# Introduction and Motivation

## General Background Information

Flight price prediction is an important area of investigation and implementation in this era of dynamic pricing. The airlines use sophisticated algorithms that calculate ticket prices, taking into consideration a great number of factors: demand, competition, seasonality, and customer behavior. To the passengers, forecasting ticket prices helps them to plan and optimize their travel costs; in the case of airlines, it helps enhance revenue management and improve customer satisfaction. With the advancement of machine learning, the clear and articulate analysis of historical data enables the construction of predictive models that forecast the prices of flight tickets.

## Motivation

This project is motivated by the hustle tourists go through just to get fairly priced tickets and the complications airlines have to endure in trying to judge the best price. Being frequent travelers, we understand the logistical headache and budgetary challenge caused by unexpectedly high ticket prices. This work aims to close this gap by applying machine learning to make flight pricing more candid and predictable.

Besides, the work with advanced regression techniques in data analysis follows my aim to study the application of machine learning to more reality-based situations.

## Project Significance

This project is important to several stakeholders:

* **Travelers and Tourists:** These people can be benefited by their savings on the cost of travel through the making of correct prior predictions thus creating more pragmatic plans.
* **Airlines: Improved** price prediction can benefit their revenue management strategies to efficiently match demand with supply.
* **Travel Agencies:** They can use the insights to design better packages and maintain competitive pricing for the customers.
* **Researchers and Data Scientists:** The project provides a use case for developing and testing machine learning models in dynamic pricing scenarios.

The present project tries to solve a practical problem, while continuing the development of wide knowledge regarding the applications of machine learning in dynamic pricing systems.

# Problem Description

## Problem Statement

The aim of this project is to provide the opportunity of developing the models that can predict flight ticket prices based on various parameters such as time of departure, destination, duration, and class. This price forecasting capability should also be in a position to make use of knowledge availability and help travellers in making intelligent choices while ensuring the optimization of travel costs. These predictions will be of great use in helping the airlines modify their pricing strategies accordingly.

## Objectives

* Understand the factors that influence ticket prices.
* Build models to predict ticket prices based on these factors.
* Test and compare multiple models to find the most accurate one.
* Provide actionable insights for travelers and businesses to improve decision-making.

## Type of Problem

This is a **regression problem** because the goal is to predict a continuous variable: flight ticket prices. A regression approach is appropriate since we are interested in numerical predictions, not categories or classifications.

## Research Questions

* What are the key factors that influence flight ticket prices?
* Which machine learning model performs best in predicting flight ticket prices?
* How does the inclusion of interaction terms and polynomial features affect model performance?
* What is the impact of regularization techniques like Lasso and Ridge on prediction accuracy?
* How well do the models generalize to new and unseen data?

# Variables

The **response variable** is price, which is **quantitative** because it represents the flight ticket price in numerical format.

**Predictors and Quantitative predictors**

**Predictors**:

* + airline (Qualitative: Factor with 6 levels)
  + source\_city (Qualitative: Factor with 6 levels)
  + departure\_time (Qualitative: Factor with 6 levels)
  + stops (Qualitative: Factor with 3 levels)
  + arrival\_time (Qualitative: Factor with 6 levels)
  + destination\_city (Qualitative: Factor with 6 levels)
  + class (Qualitative: Factor with 2 levels)
  + duration (Quantitative: Numeric)
  + days\_left (Quantitative: Integer)

**Quantitative predictors**:

* + duration
  + days\_left

## Key predictors

**Expected Key predictors**

* **Duration:** Longer flights are bound to be priced higher.
* **days\_left:** Fewer days before departure possibly means higher prices.
* **Class:** Business class tickets will be more expensive than economy class.
* **Stops:** Nonstop flights will, most of all probabilities, be costlier compared to those with at least one, or more, stops.

The key predictors can be proved by performing a linear regression model, with price being the response variable. We can check the p-values of the coefficients for each predictor. The predictors are statistically significant if their p-values are <0.05.

## Multicollinearity

A grid of blue dots

Description automatically generated

Figure 1: Collinearity plot

By referring to the correlation plot:

* Duration and days\_left are very little correlated, having a value of -0.04; hence, no collinearity.
* Qualitative predictors like stops and class, though, do not contribute directly to numerical multicollinearity but could be interpreted through regression coefficients or interaction effects.

# Overview of Data

## Data Source

This dataset is downloaded from Kaggle and is titled Flight Price Prediction. It contains various information about features in flight ticket prices such as airlines, departure and arrival time, source and destination city, flight duration, number of stops, and days left until the flight date.

## Data Cleaning

Several steps were taken to clean and prepare the dataset for analysis:

1. **Missing Values**:

In this step, I checked for missing values and no missing values were identified, so no imputation was required.

1. **Outliers**:

In this step, outliers in the price column were visually inspected using histograms and boxplots. No extreme outliers were removed as they appeared to represent legitimate flight prices.

1. **Unnecessary Columns**:

In this step, I removed non-informative columns such as X and flight

* + **Data Type Conversion**:

In this step, I converted character columns to factors to ensure proper handling during analysis

## Characteristics of data

1. **Data Structure**

A screenshot of a computer program

Description automatically generated

Figure 2: Data Structure

The dataset consists of **300,153 observations** and **10 variables**, including **2 numeric variables** (duration and days\_left) and **8 categorical variables** (airline, source\_city, departure\_time, stops, arrival\_time, destination\_city, class, and price). The target variable, price, is numeric and represents flight ticket prices. The data is clean and well-structured for analysis.

1. **Data Summary**

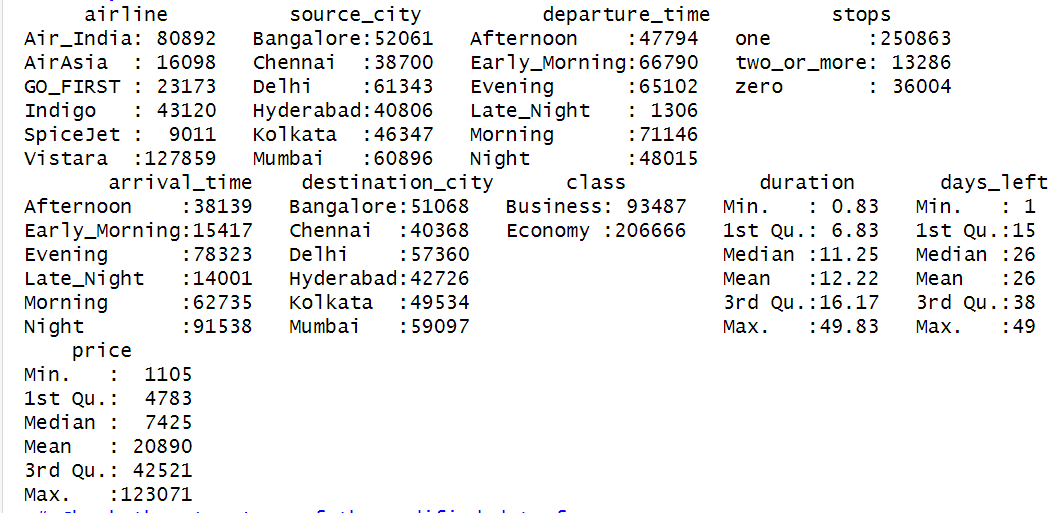


Figure 3: Data Summary

The dataset consists of 300153 observations with key characteristics as described below: The target variable price ranges from 1105 to 123071, with the overall median being 7425. The numeric predictors are duration, with a mean of 12.22 hours, and days\_left with a median of 26 days. Most flights are non-stop, numbering 250863, while Economy is the dominant class of travel with 206666 observations. The most frequent airlines are Air\_India and Vistara; the source city and destination city are Bangalore and Mumbai, respectively. Departure and arrival times are distributed over several daily timetables, peaking in the morning and evening.

1. **EDA**

Some variables are explored to understand the characteristics of the data

A graph showing a number of tickets

Description automatically generated

Figure 4: Ticket price distribution

This histogram is a right-skewed distribution of the prices of flight tickets. The biggest chunk of ticket prices falls below 20,000; there is a sharp peak at about 5,000. Then there are smaller chunks showing a big tail up toward 120,000. That would suggest that most flights are fairly reasonably priced, and really expensive flights are very rare.

A graph with orange squares

Description automatically generated

Figure 5: Airline Distribution

The bar plot shows the distribution of flights by airline. **Vistara** has the highest number of flights, followed by **Air India** and **Indigo**, while **SpiceJet** has the lowest count. This indicates that Vistara dominates the dataset, potentially reflecting a larger share of flights or data availability compared to other airlines.

A graph with a red and blue line

Description automatically generated with medium confidence

Figure 6: Ticket prices by class

This boxplot compares Business and Economy classes ticket prices. The prices for business class sit considerably well above the median, at about 50,000 points in a range of extreme outliers well over 100,000. Prices for Economy class are lower, below 10,000 at median, with some outliers but within a narrower range. This suggests that there is quite a huge price increase between the two classes.

A graph of different colored rectangular shapes

Description automatically generated

Figure 7: Average flight duration by source city

The above bar plot gives the average duration of flights according to each source city. Thus, the average duration of flights from Kolkata is the longest, followed by Hyderabad and Chennai while that from Bangalore and Mumbai are relatively shorter. That can again hint at the geographical distance and flight routes from these cities to their respective destinations.

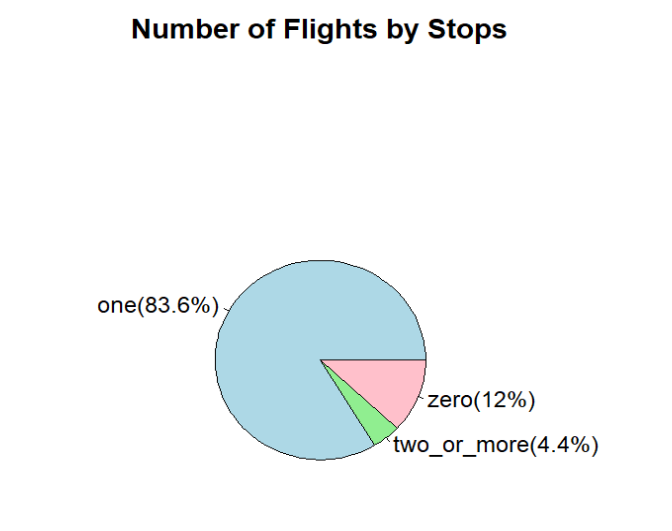


Figure 8: Number of flights by stop

The pie chart shows the distribution of flights based on the number of stops. The majority of flights (**83.6%**) have one stop, while **12%** are non-stop flights, and a small proportion (**4.4%**) have two or more stops. This indicates that most flights include a single stopover, with fewer direct or multi-stop options.

For the variables duration and days left, we examined the trends of the data to evaluate whether the effects of these variables on price could be modeled by linear terms or if they would be better served using polynomial ones. Scatter plots of the price vs. duration as well as price vs. number of days left were generated and a smoothing generative additive model was fit to the data separately for each ticket class to show the average trend.

A graph of different colored lines

Description automatically generated

Figure 9: A) Price trends of businesss and economic tickets based on duration of flight. B) Price trend of business and economic class tickets based on time until flight. Smoothing generalized additive model are fit to the data to show the trend for each class.

The relationship between the ticket price and the duration of the flight appears to differ based on the class of the ticket. While the economy class tickets show a slight non-linear effect of a more rapid rise in ticket price as the duration of the trip begins to increase before slowing down its rate of increase, this effect is much more pronounced in the business class ticket price trends. Busine class tickets appear to increase rapidly as the duration of the trip increases before leveling off at about 7-8 hours or even potentially declining slightly. This may be indicative of multi-stop flights with long wait times being in less demand. While the economy class tickets appear to be able to be approximated as linear, the trend for business class tickets is clearly non-linear and a simple linear term may be inappropriate.

The relationship between the time left until the day of the flight and the price of the ticket appear to be fairly non-linear. The pattern of the price trend appears to be somewhat similar in both economy and business class tickets, both appearing like they could be approximated crudely with a line but both more clearly resembling perhaps an exponential decay and could potentially be approximated as an exceptionally flat sloped quadratic. However, economy prices begin to rise much earlier than business class tickets do as the day of the flight nears.

# Statistical Learning Methods

For this project, we try each of the following regression techniques to best model the data:

* **Multiple Linear Regression**

In Multiple Linear Regression, the relationship between the target variable and multiple predictors is modelled by fitting a linear equation to the data. Each predictor is attributed with a certain coefficient, standing for its contribution towards the target.

Inclusion of all the factors as terms in the multiple linear regression model forms our baseline model with which to compare subsequent models..

* **Subset Selection Model**

Subset Selection identifies the most relevant predictors by exploring all possible combinations of predictors and selecting the best subset based on performance metrics like Adjusted R-squared or BIC. In situations where the model may be overfitting to the training data by including a large variety of terms in the model, subset selection methods produce potentially reduced models, such that they are resilient to noisy variation in the excluded variables with weak or spurious effects.

We utilize the forward step-wise selection algorithm which adds each subsequent term, each time selecting the one that maximally improves the adjusted R2 until the improvement in adjusted R2 is minimal it is maximized all together. In addition to potentially creating better predictive models than just multiple linear regression, the simpler model leads to easier interpretability. It is particularly useful for identifying and analyzing key contributors to ticket prices while discarding irrelevant predictors.

* **Interaction Model**

The Multiple Linear Regression method can be extended to include non-linear interaction terms. Interaction terms provide the combined effect of several predictors on the target variable. Such interaction terms could be source\_city and destination\_city that capture city-specific price effects.

Interaction terms were useful in identifying relationships between the variables not captured by additive models. Hence, this is particularly useful as an analytical framework in this research when it comes to route-specific pricing patterns.

* **Polynomial Regression**

Polynomial Regression extends the multiple linear regression method to include polynomial terms, modeling non-linear relationships between the dependent and independent variables. For example, being able to consider non-linear terms, one may be able to model phenomena such as diminishing returns or other nonlinear effects.

* **Ridge Regression**

Ridge Regression penalizes large coefficients by adding a constraint that is proportional to the sum of squared values of coefficients. Unlike Lasso, it does not get rid of predictors; it only shrinks the coefficients.

Ridge regression works effectively in multicollinearity issues of predictors, ensuring model stability and generalizing well to unseen data.

* **Lasso Regression**

Lasso is a regularization technique whereby large coefficients are penalized by adding a constraint proportional to the sum of absolute values of coefficients. This penalization term has such effect that it often reduces term coefficients completely to zero, in effect performing feature selection.

Thus in situations where the data have high-dimensional features, Lasso can potentially find the best predictive combination features all while the terms most likely to be spurious and susceptible to noisy variance are removed, preventing overfitting.

* **Cross validation**

For the penalized regression methods, the optimal penalization strength parameter lambda was chosen using 10-fold cross-validation to ensure robustness; that is the classical way to split the data into 10 subsets and exchange the sets between training and validation. In this manner, overfitting will be less likely to happen, and the results confirm consistency across different splits.

# Data Partitioning

1. **Data for training**

The training data contains 80% of the whole dataset, which was chosen randomly to represent a proper distribution of all variables. The model will be trained on a big enough subsample in this way and preserve part of it for independent testing.

Besides, to check the model robustness, and also to avoid overfitting to a plausible degree, ten-fold cross-validation was done in training.

1. **Data for Testing**

This model's performance was then tested on the remaining 20% of the dataset-the test set. In such a way, it is providing a random subset of the data for checking how the generalization of the already trained model would go for unseen data.

# Model Evaluation Metrics

Success of the models on this regression task was measured by the following metrics:

* **Mean Squared Error (MSE):** This will calculate the average of the squared differences between predicted and actual ticket prices. Lower values of MSE portray better results.
* **R-squared(R²):** This shows the percentage of deviation in a dependent variable explained by the model. The higher this value is, the better the fit; the value ranges between 0 and 1.
* **Adjusted R-squared:** Penalizes R² for the number of predictors in the model to take care of overfitting. It's a better measure of the tradeoff between fit and model complexity.

# Model Construction, Interpretation, Results, and Evaluation

1. Multiple Linear Regression (see Appendix A for full table of the model coefficients and associated statistics and p-values)

### Model Construction and Rationale

The first model we fit was the baseline multiple linear regression model with linearly modeled all terms. The ten flight and trip features being modeled included eight categorical variables, with each ranging from 2-6 levels values they could take. Using dummy variables to represent each of the levels of each of these categorical variables resulted in the construction of a multiple linear regression model with thirty variables, including the dummy variables, and one intercept term.

