



# CRACKING THE CODE OF FLIGHT PRICES

A SMART SOLUTION FOR TRAVELERS & AIRLINES

Presented By : **Group 01**

Thenuka Yatawara

Sanduni Fonseka

Vayani Kavindya

Supul Wickramasinghe



[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

# TWO PERSPECTIVES ONE CHALLENGE

Flight prices constantly fluctuate, creating confusion and uncertainty for both travelers and airlines.

- For Travelers – They struggle to find the right time to book, fearing they'll overpay or miss a better deal.
- For Airlines – They face challenges in setting the optimal ticket prices that attract customers while maximizing revenue

[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

## INTRODUCING THE PASSENGER'S PROBLEM

# MEET SARAH

## A FREQUENT BUSINESS TRAVELER

Sarah is a marketing executive who frequently travels for work and leisure.



[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

One day, she waits too long and the price jumps from \$300 to \$500. The next time, she books early, but a week later, the price drops to \$250.



## SARAH'S FRUSTRATION:

- "Why do ticket prices change so much?"
- "When is the best time to book?"
- "How can I save money without gambling on price fluctuations?"



[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

## INTRODUCING THE AIRLINE'S PROBLEM

# MEET DAVID THE AIRLINE OWNER

David owns SkyConnect Airlines, a mid-sized airline competing with larger carriers.



[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

His biggest challenge? Setting the right ticket price competing with the other airlines too high, and customers leave; too low, and he loses revenue



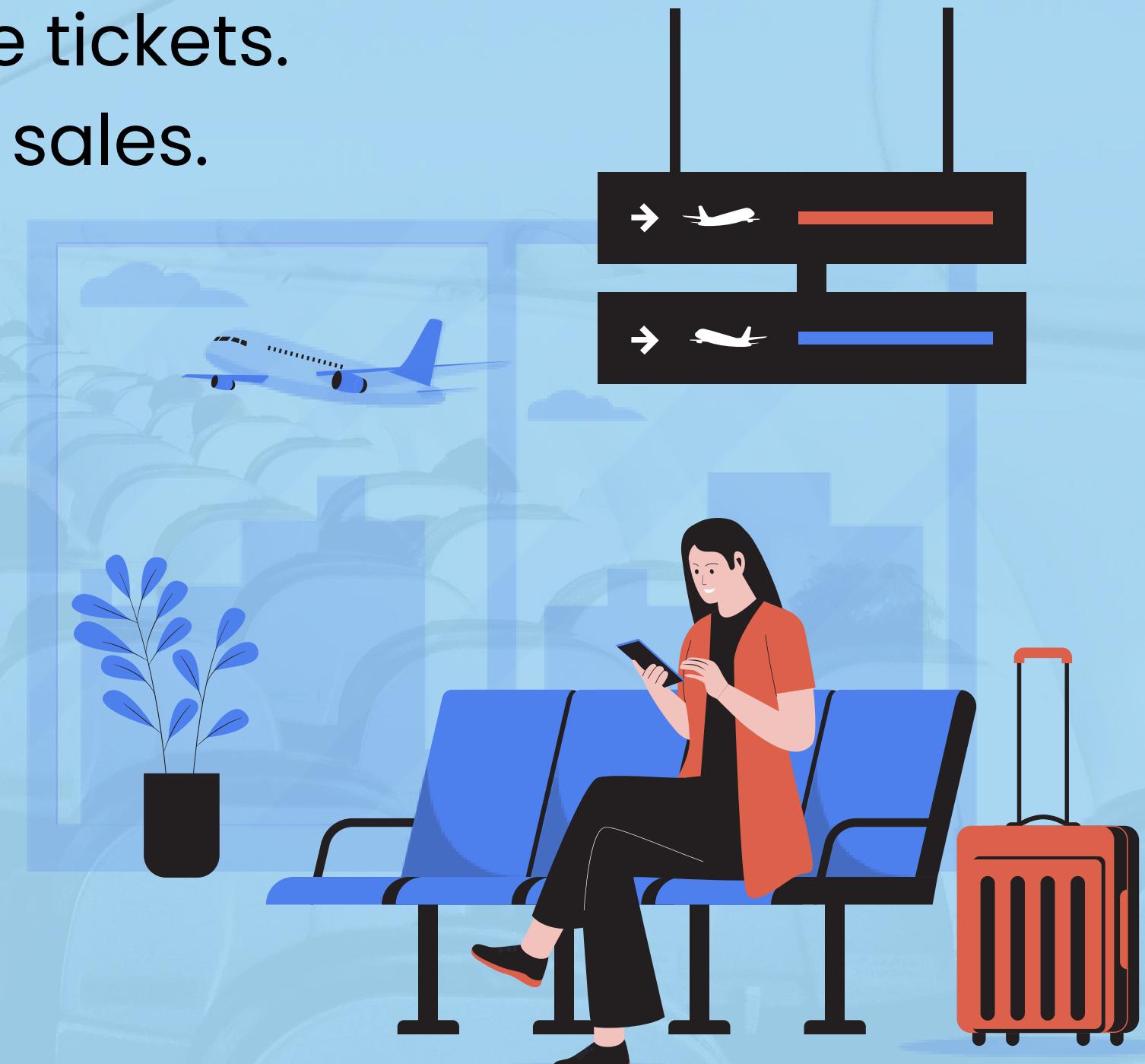
## DAVID'S STRUGGLE

- "How can I price flights to maximize revenue?"
- "What if I drop prices too soon and lose money?"
- "How do I stay competitive without undercutting my business?"



# WHAT DO SARAH & DAVID HAVE IN COMMON?

- Sarah wants affordable tickets.
- David wants profitable sales.
- But both struggle with unpredictable pricing.



