Predict Flight Prices

Rafed Abbassi, Uday Sapra,
Stephen Wu, John Yang

Agenda

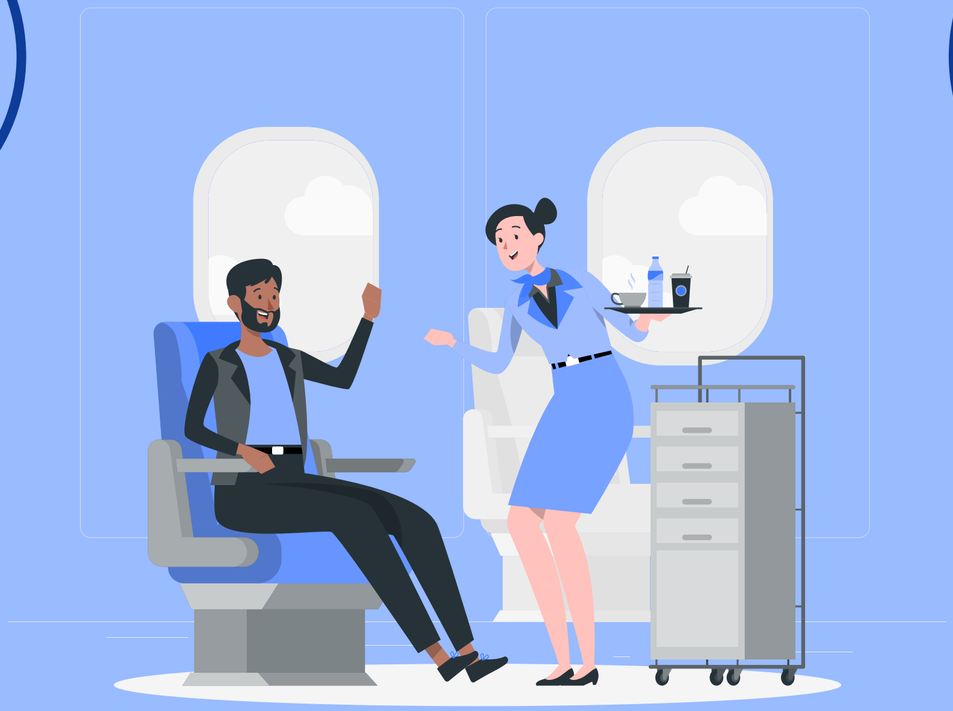
Problem and Purpose

Data Exploration

Modelling

Insights

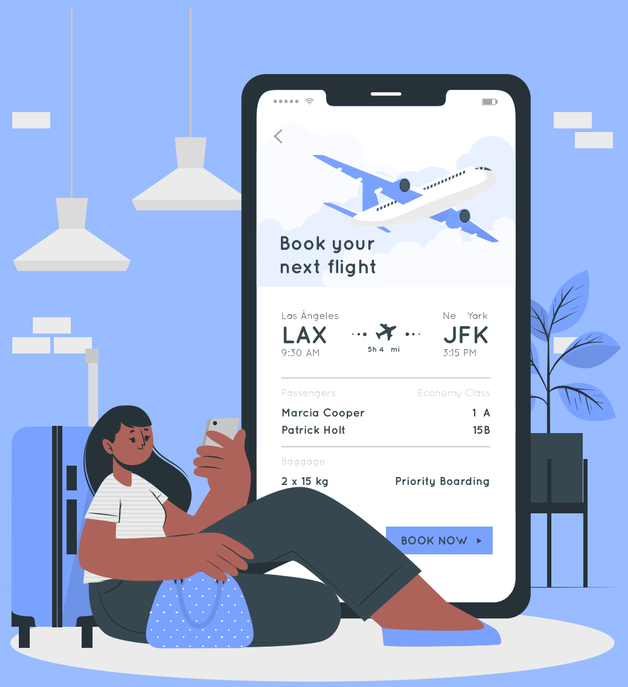
Conclusions



Problem

Purchasing flight tickets can be a very tedious process

Many factors go into ticket pricing (departure and arrival locations, season, and flight distance)