### Results and Interpretation

Fitting the model resulted in nearly every single variable reaching substantial significance. The exceptions are the dummy variable terms airlineAirAsia (which is compared to the baseline airline Air India), source\_cityChennai (which is compared to the baseline source city Bangalore), and destination\_cityMumbai (which is compared to the baseline destination city Bangalore). In essence, with these terms not reaching statistical significance, one cannot claim that the effect on price flying from Chennai has a substantially different effect from flying from Bangalore or claim that the effect on price of flying to Mumbai differs substantially from to Bangalore. Additionally, one cannot claim that flying with Air Asia has a substantially different effect on price than flying with Air India.

While the full summary of the regression results and statistics are included in Appendix A, a few of the statistically significant coefficients are highlighted below.

* **Intercept**: 52,728.60
* **Examples of statistically significant coefficients:**
  + **airlineGO\_FIRST:** 1,619.70, which means that GO\_FIRST is more expensive than the baseline airline.
  + **source\_cityDelhi:** -1409.70, which includes lower flight prices from Delhi compared to the baseline city.
  + **stopszero:** -7,600.26 (lower prices for non-stop flights compared to flights with one stop).
  + **Economy:** -44,943.27 - tickets for economy class are well over 40K rupees cheaper than business class.
  + **duration:** 43.95 (ticket prices increase by 43.95 rupees for each additional hour of flight).
  + **days\_left:** -132.18 which means 132.18 decrease in ticket price for every extra day before flying.

### Model Evaluation

To evaluate the model, we first considered its predictive proficiency, examining three metrics, the mean square error (MSE), R2, and adjusted-R2, to evaluate both goodness of fit and generalizability of the model using the training data. The MSE applied to the model’s predictions on test data forms a fourth metric. Each of these metrics for the baseline multiple linear regression model using all terms is reported in Table 1 below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Train MSE** | **Test MSE** | **R2 score (Test Data)** | **Adjusted R2 score** |
| **Multiple Linear Regression** | 46096849 | 43670434 | 0.9148625 | 0.9148193 |

Table 1: Predictive performance metrics of the baseline multiple linear regression model using all features.

Using the Adjusted R2 score to evaluate goodness of fit and potential predictive performance, the baseline multiple linear regression model appears to perform remarkably well.

However, its important to evaluate one’s predictive model not just using broad metrics of the average quality of the predictions but also using several standard evaluative plots for regression models that reveal a more detailed picture of exactly where and how well a model is working and where it may not be working, so as to aid us in understanding model weaknesses and provide ideas on how to improve the model. Figure 10 provides three plots that we used to evaluate and understand the baseline multiple linear regression model, and we continue to use these three evaluative plot types to do the same with the other models.

A graph with blue lines and a red line

Description automatically generated

Figure 10: Plots to Assess the Multiple Linear Rergression model. (left) Predicted value vs. Actual value; (center) Residuals plot (Residuals vs. Predicted Value); (right) Q-Q Plot of Residuals to evaluate normality of the residuals’ distribution

The first (left) plot in Figure 10 compares the predictions the model makes on test data to the actual measured values. The second (central) plot forms a standard residuals plot, with a red trend line generated using loess regression, show casing the residuals of the model over the range predictive values that the model can provide. Finally, the third (right) plot form a simple Q-Q plot that evaluates the goodness of fit of a Guassian distribution to the residuals.

The predictions vs. real values plot in Figure 10 reveals that despite the model achieving a fairly high adjusted R2 score, there are major gaps in the model’s performance, further confirmed by the residuals plot. The predictions vs. real values plot also shows the model divides predictions into two separated clusters, associated with the class of the ticket. In effect, the average effect of the ticket class dwarfs the average effect of any of the rest variables, creating a gap range of prices where the model is incapable of producing a prediction, while no such gap exists within the range of actual prices.

The residuals plot further shows that the residuals are not symmetrically distributed around zero throughout the range of price predictions as one would hope from a well-fit regression model. In particular, at the extreme ends of the range, the model produces residuals with a strong positive bias. This indicates that the model fits this region of the model poorly, providing unreliable predictions. The QQ plot confirms this perspective as the residuals deviate from a gaussian distribution at the extreme quantiles. Together, these plots indicate that trying to fit both the economy and business class tickets with a single model with just linear terms leads to a model with severe predictive gaps.

1. Forward Step-wise Linear Regression (see Appendix B for full table of the model coefficients and associated statistics)

### Model Construction and Rationale

The forward stepwise feature selection algorithm was used to see if using fewer features might alleviate the poor performance of the baseline model at the extreme ends of the range of the ticket price. Since there are multiple levels to the source, destination, airline, and number of stops variables, there are 30 potential variables and an intercept term in this model. The model with the lowest BIC score was chosen as the best model. After the algorithm selected the features, the final multiple linear regression model was constructed using just these features to return estimates and statistics of the regression coefficients for this reduced set of terms.

Results and Interpretation

The best model as measured by the BIC metric excluded three terms, the exact ones that failed to reach statistical significance in the baseline model: airlineAirAsia, source\_cityChennai, and source\_cityMumbai.

The regression coefficient estimates and interpretation of each of the variables did differ substantially from the baseline model (Appendix B).

Model Evaluation

The model’s predictive performance slightly lagged the baseline model using the test data, but it matched its performance on the training data, evaluated using the adjusted R2 metric (Table 2).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Train MSE** | **Test MSE** | **R2 score (Test Data)** | **Adjusted R2 score** |
| **Multiple Linear Regression** | 46096849 | 43670434 | 0.9148625 | 0.9148193 |
| **Subset Selection model** | 46097400 | 43672321 | 0.9148588 | 0.9148199 |

Table 2: Predictive performance metrics comparison of the subset selection model with the baseline multiple linear regression model using all features.

Excluding the variables that failed to reach statistical significance remarkably produced a higher F-statistic (90924.16 on 27 and 240934 degrees of freedom (DF); Appendix B) than the one the baseline model with the complete set of terms (81831.80 on 30 and 240931 DF; Appendix A)., where the F-statistic represents the comparison of each of these models separately to model with just an intercept term. Comparing the subset model to the full baseline model directly to each other, using the Anova test and evaluating using the F-statistic and associated p-value indicated that the inclusion of these excluded terms failed to improve the model to an extent that was statistically significant (Table 3).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Res.Df** | **RSS** | **DF** | **Sum of Sq** | **F-statistic** | **P-value (Pr(>F) )** |
| **Subset Selection model** | 240934 | 1.110772E+13 |  |  |  |  |
| **Multiple Linear Regression** | 240931 | 1.110759E+13 | 3 | 132812365 | 0.9603 | 0.4104 |

Table 3: Statistics of ANOVA test comparing the full baseline model to the reduced subset selection model.

Examining the array of model evaluation plots (Figure 11) reveals that the gaps in the baseline model remain in the subset selection model.

A graph with blue lines and a red line

Description automatically generated

Figure 11 Plots to Assess Subset Selection model. (left) Predicted value vs. Actual value; (center) Residuals plot (Residuals vs. Predicted Value); (right) Q-Q Plot of Residuals to evaluate normality of the residuals’ distribution

## Polynomial Regression (See Appendix C for full table of model coefficients and associated statistics)

### Model Construction and Rationale

Figure 9 revealed that the price of the tickets may have a non-linear dependence on both of the real-valued variables in the dataset: duration and days\_left. As such, we opted try a model that included polynomial versions of these features. We considered both quadratic and third order potential polynomial terms separately to capture the potential non-linear relationship between the price and these features.

### Results and Interpretation

The quadratic polynomial is an even and symmetric function, with the first order coefficient in coordination with the second order coefficient determining the duration value where its effect on the price has tapered off and is a maximum as well as the day at which the price is minimum. The coefficients of the second order terms for the quadratic model are the primary factors that determine how strongly quadratic and accelerative the relationship is between the price and the variable.

For example, examining Figure 9, modeling the effect of trip duration on prices using a second order polynomial would consider a negative quadratic, where the maximum of the quadratic is reached anywhere from 10-30 hours. The coefficient for the quadratic term most also form a compromise between modeling the accelerative and then tapering off dependence of the price on the duration of the trip and reality that the effect seems to plateau on average rather than steeply decline as quickly as it rose. Ideally, the declining portion of the quadratic will occur in a range of prices that do not generally occur.

Likewise with the other variable, looking again at Figure 9, modeling the effect of the number of days remaining until the day of the flight as a positive quadratic would likely find finding coefficients that have very slowly accelerating slope where the minimum of the quadratic is out of the range of the data.

The main advantage of the cubic polynomial over the quadratic polynomial is that it forms an odd function following the same overall directional trend as the simple linear mononomial, but still being able to model non-linearly accelerating relationships between the dependent and independent variables. Additionally, the specific combination of coefficients can either create a temporarily opposite trend within in the range of the dependent variable (potentially also allowing it the flexibility to approximate a quadratic within the typical range of the independent variable) or it can create a monotonic polynomial with just an inflection point or absolute no real critical points at all. As such, they can be powerful and flexible approximations of non-linear effects withing the typical range of the the independent variable. For a cubic polynomial of the form f(x) = ax3 + bx2 + cx + d, the value of b2 compared to the value of 3\*a\*c determines which of the two behaviors the model picks.

As we see in the model evaluation section, the model with third order terms of the two variables performs better than the model with quadratic terms. While Appendix C contains the full estimates, statistics, and p-values for both models, in this section the regression coefficients for the polynomial model are reported.

* **Intercept**: 49745.46
* **Coefficients of cubic polynomial duration terms**:
  + poly(duration, 3)3: 366659.73 (coefficient for the cubic term)
  + poly(duration, 3)2: -590613.54 (coefficient for the quadratic term)
  + poly(duration, 3)1: 504164.10 (coefficient for the linear term)
* **Coefficients of cubic polynomial days\_left terms**:
  + poly(days\_left, 3)3: -230451.69 (coefficient for the cubic term)
  + poly(days\_left, 3)2: 554713.22 (coefficient for the quadratic term)
  + poly(days\_left, 3)1: - 881938.56 (coefficient for the linear term)

For both the duration and days\_left variables, the square of the coefficient of the quadratic term, b2, is less than the three times the product of the coefficients for the cubic and linear terms, 3ac. As such both cubic polynomials that are fit to the effects of each variable on the ticket price are monotonic, with the duration cubic polynomial forming a positive one while the days\_left cubic polynomial is negative.

### Model Evaluation

Both adding only the terms for a quadratic polynomial for each of the two variables (the “Quadratic” model) and adding all the terms for a cubic polynomial for each of the two variables (the “Cubic” model) improved in predictive performance over the baseline model, as assessed by the MSE with test data as well as by the adjusted R2 score. The “Cubic” model had better predictive performance than the “Quadratic” model, even assessed by the Adjusted R2 metric, which tries to adjust for the trend the regular R2 metric to increase with the inclusion of more and more terms. Since the “Cubic” model has more terms than the “Quadratic” model, this particular metric to evaluate the performance is critical.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Train MSE** | **Test MSE** | **R2 score (Test Data)** | **Adjusted R2 score** |
| **Multiple Linear Regression** | 46096849 | 43670434 | 0.9148625 | 0.9148193 |
| **Model with Quadratic Terms** | 44576143 | 42450529 | 0.9172345 | 0.9171869 |
| **Model with Cubic Terms** | 44065918 | 41985450 | 0.9181413 | 0.9180915 |

Table 4: Predictive performance metrics comparison of two variations of the model with polynomial terms with the baseline multiple linear regression model using all features.

Additionally, we see that adding the cubic term starts to fill some of the gap in the price domain where the baseline model was incapable of making predictions (Figure 12). However, the residuals appear to perform poorly at the extreme ends of the price ranges.

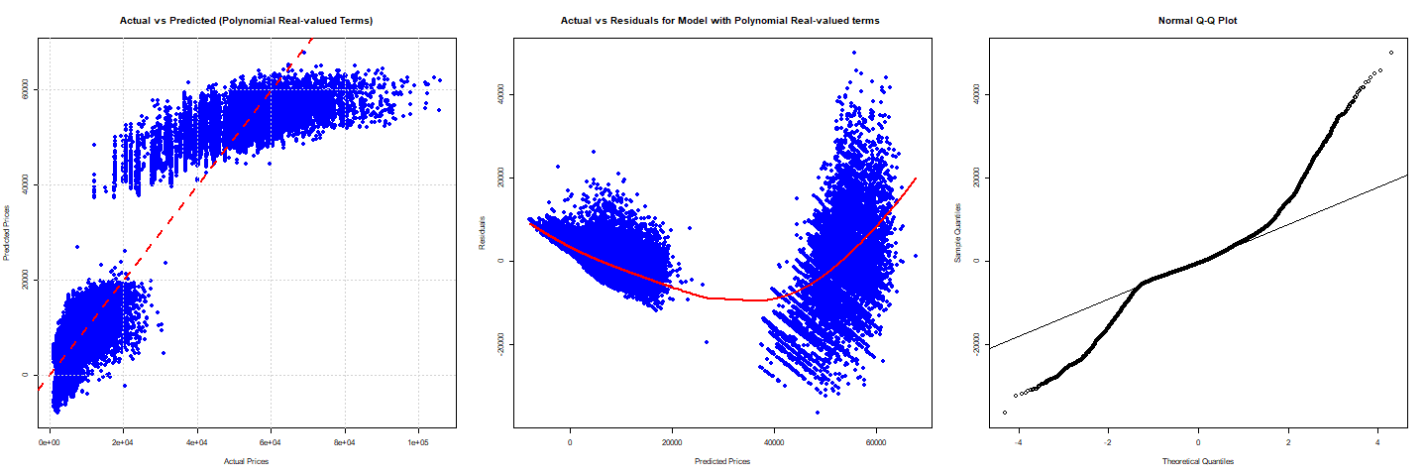


Figure 12: Plots to Assess the “Cubic" Regression model. (left) Predicted value vs. Actual value; (center) Residuals plot (Residuals vs. Predicted Value); (right) Q-Q Plot of Residuals to evaluate normality of the residuals distribution.

1. Interaction Model (See Appendix D for full table of model coefficients and associated statistics)

Model Construction and Rationale

The decision to include interactions between source\_city and destination\_city and between departure\_time and arrival\_time in the model was based on their potential to capture intuitive relationships in that influence the cost of flights:

1. **source\_city : destination\_city Interaction**:
   * **Relevance**: Flight prices are highly influenced by the route-specific demand and supply dynamics between the source and destination cities. For instance, flights between major business hubs may have higher prices compared to less popular routes, even if other factors like duration and class are the same.
   * **Insight**: Including this interaction term allows the model to account for unique price patterns associated with specific city pairs that cannot be captured by their individual effects alone.
2. **departure\_time : arrival\_time Interaction**:
   * **Relevance**: Departure and arrival times jointly influence flight prices. For example:
     + Overnight flights (late night to early morning) may be priced differently due to convenience and demand patterns.
     + Business travelers might prefer flights that depart in the morning and arrive by mid-day, which can lead to higher prices.
   * **Insight**: This interaction term captures the combined effect of departure and arrival times on pricing, enabling the model to differentiate between various time-specific travel scenarios.

Additionally, Figure 9 along with model evaluation plots that we have seen so far (Figures 10-12) indicate that each of these terms have the potential to behave differently in each of the ticket classes and that class effect is overwhelming the average effects of the other variables. By modeling the different class tickets together, the average effect of each the other terms, especially when the trends diverge in each class, is weakened in the model. However, by modeling each of these terms as interacting with ticket class, one can break up the effects of each of the variables into class specific effects. Specifically in the form chosen for this model the coefficients for the linear terms for the other variables capture business class specific effects while the class interaction coefficients essentially provide economy class specific effects relative to business class-specific effects.

The model was constructed such that cubic polynomial terms from the previous model were included for each of the real-valued variables and were allowed to interact with with the class variable. In addition to this, the two curated interactions listed above of origin and destination cities as well as the departure and arrival times are themselves allowed to interact with the ticket class, allowing for a ticket class specific interaction effect. The final model constituted a total of167 variable-dependent terms and an additional 1 intercept term.

### Results and Interpretation

The model returned singularities for 17 terms. Specifically, the each of airline and ticket class interaction terms returned singularities, suggesting that the effect of the ticket class on the price doesn’t seem to have an airline specific effect. Other singularities corresponded to impossibilities, specifically situations where the origin and destinations cities were the same or when the departure and arrival time periods (ex. Departing late at night and arriving late at night) were the same. Finally, some of the origin and destination city pairs as well as their interaction with the ticket class formed singularities.

