

# Fin-A-Lytics

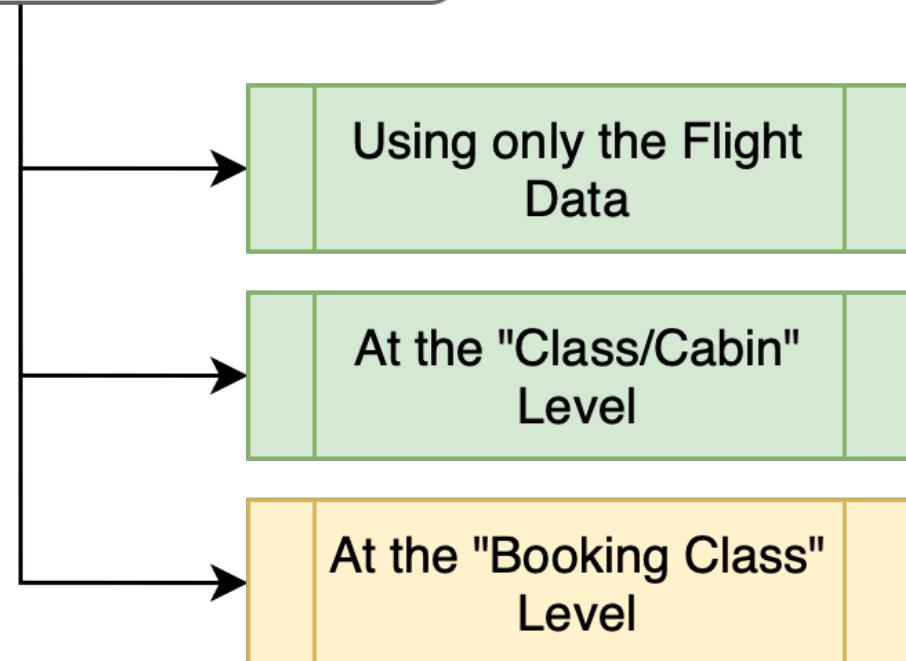
Aniruddha Mukherjee

KIIT

Team: Data Crunchers

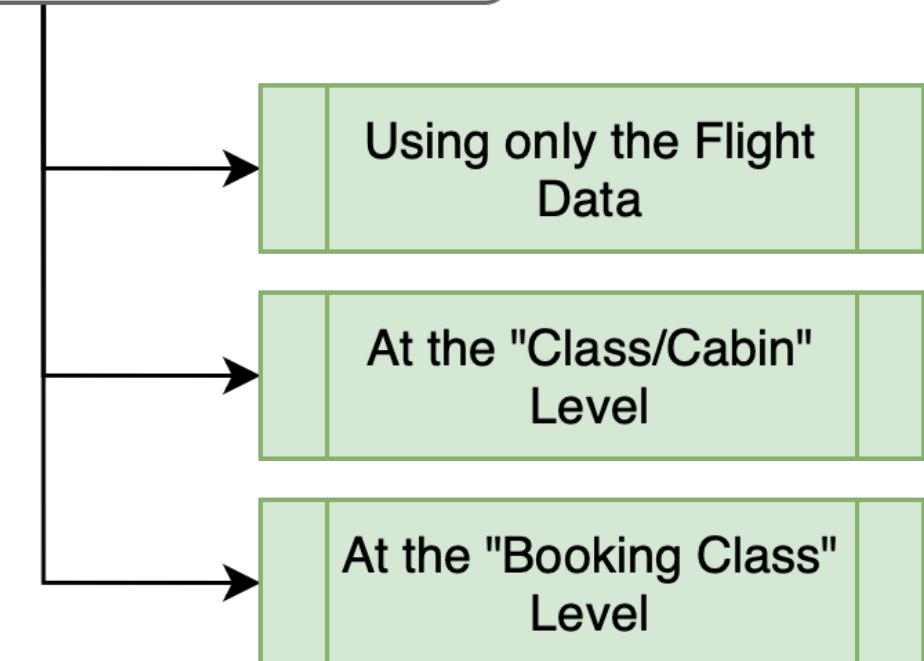
## Demand

### Booking Curve Prediction



## Revenue

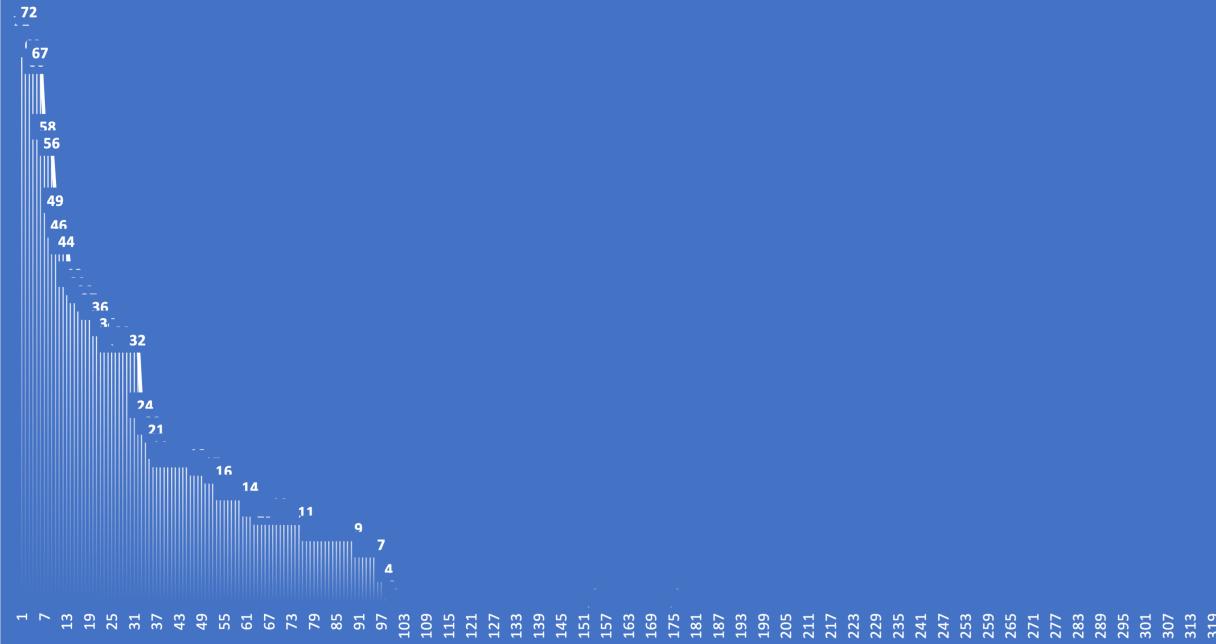
### Booking Class Fare Prediction



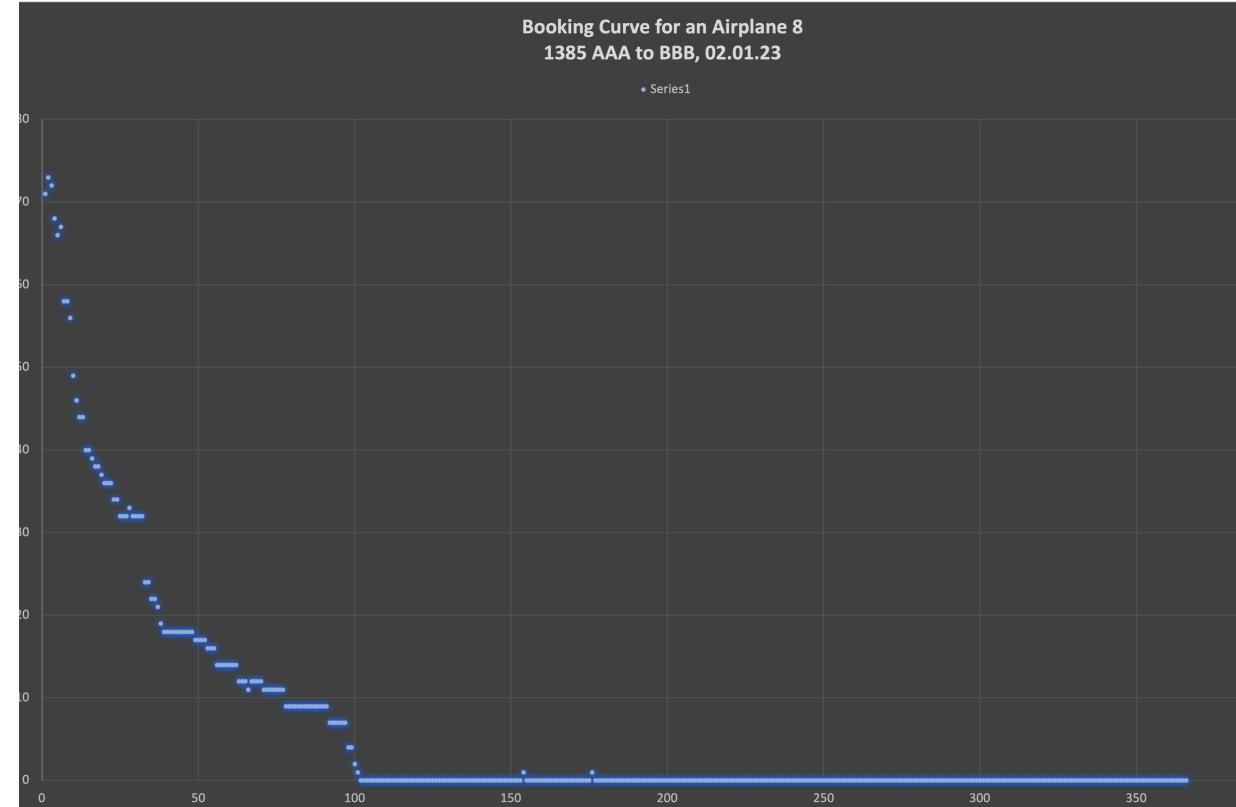
# **Booking Curve Prediction**

# Booking Curves

**Booking Curve for an Airplane 8**  
1385 AAA to BBB, 02.01.23



**Booking Curve for an Airplane 8**  
1385 AAA to BBB, 02.01.23

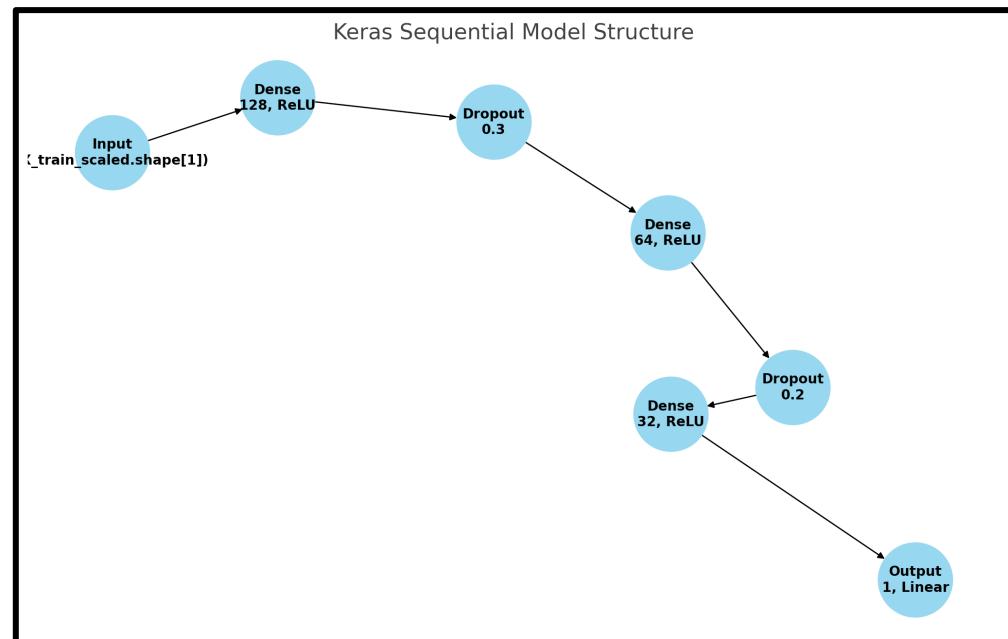


# Features Used + Engineered

| Features                        | Explanation   |
|---------------------------------|---|
| Number of Bookings              | How Many Bookings are there for a particular flight on X day  |
| Number of days Before Departure | The Day in question that the model is trying to predict   |
| Origin_AAA                      | Origin - AAA  |
| Destination_BBB                 | Destination - BBB   |
| dayofmonth                      |   |
| year                            |   |
| dayofyear                       |   |
| hour                            |   |
| minute                          |   |
| dayofweek_0                     |   |
| ...                             |   |
| dayofweek_6                     | One Hot Encoded. 0-> Sunday, 6-> Saturday   |
| quarter_1                       |   |
| ...                             |   |
| quarter_4                       | Gives the model an idea about the season  |
| month_1                         |   |
| ...                             |   |
| month_12                        | Which month the plane flew in is importatn due to Seasonal Variations. December: More bookings beforehand |

Important to note that the dates are dynamically adjusted according to the 365 day schedule for each plane individually  
To give it much more context about time

# Model Used



Used a Fully Connected Neural Network.

Has multiple hidden layers and also Dropout to add regularization & robustness to the model.