[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

# WHAT IF WE COULD

**Help Sarah** find the best time to book her flights?

**Help David** set optimal prices that maximize both sales and revenue?

**BY USING A DATA-DRIVEN SOLUTION**



[HOME](#)[ABOUT](#)[\*\*CONTENT\*\*](#)[OTHERS](#)

# ADVANCED ANALYSIS

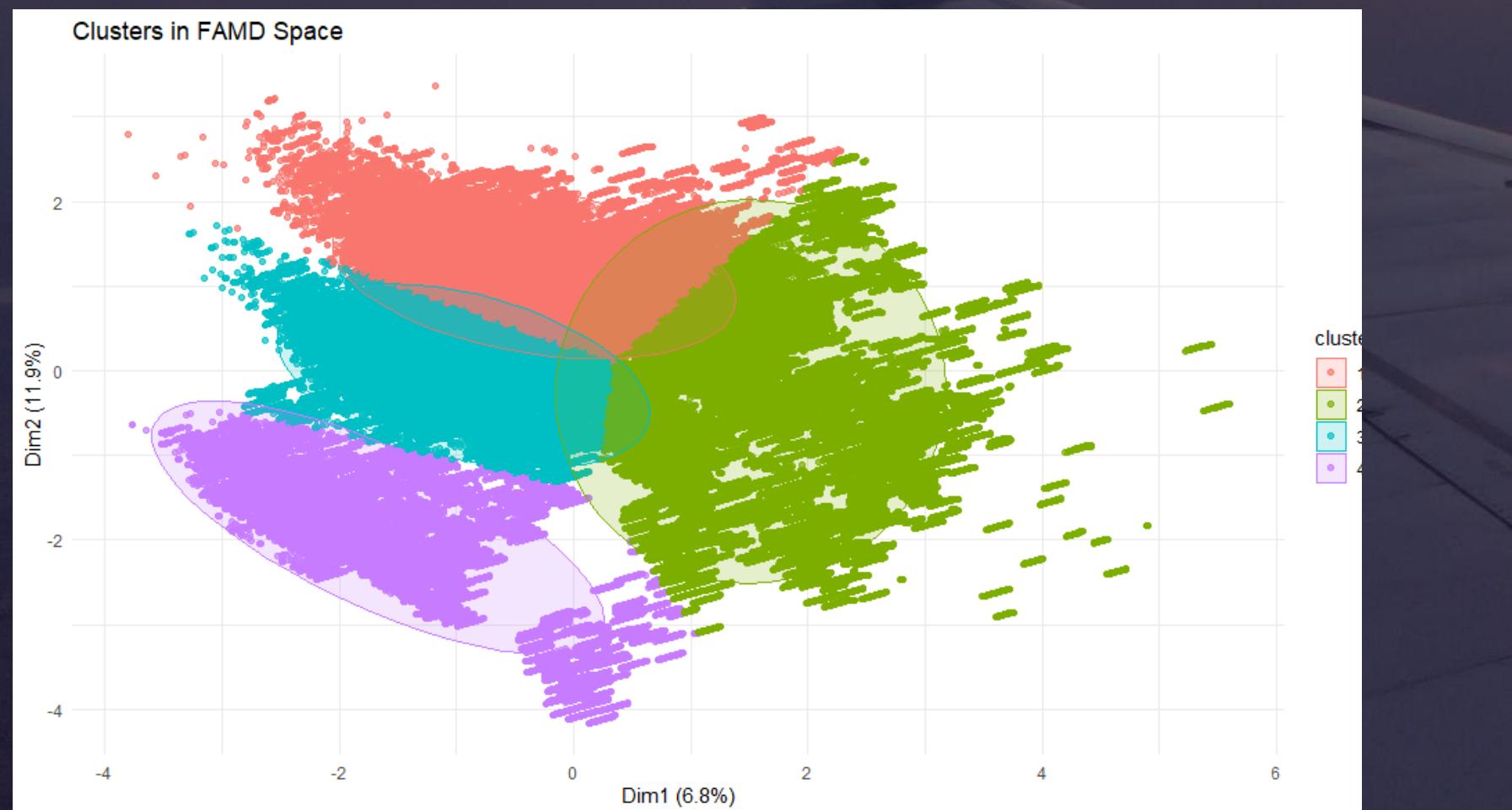
To solve the mystery of flight prices, we analyzed data with 10 key factors from Kaggle and tested different models to find the best one.

[KNOW MORE](#)

[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

# IMPORTANT RESULT FROM EDA ANALYSIS

## RESULT FROM FAMD CLUSTER ANALYSIS



MEAN SILHOUETTE WIDTH: 0.33

THERE ARE SOME  
OVERLAPPING BETWEEN  
CLUSTERS.

COULDN'T FIND OUT ANY  
KIND OF SEPECIFIC FEATURES  
IN CLUSTERS.