Some interaction terms for origin and destination cities as well as departing and arriving time of day periods as well as their match to intuition and interpretation are reported below. As stated earlier the baseline interaction effects reported correspond to business class tickets. The economy class-specific term then provides an estimate of the effect relative to the business class-specific effect. So for example, business class flights that depart late at night and then arrive the next morning are on average ~1900 rupees more expensive than business class flights that depart or arrive in the afternoon. However, we find that there is an economy class specific effect that finds the price of the red eye tickets are on average roughly 850 rupees cheaper (1909.59-2760.57). Likewise, Economy class tickets for flights that depart in the morning and arrive in the evening, tickets that are in prime demand, are on average 161.18 (5305.78 - 5144.60) more expensive than the average ticket. Appendix D contains the full set of coefficients and associated statistics and p-values.

* **Intercept**: 49734.45
* Interaction terms:
  + **source\_cityDelhi:destination\_cityChennai**: 9859.07 (higher prices for flights from Delhi to Chennai).
  + **departure\_timeLate\_Night:arrival\_timeMorning:classEconomy**: -2760.57 (Red eye economy class flights are cheaper than other flights).
  + **departure\_timeLate\_Night:arrival\_timeMorning**: 1909.59 (Prices for red eye business class flights are not
  + **departure\_timeEarly\_Morning:arrival\_timeEvening:classEconomy**: 5305.78 (higher prices for daytime flights for economy class tickets)
  + **departure\_timeEarly\_Morning:arrival\_timeEvening**: -5144.60 (The business class ticket prices appear less susceptible to this dependency, so this term counter acts this effect.

The coefficients for the cubic polynomial terms for each of the two real-valued variables and the effect of the interaction with the ticket class are reported above. We saw in Figure 9 that each ticket class appears to have a substantially different relationship with both the duration variable and the days left variable. By allowing for interaction with the ticket class, one in effect is modeling two different cubic polynomials specific to each ticket class for both of the variables. For example, the cubic polynomial terms for the effect of the trip duration (x) on the price of business class tickets are 1,500,890\*x3 - 2,165,486\*x2 + 1,186,473\*x. We find a much smaller magnitude cubic polynomial for the effect of trip duration on the price of economy class tickets. Namely the cubic polynomial terms for economy class tickets become (1,500,890-1,387,513 )\*x3 – (2,165,486-2,088,307)\*x2 + (1,186,473-1,047,388 )\*x.

The difference in scale between the cubic polynomials between each ticket class matches the difference in the scale of the effect of trip duration on the price of the tickets between each of the ticket classes in Figure 9.

* **Coefficients of cubic polynomial duration terms and interaction with ticket class**:
  + poly(duration, 3)3: 1,500,890 (coefficient for the cubic term)
  + poly(duration, 3)2: -2,165,486 (coefficient for the quadratic term)
  + poly(duration, 3)1: 1,186,473 (coefficient for the linear term)
  + classEconomy:poly(duration, 3)3: -1,387,513 (coefficient for the cubic term)
  + classEconomy:poly(duration, 3)2: 2,088,307 (coefficient for the quadratic term)
  + classEconomy:poly(duration, 3)1: -1,047,388 (coefficient for the linear term)
* **Coefficients of cubic polynomial days\_left terms**:
  + poly(days\_left, 3)3: -464146.4 (coefficient for the cubic term)
  + poly(days\_left, 3)2: 490935.5 (coefficient for the quadratic term)
  + poly(days\_left, 3)1: -625191.8 (coefficient for the linear term)
  + classEconomy:poly(days\_left, 3)3: 332083.0 (coefficient for the cubic term)
  + classEconomy:poly(days\_left, 3)2: 96284.56 (coefficient for the quadratic term)
  + classEconomy:poly(days\_left, 3)1: -392751.1 (coefficient for the linear term)

### Model Evaluation

Allowing for these interaction effects alongside the cubic polynomial terms substantially improves the predictive performance of the interaction model relative to the “Cubic” model as assessed by all four metrics in Table 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Train MSE** | **Test MSE** | **R2 score (Test Data)** | **Adjusted R2 score** |
| **Linear Regression** | 46096849 | 43670434 | 0.9148625 | 0.9148193 |
| **Model with Cubic Terms** | 44065918 | 41985450 | 0.9181413 | 0.9180915 |
| **Interaction Model** | 25541437 | 23852657 | 0.9535133 | 0.9533936 |

Table 5: Predictive performance metrics comparisons of the baseline model, the previous best "Cubic" model, and the Interaction Model.

Examining the model evaluation plots reveals the gap in the range of price predictions in the baseline model is finally filled by allowing for ticket class-specific effects by including interactions of all the terms with ticket class. Additionally, the residuals plot shows that residuals are much more symmetrically distributed around zero for the full range of price predictions than previous models, such that at each price prediction value the residuals appear to average close to zero. However, the residuals deviate much more substantially from a normal distribution.

A blue and red line graph

Description automatically generated

Figure 13: Plots to Assess the Interaction model. (left) Predicted value vs. Actual value; (center)) Residuals plot (Residuals vs. Predicted Value); (right) Q-Q Plot of Residuals to evaluate normality of the residuals distribution.

1. Ridge Regression (See Appendix E for the list of model coefficients)

Model Construction and Rationale

In addition to the hand-crafted pairwise interactions in the Interaction model, one could also consider all pairwise interactions of all the variables, including dummy variables, in the model. However, this balloons the number of terms to 414, presenting major risks of overfitting the model. One way to mitigate this risk is to use a penalized regression model that adds a penalty proportional to the square of the coefficient, Ridge Regression, which results in the shrinkage of the coefficients at rate proportion to square of the coefficient. As such it can potentially mitigate overfitting to the potentially multicollinear set of terms considered in this model.

### Results and Interpretation

A series of lambda values were considered using 10-fold cross validation, and the lambda value that minimized average prediction error in the each of the folds when trained on the other folds was considered the optimal lambda value to be used for the model trained on the complete training data. The optimal lambda value was found to be 2128.605. the magnitude of the model coefficients appeared to be reduced compared to the Interaction model. Looking at the bottom quartile of the model coefficients based on their absolute value, the maximum value has a maximum value of roughly 200. Looking at the bottom quartile of the absolute value of the Interaction model coefficients has a maximum value of about 2100. Every coefficient corresponding to each quartile of the baseline model is larger than corresponding coefficient for the Ridge model. Table 7 below in the Lasso model section provides a table comparing the distribution of model coefficients for the Ridge and Lasso regression models as represented in quartile values.

The ticket class still constituted the largest coefficient other than the intercept term. However, coefficients for other terms such as the next largest coefficient for a linear term stopszero were much more comparable in magnitude to the coefficient of the linear economy class coefficient. For example in both the baseline and the interaction models, the coefficient for the economy class term was ~40,000. For the baseline model, the coefficient for stopzero, was ~ -7600 and in the interaction model it was -11860. However, in the Ridge Regression model, economy class coefficient was -15684.33 while the coefficient for the linear stopzero term was -5191.48. The two terms are much more proportionate in the ridge regression model than either the baseline model or even the curated Interaction model, suggesting that the Ridge Regression model may be able to generate predictions in that gap seen in the baseline model.

### Model Evaluation

Allowing for all these pairwise interaction effects penalized by the L2 norm substantially improves by all four metrics the predictive performance of the Ridge Regression model compared to the baseline model (Table 6), the Subset selection model (not shown), and the “Cubic” model (not shown). However the previous Interaction model provides better predictive performance than the Ridge Regression model according to all four metrics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Train MSE** | **Test MSE** | **R2 score (Test Data)** | **Adjusted R2 score** |
| **Linear Regression** | 46096849 | 43670434 | 0.9148625 | 0.9153264 |
| **Interaction Model** | 25541437 | 23852657 | 0.9535133 | 0.9533936 |
| **Ridge Regression Model** | 32515430 | 30579377 | 0.9410538 | 0.9406386 |

Table 6: Predictive performance metrics comparisons of baseline multiple linear regression model, the previous best Interaction Model, and the Redge Regression Model

When compared to the Interaction Model and the Baseline Model when examining the Ridge Regression’s model evaluation plots, one finds that Ridge Regression does fill the gap in the range of price predictions in the baseline model, as the coefficient values examined earlier suggests was possible, but the predictions within that gap appear to be rarer from the Ridge regression model than the Interaction Model. Furthermore, the model appears to have a bias in the residuals and underpredicts the the price of tickets at the lower end of the ticket price range.

A graph with blue lines and a red line

Description automatically generated

Figure 14: Plots to Assess the Ridge Regression model. (left) Predicted value vs. Actual value; (center)) Residuals plot (Residuals vs. Predicted Value); (right) Q-Q Plot of Residuals to evaluate normality of the residuals distribution.

1. Lasso Regression (See Appendix E for the list of model coefficients)

Model Construction and Rationale

Using the same formula of terms as with the Ridge Regression model, lasso regression was performed to both perform feature selection from this large number of terms and to penalize the coefficients so as to reduce overfitting. The L1-Norm penalty works with optimization algorithms such that depending on the lambda value many of the lower value terms in non-penalized regression often go to zero, essentially being removed from the model.

### Results and Interpretation

Like with Ridge Regression, the optimal lambda value was chosen using 10-fold cross validation. The optimal lambda value was 3.719798. At this lambda value, 77 of the 414 terms had a coefficient equal to zero.

Examining the model coefficients in comparison to the ridge regression coefficients reveals that Ridge regression keeps coefficients at lower quartiles relatively high compared to Lasso regression, but Lasso regression produces coefficients that are substantially larger than the corresponding coefficients in the Ridge Regression model on the higher end.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Min.** | **1st Quartile** | **Median** | **Mean** | **3rd Quartile** | **Max.** |
| **Ridge** | 0 | 198.6 | 609.4 | 1103.1 | 1311.2 | 40375.1 |
| **Lasso** | 0.18 | 108.41 | 422.01 | 1177.56 | 1050.36 | 51088.98 |

Table 7: Summary of Quartile values of Ridge and Lasso model coefficients, excluding the 77 coefficients equal to zero in the lasso model

### Model Evaluation

The Lasso regression model became the second-best performing model closely behind the hand crafted Interaction model as evaluated by each of the four prediction performance metrics (Table 8). Interestingly, in comparison to the hand-crafted Interaction model, the Lasso model considers all pairwise interaction terms, but it does not examine the terms corresponding to pairwise interaction terms interacting with the class term. Additionally, the cubic polynomial terms for the real-valued variables were not included either. Attempts to introduce these cubic polynomial and ticket class specific first order and second order interaction terms as seen in Interaction model resulted in too many terms, overfitting the model. The MSE on the training data continued to decrease while the test MSE began to increase, indicating that the features had become too multicollinear and that the model was becoming overfit with poorer performance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Train MSE** | **Test MSE** | **R2 score (Test Data)** | **Adjusted R2 score** |
| **Linear Regression** | 46096849 | 43670434 | 0.9148625 | 0.9153264 |
| **Interaction Model** | 25541437 | 23852657 | 0.9535133 | 0.9533936 |
| **Lasso Regression Model** | 26801357 | 24980374 | 0.9513168 | 0.9509739 |

The model evaluation plots reveal that the Lasso model performed better at making predictions in the gap region where the baseline model is poor at making predictions than the Ridge Regression model in the last section. While the model continues to have a slight bias at the extreme ends of the prediction ranges, it appears to perform much better than the Ridge Regression model at these regions. The residuals appear to be distributed on either side of the zero line throughout most of the range of the price predictions, but there does seem to be a slight skew in the residual distribution throughout the range, albeit substantially less than the Ridge Regression model and much more so on the earlier models other than the Interaction model.

A graph with blue lines and a red line

Description automatically generated

# Conclusion

The following project compared various regression models that include Linear, Interaction, Polynomial, Lasso and Ridge Regression on their respective use in flight ticket price prediction, basing its work on airline, source and destination cities, time of departure and arrival, among other features. Among the models:

* The handcurated Interaction model performed the best since it essentially the effects of nearly all the terms, including curated interaction terms into an ticket class specific effects by allowing each of them to interact with the ticket class.
* Lasso regression had the second best performance, followed by Ridge Regression. Both penalized regression models reduced the magnitude of the ticket-class effect relative to the other variables, with ridge regression having a more pronounced effect, while Lasso regression eliminated terms that were likely to reduce prediction performance rather than enhance it.
* The Polynomial, Best Subsets Linear, and baseline Multiple Linear Regression models were the least performant.

In other words, this project added interaction terms, polynomial relationships drive up predictive power, and regularization helps build more robust, generalizable model.

## Possible Improvements

1. **Feature Engineering**

* Explore other derived features such as route-specific pricing trends, weekday/weekend effects or holiday seasonality.
* Encode categorical variables using more sophisticated techniques such as target encoding.

1. **Tuning Hyperparameters**

Carry out extensive grid/randomized search to optimize the regularization parameters for both Lasso and Ridge regression models.

1. **Model Experimentation**

* Advanced machine learning algorithms include the exploration of Gradient Boosting, Random Forest or XGBoost for nonlinear patterns.
* Use ensemble techniques to leverage off the strengths of disparate models for even better results.

1. **Visualization and Reporting**

Provide interactive visualizations such as SHAP values when explaining feature importance and model decisions.

These improvements will further enhance the model's predictive power and generalization capability, ensuring its applicability to real-world scenarios.

# Appendix A (Multiple Linear Regression model)

Formula: price ~ airline + source\_city + departure\_time + stops + arrival\_time + destination\_city + class + duration + days\_left

Residuals:

Min 1Q Median 3Q Max

-36340 -3144 -401 3126 64140

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 52728.598 93.109 566.313 < 2e-16 \*\*\*

airlineAirAsia -95.230 70.596 -1.349 0.177

airlineGO\_FIRST 1619.698 61.163 26.482 < 2e-16 \*\*\*

airlineIndigo 2024.944 53.009 38.200 < 2e-16 \*\*\*

airlineSpiceJet 2205.148 86.434 25.513 < 2e-16 \*\*\*

airlineVistara 3979.324 34.903 114.011 < 2e-16 \*\*\*

source\_cityChennai -54.167 51.905 -1.044 0.297

source\_cityDelhi -1409.705 47.128 -29.912 < 2e-16 \*\*\*

source\_cityHyderabad -1679.568 51.503 -32.611 < 2e-16 \*\*\*

source\_cityKolkata 1591.622 49.897 31.898 < 2e-16 \*\*\*

source\_cityMumbai -210.118 46.934 -4.477 7.58e-06 \*\*\*

departure\_timeEarly\_Morning 829.367 46.450 17.855 < 2e-16 \*\*\*

departure\_timeEvening 719.705 47.163 15.260 < 2e-16 \*\*\*

departure\_timeLate\_Night 1751.813 214.618 8.162 3.30e-16 \*\*\*

departure\_timeMorning 837.984 45.378 18.467 < 2e-16 \*\*\*

departure\_timeNight 655.134 51.185 12.799 < 2e-16 \*\*\*

stopstwo\_or\_more 2170.981 69.220 31.364 < 2e-16 \*\*\*

stopszero -7600.255 51.533 -147.484 < 2e-16 \*\*\*

arrival\_timeEarly\_Morning -764.477 74.207 -10.302 < 2e-16 \*\*\*

arrival\_timeEvening 930.051 48.072 19.347 < 2e-16 \*\*\*

arrival\_timeLate\_Night 917.513 78.371 11.707 < 2e-16 \*\*\*

arrival\_timeMorning 457.650 50.551 9.053 < 2e-16 \*\*\*

arrival\_timeNight 1142.321 47.105 24.251 < 2e-16 \*\*\*

destination\_cityChennai -212.097 51.440 -4.123 3.74e-05 \*\*\*

destination\_cityDelhi -1545.766 48.296 -32.006 < 2e-16 \*\*\*

destination\_cityHyderabad -1734.597 51.042 -33.984 < 2e-16 \*\*\*

destination\_cityKolkata 1388.386 49.222 28.206 < 2e-16 \*\*\*

destination\_cityMumbai -18.681 47.534 -0.393 0.694

classEconomy -44943.270 33.792 -1330.005 < 2e-16 \*\*\*

duration 43.950 2.628 16.723 < 2e-16 \*\*\*

days\_left -132.177 1.022 -129.349 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6790 on 240931 degrees of freedom