Loss used:

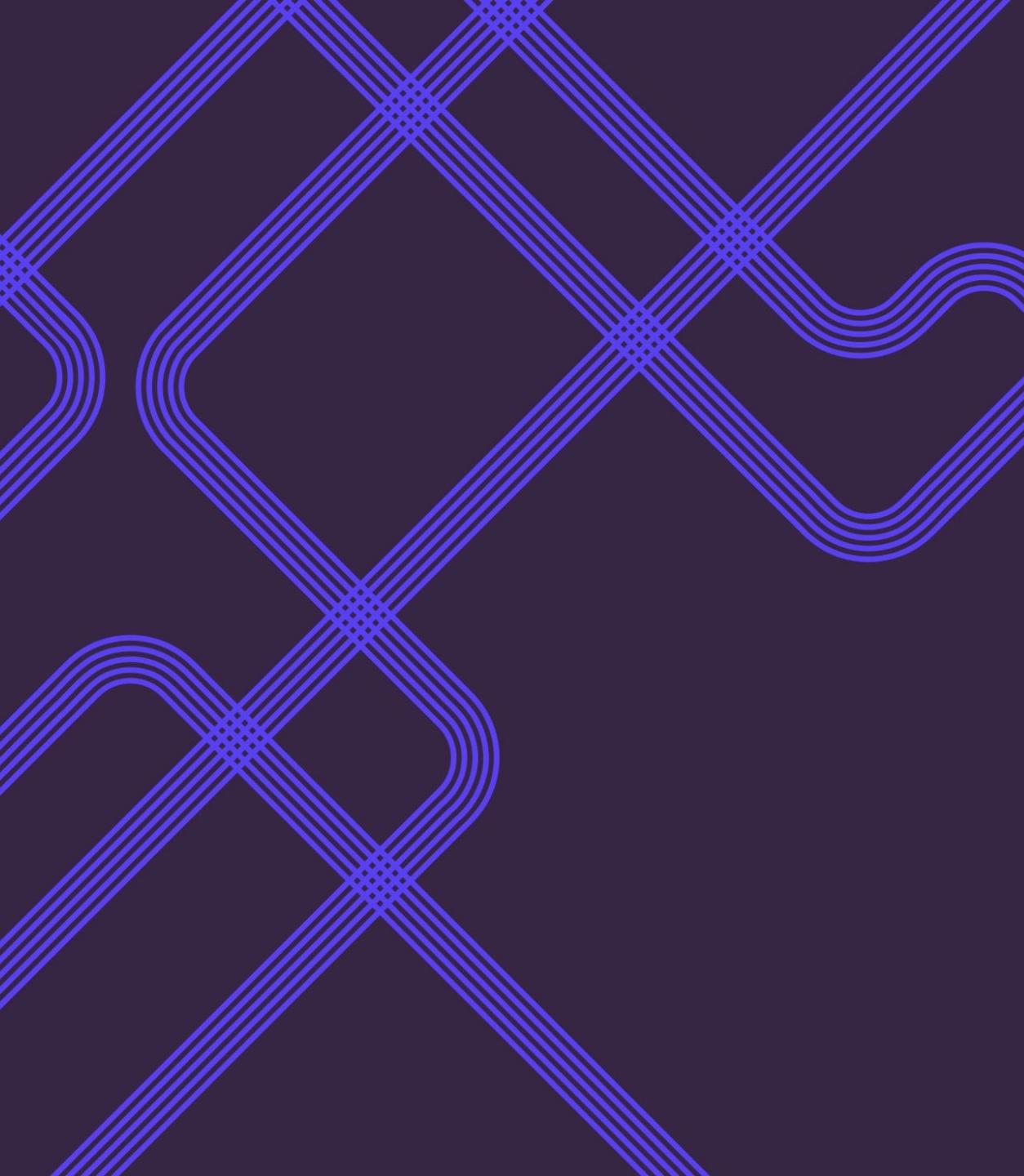
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Adam Optimizer used, with a learning rate of **0.001**

Test Mean Absolute Error:  
6.544564723968506

# Co-relation Matrix (first 10 rows)

|                    | Number of Bookings |
|--------------------|--------------------|
| Number of Bookings | 1.000000           |
| year               | 0.297504           |
| month_10           | 0.227580           |
| dayofyear          | 0.217034           |
| quarter_3          | 0.156635           |
| month_9            | 0.153367           |
| quarter_4          | 0.126598           |
| month_11           | 0.072011           |
| month_8            | 0.070916           |
| month_7            | 0.021992           |



## Model-1 [FCN]

Prediction of Booking Curve  
Using Only the Flight Data

# Data transformation

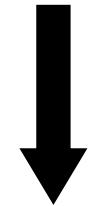
| Origin | Destination | Flight number | Cabin | Dep_Date | Dep_time | Booking Class | Booking Class Fare | B: 365   | ... | B: 0 |
|--------|-------------|---------------|-------|----------|----------|---------------|--------------------|----------|-----|------|
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | N             | 200.01             | 0        |     | 10   |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | H             | 463.65             | 0        |     | 7    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | G             | 149.83             | 0        |     | 1    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | T             | 127.29             | 0        |     | 5    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | S             | 90.92              | 0        |     | 10   |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | M             | 146.97             | 0        |     | 0    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | Y             | 681.83             | 0        |     | 3    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | L             | 309.1              | 0        |     | 44   |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | K             | 372.74             | 0        |     | 2    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | E             | 245.47             | 0        |     | 10   |
| BBB    | DDD         | 304 J         |       | 10/04/22 | 05:55    | J             | 836.38             | 0        |     | 0    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | Q             | 145.47             | 0        |     | 36   |
| BBB    | DDD         | 304 J         |       | 10/04/22 | 05:55    | D             | 386.69             | 0        |     | 1    |
| BBB    | DDD         | 304 J         |       | 10/04/22 | 05:55    | C             | 563.65             | 0        |     | 2    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | V             | 181.83             | 0        |     | 3    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | B             | 581.83             | 0        |     | 2    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | I             | 109.1              | 0        |     | 3    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | P             | 163.65             | 0        |     | 15   |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | O             | 218.2              | 0        |     | 3    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | R             | 272.74             | 0        |     | 0    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | W             | 518.2              | 0        |     | 0    |
|        |             |               |       |          |          |               | SUM: 6765.55       | SUM: 157 |     |      |

| Origin | Destination | Flight number | Dep_Date | Dep_time | Booking Class Fare USD | B: 365 | ... | B: 0 |
|--------|-------------|---------------|----------|----------|------------------------|--------|-----|------|
| BBB    | DDD         | 304           | 10/04/22 | 05:55    | 6765.55                | 0      |     | 157  |

9635 such rows

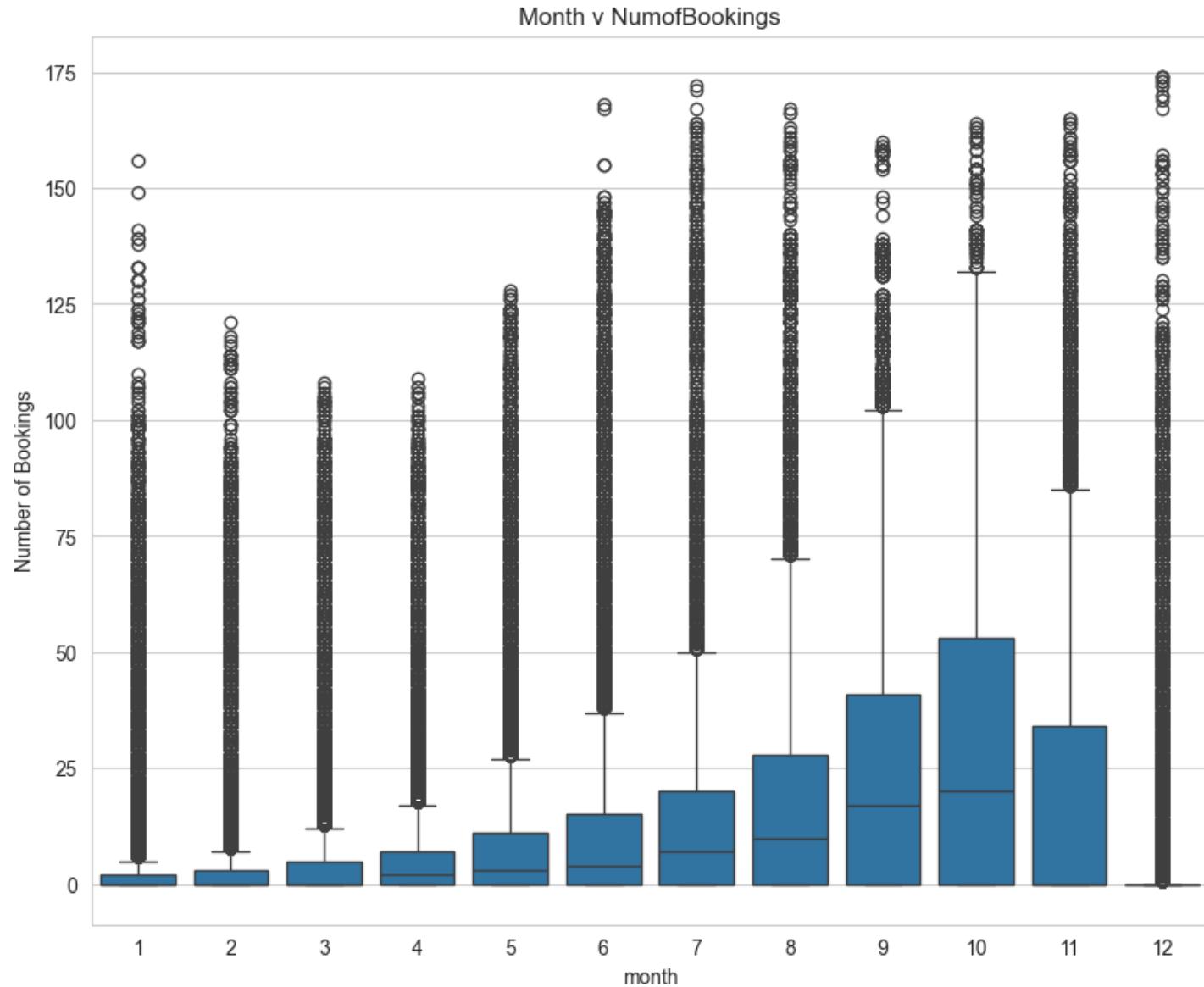
# Transforming the dataset to long form from wide-form

| ORIGIN | DESTINATION | CABIN | BOOKING CLASS | Departure Date | Departure Time | B: 0 | B: 1 | B: 2 | ... | B: 365 |
|--------|-------------|-------|---------------|----------------|----------------|------|------|------|-----|--------|
| a      | b           | c     | d             | Day X          | Time Y         | 34   | 30   | 12   |     | 0      |
| f      | g           | h     | i             | Day V          | Time J         | 56   | 55   | 32   |     | 1      |



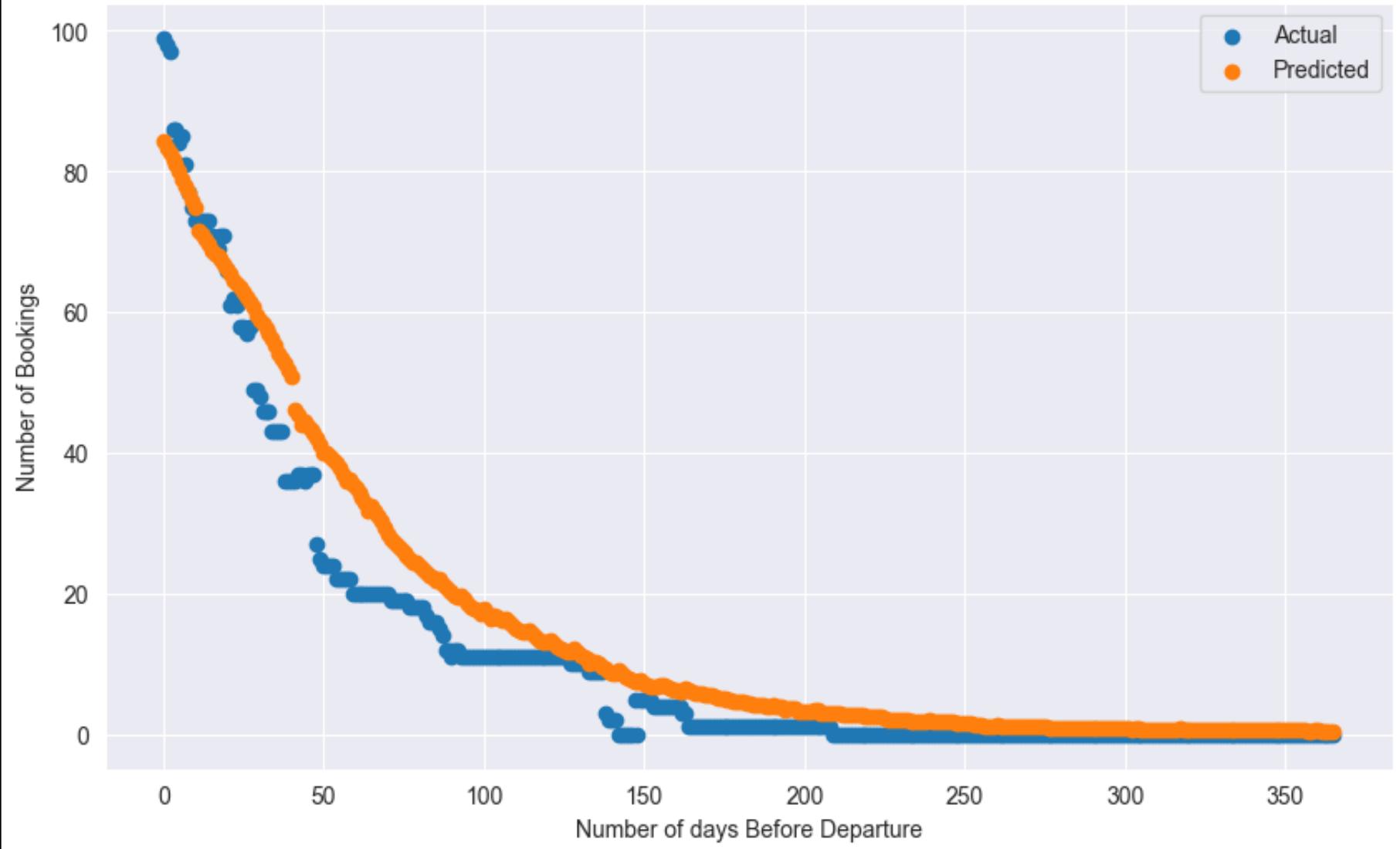
| ORIGIN | DESTINATION | CABIN | BOOKING CLASS | Departure Date | Departure Time | Number of Bookings | Number of days Before Departure | Unique_Flight_ID |
|--------|-------------|-------|---------------|----------------|----------------|--------------------|---------------------------------|------------------|
| a      | b           | c     | d             | Day X          | Time Y         | 34                 | 0                               | 1                |
| a      | b           | c     | d             | Day X-1        | Time Y         | 30                 | 1                               | 1                |
| a      | b           | c     | d             | Day X-2        | Time Y         | 12                 | 2                               | 1                |
| ...    |             |       |               |                |                |                    |                                 |                  |
| a      | b           | c     | d             | Day X-365      | Time Y         | 0                  | 365                             | 1                |
| f      | g           | h     | i             | Day V          | Time J         | 56                 | 0                               | 2                |
| f      | g           | h     | i             | Day V-1        | Time J         | 55                 | 1                               | 2                |
| f      | g           | h     | i             | Day V-2        | Time J         | 32                 | 2                               | 2                |
| ...    |             |       |               |                |                |                    |                                 |                  |
| f      | g           | h     | i             | Day V-365      | Time J         | 1                  | 365                             | 2                |

