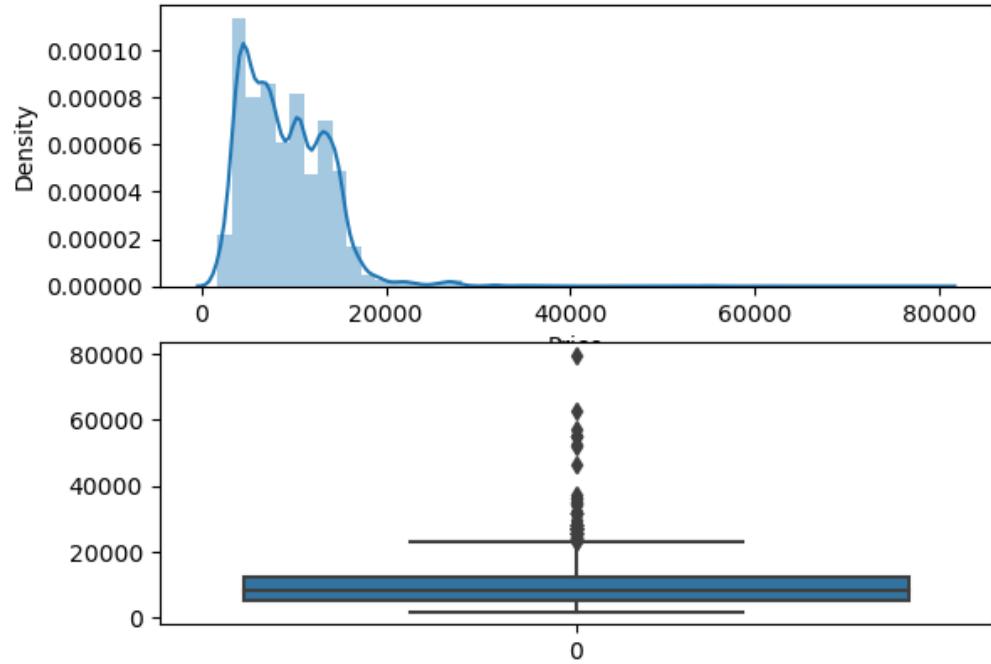
Dep. Variable:	Price	R-squared:	0.601
Model:	OLS	Adj. R-squared:	0.601
Method:	Least Squares	F-statistic:	1463.
Date:	Mon, 18 Dec 2023	Prob (F-statistic):	0.00
Time:	08:58:59	Log-Likelihood:	-2.0072e+05
No. Observations:	21364	AIC:	4.015e+05
Df Residuals:	21341	BIC:	4.017e+05
Df Model:	22		
Covariance Type:	nonrobust		