Multiple R-squared: 0.9106, Adjusted R-squared: 0.9106

F-statistic: 8.183e+04 on 30 and 240931 DF, p-value: < 2.2e-16

# Appendix B (Best Subset Regression model)

Formula: y\_train ~ airlineGO\_FIRST + airlineIndigo + airlineSpiceJet +

airlineVistara + source\_cityDelhi + source\_cityHyderabad +

source\_cityKolkata + source\_cityMumbai + departure\_timeEarly\_Morning +

departure\_timeEvening + departure\_timeLate\_Night + departure\_timeMorning +

departure\_timeNight + stopstwo\_or\_more + stopszero + arrival\_timeEarly\_Morning +

arrival\_timeEvening + arrival\_timeLate\_Night + arrival\_timeMorning +

arrival\_timeNight + destination\_cityChennai + destination\_cityDelhi +

destination\_cityHyderabad + destination\_cityKolkata + classEconomy +

duration + days\_left

Residuals:

Min 1Q Median 3Q Max

-36293 -3149 -401 3131 64135

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 52672.505 82.128 641.349 < 2e-16 \*\*\*

airlineGO\_FIRST 1646.233 58.435 28.172 < 2e-16 \*\*\*

airlineIndigo 2048.070 49.630 41.267 < 2e-16 \*\*\*

airlineSpiceJet 2223.199 85.240 26.082 < 2e-16 \*\*\*

airlineVistara 3994.173 33.150 120.487 < 2e-16 \*\*\*

source\_cityDelhi -1386.312 41.456 -33.440 < 2e-16 \*\*\*

source\_cityHyderabad -1653.563 46.086 -35.880 < 2e-16 \*\*\*

source\_cityKolkata 1615.334 44.287 36.474 < 2e-16 \*\*\*

source\_cityMumbai -181.743 40.225 -4.518 6.24e-06 \*\*\*

departure\_timeEarly\_Morning 831.270 46.409 17.912 < 2e-16 \*\*\*

departure\_timeEvening 723.302 47.013 15.385 < 2e-16 \*\*\*

departure\_timeLate\_Night 1757.650 214.481 8.195 2.52e-16 \*\*\*

departure\_timeMorning 840.823 45.339 18.545 < 2e-16 \*\*\*

departure\_timeNight 652.889 51.164 12.761 < 2e-16 \*\*\*

stopstwo\_or\_more 2163.285 68.882 31.405 < 2e-16 \*\*\*

stopszero -7596.173 51.408 -147.764 < 2e-16 \*\*\*

arrival\_timeEarly\_Morning -768.567 74.132 -10.368 < 2e-16 \*\*\*

arrival\_timeEvening 931.155 48.019 19.391 < 2e-16 \*\*\*

arrival\_timeLate\_Night 903.204 77.435 11.664 < 2e-16 \*\*\*

arrival\_timeMorning 458.743 50.504 9.083 < 2e-16 \*\*\*

arrival\_timeNight 1142.943 47.089 24.272 < 2e-16 \*\*\*

destination\_cityChennai -196.579 44.930 -4.375 1.21e-05 \*\*\*

destination\_cityDelhi -1535.316 40.991 -37.455 < 2e-16 \*\*\*

destination\_cityHyderabad -1722.039 44.354 -38.825 < 2e-16 \*\*\*

destination\_cityKolkata 1398.393 42.520 32.888 < 2e-16 \*\*\*

classEconomy -44951.920 33.157 -1355.721 < 2e-16 \*\*\*

duration 44.617 2.559 17.438 < 2e-16 \*\*\*

days\_left -132.212 1.022 -129.418 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6790 on 240934 degrees of freedom

Multiple R-squared: 0.9106, Adjusted R-squared: 0.9106

F-statistic: 9.092e+04 on 27 and 240934 DF, p-value: < 2.2e-16

# Appendix C (Polynomial Models)

## “Quadratic” Model

Formula: price ~ (airline + source\_city + departure\_time + stops + arrival\_time +

destination\_city + class + duration + days\_left) - duration -

days\_left + poly(duration, 2) + poly(days\_left, 2)

Residuals:

Min 1Q Median 3Q Max

-37032 -2985 -394 3078 61569

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.979e+04 7.918e+01 628.788 < 2e-16 \*\*\*

airlineAirAsia 1.538e+01 6.944e+01 0.221 0.824750

airlineGO\_FIRST 1.837e+03 6.020e+01 30.520 < 2e-16 \*\*\*

airlineIndigo 2.324e+03 5.312e+01 43.752 < 2e-16 \*\*\*

airlineSpiceJet 2.296e+03 8.504e+01 26.998 < 2e-16 \*\*\*

airlineVistara 3.934e+03 3.434e+01 114.562 < 2e-16 \*\*\*

source\_cityChennai 3.068e+00 5.105e+01 0.060 0.952072

source\_cityDelhi -1.395e+03 4.635e+01 -30.101 < 2e-16 \*\*\*

source\_cityHyderabad -1.625e+03 5.065e+01 -32.073 < 2e-16 \*\*\*

source\_cityKolkata 1.573e+03 4.909e+01 32.048 < 2e-16 \*\*\*

source\_cityMumbai -1.876e+02 4.616e+01 -4.064 4.82e-05 \*\*\*

departure\_timeEarly\_Morning 5.951e+02 4.613e+01 12.900 < 2e-16 \*\*\*

departure\_timeEvening 5.281e+02 4.662e+01 11.328 < 2e-16 \*\*\*

departure\_timeLate\_Night 1.775e+03 2.111e+02 8.410 < 2e-16 \*\*\*

departure\_timeMorning 6.872e+02 4.480e+01 15.341 < 2e-16 \*\*\*

departure\_timeNight 3.290e+02 5.123e+01 6.422 1.35e-10 \*\*\*

stopstwo\_or\_more 2.179e+03 6.808e+01 32.011 < 2e-16 \*\*\*

stopszero -5.874e+03 7.059e+01 -83.201 < 2e-16 \*\*\*

arrival\_timeEarly\_Morning -1.013e+03 7.336e+01 -13.808 < 2e-16 \*\*\*

arrival\_timeEvening 9.597e+02 4.727e+01 20.301 < 2e-16 \*\*\*

arrival\_timeLate\_Night 9.156e+02 7.713e+01 11.872 < 2e-16 \*\*\*

arrival\_timeMorning 1.846e+02 5.030e+01 3.670 0.000243 \*\*\*

arrival\_timeNight 1.000e+03 4.646e+01 21.533 < 2e-16 \*\*\*

destination\_cityChennai -1.708e+02 5.059e+01 -3.377 0.000733 \*\*\*

destination\_cityDelhi -1.545e+03 4.749e+01 -32.536 < 2e-16 \*\*\*

destination\_cityHyderabad -1.716e+03 5.019e+01 -34.185 < 2e-16 \*\*\*

destination\_cityKolkata 1.387e+03 4.841e+01 28.658 < 2e-16 \*\*\*

destination\_cityMumbai 2.332e+01 4.675e+01 0.499 0.617941

classEconomy -4.488e+04 3.328e+01 -1348.727 < 2e-16 \*\*\*

poly(duration, 2)1 3.128e+05 1.047e+04 29.871 < 2e-16 \*\*\*

poly(duration, 2)2 -3.620e+05 9.674e+03 -37.415 < 2e-16 \*\*\*

poly(days\_left, 2)1 -8.857e+05 6.697e+03 -132.252 < 2e-16 \*\*\*

poly(days\_left, 2)2 5.554e+05 6.698e+03 82.930 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6677 on 240929 degrees of freedom

Multiple R-squared: 0.9136, Adjusted R-squared: 0.9136

F-statistic: 7.959e+04 on 32 and 240929 DF, p-value: < 2.2e-16

## “Cubic” Model

Formulat: price ~ (airline + source\_city + departure\_time + stops + arrival\_time +

destination\_city + class + duration + days\_left) - duration -

days\_left + poly(duration, 3) + poly(days\_left, 3)

Residuals:

Min 1Q Median 3Q Max

-36539 -3000 -375 3112 60672

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 49745.46 78.73 631.853 < 2e-16 \*\*\*

airlineAirAsia 15.57 69.05 0.226 0.821537

airlineGO\_FIRST 1808.82 59.88 30.210 < 2e-16 \*\*\*

airlineIndigo 2501.66 53.01 47.195 < 2e-16 \*\*\*

airlineSpiceJet 2479.82 84.67 29.287 < 2e-16 \*\*\*

airlineVistara 3942.13 34.15 115.452 < 2e-16 \*\*\*

source\_cityChennai 9.15 50.76 0.180 0.856942

source\_cityDelhi -1384.41 46.08 -30.042 < 2e-16 \*\*\*

source\_cityHyderabad -1532.94 50.40 -30.414 < 2e-16 \*\*\*

source\_cityKolkata 1550.57 48.82 31.763 < 2e-16 \*\*\*

source\_cityMumbai -120.42 45.93 -2.622 0.008740 \*\*

departure\_timeEarly\_Morning 393.80 46.19 8.525 < 2e-16 \*\*\*

departure\_timeEvening 602.32 46.39 12.983 < 2e-16 \*\*\*

departure\_timeLate\_Night 1677.65 209.92 7.992 1.33e-15 \*\*\*

departure\_timeMorning 507.69 44.77 11.341 < 2e-16 \*\*\*

departure\_timeNight 273.98 50.97 5.376 7.64e-08 \*\*\*

stopstwo\_or\_more 1895.10 68.06 27.846 < 2e-16 \*\*\*

stopszero -3532.48 91.60 -38.563 < 2e-16 \*\*\*

arrival\_timeEarly\_Morning -1258.33 73.27 -17.175 < 2e-16 \*\*\*

arrival\_timeEvening 807.54 47.15 17.127 < 2e-16 \*\*\*

arrival\_timeLate\_Night 642.73 77.03 8.344 < 2e-16 \*\*\*

arrival\_timeMorning 10.85 50.22 0.216 0.828968

arrival\_timeNight 712.33 46.76 15.233 < 2e-16 \*\*\*

destination\_cityChennai -168.52 50.30 -3.350 0.000807 \*\*\*

destination\_cityDelhi -1598.12 47.24 -33.832 < 2e-16 \*\*\*

destination\_cityHyderabad -1655.82 49.93 -33.162 < 2e-16 \*\*\*

destination\_cityKolkata 1376.13 48.13 28.592 < 2e-16 \*\*\*

destination\_cityMumbai 18.83 46.48 0.405 0.685455

classEconomy -44894.58 33.09 -1356.899 < 2e-16 \*\*\*

poly(duration, 3)1 504164.10 11528.28 43.733 < 2e-16 \*\*\*

poly(duration, 3)2 -590613.54 11172.06 -52.865 < 2e-16 \*\*\*

poly(duration, 3)3 366659.73 9197.76 39.864 < 2e-16 \*\*\*

poly(days\_left, 3)1 -881938.56 6659.22 -132.439 < 2e-16 \*\*\*

poly(days\_left, 3)2 554713.22 6659.51 83.296 < 2e-16 \*\*\*

poly(days\_left, 3)3 -230451.69 6645.98 -34.675 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6639 on 240927 degrees of freedom

Multiple R-squared: 0.9146, Adjusted R-squared: 0.9146

F-statistic: 7.586e+04 on 34 and 240927 DF, p-value: < 2.2e-16

# Appendix D (Interaction Model)

price ~ ((airline + source\_city + departure\_time + stops + arrival\_time +

destination\_city + class + duration + days\_left) - duration -

days\_left + source\_city:destination\_city + departure\_time:arrival\_time +

poly(duration, 3) + poly(days\_left, 3)) \* class

Residuals:

Min 1Q Median 3Q Max

-82415 -1748 -269 1470 54932

Coefficients: (17 not defined because of singularities)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Term** | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| (Intercept) | 49734.45 | 253.30 | 196.345 | 0.00E+00 |
| airlineAirAsia | -2650.28 | 55.88 | -47.430 | 0.00E+00 |
| airlineGO\_FIRST | -782.28 | 49.06 | -15.946 | 3.23E-57 |
| airlineIndigo | -1267.80 | 45.23 | -28.031 | 1.31E-172 |
| airlineSpiceJet | -830.29 | 66.90 | -12.410 | 2.36E-35 |
| airlineVistara | 7986.22 | 39.86 | 200.375 | 0.00E+00 |
| source\_cityChennai | -7569.48 | 202.21 | -37.434 | 8.38E-306 |
| source\_cityDelhi | -8378.41 | 191.37 | -43.782 | 0.00E+00 |
| source\_cityHyderabad | -10153.69 | 200.63 | -50.609 | 0.00E+00 |
| source\_cityKolkata | -1754.40 | 131.27 | -13.365 | 1.00E-40 |
| source\_cityMumbai | -436.63 | 189.57 | -2.303 | 2.13E-02 |
| departure\_timeEarly\_Morning | 3818.43 | 228.86 | 16.685 | 1.84E-62 |
| departure\_timeEvening | 1695.57 | 227.45 | 7.455 | 9.04E-14 |
| departure\_timeLate\_Night | 7857.87 | 1284.92 | 6.115 | 9.64E-10 |
| departure\_timeMorning | 4367.78 | 215.28 | 20.289 | 1.91E-91 |
| departure\_timeNight | 1831.95 | 230.77 | 7.938 | 2.06E-15 |
| stopstwo\_or\_more | 11781.68 | 177.15 | 66.506 | 0.00E+00 |
| stopszero | -11860.86 | 195.24 | -60.749 | 0.00E+00 |
| arrival\_timeEarly\_Morning | -2524.85 | 337.64 | -7.478 | 7.57E-14 |
| arrival\_timeEvening | 2649.05 | 213.09 | 12.432 | 1.81E-35 |
| arrival\_timeLate\_Night | 10.06 | 366.36 | 0.027 | 9.78E-01 |
| arrival\_timeMorning | 2453.85 | 222.65 | 11.021 | 3.06E-28 |
| arrival\_timeNight | 4674.19 | 219.96 | 21.250 | 4.03E-100 |
| destination\_cityChennai | -9180.88 | 205.50 | -44.675 | 0.00E+00 |
| destination\_cityDelhi | -9301.14 | 192.84 | -48.232 | 0.00E+00 |
| destination\_cityHyderabad | -10074.83 | 197.47 | -51.019 | 0.00E+00 |
| destination\_cityKolkata | -2187.77 | 196.38 | -11.141 | 8.08E-29 |
| destination\_cityMumbai | -178.14 | 143.89 | -1.238 | 2.16E-01 |
| classEconomy | -42311.60 | 300.45 | -140.829 | 0.00E+00 |
| poly(duration, 3)1 | 1186472.63 | 24344.14 | 48.737 | 0.00E+00 |
| poly(duration, 3)2 | -2165485.72 | 33950.45 | -63.784 | 0.00E+00 |
| poly(duration, 3)3 | 1500889.85 | 25776.39 | 58.227 | 0.00E+00 |
| poly(days\_left, 3)1 | -625191.78 | 9032.00 | -69.220 | 0.00E+00 |
| poly(days\_left, 3)2 | 490935.49 | 9120.02 | 53.831 | 0.00E+00 |
| poly(days\_left, 3)3 | -464146.40 | 9189.20 | -50.510 | 0.00E+00 |
| source\_cityDelhi:destination\_cityChennai | 11475.72 | 246.33 | 46.587 | 0.00E+00 |
| source\_cityHyderabad:destination\_cityChennai | 10609.84 | 261.58 | 40.561 | 0.00E+00 |
| source\_cityKolkata:destination\_cityChennai | 8369.93 | 214.69 | 38.986 | 0.00E+00 |
| source\_cityMumbai:destination\_cityChennai | 6038.58 | 243.44 | 24.805 | 1.18E-135 |
| source\_cityChennai:destination\_cityDelhi | 11187.36 | 247.99 | 45.113 | 0.00E+00 |
| source\_cityHyderabad:destination\_cityDelhi | 7895.19 | 249.02 | 31.705 | 3.76E-220 |
| source\_cityKolkata:destination\_cityDelhi | 10182.03 | 194.37 | 52.384 | 0.00E+00 |
| source\_cityMumbai:destination\_cityDelhi | -525.28 | 226.36 | -2.321 | 2.03E-02 |
| source\_cityChennai:destination\_cityHyderabad | 9273.61 | 258.40 | 35.888 | 2.61E-281 |
| source\_cityDelhi:destination\_cityHyderabad | 8317.29 | 243.74 | 34.124 | 1.32E-254 |
| source\_cityKolkata:destination\_cityHyderabad | 7172.81 | 203.85 | 35.187 | 1.57E-270 |
| source\_cityMumbai:destination\_cityHyderabad | 2937.80 | 236.15 | 12.440 | 1.62E-35 |
| source\_cityChennai:destination\_cityKolkata | 7996.58 | 257.07 | 31.107 | 5.15E-212 |
| source\_cityDelhi:destination\_cityKolkata | 9859.07 | 237.56 | 41.501 | 0.00E+00 |
| source\_cityHyderabad:destination\_cityKolkata | 6914.61 | 252.56 | 27.379 | 8.86E-165 |
| source\_cityMumbai:destination\_cityKolkata | 404.16 | 233.32 | 1.732 | 8.32E-02 |
| source\_cityChennai:destination\_cityMumbai | 6079.34 | 209.20 | 29.060 | 2.43E-185 |
| source\_cityDelhi:destination\_cityMumbai | -933.32 | 187.42 | -4.980 | 6.37E-07 |
| source\_cityHyderabad:destination\_cityMumbai | 3871.69 | 207.04 | 18.700 | 5.59E-78 |
| departure\_timeEarly\_Morning:arrival\_timeEarly\_Morning | -563.60 | 454.86 | -1.239 | 2.15E-01 |
| departure\_timeEvening:arrival\_timeEarly\_Morning | 151.54 | 389.43 | 0.389 | 6.97E-01 |
| departure\_timeLate\_Night:arrival\_timeEarly\_Morning | -4024.01 | 1894.95 | -2.124 | 3.37E-02 |
| departure\_timeMorning:arrival\_timeEarly\_Morning | -3543.75 | 411.00 | -8.622 | 6.60E-18 |
| departure\_timeNight:arrival\_timeEarly\_Morning | -2345.66 | 420.19 | -5.582 | 2.37E-08 |
| departure\_timeEarly\_Morning:arrival\_timeEvening | -5144.60 | 267.38 | -19.241 | 1.94E-82 |
| departure\_timeEvening:arrival\_timeEvening | 2852.08 | 266.99 | 10.682 | 1.25E-26 |
| departure\_timeLate\_Night:arrival\_timeEvening | -6287.04 | 1649.40 | -3.812 | 1.38E-04 |
| departure\_timeMorning:arrival\_timeEvening | -3206.23 | 258.54 | -12.401 | 2.63E-35 |
| departure\_timeNight:arrival\_timeEvening | 743.33 | 267.40 | 2.780 | 5.44E-03 |
| departure\_timeEarly\_Morning:arrival\_timeLate\_Night | 2178.35 | 492.15 | 4.426 | 9.60E-06 |
| departure\_timeEvening:arrival\_timeLate\_Night | 4415.55 | 490.94 | 8.994 | 2.40E-19 |
| departure\_timeLate\_Night:arrival\_timeLate\_Night | 1727.70 | 911.61 | 1.895 | 5.81E-02 |
| departure\_timeMorning:arrival\_timeLate\_Night | -2888.51 | 452.29 | -6.386 | 1.70E-10 |
| departure\_timeNight:arrival\_timeLate\_Night | 7241.97 | 573.34 | 12.631 | 1.46E-36 |
| departure\_timeEarly\_Morning:arrival\_timeMorning | -1758.98 | 297.62 | -5.910 | 3.42E-09 |
| departure\_timeEvening:arrival\_timeMorning | -1136.41 | 259.18 | -4.385 | 1.16E-05 |
| departure\_timeLate\_Night:arrival\_timeMorning | 1909.59 | 1834.12 | 1.041 | 2.98E-01 |
| departure\_timeMorning:arrival\_timeMorning | -4542.40 | 271.69 | -16.719 | 1.04E-62 |
| departure\_timeNight:arrival\_timeMorning | -3369.35 | 259.29 | -12.995 | 1.35E-38 |
| departure\_timeEarly\_Morning:arrival\_timeNight | -6241.57 | 274.04 | -22.776 | 1.04E-114 |
| departure\_timeEvening:arrival\_timeNight | -80.80 | 285.13 | -0.283 | 7.77E-01 |
| departure\_timeLate\_Night:arrival\_timeNight | -8740.20 | 1510.27 | -5.787 | 7.17E-09 |
| departure\_timeMorning:arrival\_timeNight | -5608.89 | 248.24 | -22.594 | 6.44E-113 |
| departure\_timeNight:arrival\_timeNight | -1431.48 | 310.68 | -4.608 | 4.08E-06 |
| airlineVistara:classEconomy | -7462.31 | 52.88 | -141.128 | 0.00E+00 |
| source\_cityChennai:classEconomy | 7536.31 | 247.64 | 30.433 | 4.92E-203 |
| source\_cityDelhi:classEconomy | 8143.59 | 228.33 | 35.666 | 7.18E-278 |
| source\_cityHyderabad:classEconomy | 9749.06 | 242.33 | 40.230 | 0.00E+00 |
| source\_cityKolkata:classEconomy | 2551.64 | 157.95 | 16.155 | 1.13E-58 |
| source\_cityMumbai:classEconomy | 208.46 | 227.79 | 0.915 | 3.60E-01 |
| departure\_timeEarly\_Morning:classEconomy | -3933.88 | 264.72 | -14.861 | 6.26E-50 |
| departure\_timeEvening:classEconomy | -2110.17 | 274.67 | -7.682 | 1.57E-14 |
| departure\_timeLate\_Night:classEconomy | -7442.91 | 1491.16 | -4.991 | 6.00E-07 |
| departure\_timeMorning:classEconomy | -4252.75 | 250.27 | -16.992 | 1.02E-64 |
| departure\_timeNight:classEconomy | -2070.13 | 279.88 | -7.397 | 1.40E-13 |
| stopstwo\_or\_more:classEconomy | -9947.37 | 185.45 | -53.638 | 0.00E+00 |
| stopszero:classEconomy | 10491.69 | 217.45 | 48.248 | 0.00E+00 |
| arrival\_timeEarly\_Morning:classEconomy | 1856.42 | 410.21 | 4.526 | 6.03E-06 |
| arrival\_timeEvening:classEconomy | -2598.25 | 247.67 | -10.491 | 9.64E-26 |
| arrival\_timeLate\_Night:classEconomy | -584.79 | 402.68 | -1.452 | 1.46E-01 |
| arrival\_timeMorning:classEconomy | -2858.29 | 268.53 | -10.644 | 1.88E-26 |
| arrival\_timeNight:classEconomy | -4740.89 | 255.10 | -18.584 | 4.87E-77 |
| destination\_cityChennai:classEconomy | 9039.95 | 249.80 | 36.189 | 5.43E-286 |
| destination\_cityDelhi:classEconomy | 9106.78 | 229.70 | 39.646 | 0.00E+00 |
| destination\_cityHyderabad:classEconomy | 9523.94 | 238.32 | 39.962 | 0.00E+00 |
| destination\_cityKolkata:classEconomy | 2835.00 | 235.58 | 12.034 | 2.40E-33 |
| destination\_cityMumbai:classEconomy | -21.53 | 171.33 | -0.126 | 9.00E-01 |
| classEconomy:poly(duration, 3)1 | -1047387.68 | 27609.88 | -37.935 | 5.76e-314 |
| classEconomy:poly(duration, 3)2 | 2088307.29 | 38020.65 | 54.926 | 0.00E+00 |
| classEconomy:poly(duration, 3)3 | -1387513.08 | 28163.98 | -49.266 | 0.00E+00 |
| classEconomy:poly(days\_left, 3)1 | -392751.07 | 10919.15 | -35.969 | 1.45E-282 |
| classEconomy:poly(days\_left, 3)2 | 96284.56 | 10975.68 | 8.773 | 1.76E-18 |
| classEconomy:poly(days\_left, 3)3 | 332083.03 | 11011.63 | 30.157 | 2.02E-199 |
| source\_cityDelhi:destination\_cityChennai:classEconomy | -11514.46 | 296.86 | -38.788 | 0.00E+00 |
| source\_cityHyderabad:destination\_cityChennai:classEconomy | -10680.27 | 320.11 | -33.365 | 1.62E-243 |
| source\_cityKolkata:destination\_cityChennai:classEconomy | -7885.70 | 261.99 | -30.100 | 1.15E-198 |
| source\_cityMumbai:destination\_cityChennai:classEconomy | -6207.83 | 296.35 | -20.948 | 2.41E-97 |
| source\_cityChennai:destination\_cityDelhi:classEconomy | -11254.52 | 299.37 | -37.594 | 0.00E+00 |
| source\_cityHyderabad:destination\_cityDelhi:classEconomy | -7611.48 | 298.25 | -25.520 | 1.83E-143 |
| source\_cityKolkata:destination\_cityDelhi:classEconomy | -9987.80 | 229.79 | -43.465 | 0.00E+00 |
| source\_cityMumbai:destination\_cityDelhi:classEconomy | 579.43 | 271.48 | 2.134 | 3.28E-02 |
| source\_cityChennai:destination\_cityHyderabad:classEconomy | -9714.25 | 317.37 | -30.609 | 2.33E-205 |
| source\_cityDelhi:destination\_cityHyderabad:classEconomy | -7977.67 | 291.99 | -27.322 | 4.18E-164 |
| source\_cityKolkata:destination\_cityHyderabad:classEconomy | -6880.48 | 245.86 | -27.986 | 4.60E-172 |
| source\_cityMumbai:destination\_cityHyderabad:classEconomy | -3504.75 | 286.01 | -12.254 | 1.64E-34 |
| source\_cityChennai:destination\_cityKolkata:classEconomy | -7851.09 | 312.54 | -25.120 | 4.55E-139 |
| source\_cityDelhi:destination\_cityKolkata:classEconomy | -9871.66 | 283.31 | -34.844 | 2.42E-265 |
| source\_cityHyderabad:destination\_cityKolkata:classEconomy | -7015.27 | 304.16 | -23.064 | 1.42E-117 |
| source\_cityMumbai:destination\_cityKolkata:classEconomy | -426.86 | 280.35 | -1.523 | 1.28E-01 |
| source\_cityChennai:destination\_cityMumbai:classEconomy | -6002.55 | 256.10 | -23.439 | 2.36E-121 |
| source\_cityDelhi:destination\_cityMumbai:classEconomy | 1157.26 | 223.94 | 5.168 | 2.37E-07 |
| source\_cityHyderabad:destination\_cityMumbai:classEconomy | -4098.59 | 249.48 | -16.429 | 1.29E-60 |
| departure\_timeEarly\_Morning:arrival\_timeEarly\_Morning:classEconomy | 921.86 | 533.44 | 1.728 | 8.40E-02 |
| departure\_timeEvening:arrival\_timeEarly\_Morning:classEconomy | -29.25 | 473.49 | -0.062 | 9.51E-01 |
| departure\_timeLate\_Night:arrival\_timeEarly\_Morning:classEconomy | 4122.27 | 2067.46 | 1.994 | 4.62E-02 |
| departure\_timeMorning:arrival\_timeEarly\_Morning:classEconomy | 3246.81 | 509.05 | 6.378 | 1.80E-10 |
| departure\_timeNight:arrival\_timeEarly\_Morning:classEconomy | 2098.15 | 501.42 | 4.184 | 2.86E-05 |
| departure\_timeEarly\_Morning:arrival\_timeEvening:classEconomy | 5305.78 | 307.96 | 17.229 | 1.77E-66 |
| departure\_timeEvening:arrival\_timeEvening:classEconomy | -2544.94 | 321.39 | -7.918 | 2.41E-15 |
| departure\_timeLate\_Night:arrival\_timeEvening:classEconomy | 6519.85 | 1948.84 | 3.346 | 8.21E-04 |
| departure\_timeMorning:arrival\_timeEvening:classEconomy | 3320.17 | 297.72 | 11.152 | 7.11E-29 |
| departure\_timeNight:arrival\_timeEvening:classEconomy | -545.32 | 326.46 | -1.670 | 9.48E-02 |
| departure\_timeEarly\_Morning:arrival\_timeLate\_Night:classEconomy | -999.96 | 592.56 | -1.688 | 9.15E-02 |
| departure\_timeEvening:arrival\_timeLate\_Night:classEconomy | -3654.81 | 534.66 | -6.836 | 8.18E-12 |
| departure\_timeMorning:arrival\_timeLate\_Night:classEconomy | 2806.62 | 519.22 | 5.405 | 6.47E-08 |
| departure\_timeNight:arrival\_timeLate\_Night:classEconomy | -6211.42 | 622.59 | -9.977 | 1.95E-23 |
| departure\_timeEarly\_Morning:arrival\_timeMorning:classEconomy | 1885.49 | 349.15 | 5.400 | 6.66E-08 |
| departure\_timeEvening:arrival\_timeMorning:classEconomy | 1437.03 | 319.93 | 4.492 | 7.07E-06 |
| departure\_timeLate\_Night:arrival\_timeMorning:classEconomy | -2760.56 | 2007.73 | -1.375 | 1.69E-01 |
| departure\_timeMorning:arrival\_timeMorning:classEconomy | 4446.03 | 325.03 | 13.679 | 1.41E-42 |
| departure\_timeNight:arrival\_timeMorning:classEconomy | 3363.94 | 321.44 | 10.465 | 1.26E-25 |
| departure\_timeEarly\_Morning:arrival\_timeNight:classEconomy | 6083.99 | 315.79 | 19.266 | 1.20E-82 |
| departure\_timeEvening:arrival\_timeNight:classEconomy | 220.32 | 336.90 | 0.654 | 5.13E-01 |
| departure\_timeLate\_Night:arrival\_timeNight:classEconomy | 8999.74 | 1870.27 | 4.812 | 1.50E-06 |
| departure\_timeMorning:arrival\_timeNight:classEconomy | 5682.52 | 288.91 | 19.669 | 4.64E-86 |
| departure\_timeNight:arrival\_timeNight:classEconomy | 1280.01 | 369.12 | 3.468 | 5.25E-04 |