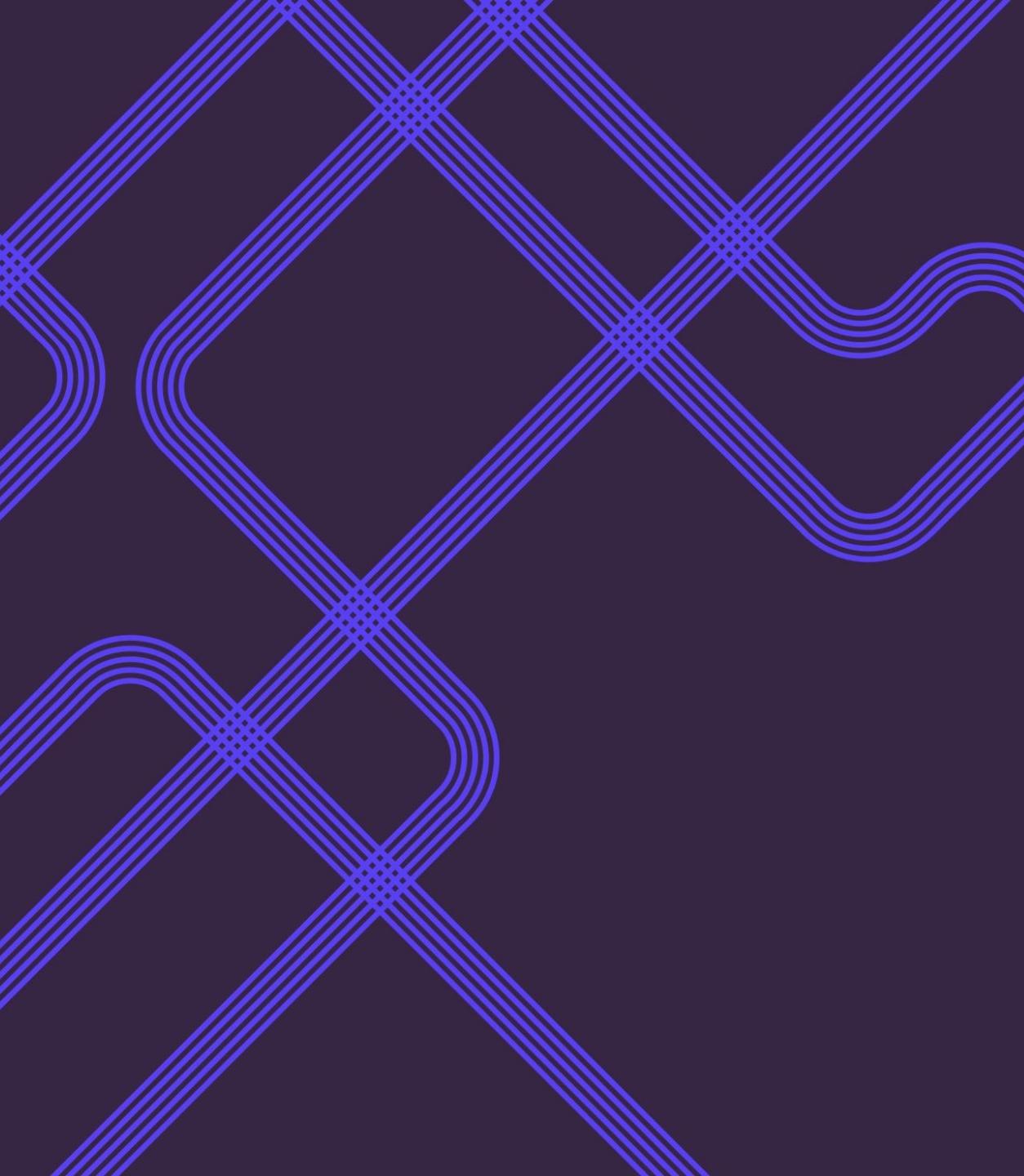
# EDA



Boxplot of the Month vs Number of Bookings in my sample of 1000 flights.

Actual vs Predicted Bookings for Unique\_Flight\_ID 530



The left side of the slide features a dark purple background with a complex, abstract pattern of blue lines. These lines form several large, stylized letters: an 'R' at the bottom left, a 'G' in the center, and a 'Z' at the top right. The lines are thin and light blue, creating a sense of depth and motion.

## Model-2 [FCN]

Prediction Using Booking  
Class

[Y: Economy or J: Business]

# Transforming/Re-Formatting the data

| Origin       | Destination | Flight number | Cabin | Dep_Date | Dep_time | Booking Class | Booking Class Fare | B: 365 | ... | B: 0 |
|--------------|-------------|---------------|-------|----------|----------|---------------|--------------------|--------|-----|------|
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | N             | 200.01             | 0      |     | 10   |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | H             | 463.65             | 0      |     | 7    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | G             | 149.83             | 0      |     | 1    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | T             | 127.29             | 0      |     | 5    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | S             | 90.92              | 0      |     | 10   |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | M             | 146.97             | 0      |     | 0    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | Y             | 681.83             | 0      |     | 3    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | L             | 309.1              | 0      |     | 44   |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | K             | 372.74             | 0      |     | 2    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | E             | 245.47             | 0      |     | 10   |
| BBB          | DDD         | 304 J         |       | 10/04/22 | 05:55    | J             | 836.38             | 0      |     | 0    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | Q             | 145.47             | 0      |     | 36   |
| BBB          | DDD         | 304 J         |       | 10/04/22 | 05:55    | D             | 386.69             | 0      |     | 1    |
| BBB          | DDD         | 304 J         |       | 10/04/22 | 05:55    | C             | 563.65             | 0      |     | 2    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | V             | 181.83             | 0      |     | 3    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | B             | 581.83             | 0      |     | 2    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | I             | 109.1              | 0      |     | 3    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | P             | 163.65             | 0      |     | 15   |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | O             | 218.2              | 0      |     | 3    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | R             | 272.74             | 0      |     | 0    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | W             | 518.2              | 0      |     | 0    |
| SUM: 6765.55 |             |               |       |          |          |               | SUM: 157           |        |     |      |

| Origin       | Destination | Flight number | Cabin | Dep_Date | Dep_time | Booking Class Fare USD | B: 365 | B: 364 | ... | B: 0     |
|--------------|-------------|---------------|-------|----------|----------|------------------------|--------|--------|-----|----------|
| BBB          | DDD         | 304           | Y     | 44661    | 05:55    | 4978.83                | 0      | 0      |     | 154      |
| BBB          | DDD         | 304           | J     | 44661    | 05:55    | 1786.72                | 0      | 0      |     | 3        |
| SUM: 6765.55 |             |               |       |          |          |                        |        |        |     | SUM: 157 |

18574 such rows

# Transforming the dataset to long-form from wide-form

| ORIGIN | DESTINATION | CABIN | BOOKING CLASS | Departure Date | Departure Time | B: 0 | B: 1 | B: 2 | ... | B: 365 |
|--------|-------------|-------|---------------|----------------|----------------|------|------|------|-----|--------|
| a      | b           | c     | d             | Day X          | Time Y         | 34   | 30   | 12   |     | 0      |
| f      | g           | h     | i             | Day V          | Time J         | 56   | 55   | 32   |     | 1      |

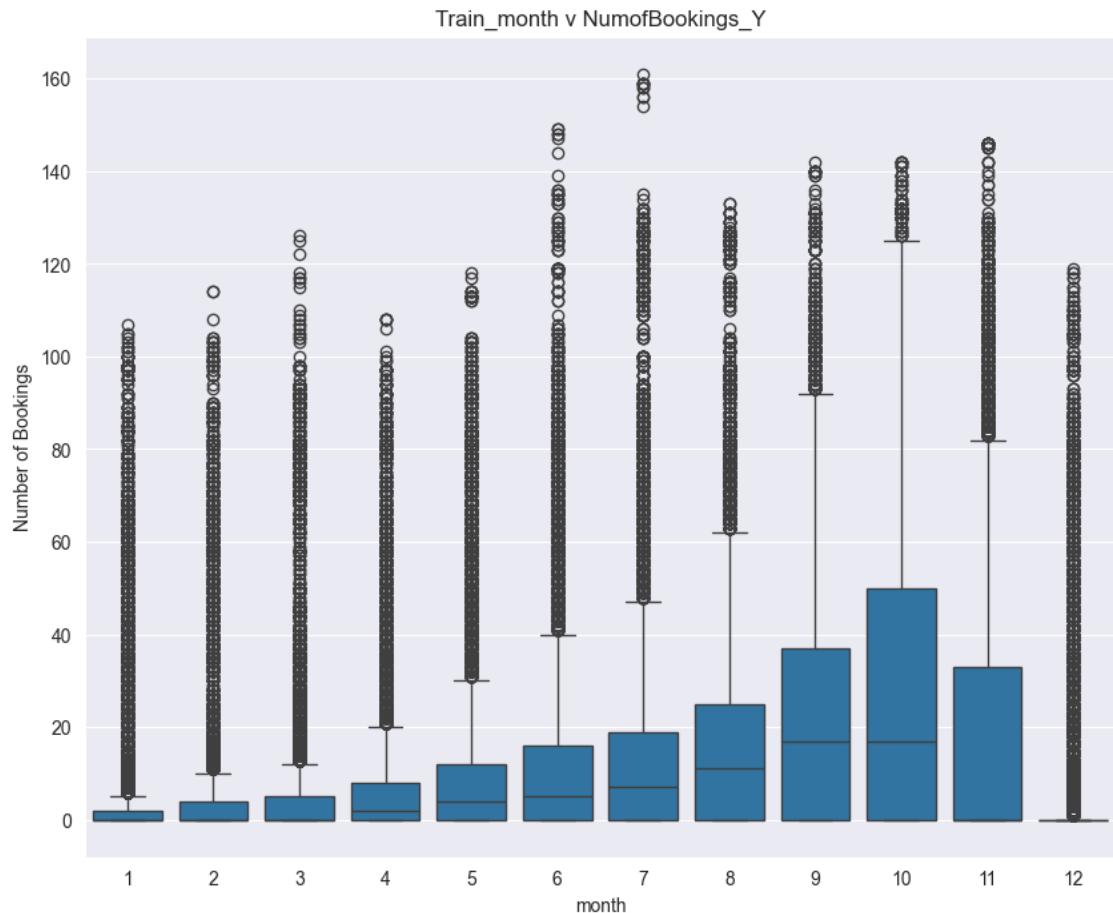


| ORIGIN | DESTINATION | CABIN | BOOKING CLASS | Departure Date | Departure Time | Number of Bookings | Number of days Before Departure | Unique_Flight_ID |
|--------|-------------|-------|---------------|----------------|----------------|--------------------|---------------------------------|------------------|
| a      | b           | c     | d             | Day X          | Time Y         | 34                 | 0                               | 1                |
| a      | b           | c     | d             | Day X-1        | Time Y         | 30                 | 1                               | 1                |
| a      | b           | c     | d             | Day X-2        | Time Y         | 12                 | 2                               | 1                |
| ...    |             |       |               |                |                |                    |                                 |                  |
| a      | b           | c     | d             | Day X-365      | Time Y         | 0                  | 365                             | 1                |
| f      | g           | h     | i             | Day V          | Time J         | 56                 | 0                               | 2                |
| f      | g           | h     | i             | Day V-1        | Time J         | 55                 | 1                               | 2                |
| f      | g           | h     | i             | Day V-2        | Time J         | 32                 | 2                               | 2                |
| ...    |             |       |               |                |                |                    |                                 |                  |
| f      | g           | h     | i             | Day V-365      | Time J         | 1                  | 365                             | 2                |

