Flight Fare Prediction

Group members :-

Dhru Prajapati (C0867085)

Malav Shah (C0870066)

Dhruvraj Chavda (C0867457)

Sneh Patel (C0869367)

Nandiniben Patel (C0869773)

Introduction

- Flight fare prediction is the process of using various techniques to estimate the cost of air travel based on various factors such as time of year, day of the week, demand for tickets, competition among airlines, fuel prices, and other relevant data.
- With the help of flight fare prediction, travelers can plan and book their trips more effectively, saving both time and money.
- Predicting flight fares accurately is crucial in today's travel industry, as prices can fluctuate rapidly and unexpectedly.
- This presentation will explore the importance of flight fare prediction, the factors that affect flight fares, techniques for predicting flight fares, and examples of flight fare prediction tools and websites that travelers can use to save money on air travel.

Why flight fare prediction is important

1. **Saves Money** : By using flight fare prediction tools or techniques, travelers can compare prices from different airlines, find the best deals, and book their tickets at the right time to avoid paying high prices.

2. **Helps Plan Travel**: Flight fare prediction can help travelers plan their trips more effectively. By predicting flight fares, travelers can decide on the best time to book their flights, which airlines to choose, and how to allocate their travel budget.

3. **Increases Revenue**: Accurate flight fare prediction can help airlines and travel companies optimize their pricing strategies and increase revenue. By adjusting prices based on demand and other relevant factors, airlines can sell more seats and maximize profits.

4. **Enhances Customer Experience**: Travelers can have more confidence in their travel plans and feel more in control of their budget. This can result in increased customer loyalty and positive word-of-mouth recommendations.

Factors that affect flight fares

1. **Time of Year**: Flight fares tend to be higher during peak travel seasons such as holidays, school breaks, and summer vacations. Conversely, fares may be lower during off-peak seasons when demand is lower.
2. **Day of the Week**: Fares may be lower on certain days of the week, such as Tuesdays and Wednesdays, which are typically less busy travel days.
3. **Advance Booking**: Booking in advance can sometimes result in lower fares. However, booking too far in advance may result in higher fares, as airlines adjust prices based on demand and availability.

Continuation of Factors that affect flight fares

4.**Competition**: Competition among airlines can affect fares. When multiple airlines offer flights on the same route, fares may be lower due to increased competition.

5.**Fuel Prices**: Fluctuations in fuel prices can affect flight fares. When fuel prices rise, airlines may pass on the cost to consumers through higher fares.

6.**Route and Destination**: Fares can vary depending on the route and destination. Popular routes and destinations may have higher fares due to increased demand.

7.**Airline and Class**: Fares can also vary depending on the airline and class of travel. Some airlines may offer lower fares for economy class, while others may focus on premium class travel

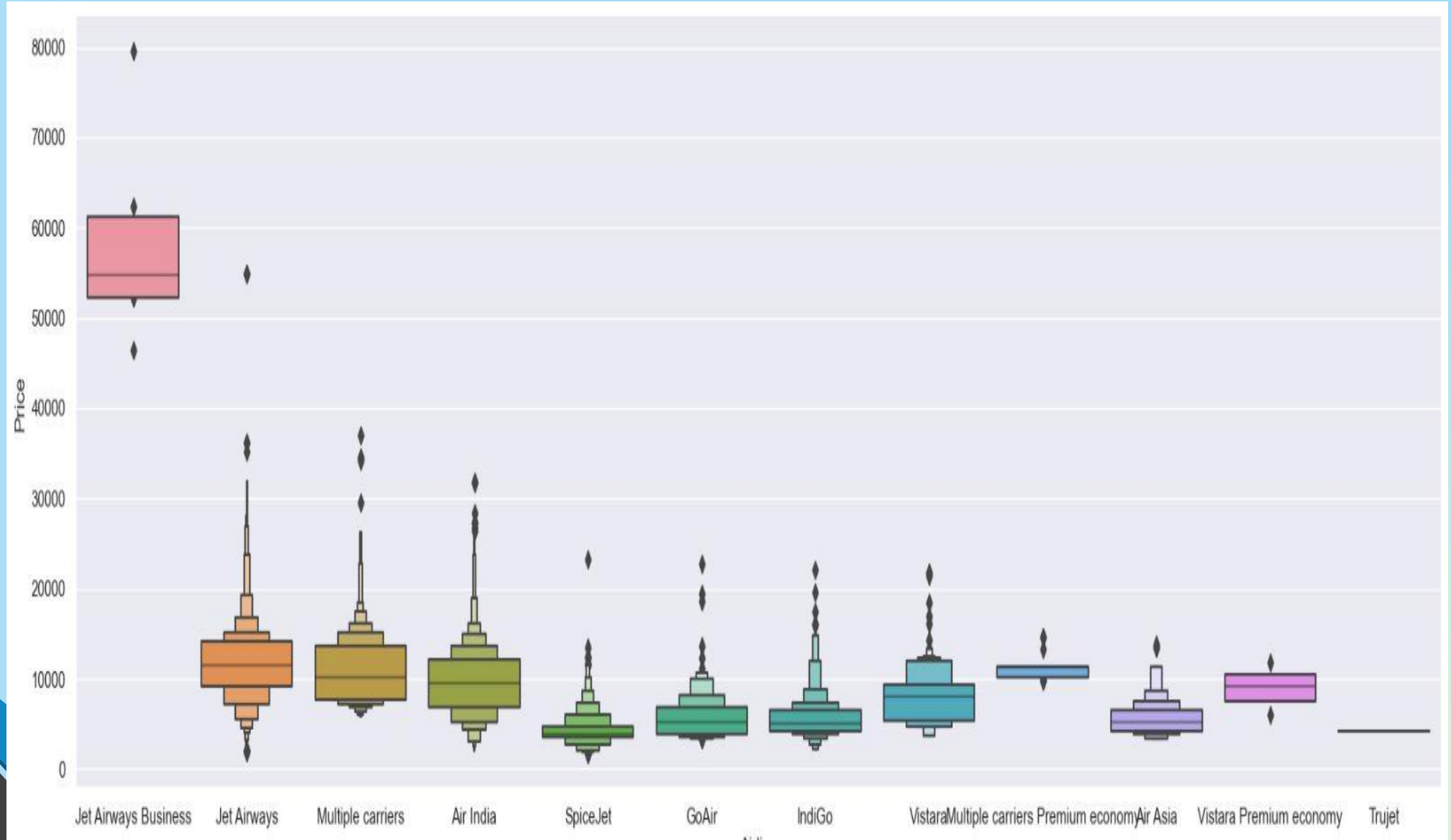
Factors Affecting Domestic Flight Pricing in the US