# Appendix E (Ridge Regression Model)

|  |  |  |  |
| --- | --- | --- | --- |
| **Term** | **Estimate** | **Term** | **Estimate** |
| (Intercept) | 40375.14 | source\_cityHyderabad:arrival\_timeMorning | -538.51 |
| airlineAirAsia | -2569.42 | source\_cityKolkata:arrival\_timeMorning | 877.88 |
| airlineGO\_FIRST | -2560.62 | source\_cityMumbai:arrival\_timeMorning | 926.52 |
| airlineIndigo | -2437.73 | source\_cityChennai:arrival\_timeNight | 443.68 |
| airlineSpiceJet | -1823.08 | source\_cityDelhi:arrival\_timeNight | 253.12 |
| airlineVistara | 3487.51 | source\_cityHyderabad:arrival\_timeNight | 392.44 |
| source\_cityChennai | 615.57 | source\_cityKolkata:arrival\_timeNight | 1141.51 |
| source\_cityDelhi | -1028.61 | source\_cityMumbai:arrival\_timeNight | 1272.84 |
| source\_cityHyderabad | -522.17 | source\_cityChennai:destination\_cityChennai | 0.00 |
| source\_cityKolkata | 1292.06 | source\_cityDelhi:destination\_cityChennai | 1455.60 |
| source\_cityMumbai | 568.70 | source\_cityHyderabad:destination\_cityChennai | 1076.10 |
| departure\_timeEarly\_Morning | 555.85 | source\_cityKolkata:destination\_cityChennai | 1271.92 |
| departure\_timeEvening | 569.01 | source\_cityMumbai:destination\_cityChennai | 485.57 |
| departure\_timeLate\_Night | -342.87 | source\_cityChennai:destination\_cityDelhi | 1843.35 |
| departure\_timeMorning | 357.93 | source\_cityDelhi:destination\_cityDelhi | 0.00 |
| departure\_timeNight | 944.49 | source\_cityHyderabad:destination\_cityDelhi | 659.29 |
| stopstwo\_or\_more | 1259.78 | source\_cityKolkata:destination\_cityDelhi | 1738.95 |
| stopszero | -5191.45 | source\_cityMumbai:destination\_cityDelhi | -1532.23 |
| arrival\_timeEarly\_Morning | -1028.97 | source\_cityChennai:destination\_cityHyderabad | 460.40 |
| arrival\_timeEvening | 375.53 | source\_cityDelhi:destination\_cityHyderabad | 938.81 |
| arrival\_timeLate\_Night | 397.42 | source\_cityHyderabad:destination\_cityHyderabad | 0.00 |
| arrival\_timeMorning | 508.77 | source\_cityKolkata:destination\_cityHyderabad | 672.64 |
| arrival\_timeNight | 1367.57 | source\_cityMumbai:destination\_cityHyderabad | -902.16 |
| destination\_cityChennai | 509.51 | source\_cityChennai:destination\_cityKolkata | 903.59 |
| destination\_cityDelhi | -823.61 | source\_cityDelhi:destination\_cityKolkata | 1402.58 |
| destination\_cityHyderabad | -647.87 | source\_cityHyderabad:destination\_cityKolkata | 454.85 |
| destination\_cityKolkata | 1078.82 | source\_cityKolkata:destination\_cityKolkata | 0.00 |
| destination\_cityMumbai | 609.64 | source\_cityMumbai:destination\_cityKolkata | -194.78 |
| classEconomy | -15684.33 | source\_cityChennai:destination\_cityMumbai | 499.20 |
| duration | 159.42 | source\_cityDelhi:destination\_cityMumbai | -1285.63 |
| days\_left | -17.99 | source\_cityHyderabad:destination\_cityMumbai | -487.39 |
| airlineAirAsia:source\_cityChennai | -1254.20 | source\_cityKolkata:destination\_cityMumbai | -469.78 |
| airlineGO\_FIRST:source\_cityChennai | -519.48 | source\_cityMumbai:destination\_cityMumbai | 0.00 |
| airlineIndigo:source\_cityChennai | -393.21 | source\_cityChennai:classEconomy | -3868.49 |
| airlineSpiceJet:source\_cityChennai | -640.66 | source\_cityDelhi:classEconomy | -1986.21 |
| airlineVistara:source\_cityChennai | 410.07 | source\_cityHyderabad:classEconomy | -1732.63 |
| airlineAirAsia:source\_cityDelhi | 282.36 | source\_cityKolkata:classEconomy | -4947.21 |
| airlineGO\_FIRST:source\_cityDelhi | 419.21 | source\_cityMumbai:classEconomy | -4598.66 |
| airlineIndigo:source\_cityDelhi | 783.95 | source\_cityChennai:duration | 46.65 |
| airlineSpiceJet:source\_cityDelhi | 293.66 | source\_cityDelhi:duration | -12.45 |
| airlineVistara:source\_cityDelhi | 1955.98 | source\_cityHyderabad:duration | -10.19 |
| airlineAirAsia:source\_cityHyderabad | -1214.89 | source\_cityKolkata:duration | 55.23 |
| airlineGO\_FIRST:source\_cityHyderabad | -58.27 | source\_cityMumbai:duration | 42.51 |
| airlineIndigo:source\_cityHyderabad | 214.25 | source\_cityChennai:days\_left | -16.43 |
| airlineSpiceJet:source\_cityHyderabad | 86.39 | source\_cityDelhi:days\_left | -9.98 |
| airlineVistara:source\_cityHyderabad | 242.12 | source\_cityHyderabad:days\_left | 5.72 |
| airlineAirAsia:source\_cityKolkata | -164.40 | source\_cityKolkata:days\_left | 10.02 |
| airlineGO\_FIRST:source\_cityKolkata | 617.58 | source\_cityMumbai:days\_left | -0.40 |
| airlineIndigo:source\_cityKolkata | 896.53 | departure\_timeEarly\_Morning:stopstwo\_or\_more | 768.98 |
| airlineSpiceJet:source\_cityKolkata | -219.21 | departure\_timeEvening:stopstwo\_or\_more | 20.66 |
| airlineVistara:source\_cityKolkata | 2077.74 | departure\_timeLate\_Night:stopstwo\_or\_more | -1166.24 |
| airlineAirAsia:source\_cityMumbai | 997.29 | departure\_timeMorning:stopstwo\_or\_more | 542.09 |
| airlineGO\_FIRST:source\_cityMumbai | 1278.79 | departure\_timeNight:stopstwo\_or\_more | 565.68 |
| airlineIndigo:source\_cityMumbai | 1834.93 | departure\_timeEarly\_Morning:stopszero | -1241.76 |
| airlineSpiceJet:source\_cityMumbai | 1398.47 | departure\_timeEvening:stopszero | -1081.16 |
| airlineVistara:source\_cityMumbai | 2862.66 | departure\_timeLate\_Night:stopszero | -1690.00 |
| airlineAirAsia:departure\_timeEarly\_Morning | 181.49 | departure\_timeMorning:stopszero | -2270.42 |
| airlineGO\_FIRST:departure\_timeEarly\_Morning | 493.54 | departure\_timeNight:stopszero | -1239.63 |
| airlineIndigo:departure\_timeEarly\_Morning | 291.36 | departure\_timeEarly\_Morning:arrival\_timeEarly\_Morning | -809.85 |
| airlineSpiceJet:departure\_timeEarly\_Morning | -390.32 | departure\_timeEvening:arrival\_timeEarly\_Morning | 62.30 |
| airlineVistara:departure\_timeEarly\_Morning | 1270.19 | departure\_timeLate\_Night:arrival\_timeEarly\_Morning | -1898.61 |
| airlineAirAsia:departure\_timeEvening | 650.43 | departure\_timeMorning:arrival\_timeEarly\_Morning | -791.24 |
| airlineGO\_FIRST:departure\_timeEvening | 1008.79 | departure\_timeNight:arrival\_timeEarly\_Morning | -1050.80 |
| airlineIndigo:departure\_timeEvening | 1273.05 | departure\_timeEarly\_Morning:arrival\_timeEvening | 1957.92 |
| airlineSpiceJet:departure\_timeEvening | -333.87 | departure\_timeEvening:arrival\_timeEvening | 49.27 |
| airlineVistara:departure\_timeEvening | 1782.04 | departure\_timeLate\_Night:arrival\_timeEvening | 2460.79 |
| airlineAirAsia:departure\_timeLate\_Night | -475.76 | departure\_timeMorning:arrival\_timeEvening | 2453.51 |
| airlineGO\_FIRST:departure\_timeLate\_Night | 510.91 | departure\_timeNight:arrival\_timeEvening | -167.81 |
| airlineIndigo:departure\_timeLate\_Night | 147.78 | departure\_timeEarly\_Morning:arrival\_timeLate\_Night | 1646.09 |
| airlineSpiceJet:departure\_timeLate\_Night | 0.00 | departure\_timeEvening:arrival\_timeLate\_Night | 214.63 |
| airlineVistara:departure\_timeLate\_Night | 0.00 | departure\_timeLate\_Night:arrival\_timeLate\_Night | -270.48 |
| airlineAirAsia:departure\_timeMorning | 97.14 | departure\_timeMorning:arrival\_timeLate\_Night | 1662.05 |
| airlineGO\_FIRST:departure\_timeMorning | 1157.71 | departure\_timeNight:arrival\_timeLate\_Night | -366.74 |
| airlineIndigo:departure\_timeMorning | 414.78 | departure\_timeEarly\_Morning:arrival\_timeMorning | -889.62 |
| airlineSpiceJet:departure\_timeMorning | -116.29 | departure\_timeEvening:arrival\_timeMorning | 1165.59 |
| airlineVistara:departure\_timeMorning | 1382.34 | departure\_timeLate\_Night:arrival\_timeMorning | 367.52 |
| airlineAirAsia:departure\_timeNight | -745.89 | departure\_timeMorning:arrival\_timeMorning | -565.47 |
| airlineGO\_FIRST:departure\_timeNight | -324.00 | departure\_timeNight:arrival\_timeMorning | 495.51 |
| airlineIndigo:departure\_timeNight | 175.42 | departure\_timeEarly\_Morning:arrival\_timeNight | 514.06 |
| airlineSpiceJet:departure\_timeNight | -1588.60 | departure\_timeEvening:arrival\_timeNight | -928.33 |
| airlineVistara:departure\_timeNight | 165.85 | departure\_timeLate\_Night:arrival\_timeNight | 748.45 |
| airlineAirAsia:stopstwo\_or\_more | -782.97 | departure\_timeMorning:arrival\_timeNight | 1698.28 |
| airlineGO\_FIRST:stopstwo\_or\_more | -1049.99 | departure\_timeNight:arrival\_timeNight | -1768.04 |
| airlineIndigo:stopstwo\_or\_more | 1666.97 | departure\_timeEarly\_Morning:destination\_cityChennai | -24.68 |
| airlineSpiceJet:stopstwo\_or\_more | 0.00 | departure\_timeEvening:destination\_cityChennai | -311.66 |
| airlineVistara:stopstwo\_or\_more | 1549.78 | departure\_timeLate\_Night:destination\_cityChennai | -888.97 |
| airlineAirAsia:stopszero | 1167.69 | departure\_timeMorning:destination\_cityChennai | 151.39 |
| airlineGO\_FIRST:stopszero | 635.26 | departure\_timeNight:destination\_cityChennai | -30.31 |
| airlineIndigo:stopszero | 1599.73 | departure\_timeEarly\_Morning:destination\_cityDelhi | -370.27 |
| airlineSpiceJet:stopszero | 2079.42 | departure\_timeEvening:destination\_cityDelhi | -240.77 |
| airlineVistara:stopszero | -4556.69 | departure\_timeLate\_Night:destination\_cityDelhi | 2054.60 |
| airlineAirAsia:arrival\_timeEarly\_Morning | -339.03 | departure\_timeMorning:destination\_cityDelhi | -47.83 |
| airlineGO\_FIRST:arrival\_timeEarly\_Morning | -292.06 | departure\_timeNight:destination\_cityDelhi | 800.62 |
| airlineIndigo:arrival\_timeEarly\_Morning | 123.20 | departure\_timeEarly\_Morning:destination\_cityHyderabad | -377.49 |
| airlineSpiceJet:arrival\_timeEarly\_Morning | -2039.50 | departure\_timeEvening:destination\_cityHyderabad | 27.56 |
| airlineVistara:arrival\_timeEarly\_Morning | 1167.74 | departure\_timeLate\_Night:destination\_cityHyderabad | -284.36 |
| airlineAirAsia:arrival\_timeEvening | 215.31 | departure\_timeMorning:destination\_cityHyderabad | 480.34 |
| airlineGO\_FIRST:arrival\_timeEvening | 497.25 | departure\_timeNight:destination\_cityHyderabad | 55.90 |
| airlineIndigo:arrival\_timeEvening | 566.19 | departure\_timeEarly\_Morning:destination\_cityKolkata | 111.57 |
| airlineSpiceJet:arrival\_timeEvening | -50.26 | departure\_timeEvening:destination\_cityKolkata | 533.76 |
| airlineVistara:arrival\_timeEvening | 1423.78 | departure\_timeLate\_Night:destination\_cityKolkata | 609.38 |
| airlineAirAsia:arrival\_timeLate\_Night | -684.99 | departure\_timeMorning:destination\_cityKolkata | 812.20 |
| airlineGO\_FIRST:arrival\_timeLate\_Night | -1276.16 | departure\_timeNight:destination\_cityKolkata | 548.41 |
| airlineIndigo:arrival\_timeLate\_Night | -1133.50 | departure\_timeEarly\_Morning:destination\_cityMumbai | 856.35 |