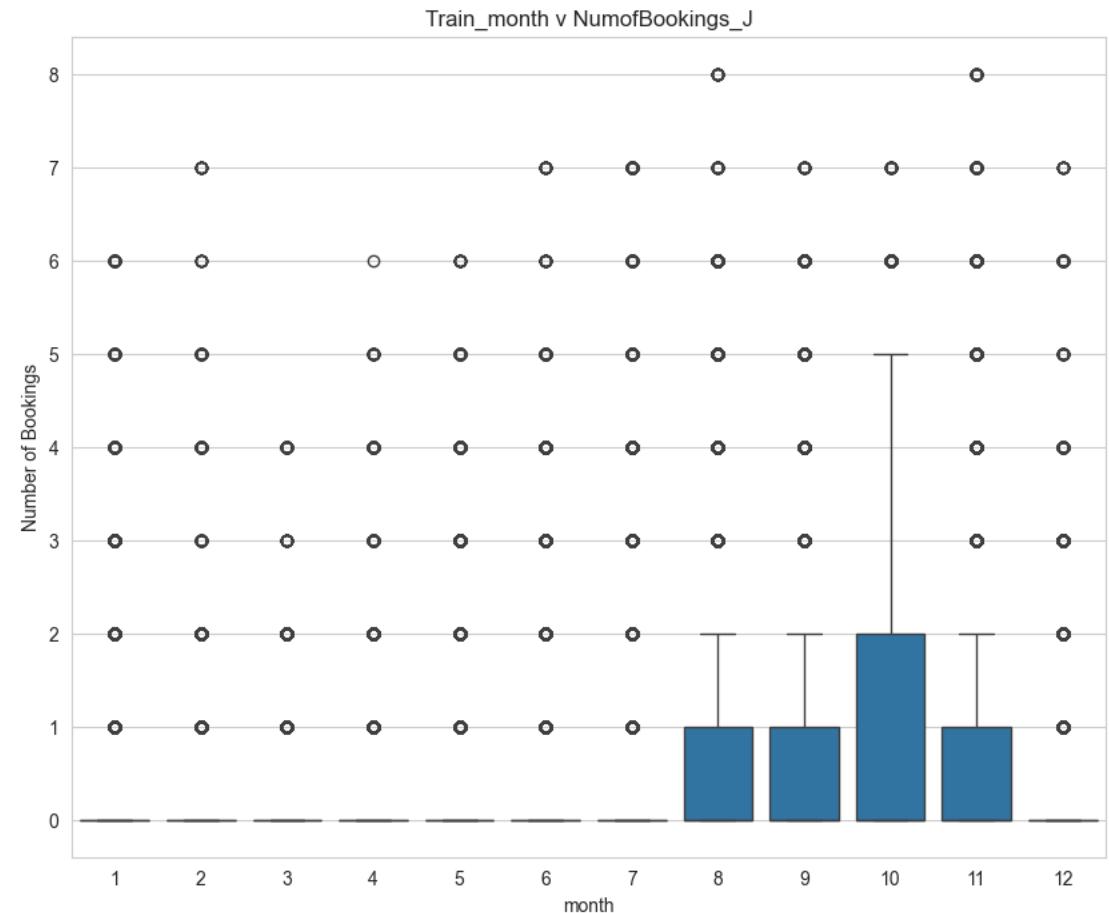
Similar “melting” or unpivoting of the set was conducted. Addition of Y and J col is the only change.

# EDA

## Economy

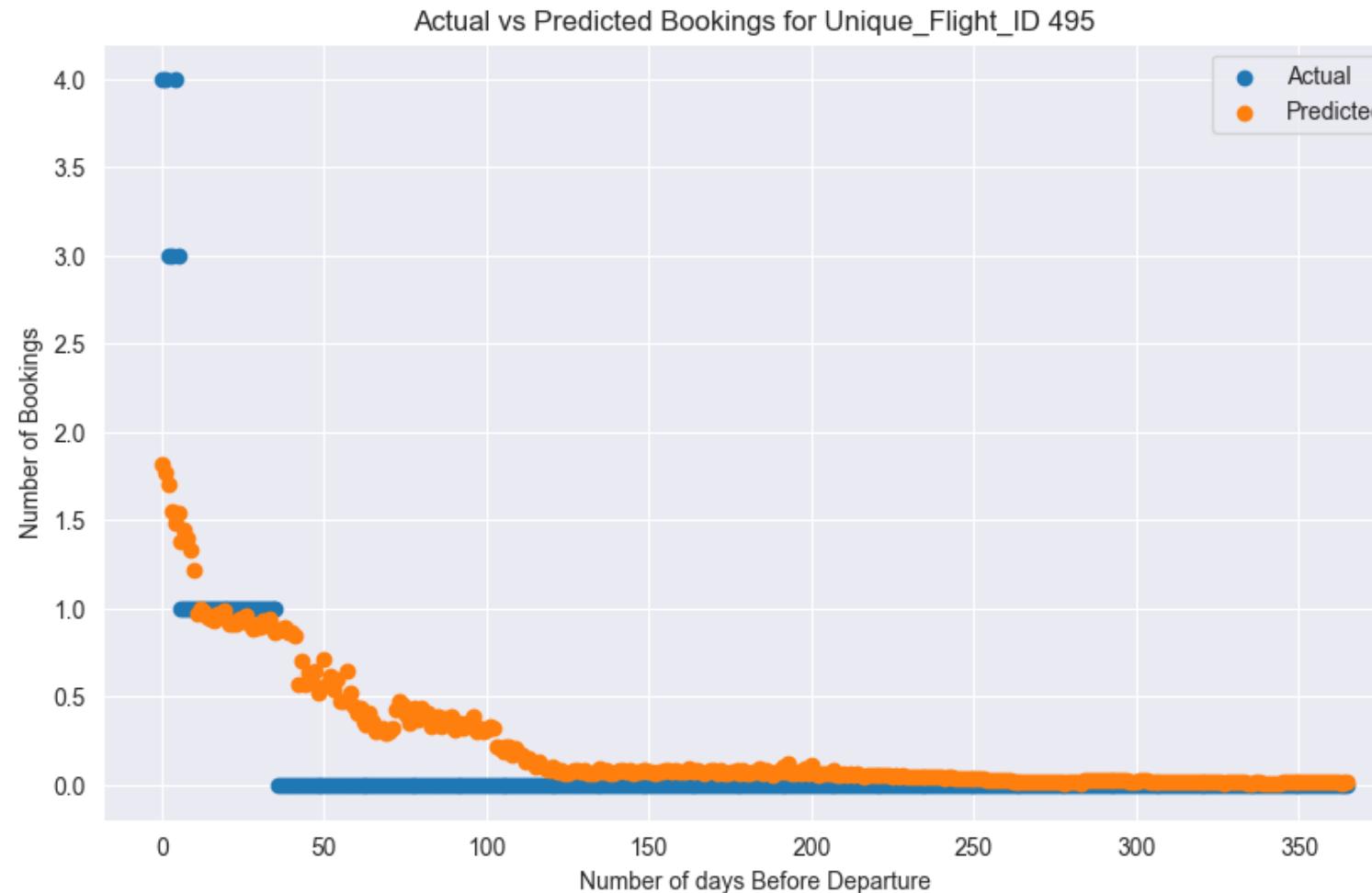


## Business



Month of Booking vs Number of Bookings

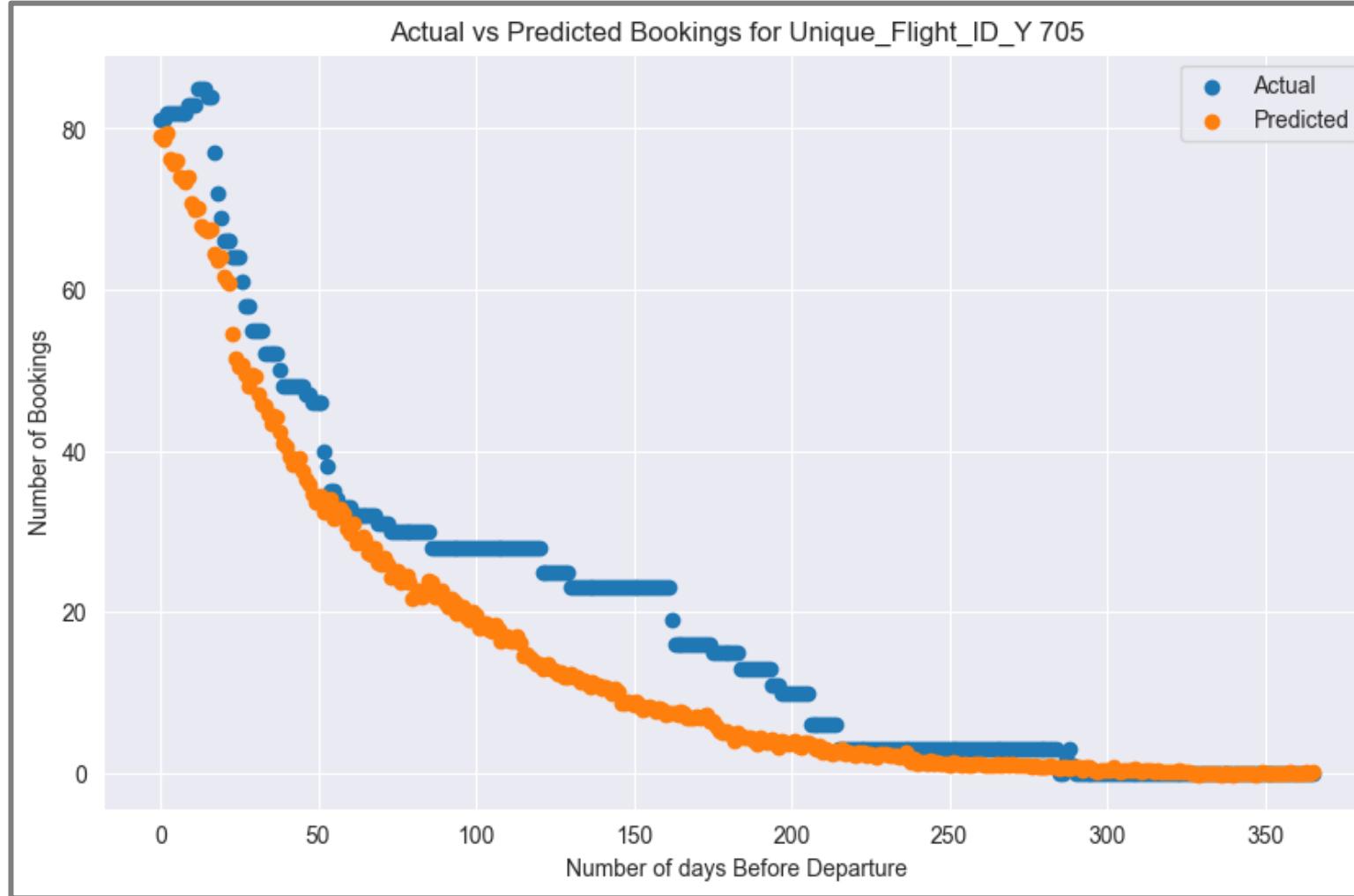
# Prediction Using Business Class



Flat curve.  
Fewer updates in the  
Booking Number all year  
round.

Prediction curve for Business Class. Trained a Model specifically on the Subset of Business Class booking data  
Used a fully connected dense neural network as the regressor.

# Prediction Using Economy Class



Curvilinear curve.  
Much more updates in the  
Booking Number all year round  
due to cheaper cost

Prediction curve for Economy Class. Trained a Model specifically on the Subset of Economy Class booking data  
Used a fully connected dense neural network as the regressor.