- We base our research on a real-world dataset that includes data on flight times, destinations, and costs.
- We concentrate on domestic travel within the United States and take into account a number of variables, including seasonality, competition, demand-supply dynamics, and operating expenses, that have an impact on airline pricing.

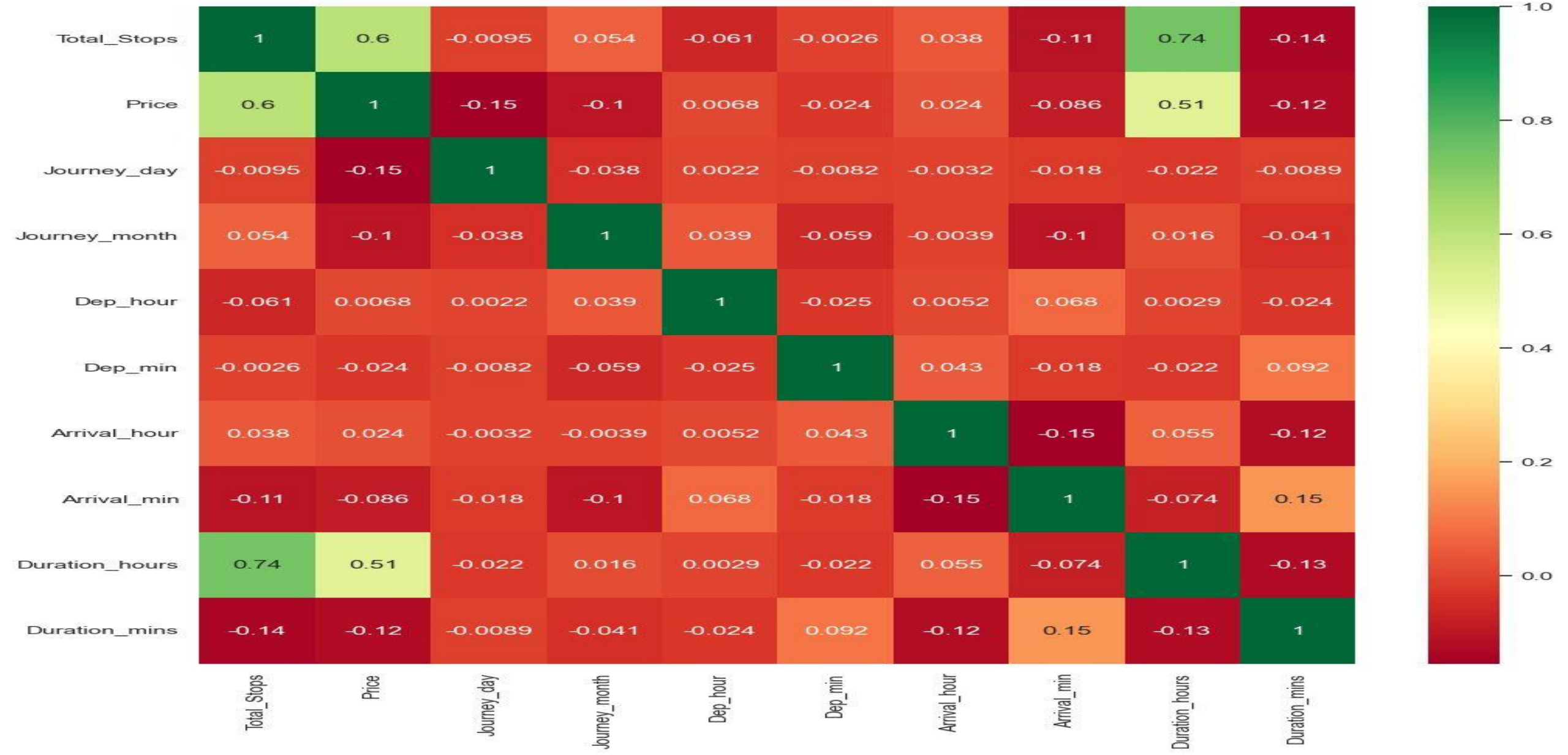
Using Machine Learning to Forecast Flight Prices