Project Goals

Goals:

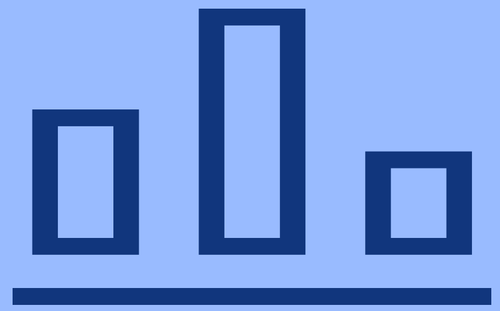
Construct different models that can predict the prices of different flights with a high degree of accuracy.

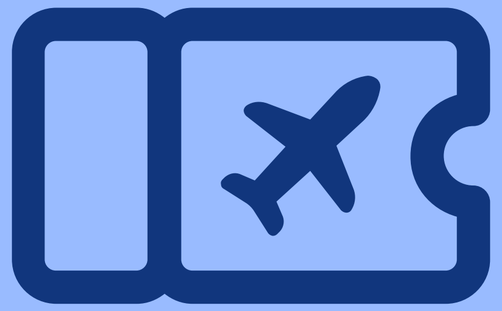
Identify which factors have the largest effect on flight tickets' prices.

Test the accuracy of different models and identify which model was the most accurate.

Data Overview

kaggle





Dataset found using Kaggle with 300,000 rows and 11 features

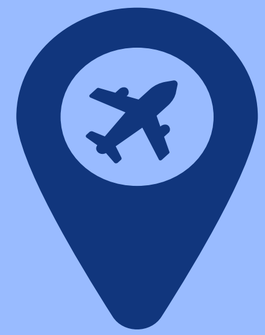
Contains data from the Indian Markets - Q1 2022

Predictor variables include airline, duration, stops, airport, etc.