[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

# MULTIPLE LINEAR REGRESSION

Training

Testing

RMSE

6752.55

6760.42

$R^2$

0.9115

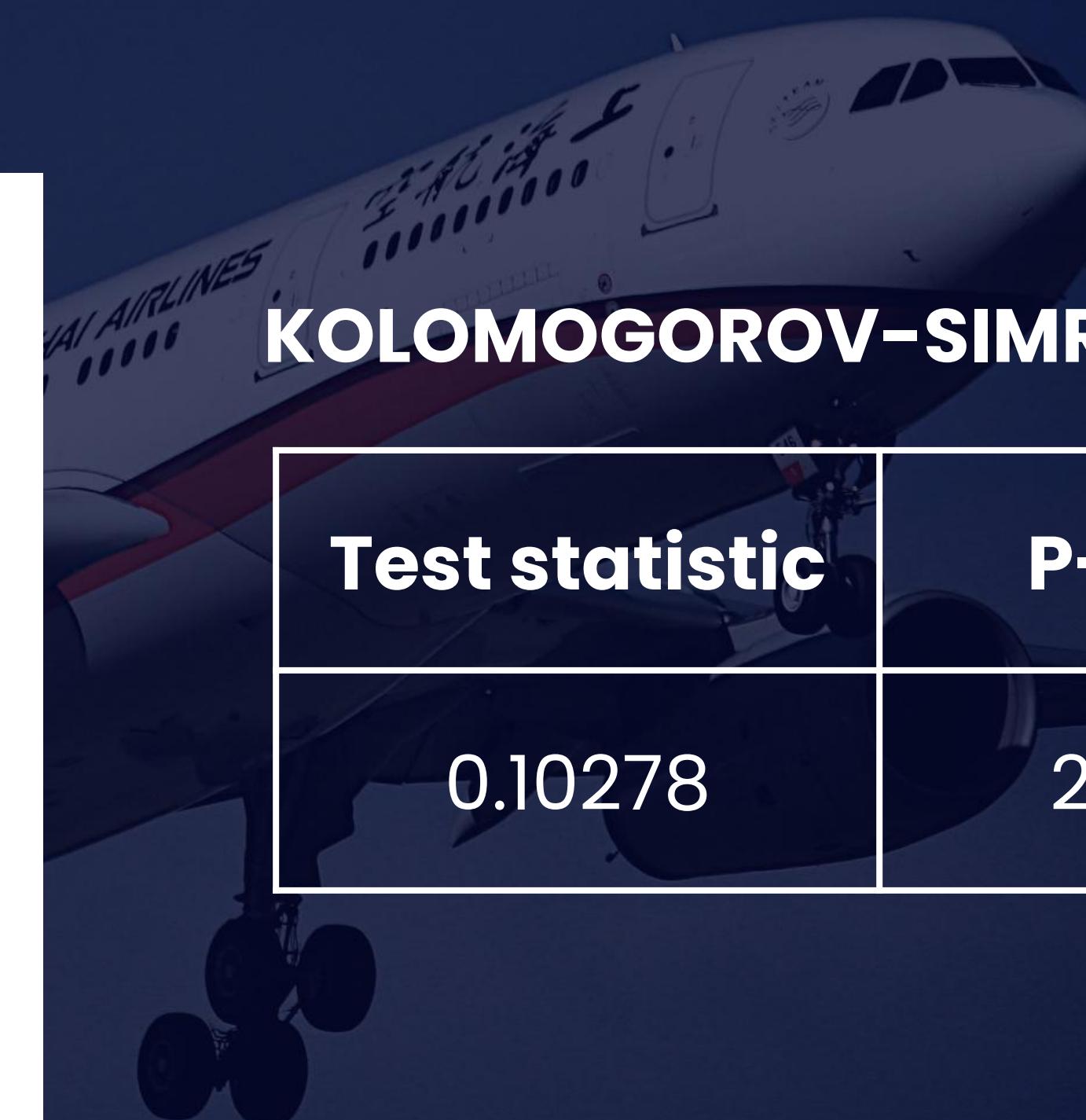
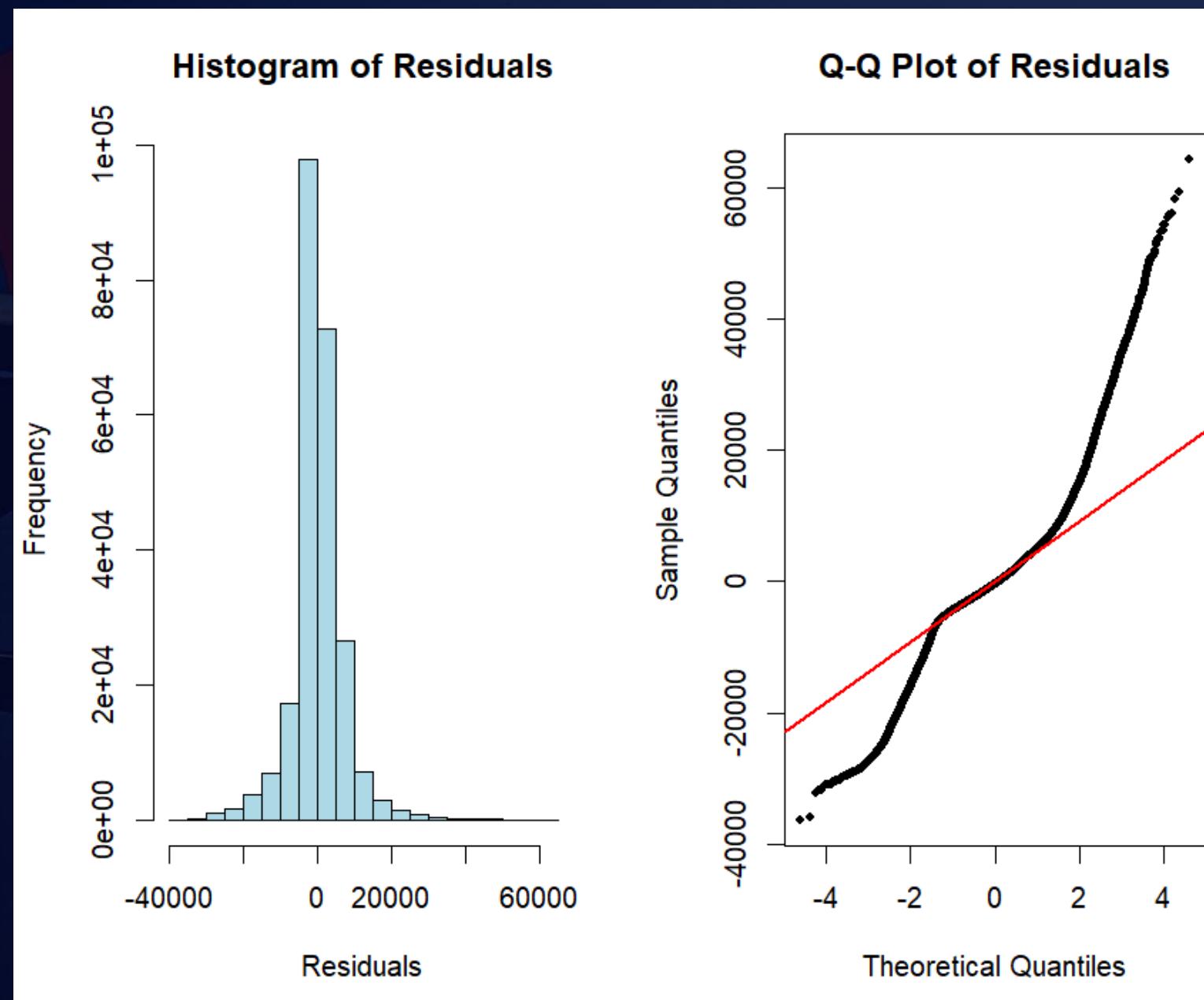
0.9113