| airlineSpiceJet:arrival\_timeLate\_Night | -2133.18 | departure\_timeEvening:destination\_cityMumbai | 477.30 |
| airlineVistara:arrival\_timeLate\_Night | -1520.04 | departure\_timeLate\_Night:destination\_cityMumbai | 1231.04 |
| airlineAirAsia:arrival\_timeMorning | 925.13 | departure\_timeMorning:destination\_cityMumbai | 1416.16 |
| airlineGO\_FIRST:arrival\_timeMorning | 492.88 | departure\_timeNight:destination\_cityMumbai | 1433.02 |
| airlineIndigo:arrival\_timeMorning | 1587.69 | departure\_timeEarly\_Morning:classEconomy | -4078.95 |
| airlineSpiceJet:arrival\_timeMorning | -307.58 | departure\_timeEvening:classEconomy | -3806.93 |
| airlineVistara:arrival\_timeMorning | 1456.11 | departure\_timeLate\_Night:classEconomy | -4036.75 |
| airlineAirAsia:arrival\_timeNight | 656.96 | departure\_timeMorning:classEconomy | -4170.30 |
| airlineGO\_FIRST:arrival\_timeNight | 373.73 | departure\_timeNight:classEconomy | -3383.21 |
| airlineIndigo:arrival\_timeNight | 288.82 | departure\_timeEarly\_Morning:duration | 50.83 |
| airlineSpiceJet:arrival\_timeNight | -403.29 | departure\_timeEvening:duration | 58.09 |
| airlineVistara:arrival\_timeNight | 1903.77 | departure\_timeLate\_Night:duration | 35.84 |
| airlineAirAsia:destination\_cityChennai | -986.71 | departure\_timeMorning:duration | 29.94 |
| airlineGO\_FIRST:destination\_cityChennai | 197.42 | departure\_timeNight:duration | 74.96 |
| airlineIndigo:destination\_cityChennai | -284.99 | departure\_timeEarly\_Morning:days\_left | 12.53 |
| airlineSpiceJet:destination\_cityChennai | 16.45 | departure\_timeEvening:days\_left | -2.17 |
| airlineVistara:destination\_cityChennai | -624.74 | departure\_timeLate\_Night:days\_left | 42.49 |
| airlineAirAsia:destination\_cityDelhi | 692.92 | departure\_timeMorning:days\_left | 0.68 |
| airlineGO\_FIRST:destination\_cityDelhi | 898.59 | departure\_timeNight:days\_left | 14.27 |
| airlineIndigo:destination\_cityDelhi | 1078.22 | stopstwo\_or\_more:arrival\_timeEarly\_Morning | 1887.94 |
| airlineSpiceJet:destination\_cityDelhi | 199.67 | stopszero:arrival\_timeEarly\_Morning | -1273.12 |
| airlineVistara:destination\_cityDelhi | 2093.96 | stopstwo\_or\_more:arrival\_timeEvening | 935.55 |
| airlineAirAsia:destination\_cityHyderabad | -427.12 | stopszero:arrival\_timeEvening | -2053.13 |
| airlineGO\_FIRST:destination\_cityHyderabad | -205.98 | stopstwo\_or\_more:arrival\_timeLate\_Night | 941.41 |
| airlineIndigo:destination\_cityHyderabad | 266.65 | stopszero:arrival\_timeLate\_Night | -532.69 |
| airlineSpiceJet:destination\_cityHyderabad | -1130.75 | stopstwo\_or\_more:arrival\_timeMorning | 908.38 |
| airlineVistara:destination\_cityHyderabad | -59.64 | stopszero:arrival\_timeMorning | -955.87 |
| airlineAirAsia:destination\_cityKolkata | -355.71 | stopstwo\_or\_more:arrival\_timeNight | 433.83 |
| airlineGO\_FIRST:destination\_cityKolkata | 355.78 | stopszero:arrival\_timeNight | -1138.10 |
| airlineIndigo:destination\_cityKolkata | 1144.86 | stopstwo\_or\_more:destination\_cityChennai | 2873.05 |
| airlineSpiceJet:destination\_cityKolkata | -197.29 | stopszero:destination\_cityChennai | -1890.06 |
| airlineVistara:destination\_cityKolkata | 2119.26 | stopstwo\_or\_more:destination\_cityDelhi | 1000.66 |
| airlineAirAsia:destination\_cityMumbai | 214.58 | stopszero:destination\_cityDelhi | 313.56 |
| airlineGO\_FIRST:destination\_cityMumbai | 1410.24 | stopstwo\_or\_more:destination\_cityHyderabad | 1732.44 |
| airlineIndigo:destination\_cityMumbai | 1783.41 | stopszero:destination\_cityHyderabad | -3491.94 |
| airlineSpiceJet:destination\_cityMumbai | 1376.15 | stopstwo\_or\_more:destination\_cityKolkata | 990.89 |
| airlineVistara:destination\_cityMumbai | 2939.36 | stopszero:destination\_cityKolkata | -1220.36 |
| airlineAirAsia:classEconomy | -2456.31 | stopstwo\_or\_more:destination\_cityMumbai | 1416.36 |
| airlineGO\_FIRST:classEconomy | -2449.21 | stopszero:destination\_cityMumbai | -2522.96 |
| airlineIndigo:classEconomy | -2279.39 | stopstwo\_or\_more:classEconomy | -4544.08 |
| airlineSpiceJet:classEconomy | -1642.70 | stopszero:classEconomy | 8728.55 |
| airlineVistara:classEconomy | -8494.50 | stopstwo\_or\_more:duration | 82.73 |
| airlineAirAsia:duration | -97.16 | stopszero:duration | -1265.00 |
| airlineGO\_FIRST:duration | -117.56 | stopstwo\_or\_more:days\_left | -6.92 |
| airlineIndigo:duration | -208.51 | stopszero:days\_left | -8.10 |
| airlineSpiceJet:duration | 56.15 | arrival\_timeEarly\_Morning:destination\_cityChennai | 128.18 |
| airlineVistara:duration | 20.83 | arrival\_timeEvening:destination\_cityChennai | 768.31 |
| airlineAirAsia:days\_left | 24.72 | arrival\_timeLate\_Night:destination\_cityChennai | 643.84 |
| airlineGO\_FIRST:days\_left | 56.34 | arrival\_timeMorning:destination\_cityChennai | 512.01 |
| airlineIndigo:days\_left | 11.67 | arrival\_timeNight:destination\_cityChennai | 742.98 |
| airlineSpiceJet:days\_left | 64.45 | arrival\_timeEarly\_Morning:destination\_cityDelhi | 263.81 |
| airlineVistara:days\_left | -4.24 | arrival\_timeEvening:destination\_cityDelhi | 165.20 |
| source\_cityChennai:departure\_timeEarly\_Morning | 911.44 | arrival\_timeLate\_Night:destination\_cityDelhi | -247.12 |
| source\_cityDelhi:departure\_timeEarly\_Morning | 760.57 | arrival\_timeMorning:destination\_cityDelhi | -411.82 |
| source\_cityHyderabad:departure\_timeEarly\_Morning | 351.55 | arrival\_timeNight:destination\_cityDelhi | 377.19 |
| source\_cityKolkata:departure\_timeEarly\_Morning | 456.42 | arrival\_timeEarly\_Morning:destination\_cityHyderabad | -1356.47 |
| source\_cityMumbai:departure\_timeEarly\_Morning | 299.57 | arrival\_timeEvening:destination\_cityHyderabad | 1295.62 |
| source\_cityChennai:departure\_timeEvening | 2574.78 | arrival\_timeLate\_Night:destination\_cityHyderabad | 167.06 |
| source\_cityDelhi:departure\_timeEvening | 478.70 | arrival\_timeMorning:destination\_cityHyderabad | 659.38 |
| source\_cityHyderabad:departure\_timeEvening | 1400.19 | arrival\_timeNight:destination\_cityHyderabad | -346.32 |
| source\_cityKolkata:departure\_timeEvening | 127.28 | arrival\_timeEarly\_Morning:destination\_cityKolkata | 1325.81 |
| source\_cityMumbai:departure\_timeEvening | 781.59 | arrival\_timeEvening:destination\_cityKolkata | 848.87 |
| source\_cityChennai:departure\_timeLate\_Night | 457.73 | arrival\_timeLate\_Night:destination\_cityKolkata | 423.76 |
| source\_cityDelhi:departure\_timeLate\_Night | -66.34 | arrival\_timeMorning:destination\_cityKolkata | 637.74 |
| source\_cityHyderabad:departure\_timeLate\_Night | 3476.15 | arrival\_timeNight:destination\_cityKolkata | 927.52 |
| source\_cityKolkata:departure\_timeLate\_Night | 415.46 | arrival\_timeEarly\_Morning:destination\_cityMumbai | 2583.39 |
| source\_cityMumbai:departure\_timeLate\_Night | 372.88 | arrival\_timeEvening:destination\_cityMumbai | 439.72 |
| source\_cityChennai:departure\_timeMorning | 896.04 | arrival\_timeLate\_Night:destination\_cityMumbai | 637.05 |
| source\_cityDelhi:departure\_timeMorning | -184.73 | arrival\_timeMorning:destination\_cityMumbai | 125.19 |
| source\_cityHyderabad:departure\_timeMorning | -471.18 | arrival\_timeNight:destination\_cityMumbai | 886.76 |
| source\_cityKolkata:departure\_timeMorning | 1296.57 | arrival\_timeEarly\_Morning:classEconomy | -1503.77 |
| source\_cityMumbai:departure\_timeMorning | 769.13 | arrival\_timeEvening:classEconomy | -4517.84 |
| source\_cityChennai:departure\_timeNight | 758.20 | arrival\_timeLate\_Night:classEconomy | -4179.43 |
| source\_cityDelhi:departure\_timeNight | -454.93 | arrival\_timeMorning:classEconomy | -3912.87 |
| source\_cityHyderabad:departure\_timeNight | -518.97 | arrival\_timeNight:classEconomy | -5333.73 |
| source\_cityKolkata:departure\_timeNight | 78.89 | arrival\_timeEarly\_Morning:duration | -41.67 |
| source\_cityMumbai:departure\_timeNight | 233.82 | arrival\_timeEvening:duration | 81.08 |
| source\_cityChennai:stopstwo\_or\_more | 755.96 | arrival\_timeLate\_Night:duration | 78.03 |
| source\_cityDelhi:stopstwo\_or\_more | 2393.75 | arrival\_timeMorning:duration | 52.65 |
| source\_cityHyderabad:stopstwo\_or\_more | 750.59 | arrival\_timeNight:duration | 61.18 |
| source\_cityKolkata:stopstwo\_or\_more | -205.48 | arrival\_timeEarly\_Morning:days\_left | 23.58 |
| source\_cityMumbai:stopstwo\_or\_more | 1446.43 | arrival\_timeEvening:days\_left | -4.19 |
| source\_cityChennai:stopszero | -1894.43 | arrival\_timeLate\_Night:days\_left | 32.84 |
| source\_cityDelhi:stopszero | 317.55 | arrival\_timeMorning:days\_left | 10.93 |
| source\_cityHyderabad:stopszero | -3221.16 | arrival\_timeNight:days\_left | 14.33 |
| source\_cityKolkata:stopszero | -1188.64 | destination\_cityChennai:classEconomy | -3441.26 |
| source\_cityMumbai:stopszero | -2418.67 | destination\_cityDelhi:classEconomy | -1742.54 |
| source\_cityChennai:arrival\_timeEarly\_Morning | -368.72 | destination\_cityHyderabad:classEconomy | -1739.37 |
| source\_cityDelhi:arrival\_timeEarly\_Morning | 2220.43 | destination\_cityKolkata:classEconomy | -5006.84 |
| source\_cityHyderabad:arrival\_timeEarly\_Morning | -68.92 | destination\_cityMumbai:classEconomy | -4749.41 |
| source\_cityKolkata:arrival\_timeEarly\_Morning | 1649.32 | destination\_cityChennai:duration | 23.91 |
| source\_cityMumbai:arrival\_timeEarly\_Morning | 1136.50 | destination\_cityDelhi:duration | -23.21 |
| source\_cityChennai:arrival\_timeEvening | -84.49 | destination\_cityHyderabad:duration | -13.60 |
| source\_cityDelhi:arrival\_timeEvening | 520.09 | destination\_cityKolkata:duration | 57.42 |
| source\_cityHyderabad:arrival\_timeEvening | -414.86 | destination\_cityMumbai:duration | 42.23 |
| source\_cityKolkata:arrival\_timeEvening | 54.77 | destination\_cityChennai:days\_left | 0.83 |
| source\_cityMumbai:arrival\_timeEvening | 316.17 | destination\_cityDelhi:days\_left | -11.58 |
| source\_cityChennai:arrival\_timeLate\_Night | 1411.89 | destination\_cityHyderabad:days\_left | 7.86 |
| source\_cityDelhi:arrival\_timeLate\_Night | 967.98 | destination\_cityKolkata:days\_left | 10.85 |
| source\_cityHyderabad:arrival\_timeLate\_Night | 1035.62 | destination\_cityMumbai:days\_left | -2.06 |
| source\_cityKolkata:arrival\_timeLate\_Night | 2114.67 | classEconomy:duration | -385.79 |
| source\_cityMumbai:arrival\_timeLate\_Night | 1336.83 | classEconomy:days\_left | -177.41 |
| source\_cityChennai:arrival\_timeMorning | 168.71 | duration:days\_left | -0.83 |
| source\_cityDelhi:arrival\_timeMorning | 99.64 |  |  |