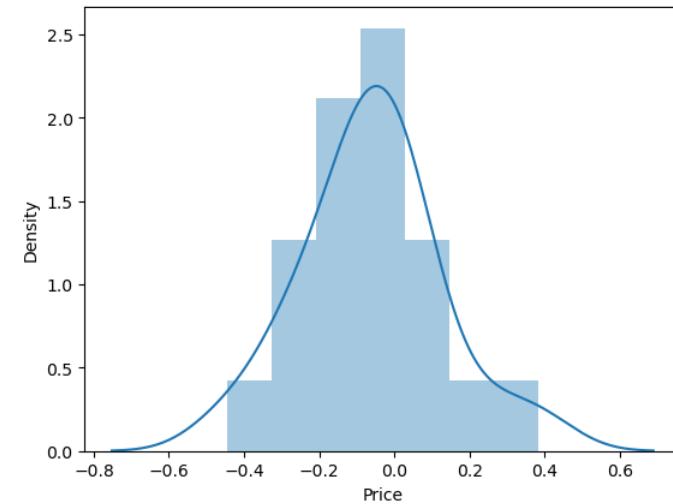
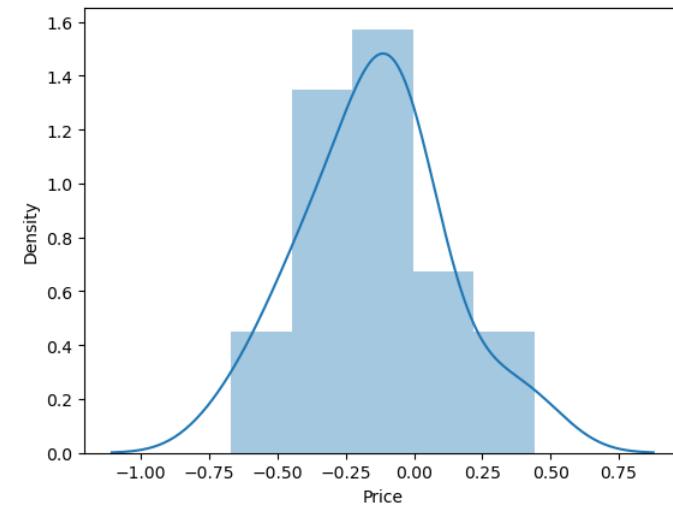
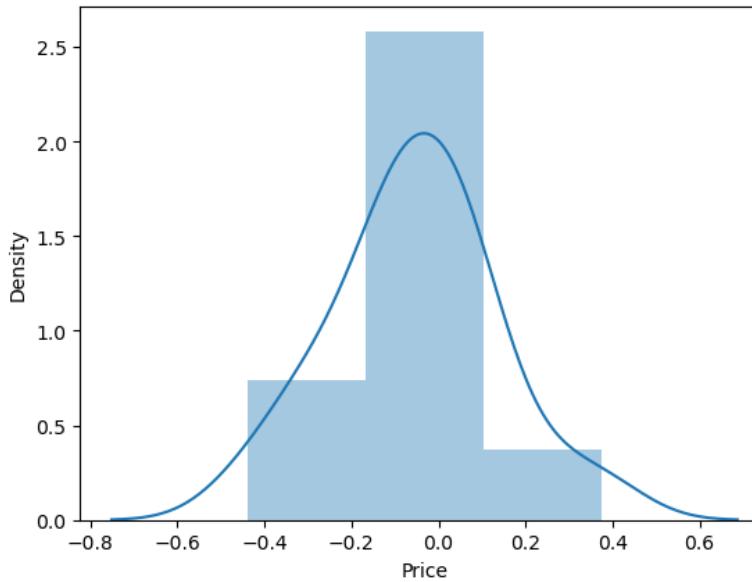
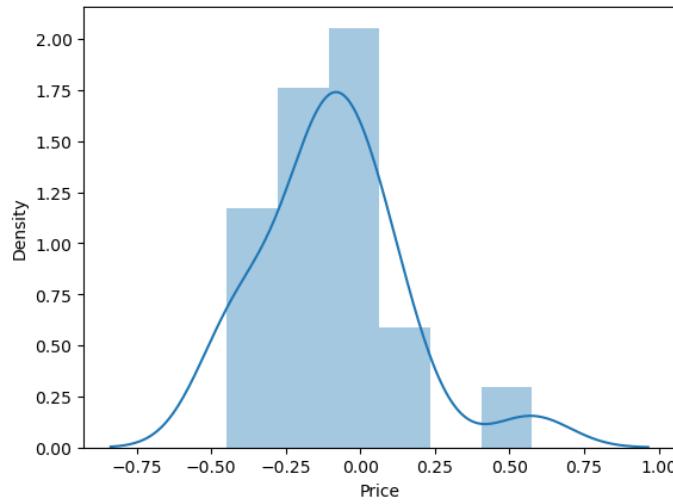
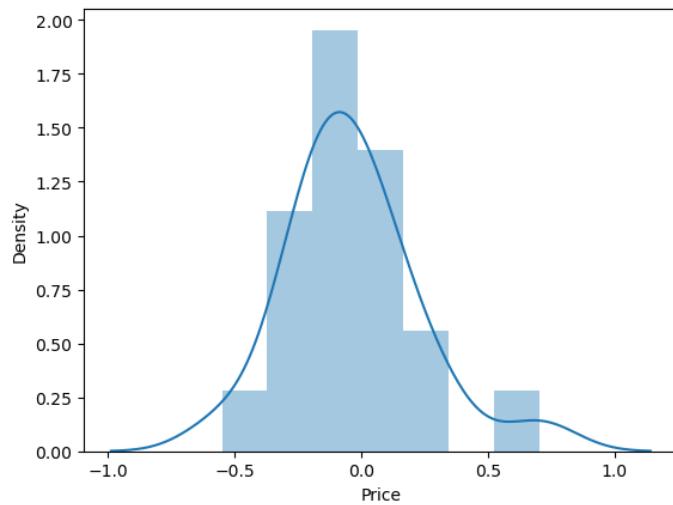