BEST SUBSET

[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

# CHECKING ASSUMPTIONS

## NORMALITY



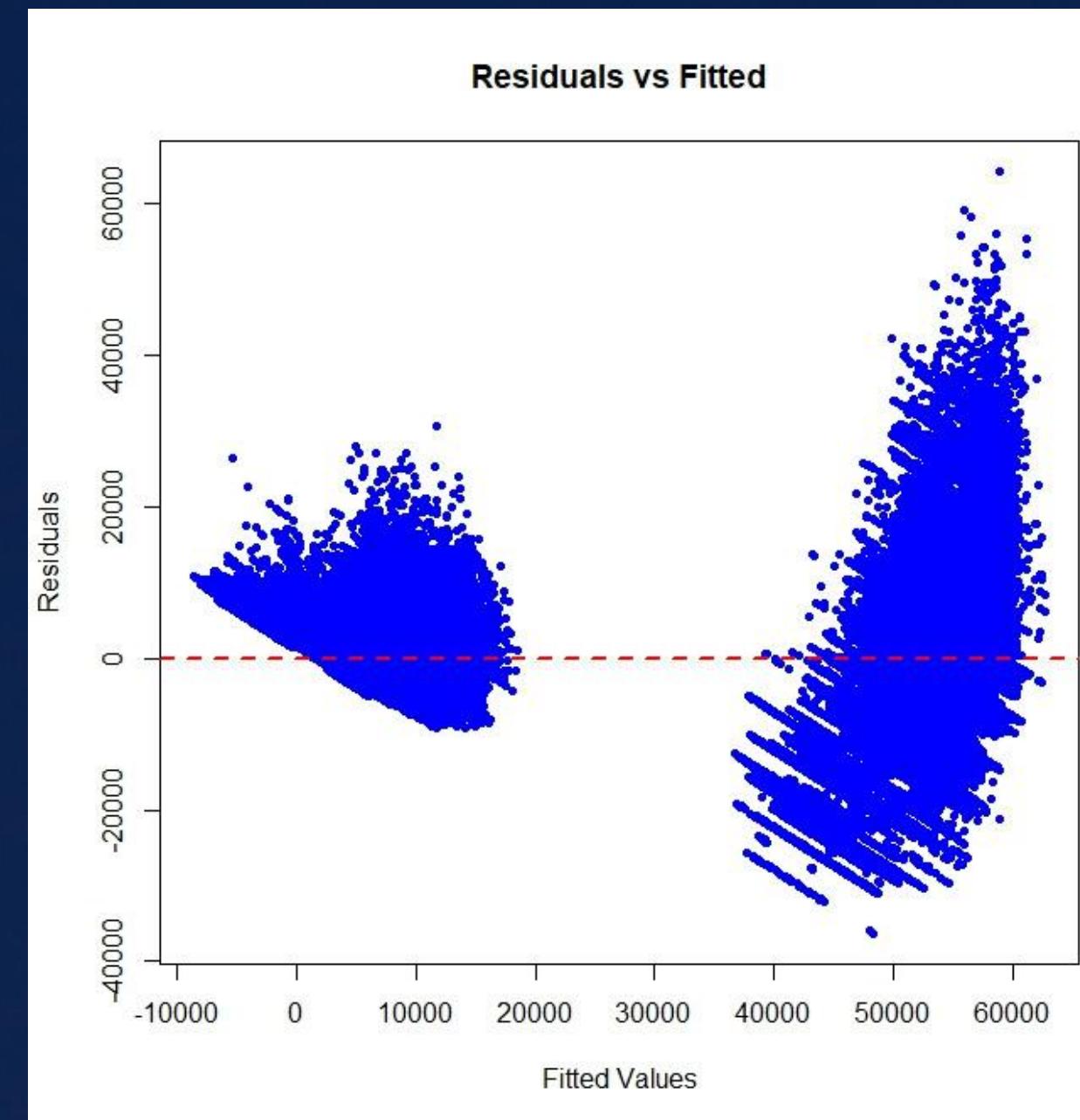
## KOLOMOGOROV-SIMRNOV TEST

Test statistic	P-value
0.10278	2.2e-16

[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

# CHECKING ASSUMPTIONS

## LINEARITY & HOMOSCEDASTICITY



## INDEPENDENCY

Test statistic	P-value
0.39189	2.2e-16

## DURBIN-WATSON TEST



# LASSO REGRESSION

## WHY WE NEED LASSO?

- BETTER HANDLING OF NON-NORMAL RESIDUALS.
- IMPROVES MODEL INTERPRETABILITY AND FLEXIBILITY.

	Training	Testing
RMSE	6754.84	6762.54
$R^2$	0.9114	0.9113

[HOME](#)[ABOUT](#)[\*\*CONTENT\*\*](#)[SOLUTION](#)

# WHY WE NEED TREE-BASED METHODS

- **CAPTURING NON-LINEAR RELATIONSHIPS.**
- **MORE ROBUST TO OUTLIERS.**
- **SUPERIOR PREDICTIVE POWER.**

[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

# TREE-BASED MODELS

Model	Training		Testing	
	RMSE	$R^2$	RMSE	$R^2$
<b>Rando m Forest</b>	3517.57	0.9760	3636.75	0.9742
<b>XG Boost</b>	3408.66	0.9774	3501.30	0.9762



# MODEL COMPARISON

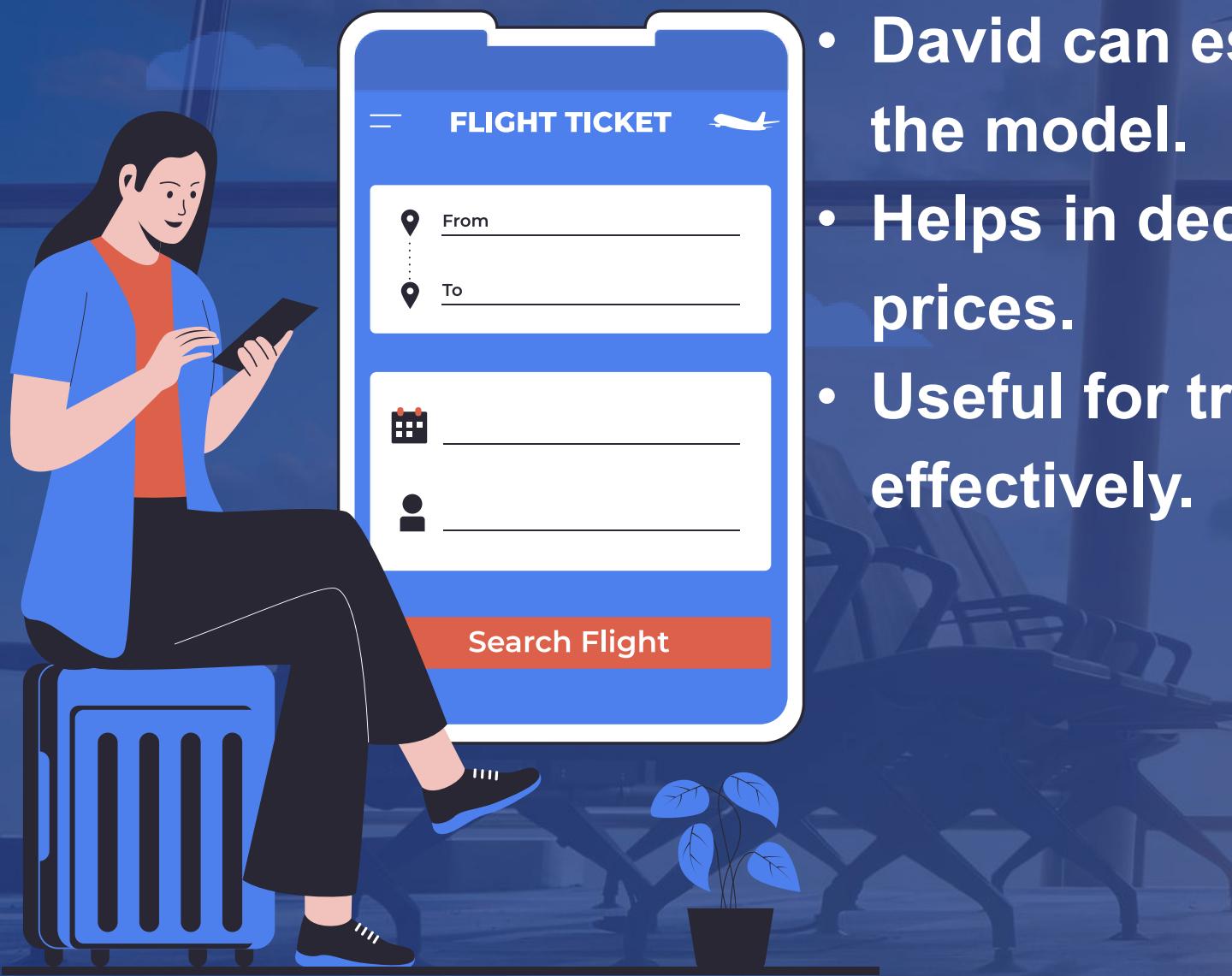
	Training		Testing	
	RMSE	$R^2$	RMSE	$R^2$
<b>MLR</b>	6752.55	0.9115	6760.42	0.9113
<b>Lasso</b>	6754.84	0.9114	6762.54	0.9113
<b>Random Forest</b>	3517.57	0.9760	3636.75	0.9742
<b>XGBoost</b>	3408.66	0.9774	3501.30	0.9762