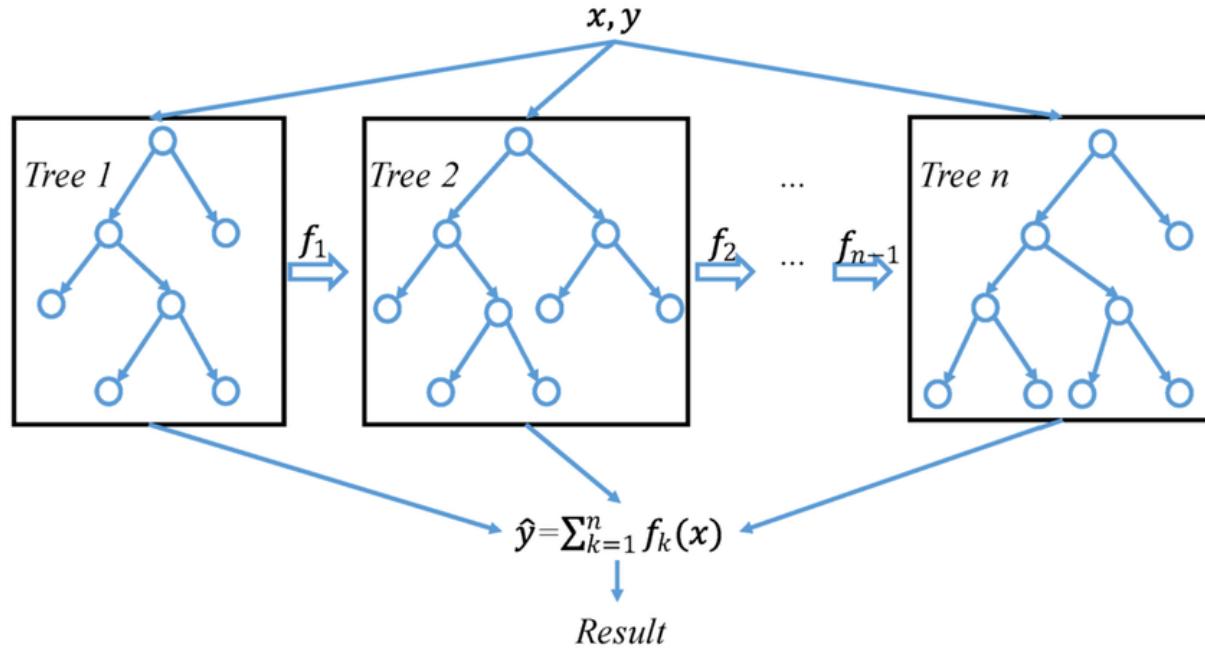
# **Booking Class Fare Prediction**

| y_test  | y_pred      | actual_vs_prd |
|---------|-------------|---------------|
| 522.96  | 523.568237  | -0.608237     |
| 565.79  | 555.228577  | 10.561423     |
| 276.28  | 282.207184  | -5.927184     |
| 233.55  | 235.611893  | -2.061893     |
| 245.47  | 243.585754  | 1.884246      |
| 179.8   | 181.19873   | -1.39873      |
| 567.19  | 567.44812   | -0.25812      |
| 404.78  | 403.30481   | 1.47519       |
| 305.47  | 297.259735  | 8.210265      |
| 1595.69 | 1583.772705 | 11.917295     |

# Features Used + Engineered

| Booking Class Fare USD | Avg Fare given by customer to book the ticket  |
|------------------------|--|
| B: 0                   |  |
| ...                    |  |
| B: 365                 | B: x : Number of Bookings x days before the flight leaves  |
| Origin_AAA             |  |
| ...                    |  |
| Origin_ZZZ             | Origin of the Flight   |
| Destination_AAA        |  |
| ...                    |  |
| Destination_YYY        | Destination of the flight  |
| dayofmonth             |  |
| year                   |  |
| dayofyear              |  |
| hour                   |  |
| minute                 |  |
| dayofweek_0            |  |
| ...                    |  |
| dayofweek_6            | One Hot Encoded. 0-> Sunday, 6 -> Saturday   |
| quarter_1              |  |
| ...                    |  |
| quarter_4              | Also OneHot Encoded. Gives the Model an Idea about the seasons.  |
| month_1                |  |
| ...                    |  |
| month_12               | Which month the plane flew in is importatn due to Seasonal Variations. December: More bookings beorehand |

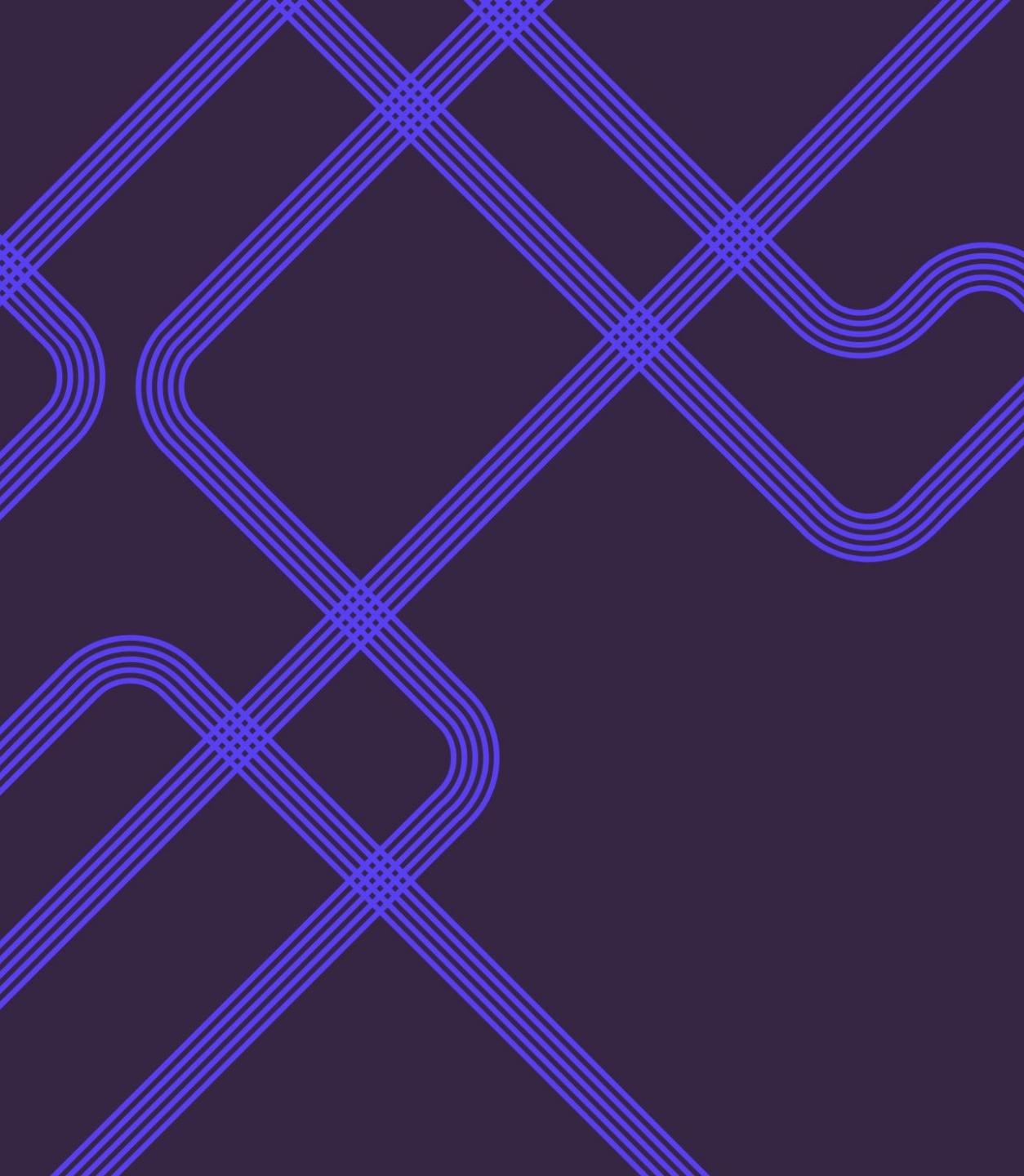
# Model Used



- Used XGBoost Regressor
- Loss/Objective: `reg:squarederror`
- `n_estimators=100`
- `learning_rate=0.1`

# Correlation Matrix

|                        | Booking Class Fare USD |
|------------------------|------------------------|
| Booking Class Fare USD | 1.000000               |
| Origin_ZZZ             | 0.765534               |
| Destination_YYY        | 0.765534               |
| B: 84                  | 0.509963               |
| B: 83                  | 0.509629               |
| B: 85                  | 0.509272               |
| B: 86                  | 0.508781               |
| B: 82                  | 0.508766               |
| B: 87                  | 0.508747               |
| B: 88                  | 0.508484               |



## Model-3 [*XGBR*]

Prediction of “Booking Class Fare USD” using Only the Flight Data

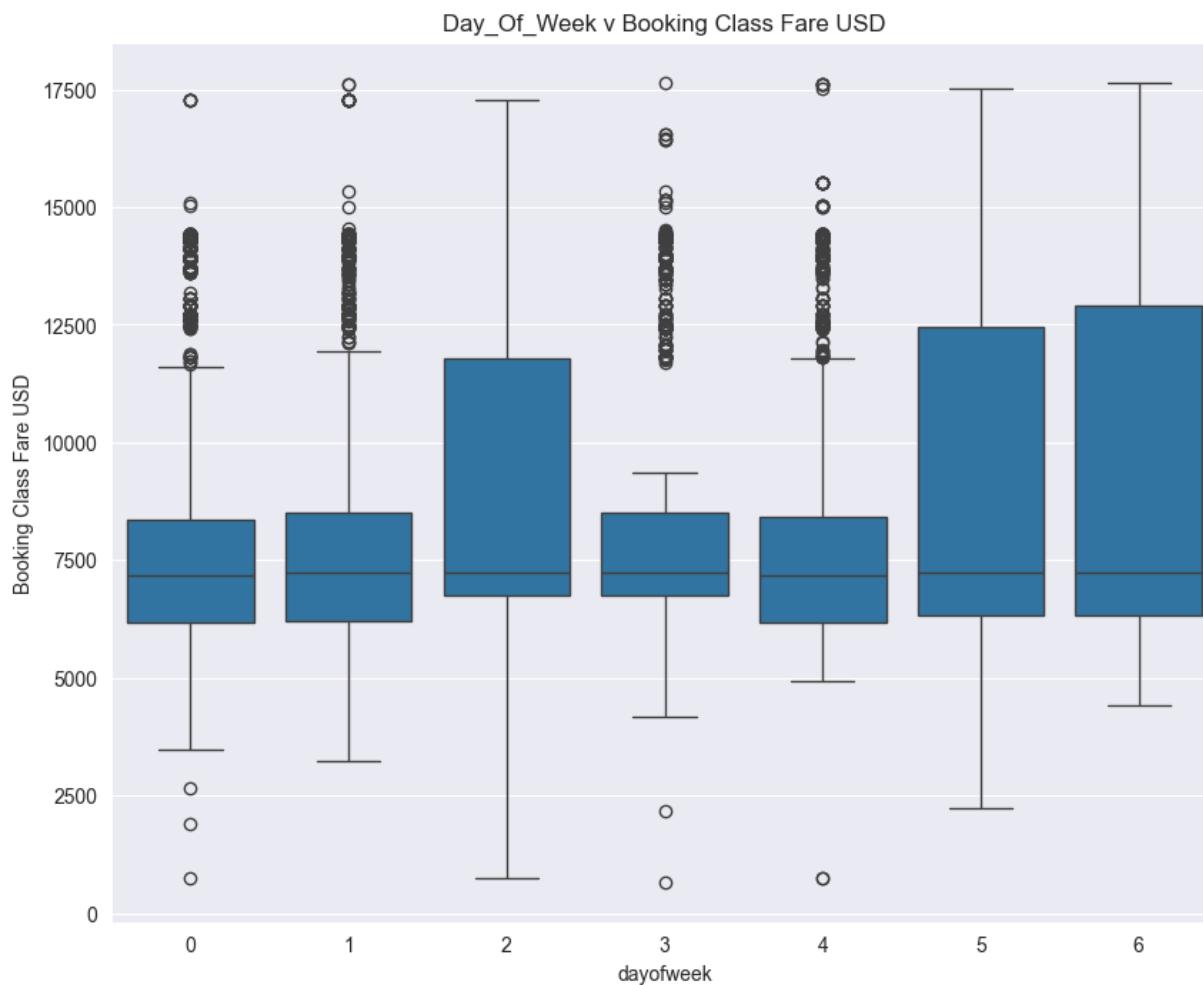
# Condensation/Transformation of Dataset

| Origin | Destination | Flight number | Cabin | Dep_Date | Dep_time | Booking Class | Booking Class Fare | B: 365   | ... | B: 0 |
|--------|-------------|---------------|-------|----------|----------|---------------|--------------------|----------|-----|------|
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | N             | 200.01             | 0        |     | 10   |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | H             | 463.65             | 0        |     | 7    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | G             | 149.83             | 0        |     | 1    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | T             | 127.29             | 0        |     | 5    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | S             | 90.92              | 0        |     | 10   |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | M             | 146.97             | 0        |     | 0    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | Y             | 681.83             | 0        |     | 3    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | L             | 309.1              | 0        |     | 44   |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | K             | 372.74             | 0        |     | 2    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | E             | 245.47             | 0        |     | 10   |
| BBB    | DDD         | 304 J         |       | 10/04/22 | 05:55    | J             | 836.38             | 0        |     | 0    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | Q             | 145.47             | 0        |     | 36   |
| BBB    | DDD         | 304 J         |       | 10/04/22 | 05:55    | D             | 386.69             | 0        |     | 1    |
| BBB    | DDD         | 304 J         |       | 10/04/22 | 05:55    | C             | 563.65             | 0        |     | 2    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | V             | 181.83             | 0        |     | 3    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | B             | 581.83             | 0        |     | 2    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | I             | 109.1              | 0        |     | 3    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | P             | 163.65             | 0        |     | 15   |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | O             | 218.2              | 0        |     | 3    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | R             | 272.74             | 0        |     | 0    |
| BBB    | DDD         | 304 Y         |       | 10/04/22 | 05:55    | W             | 518.2              | 0        |     | 0    |
|        |             |               |       |          |          |               | SUM: 6765.55       | SUM: 157 |     |      |