DATA & METHODS

- The visualizations depict the distribution of flight prices based on total stops, source cities, destinations, and airlines. Key observations include a decreasing median price with increasing stops, varied city-wise median prices, destination-specific patterns, and diverse airline behaviors. These visuals offer valuable insights into price trends and potential outliers.

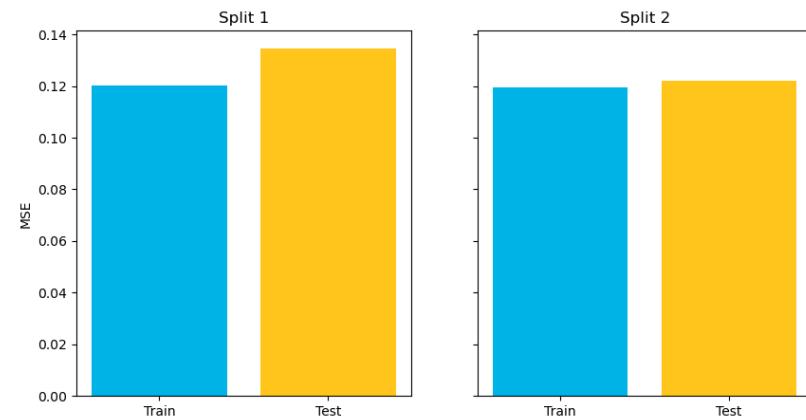






OLS Regression Results

Dep. Variable:	Price	R-squared:	0.586
Model:	OLS	Adj. R-squared:	0.585
Method:	Least Squares	F-statistic:	686.7
Date:	Mon, 18 Dec 2023	Prob (F-statistic):	0.00
Time:	10:37:54	Log-Likelihood:	-1.0003e+05
No. Observations:	10682	AIC:	2.001e+05
Df Residuals:	10659	BIC:	2.003e+05
Df Model:	22		
Covariance Type:	nonrobust		



```
r2_score(y_test,prediction)
```

0.8566013690075043



RESULT

- **Duration_total_minutes:**
- Given its substantial importance, prioritize strategies and optimizations related to Duration_total_minutes, as it strongly influences the prediction of the target variable.
- **Airline:**
- Acknowledge the significant impact of the Airline feature and consider tailored approaches for airlines to enhance predictive outcomes.
- **Destination:**
- Recognize the importance of the Destination feature; tailor marketing or pricing strategies based on destination-specific insights to improve overall predictive accuracy.



RESULT

- **Total_Stops:**
 - Emphasize the relevance of Total_Stops in predicting the target variable, integrating insights related to the number of stops into decision-making processes.
- **Source:**
 - Acknowledge the moderate importance of the Source feature; consider optimizing strategies related to source locations to contribute to predictive accuracy.
- **journey_month:**
 - While slightly less influential, journey_month remains informative. Leverage this feature in conjunction with others for a comprehensive understanding and strategic decision-making.
- These recommendations aim to guide business actions based on the importance of each feature in predicting the target variable. Prioritizing efforts in line with feature importance can enhance the overall effectiveness of predictive models.



CONCLUSION

- The initial analysis and modeling provided valuable insights into the factors influencing the flight prices.
- Feature importance analysis highlighted the critical role of 'Duration_total_minutes,' 'Airline,' and 'Destination' in predicting prices.
- The Random Forest model, after hyperparameter tuning, demonstrated enhanced predictive performance.
- Ongoing evaluation and refinement of models are crucial for ensuring accurate predictions and addressing potential overfitting.





THANK YOU!

Email: mariferamoran@gmail.com

Github: [@marifermrn](https://github.com/marifermrn)

LinkedIn: www.linkedin.com/in/marifermrn