BEST MODEL XG BOOST

[HOME](#)[ABOUT](#)[CONTENT](#)[SOLUTION](#)

# NOW USING THIS BEST MODEL WE CAN HELP BOTH SARAH AND DAVID

- Sarah can use the model to set feature values and predict her flight price.
- David can estimate competitors' flight prices using the model.
- Helps in decision-making based on predicted prices.
- Useful for travelers and businesses to plan effectively.





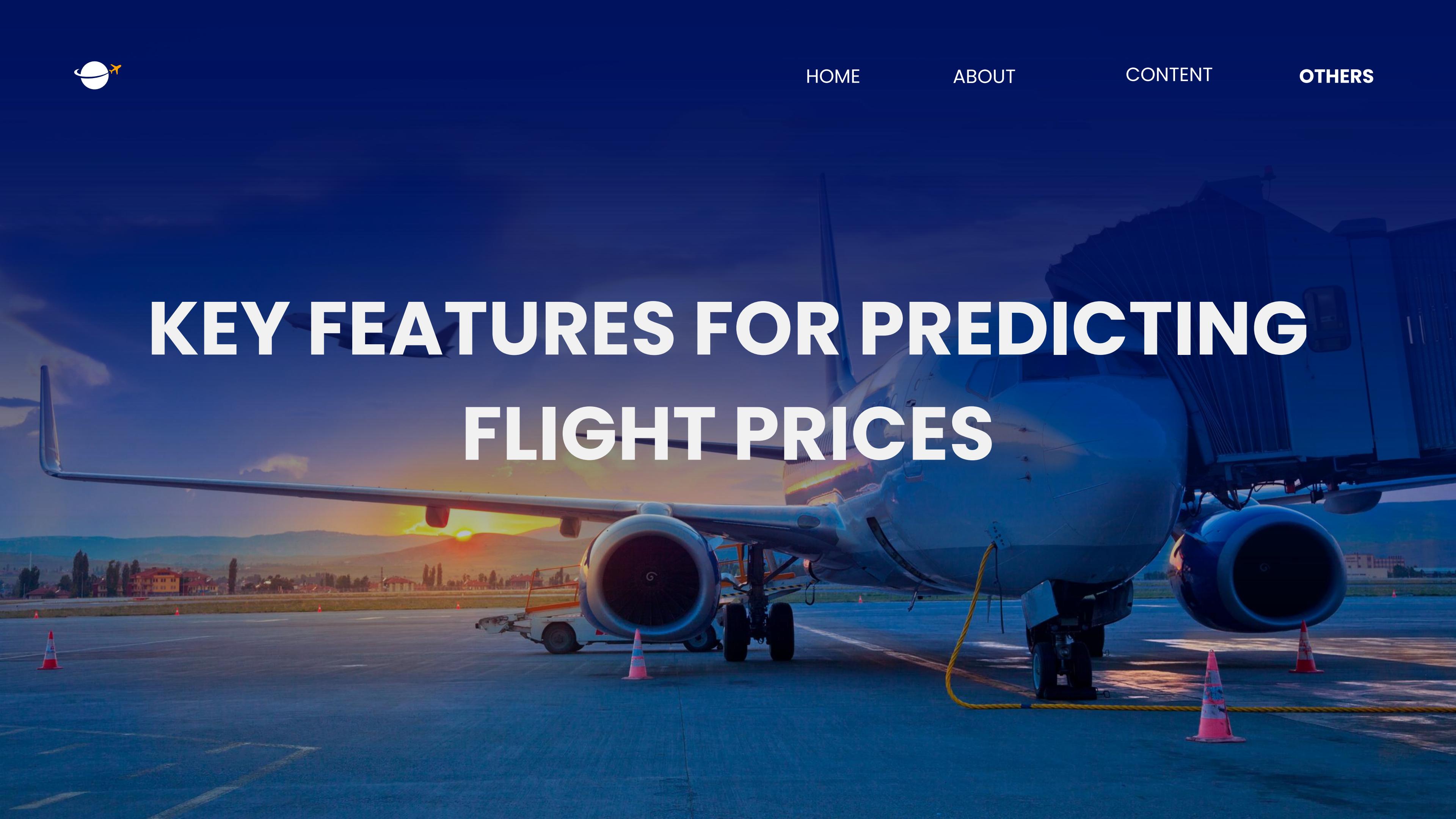