- We employ machine learning-based methodologies to forecast flight prices in our study's quantitative research design.
- In order to compare the performance of several models, we separate the dataset into training and testing sets.

Price vs airline



Heat map


Total_Stops 1 0.6 -0.0095 0.054 -0.061 -0.0026 0.038 -0.11 0.74 -0.14
Price 0.6 1 -0.15 -0.1 0.0068 -0.024 0.024 -0.086 0.51 -0.12
Journey_day -0.0095 -0.15 1 -0.038 0.0022 -0.0082 -0.0032 -0.018 -0.022 -0.0089
Journey_month 0.054 -0.1 -0.038 1 0.039 -0.059 -0.0039 -0.1 0.016 -0.041
Dep_hour -0.061 0.0068 0.0022 0.039 1 -0.025 0.0052 0.068 0.0029 -0.024
Dep_min -0.0026 -0.024 -0.0082 -0.059 -0.025 1 0.043 -0.018 -0.022 0.092
Arrival_hour 0.038 0.024 -0.0032 -0.0039 0.0052 0.043 1 -0.15 0.055 -0.12
Arrival_min -0.11 -0.086 -0.018 -0.1 0.068 -0.018 -0.15 1 -0.074 0.15
Duration_hours 0.74 0.51 -0.022 0.016 0.0029 -0.022 0.055 -0.074 1 -0.13
Duration_mins -0.14 -0.12 -0.0089 -0.041 -0.024 0.092 -0.12 0.15 -0.13 1
Total_Stops
Price
Journey_day
Journey_month
Dep_hour
Dep_min
Arrival_hour
Arrival_min
Duration_hours
Duration_mins
1.0
0.8
0.6
0.4
0.2
0.0

From where we obtained our data

- We got the data from one of the most prominent sites called Kaggle
- Over 10682 rows and columns with 11 characteristics are included in our dataset. The dataset was gathered from freely accessible sources, such as airline and travel agency websites.

List of features

- 1. Airline
- 2. Date of Journey
- 3. Source
- 4. Destination
- 5. Route
- 6. Dep Time
- 7. Arrival Time
- 8. Duration
- 9. Total Stops
- 10. Additional Info
- 11. Cost

The strategy that was unsuccessful

- Given complicated and nonlinear interactions, linear regression was not performing well because it implies a linear relationship between the input features and the output variable.
- Neural networks were capturing intricate relationships between information, but they experienced overfitting and need a lot of hyperparameter tuning.

The strategy that was unsuccessful

- The random forest approach outperformed the other machine learning models we tested, scoring 79% accuracy.
- An ensemble learning technique called random forest combines several decision trees to increase the reliability and accuracy of predictions. It handles high-dimensional datasets with intricate feature interactions exceptionally well.

Result

- In order to forecast flight costs, we trained a random forest model.The model outperformed other machine learning-based models, including linear regression and neural networks, with an accuracy score of over 79% on the test set.
- We were able to get important insights into the underlying elements that affect airline fares by employing machine learning techniques, a meticulous and rigorous data analysis procedure, and other methods.
- this information can help airlines develop price and revenue management strategies, which could ultimately result in higher profitability and happier customers.

Future endeavours

- Investigate how environmental rules, fuel prices, and global events affect airline ticket prices.
- It may be possible to tailor pricing tactics further and boost customer satisfaction by incorporating more detailed information on client behavior, such as booking history and preferences.
- The dynamics of the airline business may be better understood by examining how variables like weather and world events affect flight costs.

References

- Gulati, A. P. (2022, August 31). Flight Fare Prediction Using Machine Learning. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2022/01/flight-fare-prediction-using-machine-learning/
- C. (2022, September 6). EDA+Forest Price Prediction. Kaggle. https://www.kaggle.com/code/cbhavik/eda-forest-price-prediction/notebook

Thank you