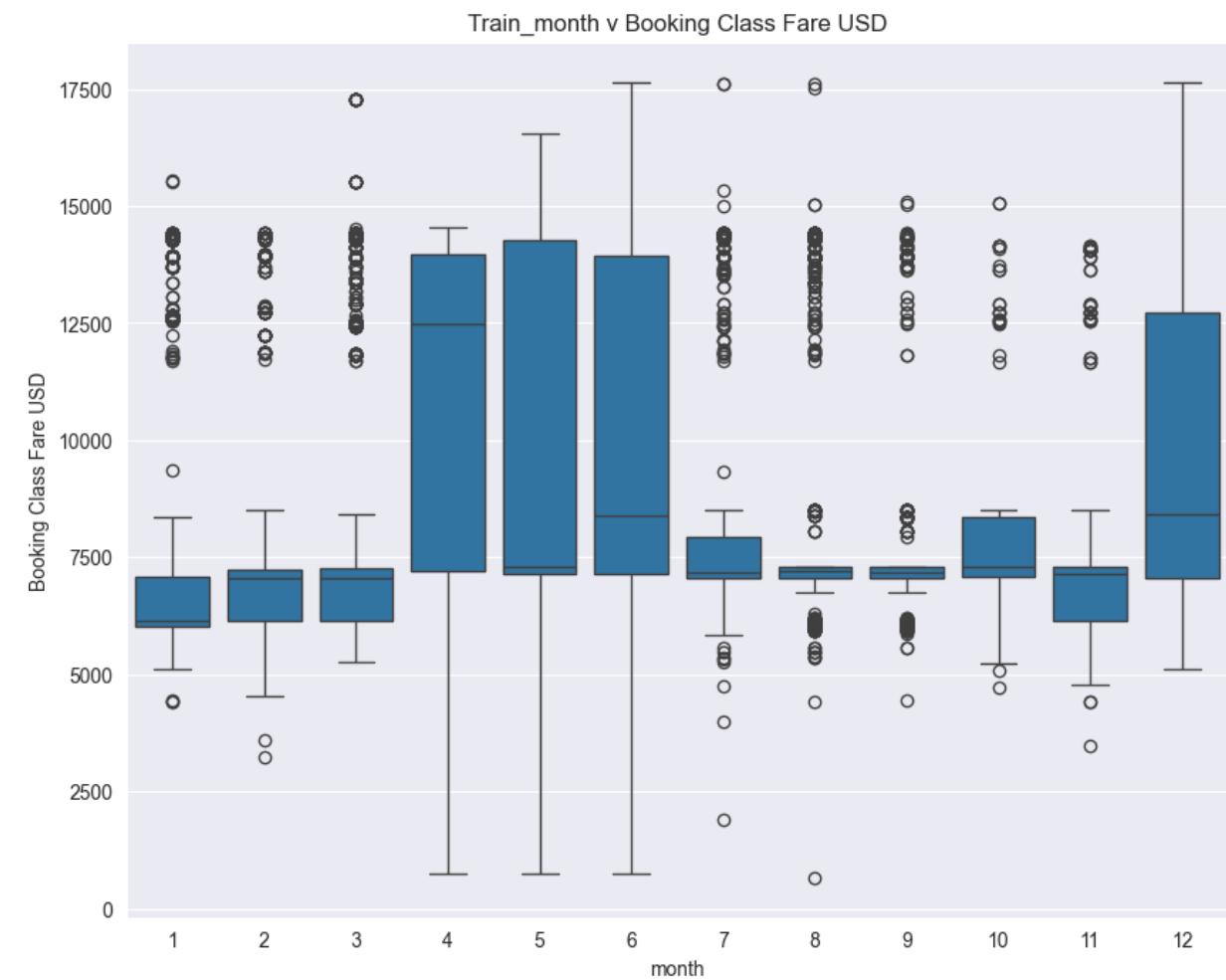
| Origin | Destination | Flight number | Dep_Date | Dep_time | Booking Class Fare USD | B: 365 | ... | B: 0 |
|--------|-------------|---------------|----------|----------|------------------------|--------|-----|------|
| BBB    | DDD         | 304           | 10/04/22 | 05:55    | 6765.55                | 0      |     | 157  |

9635 such rows

# Prediction Using Only the Flight Data: EDA

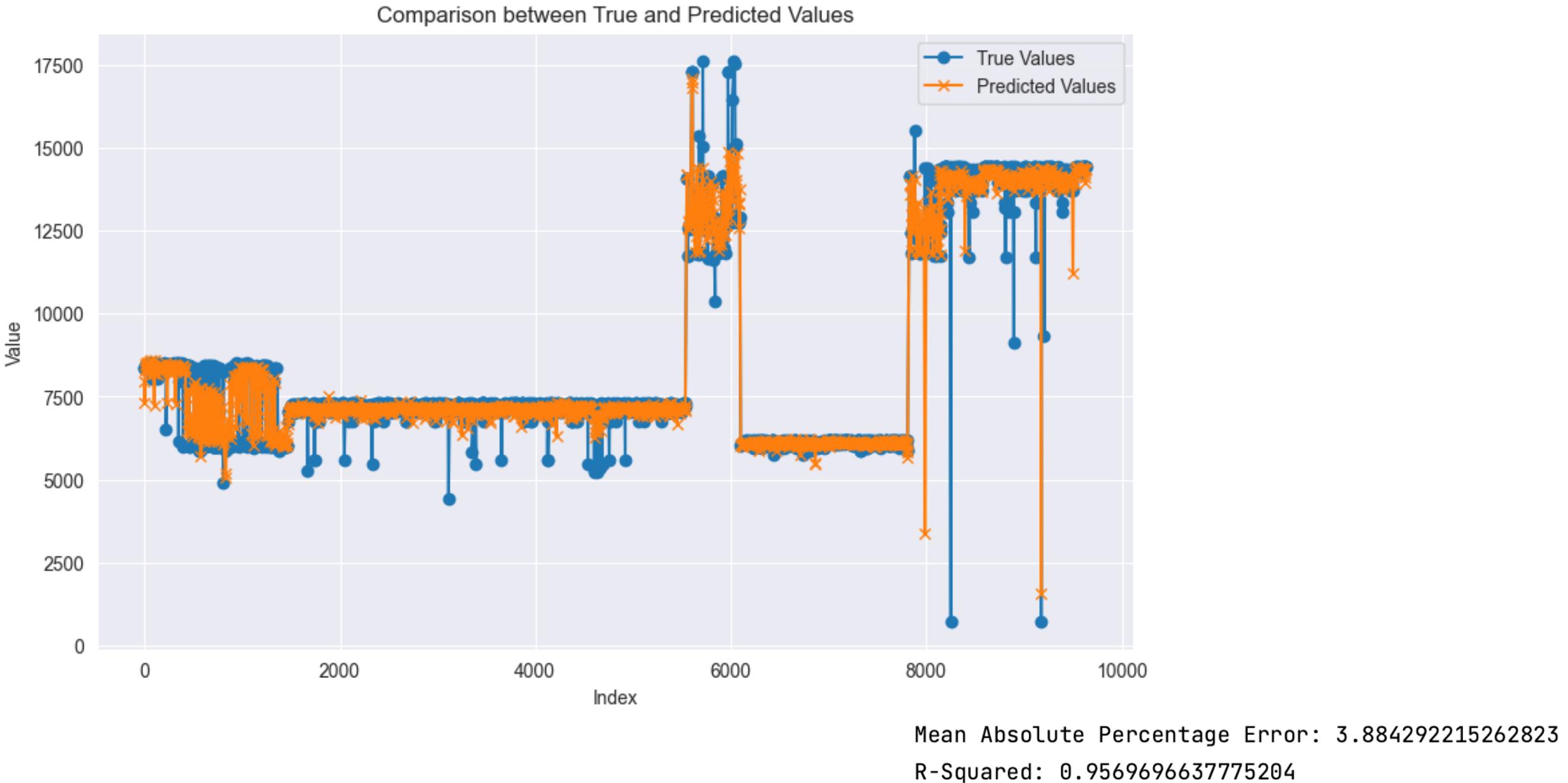


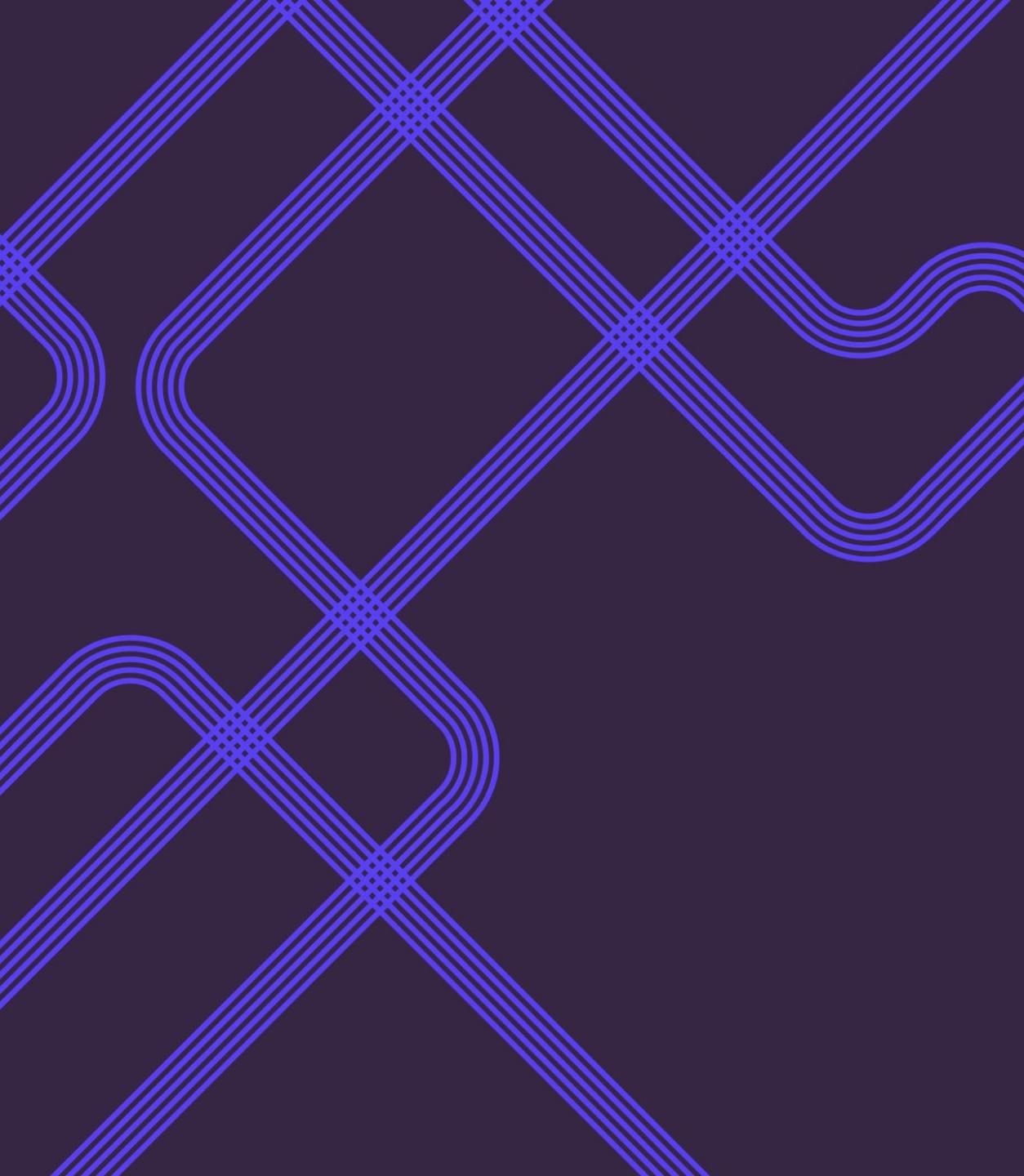
Day of Week vs Booking Class Fare



Month vs Booking Class Fare

# Prediction Using Only the Flight Data: EDA





## Model-4 *[XGBR]*

Predicts Booking Price Uses  
Y (Economy)  
&  
J (Business)