HOME

ABOUT

CONTENT

OTHERS

# KEY FEATURES FOR PREDICTING FLIGHT PRICES

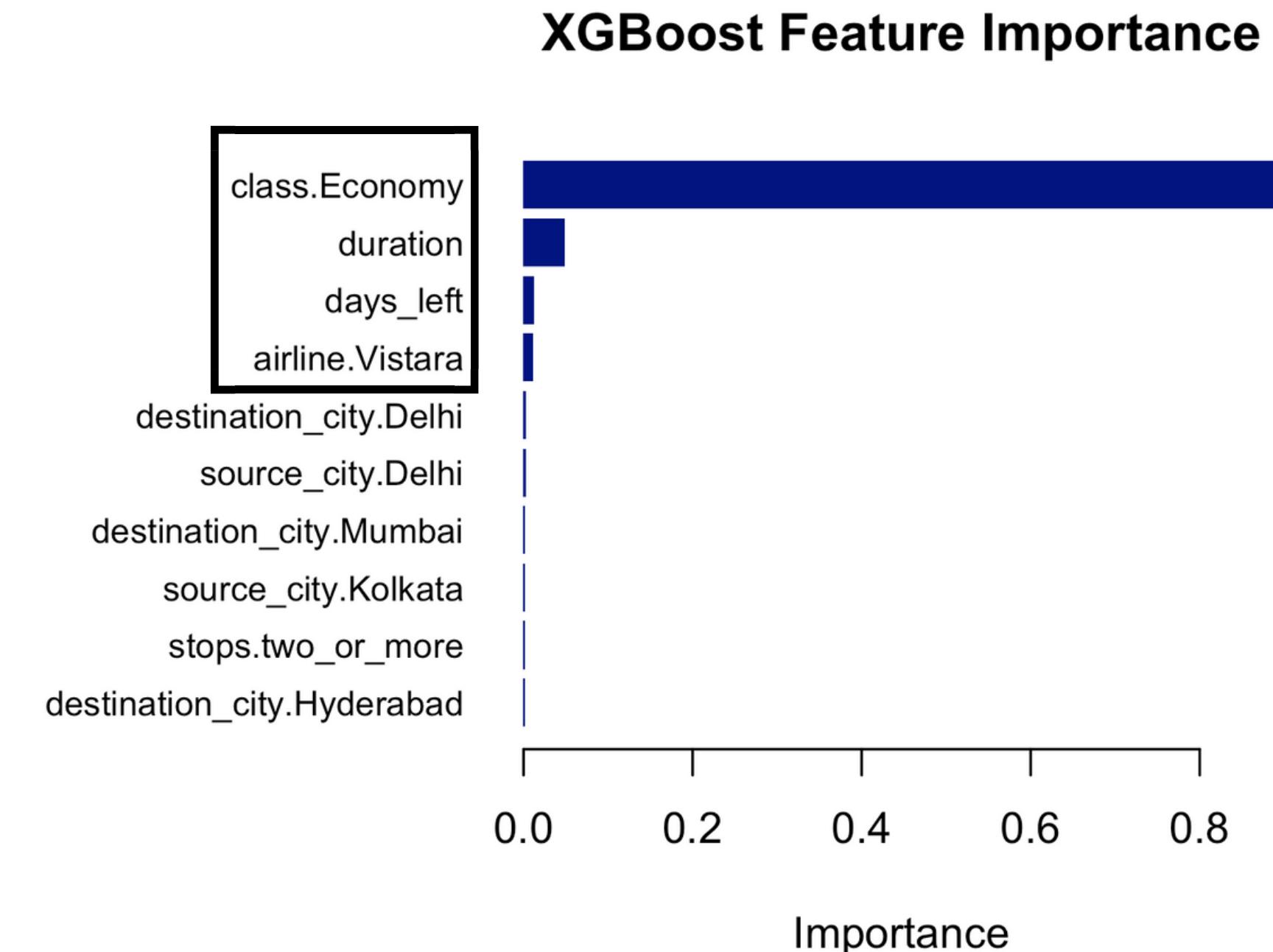


[HOME](#)[ABOUT](#)[CONTENT](#)[OTHERS](#)

# XG BOOST VARIABLE IMPORTANCE

MORE INFLUENTIAL  
FACTORS

**class.Economy**  
**duration**  
**days\_left**  
**airline.Vistara**



[HOME](#)[ABOUT](#)[CONTENT](#)[OTHERS](#)

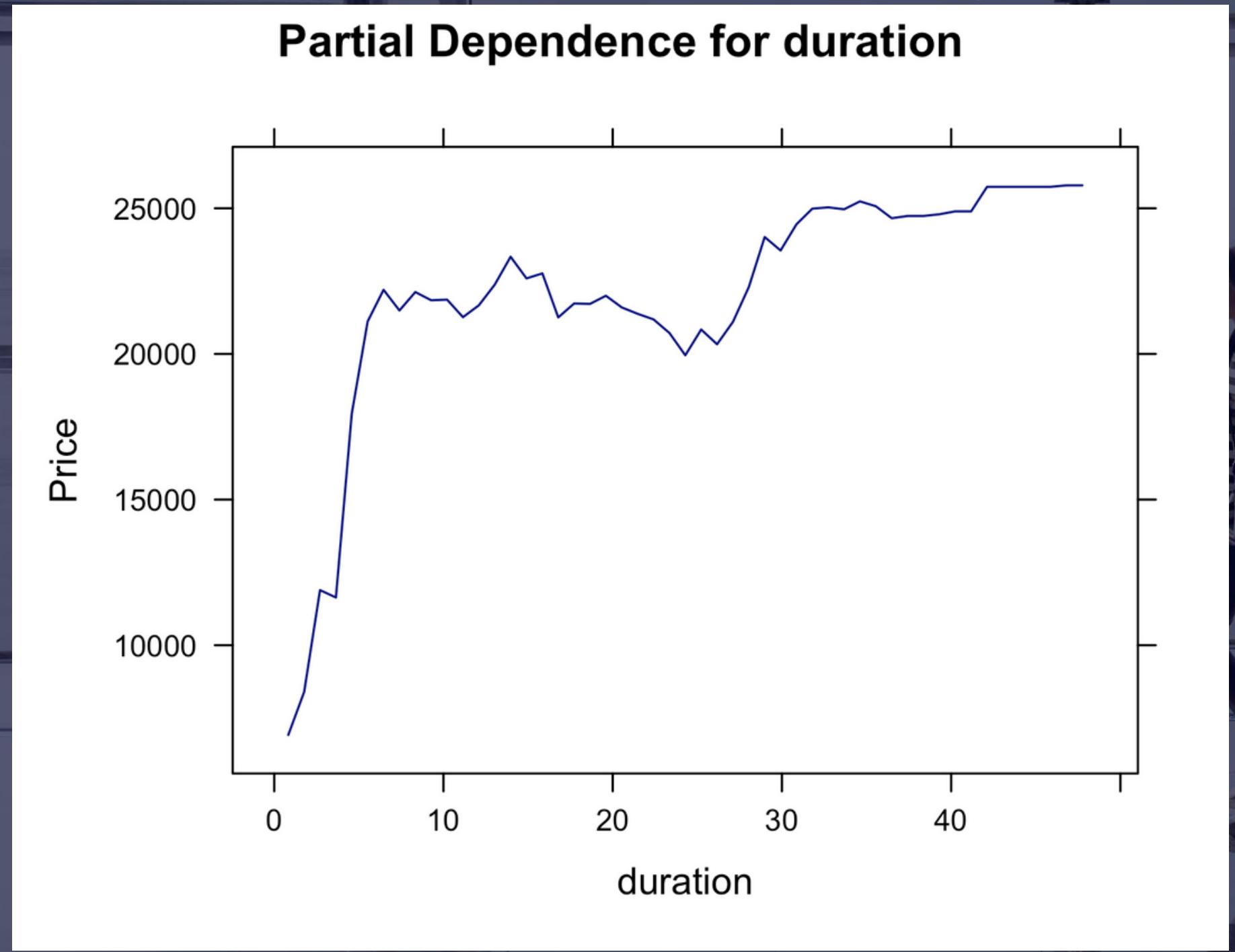
# HOW TO EXPLAIN THESE RESULTS?