Data Cleaning

Chose top 5 time-frames based on preference

Top 5 airports - departure and landing



Top 5 airlines by traffic

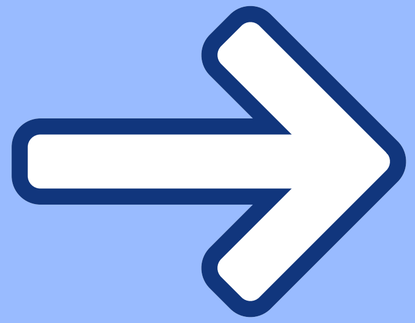
Cleaned Dataset

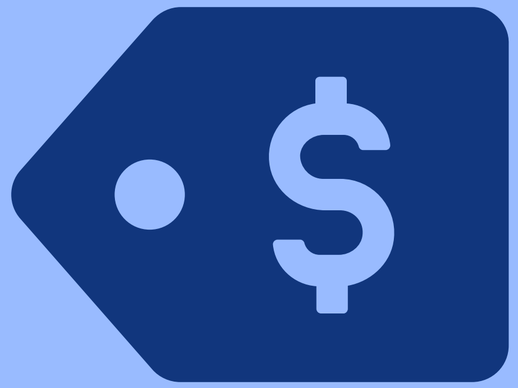
Seat Class

Airline

Length

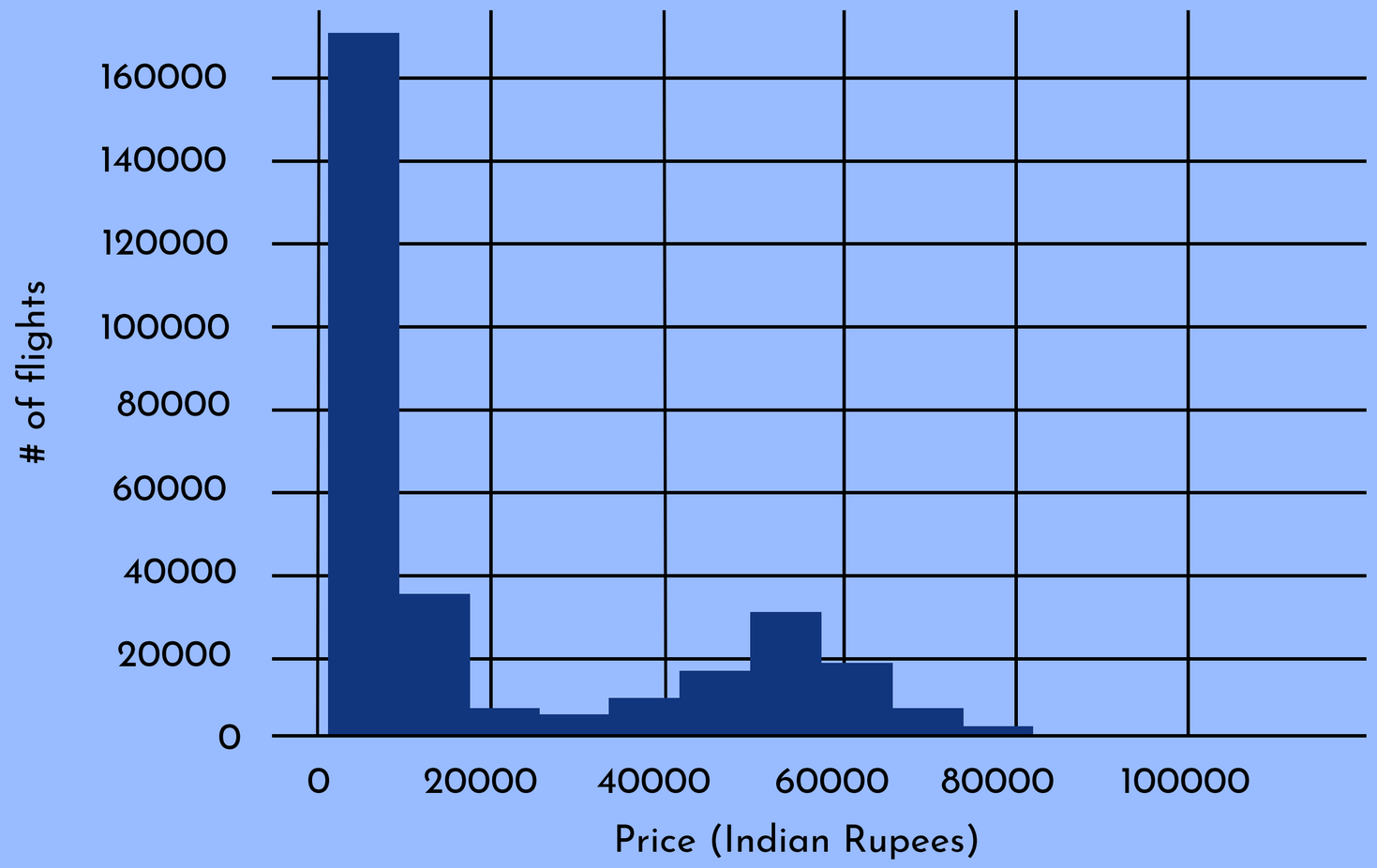
Departure



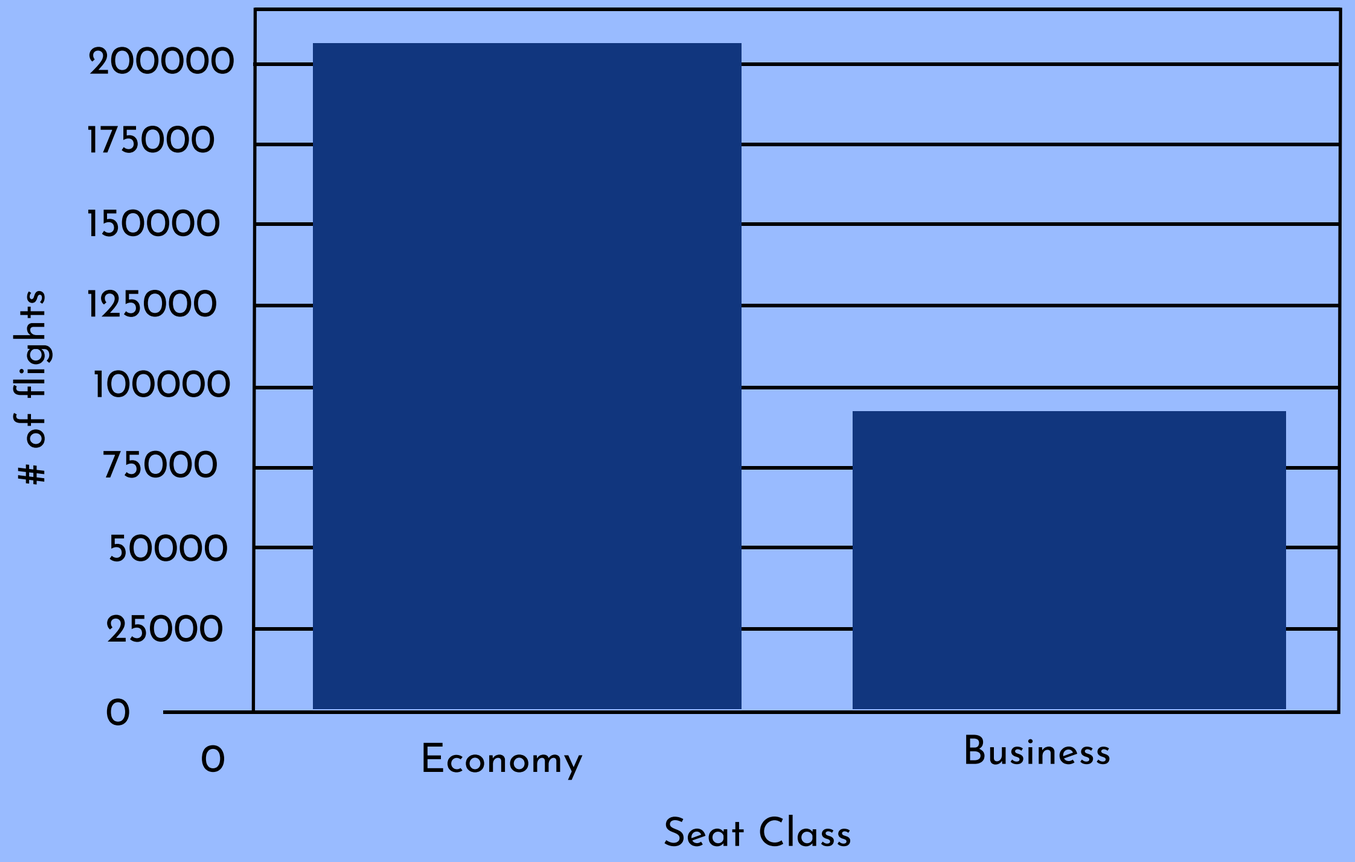


Price

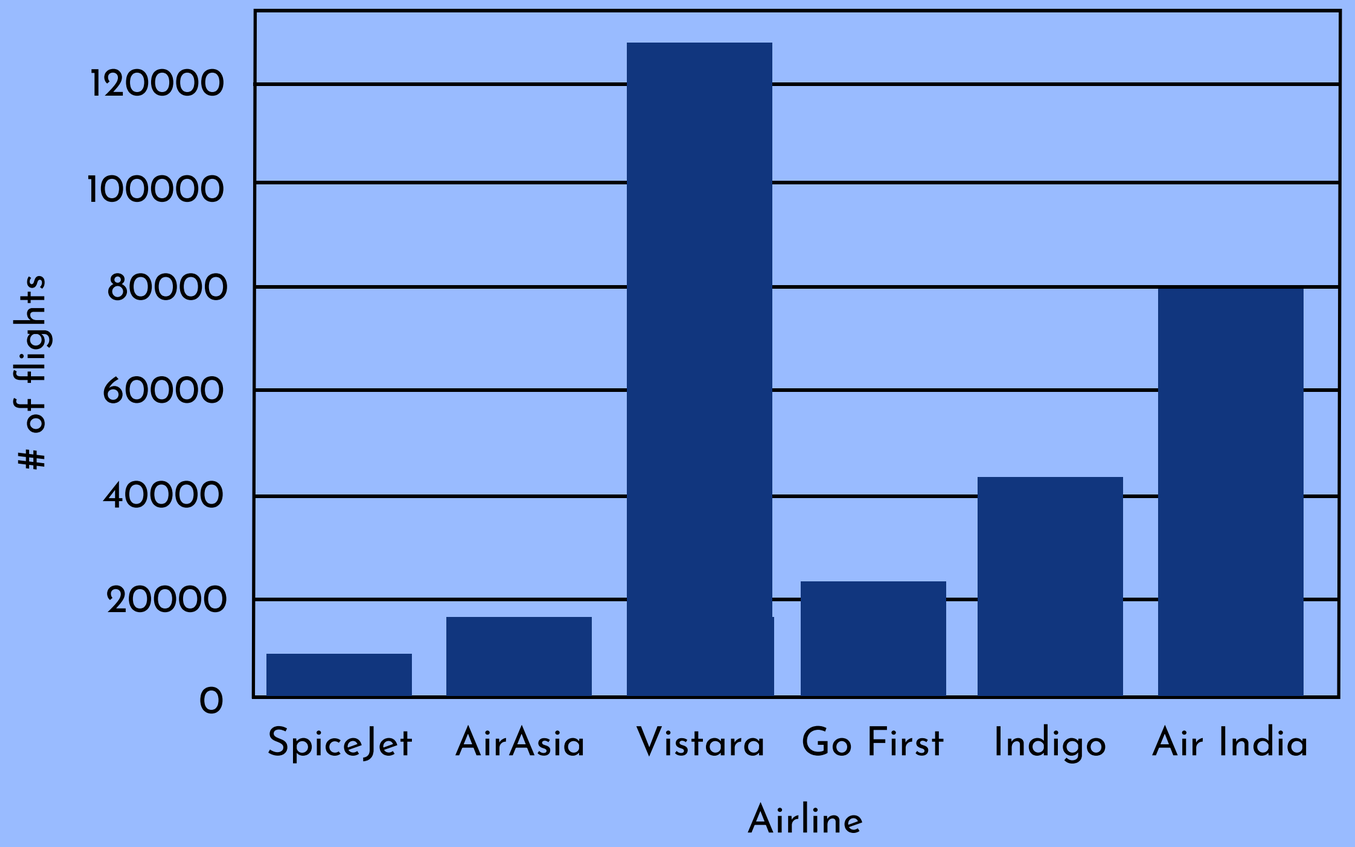
Price Distribution



Economy vs Business Class



Airline Distribution

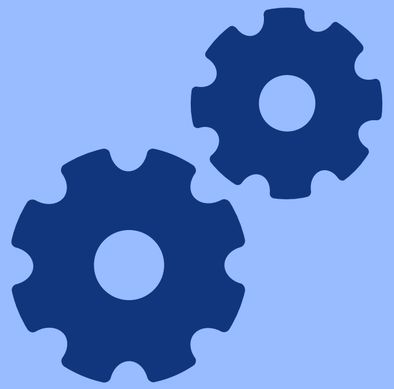


Model Evaluation

	Linear Regression	Decision Trees	Random Forest
Adjusted R^2	0.830	0.873	0.971
RMSE	6754	5797	2736

Model Accuracy

Flight from Vistara to Bombay
2 hours and 10 minutes
0 stops
2 days form departure
Actual Price: 5415 INR



Predicted Price: 6023 INR
Accuracy: 88.77%

Variable Analysis

- Ticket class had the highest influence on ticket price.
- Length of the flight signifcantly impacted price.
- Days until departure largely affected ticket price.

Limitations

- Changing datasets from a US dataset to an Indian dataset.
- Data is in INR not USD, causing increased variance and exaggeration of RMSE.
- Data only for Q1 2022, and therefore does not reflect flight prices year-round.

Key Takeaways

1. Ticket class was the biggest contributor for ticket price (Business vs Economy tickets).
2. Holding this constant, the length of the flight was the 2nd biggest predictor.
3. After running multiple models, we had the highest success with the Random Forest.
4. There is more pricing uniformity in the Indian markets compared to the US.

Questions?