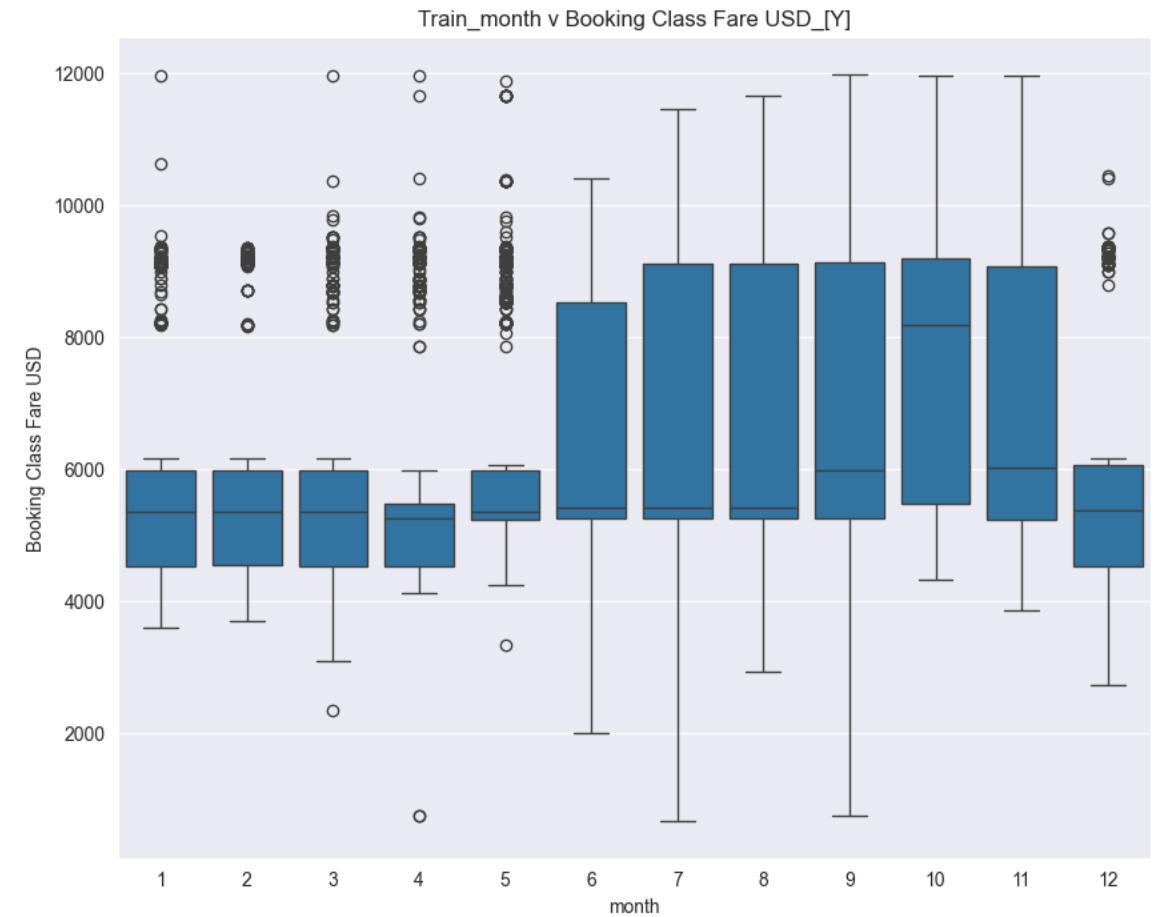
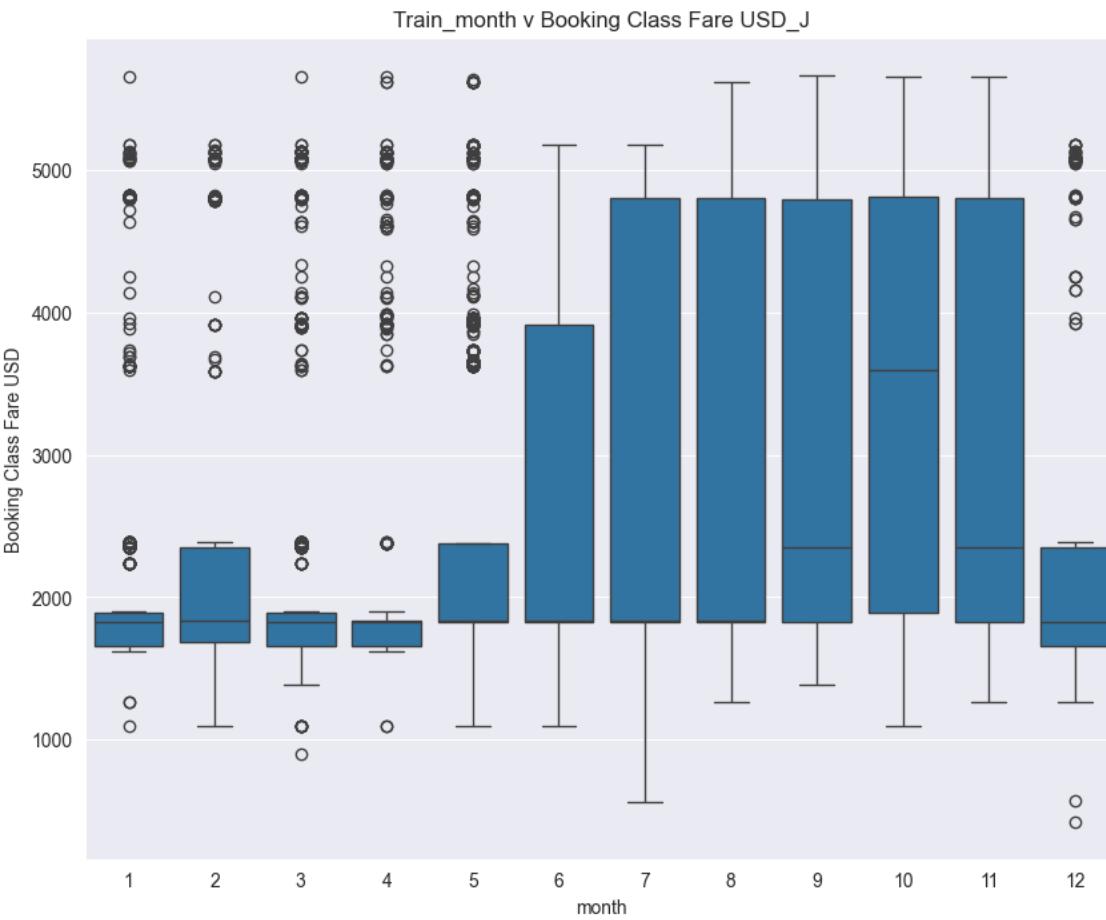
# Condensation of the Data

| Origin       | Destination | Flight number | Cabin | Dep_Date | Dep_time | Booking Class | Booking Class Fare | B: 365 | ... | B: 0 |
|--------------|-------------|---------------|-------|----------|----------|---------------|--------------------|--------|-----|------|
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | N             | 200.01             | 0      |     | 10   |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | H             | 463.65             | 0      |     | 7    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | G             | 149.83             | 0      |     | 1    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | T             | 127.29             | 0      |     | 5    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | S             | 90.92              | 0      |     | 10   |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | M             | 146.97             | 0      |     | 0    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | Y             | 681.83             | 0      |     | 3    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | L             | 309.1              | 0      |     | 44   |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | K             | 372.74             | 0      |     | 2    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | E             | 245.47             | 0      |     | 10   |
| BBB          | DDD         | 304 J         |       | 10/04/22 | 05:55    | J             | 836.38             | 0      |     | 0    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | Q             | 145.47             | 0      |     | 36   |
| BBB          | DDD         | 304 J         |       | 10/04/22 | 05:55    | D             | 386.69             | 0      |     | 1    |
| BBB          | DDD         | 304 J         |       | 10/04/22 | 05:55    | C             | 563.65             | 0      |     | 2    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | V             | 181.83             | 0      |     | 3    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | B             | 581.83             | 0      |     | 2    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | I             | 109.1              | 0      |     | 3    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | P             | 163.65             | 0      |     | 15   |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | O             | 218.2              | 0      |     | 3    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | R             | 272.74             | 0      |     | 0    |
| BBB          | DDD         | 304 Y         |       | 10/04/22 | 05:55    | W             | 518.2              | 0      |     | 0    |
| SUM: 6765.55 |             |               |       |          |          |               | SUM: 157           |        |     |      |

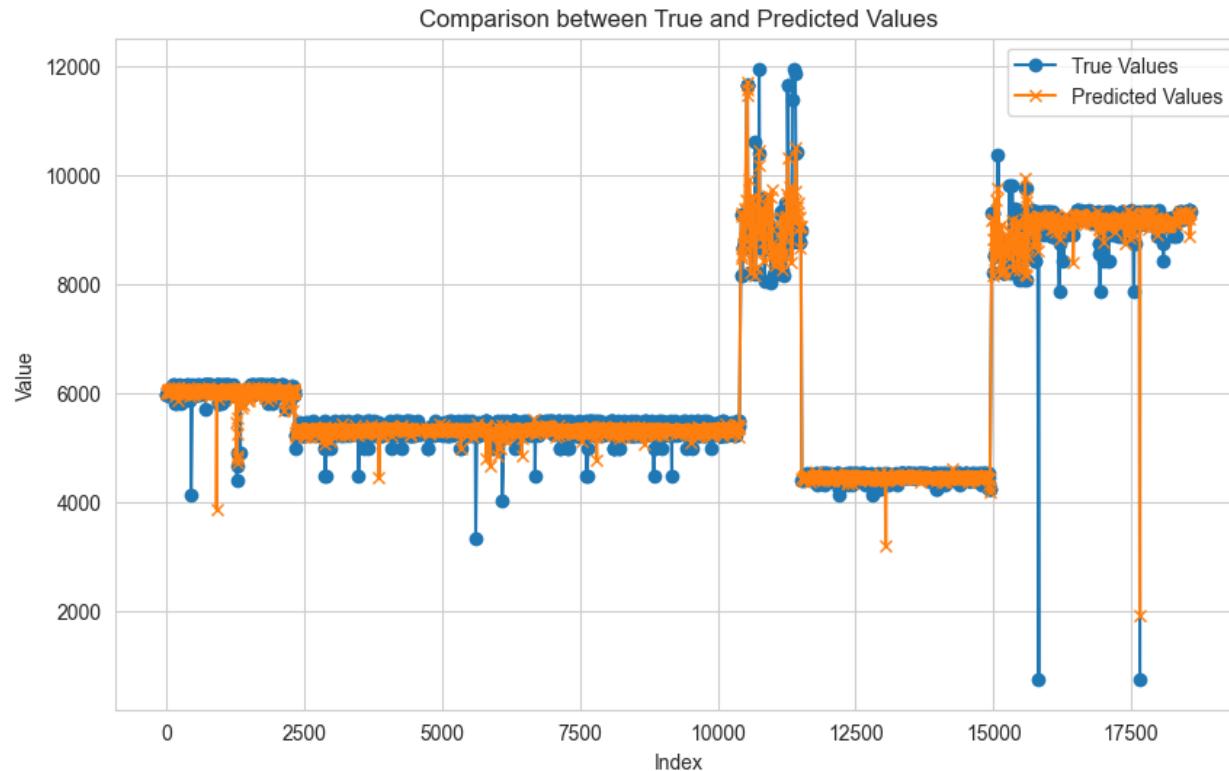
| Origin       | Destination | Flight number | Cabin | Dep_Date | Dep_time | Booking Class Fare USD | B: 365 | B: 364 | ... | B: 0     |
|--------------|-------------|---------------|-------|----------|----------|------------------------|--------|--------|-----|----------|
| BBB          | DDD         | 304           | Y     | 44661    | 05:55    | 4978.83                | 0      | 0      |     | 154      |
| BBB          | DDD         | 304           | J     | 44661    | 05:55    | 1786.72                | 0      | 0      |     | 3        |
| SUM: 6765.55 |             |               |       |          |          |                        |        |        |     | SUM: 157 |