**PARTIAL  
DEPENDENCIES PLOTS**



# PARTIAL DEPENDENCY PLOT OF DURATION

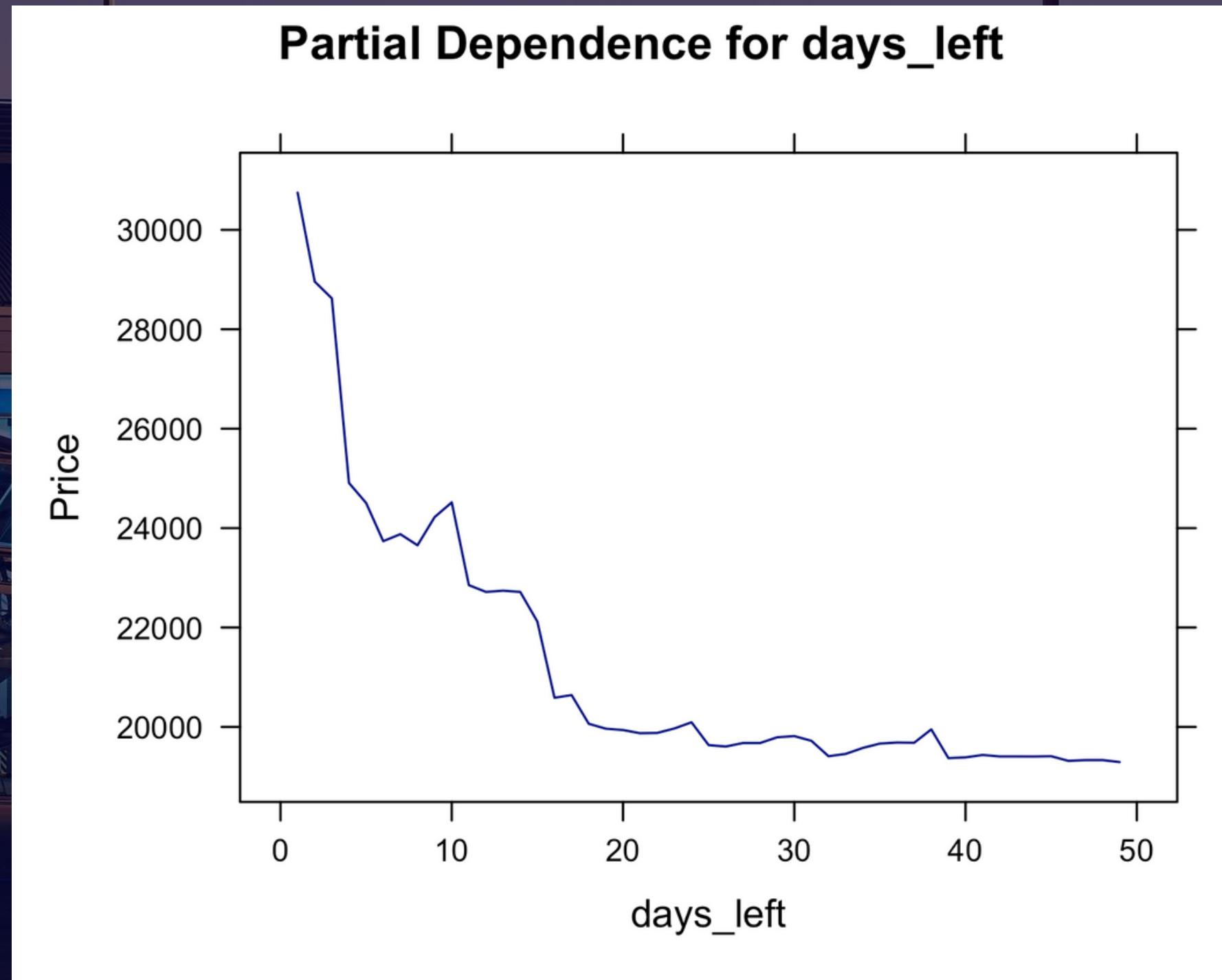


## KEY POINTS

Until the duration reaches 10, the price increases significantly. However, when the flight duration exceeds 10, the rate of price increase slows down, and the price stabilizes.

[HOME](#)[ABOUT](#)[CONTENT](#)[OTHERS](#)

# PARTIAL DEPENDENCY PLOT OF DAYS LEFT



## KEY POINTS

AROUND 7–8 DAYS LEFT, THERE IS A SMALL SPIKE WHERE THE PRICE TEMPORARILY INCREASES.

After about 20 days left, the price stabilizes around 20,000 and fluctuates only slightly.

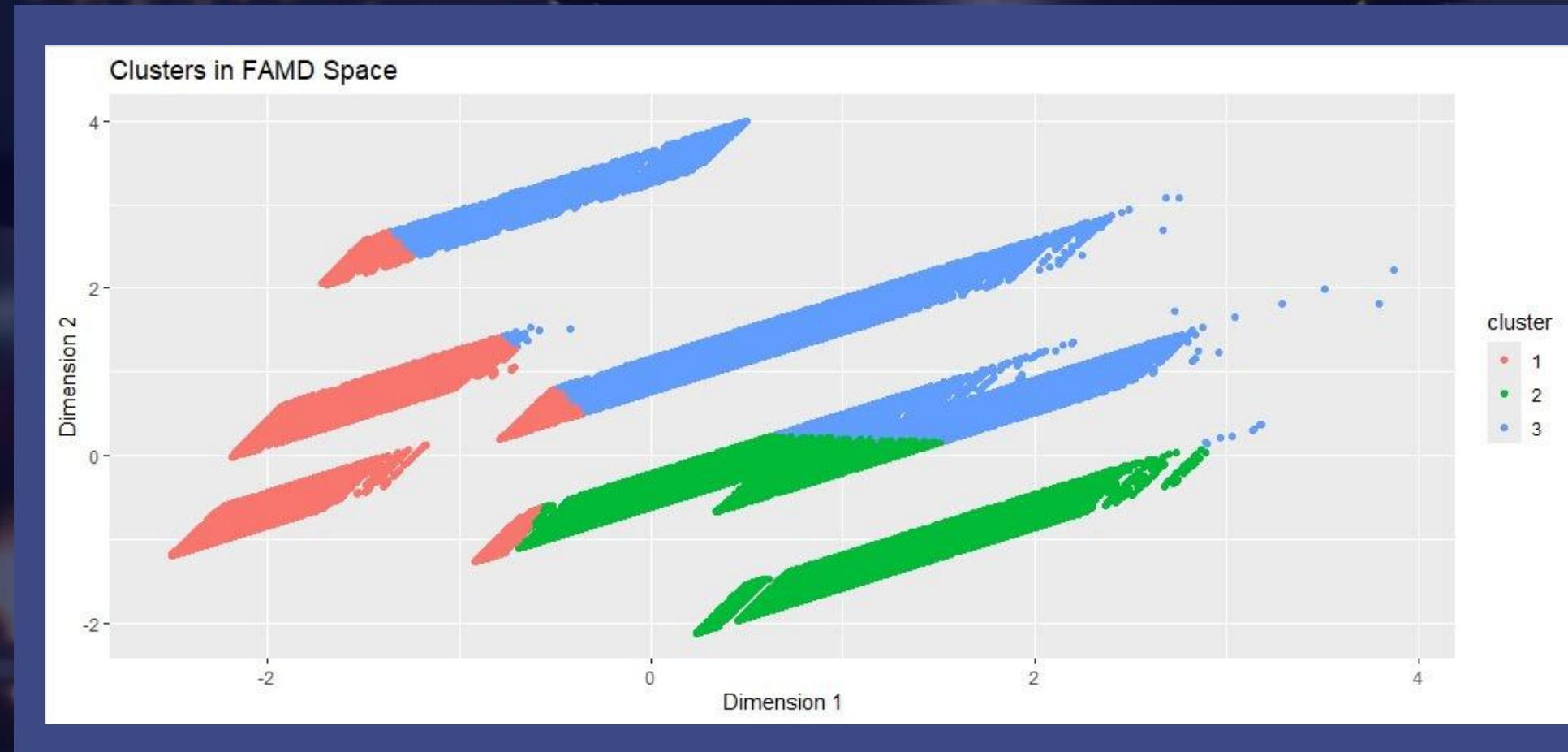
[HOME](#)[ABOUT](#)[CONTENT](#)[OTHERS](#)

# QUESTION ?

I have time, and I can go on vacations. I am not a busy woman. But do I have enough money to choose this options? or I can only book economy class. I can never book business class.

[HOME](#)[ABOUT](#)[CONTENT](#)[OTHERS](#)

# CLUSTERING BASED ON IMPORTANT FEATURES



3 CLUSTERS  
MEAN SILHOUETTE  
WIDTH: 0.52



## CLUSTER

1

- Median price: 4589
- Median duration: 6.2 hours
- Median days left: 27
- Dominantly Economy class tickets
- Primarily indigo airline

**Budget-conscious travelers**

Indigo awarded as the “Best Low-Cost Airline in India & South Asia” by SkyTrax in 2024, for the 14th time in a row.

-IndiGo's Official Website

## CLUSTER

2

- Median price: 8624
- Median duration: 17.5 hours
- Median days left: 23
- Dominantly Economy class tickets
- Primarily AIR India airline

**Flexible travelers**

For passengers seeking enhanced comfort, Air India offers a Premium Economy option on select domestic routes.

-Air India's Official Website

## CLUSTER

3

- Median price: 39161
- Median duration: 12 hours
- Median days left: 27
- Dominantly business class tickets
- Primarily vistara airline

**Premium travelers**

Vistara charges premium prices for its services both on-board and on-ground. The company seeks to provide a good and memorable flying experience to its travellers at value-driven price points.

-South Asian Journal of Business and Management Cases



# Problems encountered

- Moderate clustering performance ( $\text{silhouette} = 0.52$ ) suggests overlapping traveler segments.
- Short dataset duration (50 days) restricted analysis of long-term trends and seasonality.

- To address seasonal trends and improve clustering accuracy, expand the dataset to 12+ months, capturing variations like holiday demand.
- Refine feature engineering by adding booking window categories and airline-specific promotions to improve cluster distinctiveness.
- Integrate real-time flight updates, external factors, and anonymized passenger demographics to add contextual depth.

# Suggestions for improvement

[HOME](#)[ABOUT](#)[CONTENT](#)[OTHERS](#)

# REFERENCE

Search from hundreds of flights daily to all destinations

- <https://wowfare.com/blog/factors-that-affect-flight-ticket-prices/>
- <https://arxiv.org/pdf/1804.03515.pdf>
- <https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>
- <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html>
- <https://medium.com/@heyamit10/xgboost-in-r-a-practical-guide-f14b722866c1>
- <https://xgboost.readthedocs.io/en/latest/parameter.html>
- <https://www.apppsilon.com/post/rxgboost#:~:text=boosting%20algorithm%20%2D%20XGBoost,-Introduction%20to%20R%20XGBoost,as%20a%20go%2Dto%20algorithm.>
- <https://www.statology.org/lasso-regression-in-r/>
- <https://kth.diva-portal.org/smash/get/diva2%3A1894641/FULLTEXT01.pdf>
- [https://www.researchgate.net/publication/373531278\\_An\\_Investigation\\_into\\_Predicting\\_Flight\\_Fares\\_in\\_India\\_Using\\_Machine\\_Learning\\_Models](https://www.researchgate.net/publication/373531278_An_Investigation_into_Predicting_Flight_Fares_in_India_Using_Machine_Learning_Models)

Search

Search from hundreds of destinations, with various airlines and travel partners, to find the best price. Include your flights, hotels and car hire to get the best deal.



HOME

ABOUT

CONTENT

OTHERS

# THANK YOU!