18574 such rows

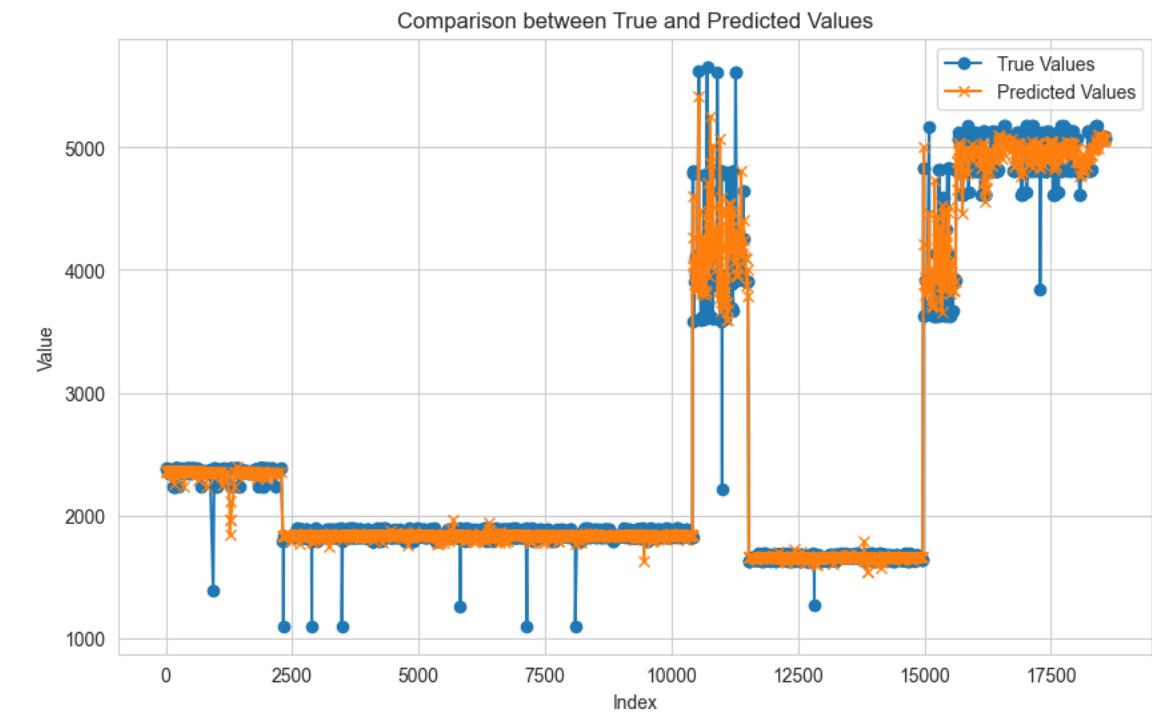
# EDA



# Prediction Using Booking Class [Y: Economy or J: Business]



Economy

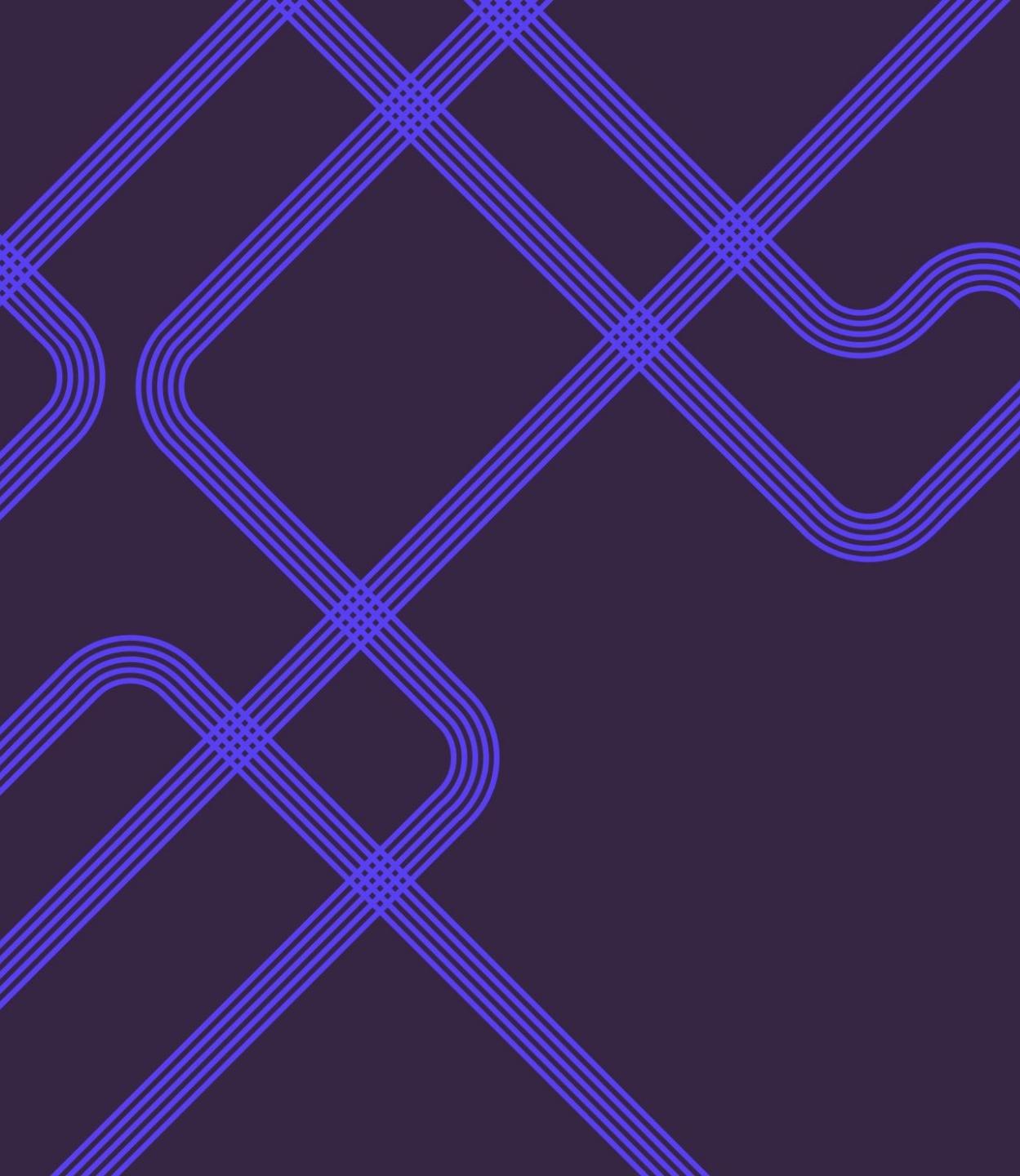


Business

Models trained for Economy & Business perform accurately.

Mean Absolute Percentage Error: 2.4636589250020116  
R-Squared: 0.9705404089251191

Mean Absolute Percentage Error: 2.331472896155585  
R-Squared: 0.983082294132624



## Model-3

**[XGBR]**

Predicts Booking Price  
(by *both* cabin +  
booking class)

# Input Features

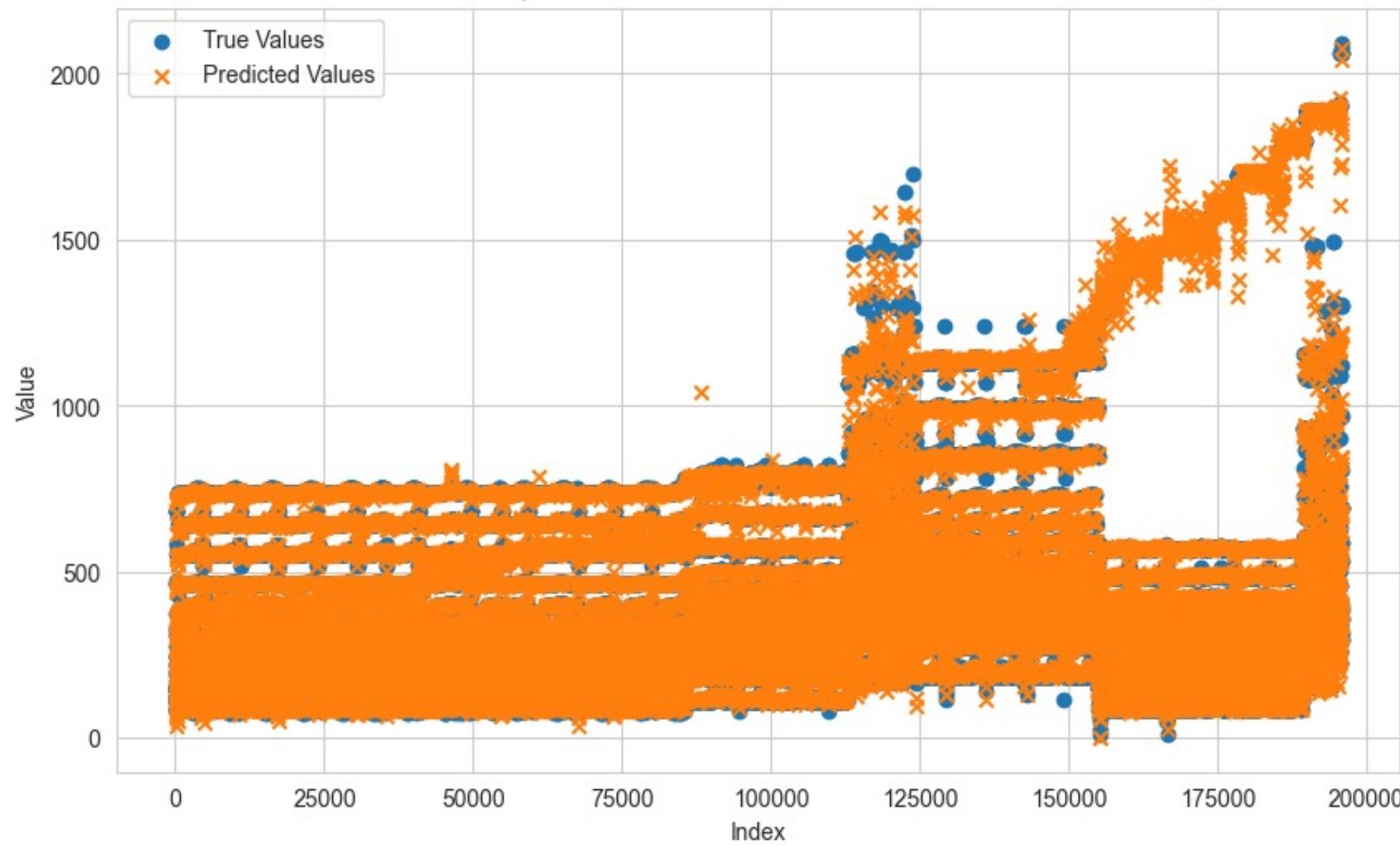
| Name of Feature             | Explanation/Utility   |
|-----------------------------|---|
| Booking Class Fare USD      | Target  |
| num of bookings before: 365 | Same as "Class RDay Res Hold Total RDay: x" Which is basically the number of booking at x days before departure |
| num of bookings before: 364 |   |
| num of bookings before: 363 |   |
| ...                         |   |
| num of bookings before: 3   |   |
| num of bookings before: 2   |   |
| num of bookings before: 1   |   |
| num of bookings before: 0   |   |
| Day                         |   |
| Month                       |   |
| Year                        | Date of Booking   |
| Hour                        |   |
| Minute                      |   |
| Origin_AAA                  |   |
| ...                         | One-Hot Encoding of the Origin City of the Flight   |
| Origin_ZZZ                  |   |
| Destination_AAA             |   |
| ...                         |   |
| Destination_YYY             | One-Hot Encoding of the Origin City of the Flight   |
| Cabin_J                     |   |
| Cabin_Y                     |   |
| Booking Class_A             |   |
| ...                         | Type of Cabin   |
| Booking Class_Z             |   |
| Airline_AA                  |   |
| Flight number_65            | One-Hot encoding of the flight number   |
| ...                         |   |
| Flight number_9538          |   |
| Day_of_Week_Friday          |   |
| ...                         | Feature engineered the Day of the week, as it might influence the price   |
| Day_of_Week_Wednesday       |   |

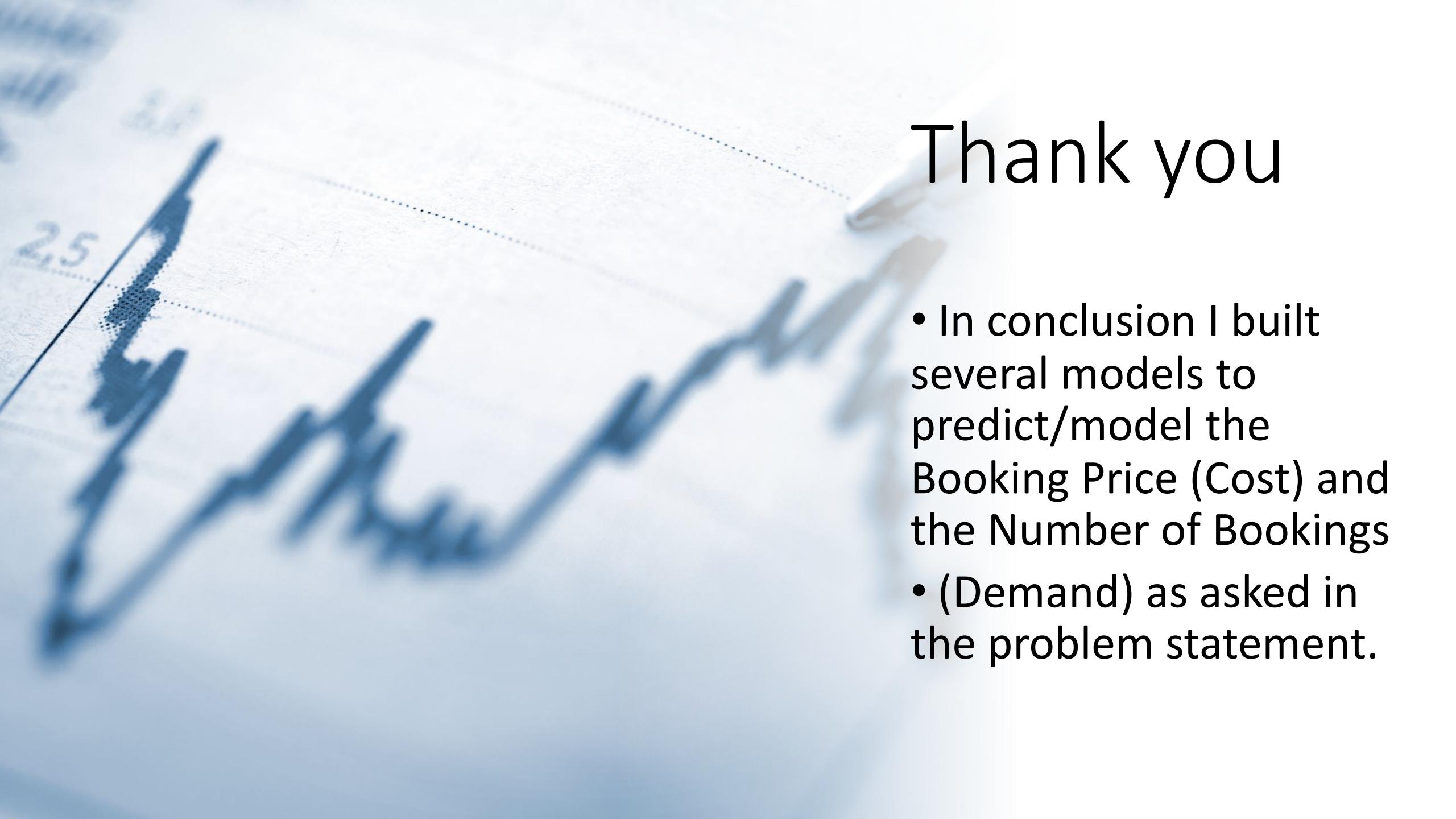
# Correlation of the Features to the Output

< < 10 rows <|> Length: 10, dtype: float64 pd.Series ↗

|                        | Booking Class Fare USD |
|------------------------|------------------------|
| Booking Class Fare USD | 1.000000               |
| Cabin_J                | 0.549081               |
| Booking Class_J        | 0.450546               |
| Origin_ZZZ             | 0.338442               |
| Destination_YYY        | 0.338442               |
| Booking Class_C        | 0.283994               |
| Booking Class_Y        | 0.282234               |
| Booking Class_B        | 0.200664               |
| Origin_CCC             | 0.186830               |
| Booking Class_D        | 0.169914               |

Comparison between True and Predicted Values





# Thank you

- In conclusion I built several models to predict/model the Booking Price (Cost) and the Number of Bookings
- (Demand) as asked in the problem statement.