# Appendix F (Lasso Regression Model)

|  |  |  |  |
| --- | --- | --- | --- |
| **Term** | **Estimate** | **Term** | **Estimate** |
| (Intercept) | 51088.98 | source\_cityHyderabad:arrival\_timeMorning | -548.78 |
| airlineAirAsia | -1193.72 | source\_cityKolkata:arrival\_timeMorning | 151.27 |
| airlineGO\_FIRST | -1237.51 | source\_cityMumbai:arrival\_timeMorning | 630.85 |
| airlineIndigo | -137.90 | source\_cityChennai:arrival\_timeNight | -190.21 |
| airlineSpiceJet | -33.23 | source\_cityDelhi:arrival\_timeNight | -26.44 |
| airlineVistara | 6305.01 | source\_cityHyderabad:arrival\_timeNight | -33.66 |
| source\_cityChennai | 0.00 | source\_cityKolkata:arrival\_timeNight | 57.47 |
| source\_cityDelhi | -3589.40 | source\_cityMumbai:arrival\_timeNight | 490.87 |
| source\_cityHyderabad | -2748.38 | source\_cityChennai:destination\_cityChennai | 0.00 |
| source\_cityKolkata | 2340.56 | source\_cityDelhi:destination\_cityChennai | 803.92 |
| source\_cityMumbai | 0.00 | source\_cityHyderabad:destination\_cityChennai | 660.30 |
| departure\_timeEarly\_Morning | 203.92 | source\_cityKolkata:destination\_cityChennai | 456.46 |
| departure\_timeEvening | 0.00 | source\_cityMumbai:destination\_cityChennai | -482.23 |
| departure\_timeLate\_Night | 0.00 | source\_cityChennai:destination\_cityDelhi | 1251.60 |
| departure\_timeMorning | 0.00 | source\_cityDelhi:destination\_cityDelhi | 0.00 |
| departure\_timeNight | 304.77 | source\_cityHyderabad:destination\_cityDelhi | 0.00 |
| stopstwo\_or\_more | 10071.49 | source\_cityKolkata:destination\_cityDelhi | 1069.52 |
| stopszero | -26568.79 | source\_cityMumbai:destination\_cityDelhi | -2666.14 |
| arrival\_timeEarly\_Morning | -2450.98 | source\_cityChennai:destination\_cityHyderabad | 206.76 |
| arrival\_timeEvening | -13.47 | source\_cityDelhi:destination\_cityHyderabad | 224.40 |
| arrival\_timeLate\_Night | 968.11 | source\_cityHyderabad:destination\_cityHyderabad | 0.00 |
| arrival\_timeMorning | 201.20 | source\_cityKolkata:destination\_cityHyderabad | 0.00 |
| arrival\_timeNight | 2512.83 | source\_cityMumbai:destination\_cityHyderabad | -1910.55 |
| destination\_cityChennai | 0.00 | source\_cityChennai:destination\_cityKolkata | 389.59 |
| destination\_cityDelhi | -3070.81 | source\_cityDelhi:destination\_cityKolkata | 846.30 |
| destination\_cityHyderabad | -3304.00 | source\_cityHyderabad:destination\_cityKolkata | -194.32 |
| destination\_cityKolkata | 2223.00 | source\_cityKolkata:destination\_cityKolkata | 0.00 |
| destination\_cityMumbai | 0.00 | source\_cityMumbai:destination\_cityKolkata | -1628.27 |
| classEconomy | -40149.48 | source\_cityChennai:destination\_cityMumbai | -199.26 |
| duration | 0.00 | source\_cityDelhi:destination\_cityMumbai | -2422.24 |
| days\_left | -55.05 | source\_cityHyderabad:destination\_cityMumbai | -1484.22 |
| airlineAirAsia:source\_cityChennai | -1327.04 | source\_cityKolkata:destination\_cityMumbai | -1997.46 |
| airlineGO\_FIRST:source\_cityChennai | -145.67 | source\_cityMumbai:destination\_cityMumbai | 0.00 |
| airlineIndigo:source\_cityChennai | -854.68 | source\_cityChennai:classEconomy | -364.57 |
| airlineSpiceJet:source\_cityChennai | -1055.08 | source\_cityDelhi:classEconomy | 2429.69 |
| airlineVistara:source\_cityChennai | 320.35 | source\_cityHyderabad:classEconomy | 3032.16 |
| airlineAirAsia:source\_cityDelhi | 0.00 | source\_cityKolkata:classEconomy | -2263.63 |
| airlineGO\_FIRST:source\_cityDelhi | 317.89 | source\_cityMumbai:classEconomy | -907.17 |
| airlineIndigo:source\_cityDelhi | 268.36 | source\_cityChennai:duration | 10.79 |
| airlineSpiceJet:source\_cityDelhi | -61.54 | source\_cityDelhi:duration | -1.44 |
| airlineVistara:source\_cityDelhi | 1992.70 | source\_cityHyderabad:duration | -27.57 |
| airlineAirAsia:source\_cityHyderabad | -1204.56 | source\_cityKolkata:duration | 14.30 |
| airlineGO\_FIRST:source\_cityHyderabad | -621.60 | source\_cityMumbai:duration | 20.07 |
| airlineIndigo:source\_cityHyderabad | -300.70 | source\_cityChennai:days\_left | -22.33 |
| airlineSpiceJet:source\_cityHyderabad | -519.82 | source\_cityDelhi:days\_left | -9.68 |
| airlineVistara:source\_cityHyderabad | 189.04 | source\_cityHyderabad:days\_left | 6.43 |
| airlineAirAsia:source\_cityKolkata | -235.45 | source\_cityKolkata:days\_left | 0.00 |
| airlineGO\_FIRST:source\_cityKolkata | 716.85 | source\_cityMumbai:days\_left | -6.94 |
| airlineIndigo:source\_cityKolkata | 582.51 | departure\_timeEarly\_Morning:stopstwo\_or\_more | 54.65 |
| airlineSpiceJet:source\_cityKolkata | 0.00 | departure\_timeEvening:stopstwo\_or\_more | -345.81 |
| airlineVistara:source\_cityKolkata | 2086.15 | departure\_timeLate\_Night:stopstwo\_or\_more | -795.59 |
| airlineAirAsia:source\_cityMumbai | 1051.36 | departure\_timeMorning:stopstwo\_or\_more | 64.64 |
| airlineGO\_FIRST:source\_cityMumbai | 1600.36 | departure\_timeNight:stopstwo\_or\_more | 245.87 |
| airlineIndigo:source\_cityMumbai | 2092.69 | departure\_timeEarly\_Morning:stopszero | 0.00 |
| airlineSpiceJet:source\_cityMumbai | 1326.72 | departure\_timeEvening:stopszero | 0.00 |
| airlineVistara:source\_cityMumbai | 3115.90 | departure\_timeLate\_Night:stopszero | -1612.01 |
| airlineAirAsia:departure\_timeEarly\_Morning | -94.16 | departure\_timeMorning:stopszero | 0.00 |
| airlineGO\_FIRST:departure\_timeEarly\_Morning | 0.00 | departure\_timeNight:stopszero | 0.00 |
| airlineIndigo:departure\_timeEarly\_Morning | -43.11 | departure\_timeEarly\_Morning:arrival\_timeEarly\_Morning | -840.91 |
| airlineSpiceJet:departure\_timeEarly\_Morning | -811.82 | departure\_timeEvening:arrival\_timeEarly\_Morning | 0.00 |
| airlineVistara:departure\_timeEarly\_Morning | 452.79 | departure\_timeLate\_Night:arrival\_timeEarly\_Morning | -1241.83 |
| airlineAirAsia:departure\_timeEvening | 471.60 | departure\_timeMorning:arrival\_timeEarly\_Morning | -439.30 |
| airlineGO\_FIRST:departure\_timeEvening | 828.00 | departure\_timeNight:arrival\_timeEarly\_Morning | -931.59 |
| airlineIndigo:departure\_timeEvening | 567.97 | departure\_timeEarly\_Morning:arrival\_timeEvening | 1402.88 |
| airlineSpiceJet:departure\_timeEvening | -447.48 | departure\_timeEvening:arrival\_timeEvening | 0.00 |
| airlineVistara:departure\_timeEvening | 1271.79 | departure\_timeLate\_Night:arrival\_timeEvening | 1106.35 |
| airlineAirAsia:departure\_timeLate\_Night | -397.37 | departure\_timeMorning:arrival\_timeEvening | 1608.99 |
| airlineGO\_FIRST:departure\_timeLate\_Night | -197.18 | departure\_timeNight:arrival\_timeEvening | 0.00 |
| airlineIndigo:departure\_timeLate\_Night | 0.00 | departure\_timeEarly\_Morning:arrival\_timeLate\_Night | 440.00 |
| airlineSpiceJet:departure\_timeLate\_Night | 0.00 | departure\_timeEvening:arrival\_timeLate\_Night | 118.10 |
| airlineVistara:departure\_timeLate\_Night | 0.00 | departure\_timeLate\_Night:arrival\_timeLate\_Night | 0.00 |
| airlineAirAsia:departure\_timeMorning | -434.71 | departure\_timeMorning:arrival\_timeLate\_Night | 415.56 |
| airlineGO\_FIRST:departure\_timeMorning | 762.09 | departure\_timeNight:arrival\_timeLate\_Night | -138.05 |
| airlineIndigo:departure\_timeMorning | -200.41 | departure\_timeEarly\_Morning:arrival\_timeMorning | -643.78 |
| airlineSpiceJet:departure\_timeMorning | -69.40 | departure\_timeEvening:arrival\_timeMorning | 778.59 |
| airlineVistara:departure\_timeMorning | 607.63 | departure\_timeLate\_Night:arrival\_timeMorning | 0.00 |
| airlineAirAsia:departure\_timeNight | -616.64 | departure\_timeMorning:arrival\_timeMorning | -688.71 |
| airlineGO\_FIRST:departure\_timeNight | -410.51 | departure\_timeNight:arrival\_timeMorning | 128.20 |
| airlineIndigo:departure\_timeNight | 0.00 | departure\_timeEarly\_Morning:arrival\_timeNight | 79.79 |
| airlineSpiceJet:departure\_timeNight | -1611.23 | departure\_timeEvening:arrival\_timeNight | -1190.06 |
| airlineVistara:departure\_timeNight | -445.79 | departure\_timeLate\_Night:arrival\_timeNight | -3.41 |
| airlineAirAsia:stopstwo\_or\_more | -1092.32 | departure\_timeMorning:arrival\_timeNight | 832.72 |
| airlineGO\_FIRST:stopstwo\_or\_more | -1047.36 | departure\_timeNight:arrival\_timeNight | -1909.29 |
| airlineIndigo:stopstwo\_or\_more | 0.00 | departure\_timeEarly\_Morning:destination\_cityChennai | -377.24 |
| airlineSpiceJet:stopstwo\_or\_more | 0.00 | departure\_timeEvening:destination\_cityChennai | -335.10 |
| airlineVistara:stopstwo\_or\_more | 1331.36 | departure\_timeLate\_Night:destination\_cityChennai | -1157.12 |
| airlineAirAsia:stopszero | 326.10 | departure\_timeMorning:destination\_cityChennai | 0.00 |
| airlineGO\_FIRST:stopszero | -1667.27 | departure\_timeNight:destination\_cityChennai | -114.15 |
| airlineIndigo:stopszero | -159.07 | departure\_timeEarly\_Morning:destination\_cityDelhi | -371.76 |
| airlineSpiceJet:stopszero | 0.00 | departure\_timeEvening:destination\_cityDelhi | -435.94 |
| airlineVistara:stopszero | -3708.06 | departure\_timeLate\_Night:destination\_cityDelhi | 915.50 |
| airlineAirAsia:arrival\_timeEarly\_Morning | -475.32 | departure\_timeMorning:destination\_cityDelhi | 0.00 |
| airlineGO\_FIRST:arrival\_timeEarly\_Morning | -38.86 | departure\_timeNight:destination\_cityDelhi | 717.26 |
| airlineIndigo:arrival\_timeEarly\_Morning | -214.94 | departure\_timeEarly\_Morning:destination\_cityHyderabad | -467.36 |
| airlineSpiceJet:arrival\_timeEarly\_Morning | -858.92 | departure\_timeEvening:destination\_cityHyderabad | 0.00 |
| airlineVistara:arrival\_timeEarly\_Morning | 308.92 | departure\_timeLate\_Night:destination\_cityHyderabad | -619.52 |
| airlineAirAsia:arrival\_timeEvening | 0.00 | departure\_timeMorning:destination\_cityHyderabad | 395.21 |
| airlineGO\_FIRST:arrival\_timeEvening | 102.84 | departure\_timeNight:destination\_cityHyderabad | 0.00 |
| airlineIndigo:arrival\_timeEvening | 243.57 | departure\_timeEarly\_Morning:destination\_cityKolkata | -561.73 |
| airlineSpiceJet:arrival\_timeEvening | 46.06 | departure\_timeEvening:destination\_cityKolkata | 0.00 |
| airlineVistara:arrival\_timeEvening | 106.49 | departure\_timeLate\_Night:destination\_cityKolkata | 0.00 |
| airlineAirAsia:arrival\_timeLate\_Night | -829.32 | departure\_timeMorning:destination\_cityKolkata | 48.16 |
| airlineGO\_FIRST:arrival\_timeLate\_Night | -1144.95 | departure\_timeNight:destination\_cityKolkata | 0.00 |
| airlineIndigo:arrival\_timeLate\_Night | -1103.95 | departure\_timeEarly\_Morning:destination\_cityMumbai | 408.53 |
| airlineSpiceJet:arrival\_timeLate\_Night | -1432.50 | departure\_timeEvening:destination\_cityMumbai | 271.87 |
| airlineVistara:arrival\_timeLate\_Night | -3056.61 | departure\_timeLate\_Night:destination\_cityMumbai | 0.00 |
| airlineAirAsia:arrival\_timeMorning | 0.00 | departure\_timeMorning:destination\_cityMumbai | 995.44 |
| airlineGO\_FIRST:arrival\_timeMorning | -65.72 | departure\_timeNight:destination\_cityMumbai | 1265.93 |
| airlineIndigo:arrival\_timeMorning | 71.07 | departure\_timeEarly\_Morning:classEconomy | -482.08 |
| airlineSpiceJet:arrival\_timeMorning | -272.51 | departure\_timeEvening:classEconomy | -1629.01 |
| airlineVistara:arrival\_timeMorning | 224.84 | departure\_timeLate\_Night:classEconomy | -183.45 |
| airlineAirAsia:arrival\_timeNight | 64.17 | departure\_timeMorning:classEconomy | -490.01 |
| airlineGO\_FIRST:arrival\_timeNight | 0.00 | departure\_timeNight:classEconomy | -1181.53 |
| airlineIndigo:arrival\_timeNight | -130.62 | departure\_timeEarly\_Morning:duration | 0.00 |
| airlineSpiceJet:arrival\_timeNight | -465.25 | departure\_timeEvening:duration | 53.80 |
| airlineVistara:arrival\_timeNight | 303.64 | departure\_timeLate\_Night:duration | 0.00 |
| airlineAirAsia:destination\_cityChennai | -975.70 | departure\_timeMorning:duration | -15.76 |
| airlineGO\_FIRST:destination\_cityChennai | -130.85 | departure\_timeNight:duration | 81.25 |
| airlineIndigo:destination\_cityChennai | -637.15 | departure\_timeEarly\_Morning:days\_left | 5.91 |
| airlineSpiceJet:destination\_cityChennai | -471.57 | departure\_timeEvening:days\_left | -6.88 |
| airlineVistara:destination\_cityChennai | -593.22 | departure\_timeLate\_Night:days\_left | 13.05 |
| airlineAirAsia:destination\_cityDelhi | 181.76 | departure\_timeMorning:days\_left | -3.41 |
| airlineGO\_FIRST:destination\_cityDelhi | 438.32 | departure\_timeNight:days\_left | 11.94 |
| airlineIndigo:destination\_cityDelhi | 243.49 | stopstwo\_or\_more:arrival\_timeEarly\_Morning | 114.69 |
| airlineSpiceJet:destination\_cityDelhi | -157.89 | stopszero:arrival\_timeEarly\_Morning | -204.44 |
| airlineVistara:destination\_cityDelhi | 2133.28 | stopstwo\_or\_more:arrival\_timeEvening | 0.00 |
| airlineAirAsia:destination\_cityHyderabad | -212.26 | stopszero:arrival\_timeEvening | 166.63 |
| airlineGO\_FIRST:destination\_cityHyderabad | -205.69 | stopstwo\_or\_more:arrival\_timeLate\_Night | 0.00 |
| airlineIndigo:destination\_cityHyderabad | 1.04 | stopszero:arrival\_timeLate\_Night | -61.35 |
| airlineSpiceJet:destination\_cityHyderabad | -427.87 | stopstwo\_or\_more:arrival\_timeMorning | 42.92 |
| airlineVistara:destination\_cityHyderabad | 0.63 | stopszero:arrival\_timeMorning | 0.00 |
| airlineAirAsia:destination\_cityKolkata | -251.35 | stopstwo\_or\_more:arrival\_timeNight | -308.38 |
| airlineGO\_FIRST:destination\_cityKolkata | 316.73 | stopszero:arrival\_timeNight | 203.53 |
| airlineIndigo:destination\_cityKolkata | 1298.68 | stopstwo\_or\_more:destination\_cityChennai | 1742.94 |
| airlineSpiceJet:destination\_cityKolkata | 45.83 | stopszero:destination\_cityChennai | -65.19 |
| airlineVistara:destination\_cityKolkata | 2409.67 | stopstwo\_or\_more:destination\_cityDelhi | -685.93 |
| airlineAirAsia:destination\_cityMumbai | 43.78 | stopszero:destination\_cityDelhi | 2330.82 |
| airlineGO\_FIRST:destination\_cityMumbai | 1454.60 | stopstwo\_or\_more:destination\_cityHyderabad | 521.50 |
| airlineIndigo:destination\_cityMumbai | 1861.46 | stopszero:destination\_cityHyderabad | -729.59 |
| airlineSpiceJet:destination\_cityMumbai | 1427.66 | stopstwo\_or\_more:destination\_cityKolkata | 0.00 |
| airlineVistara:destination\_cityMumbai | 3281.92 | stopszero:destination\_cityKolkata | -260.37 |
| airlineAirAsia:classEconomy | 0.00 | stopstwo\_or\_more:destination\_cityMumbai | 157.69 |
| airlineGO\_FIRST:classEconomy | -297.02 | stopszero:destination\_cityMumbai | -610.14 |
| airlineIndigo:classEconomy | -570.69 | stopstwo\_or\_more:classEconomy | -8164.03 |
| airlineSpiceJet:classEconomy | -458.74 | stopszero:classEconomy | 22992.78 |
| airlineVistara:classEconomy | -7365.38 | stopstwo\_or\_more:duration | 0.00 |
| airlineAirAsia:duration | -154.22 | stopszero:duration | 623.48 |
| airlineGO\_FIRST:duration | -202.91 | stopstwo\_or\_more:days\_left | -20.00 |
| airlineIndigo:duration | -149.39 | stopszero:days\_left | 32.35 |
| airlineSpiceJet:duration | -31.76 | arrival\_timeEarly\_Morning:destination\_cityChennai | 26.95 |
| airlineVistara:duration | -63.99 | arrival\_timeEvening:destination\_cityChennai | 13.79 |
| airlineAirAsia:days\_left | 16.65 | arrival\_timeLate\_Night:destination\_cityChennai | 0.00 |
| airlineGO\_FIRST:days\_left | 69.78 | arrival\_timeMorning:destination\_cityChennai | 0.00 |
| airlineIndigo:days\_left | -3.55 | arrival\_timeNight:destination\_cityChennai | -50.94 |
| airlineSpiceJet:days\_left | 49.92 | arrival\_timeEarly\_Morning:destination\_cityDelhi | 223.56 |
| airlineVistara:days\_left | -22.59 | arrival\_timeEvening:destination\_cityDelhi | 153.93 |
| source\_cityChennai:departure\_timeEarly\_Morning | 582.75 | arrival\_timeLate\_Night:destination\_cityDelhi | -3.61 |
| source\_cityDelhi:departure\_timeEarly\_Morning | 956.93 | arrival\_timeMorning:destination\_cityDelhi | -358.67 |
| source\_cityHyderabad:departure\_timeEarly\_Morning | 455.22 | arrival\_timeNight:destination\_cityDelhi | 357.44 |
| source\_cityKolkata:departure\_timeEarly\_Morning | 0.00 | arrival\_timeEarly\_Morning:destination\_cityHyderabad | -170.93 |
| source\_cityMumbai:departure\_timeEarly\_Morning | -46.65 | arrival\_timeEvening:destination\_cityHyderabad | 1746.99 |
| source\_cityChennai:departure\_timeEvening | 2456.67 | arrival\_timeLate\_Night:destination\_cityHyderabad | 37.73 |
| source\_cityDelhi:departure\_timeEvening | 565.05 | arrival\_timeMorning:destination\_cityHyderabad | 1092.64 |
| source\_cityHyderabad:departure\_timeEvening | 1596.92 | arrival\_timeNight:destination\_cityHyderabad | -0.50 |
| source\_cityKolkata:departure\_timeEvening | -81.08 | arrival\_timeEarly\_Morning:destination\_cityKolkata | 1079.95 |
| source\_cityMumbai:departure\_timeEvening | 747.08 | arrival\_timeEvening:destination\_cityKolkata | 0.00 |
| source\_cityChennai:departure\_timeLate\_Night | 0.00 | arrival\_timeLate\_Night:destination\_cityKolkata | -166.07 |
| source\_cityDelhi:departure\_timeLate\_Night | -213.51 | arrival\_timeMorning:destination\_cityKolkata | 0.00 |
| source\_cityHyderabad:departure\_timeLate\_Night | 982.71 | arrival\_timeNight:destination\_cityKolkata | 0.00 |
| source\_cityKolkata:departure\_timeLate\_Night | 0.00 | arrival\_timeEarly\_Morning:destination\_cityMumbai | 2713.96 |
| source\_cityMumbai:departure\_timeLate\_Night | 0.00 | arrival\_timeEvening:destination\_cityMumbai | 131.21 |
| source\_cityChennai:departure\_timeMorning | 504.22 | arrival\_timeLate\_Night:destination\_cityMumbai | 345.31 |
| source\_cityDelhi:departure\_timeMorning | 0.00 | arrival\_timeMorning:destination\_cityMumbai | 0.00 |
| source\_cityHyderabad:departure\_timeMorning | -369.36 | arrival\_timeNight:destination\_cityMumbai | 419.69 |
| source\_cityKolkata:departure\_timeMorning | 827.86 | arrival\_timeEarly\_Morning:classEconomy | 1398.58 |
| source\_cityMumbai:departure\_timeMorning | 373.84 | arrival\_timeEvening:classEconomy | -977.69 |
| source\_cityChennai:departure\_timeNight | 424.32 | arrival\_timeLate\_Night:classEconomy | -1677.29 |
| source\_cityDelhi:departure\_timeNight | -392.45 | arrival\_timeMorning:classEconomy | -967.89 |
| source\_cityHyderabad:departure\_timeNight | -430.79 | arrival\_timeNight:classEconomy | -2207.71 |
| source\_cityKolkata:departure\_timeNight | -314.33 | arrival\_timeEarly\_Morning:duration | -51.01 |
| source\_cityMumbai:departure\_timeNight | 0.00 | arrival\_timeEvening:duration | 52.13 |
| source\_cityChennai:stopstwo\_or\_more | 0.00 | arrival\_timeLate\_Night:duration | 100.05 |
| source\_cityDelhi:stopstwo\_or\_more | 638.39 | arrival\_timeMorning:duration | 26.79 |
| source\_cityHyderabad:stopstwo\_or\_more | 0.00 | arrival\_timeNight:duration | 0.00 |
| source\_cityKolkata:stopstwo\_or\_more | -1112.83 | arrival\_timeEarly\_Morning:days\_left | 15.24 |
| source\_cityMumbai:stopstwo\_or\_more | 0.00 | arrival\_timeEvening:days\_left | -11.58 |
| source\_cityChennai:stopszero | -19.77 | arrival\_timeLate\_Night:days\_left | 18.00 |
| source\_cityDelhi:stopszero | 2638.03 | arrival\_timeMorning:days\_left | 3.09 |
| source\_cityHyderabad:stopszero | -353.83 | arrival\_timeNight:days\_left | 0.00 |
| source\_cityKolkata:stopszero | -271.92 | destination\_cityChennai:classEconomy | 0.00 |
| source\_cityMumbai:stopszero | -75.77 | destination\_cityDelhi:classEconomy | 3009.81 |
| source\_cityChennai:arrival\_timeEarly\_Morning | -662.87 | destination\_cityHyderabad:classEconomy | 2809.03 |
| source\_cityDelhi:arrival\_timeEarly\_Morning | 2224.93 | destination\_cityKolkata:classEconomy | -2430.52 |
| source\_cityHyderabad:arrival\_timeEarly\_Morning | -386.44 | destination\_cityMumbai:classEconomy | -846.80 |
| source\_cityKolkata:arrival\_timeEarly\_Morning | 776.32 | destination\_cityChennai:duration | 0.00 |
| source\_cityMumbai:arrival\_timeEarly\_Morning | 1098.56 | destination\_cityDelhi:duration | -39.58 |
| source\_cityChennai:arrival\_timeEvening | -434.92 | destination\_cityHyderabad:duration | -38.10 |
| source\_cityDelhi:arrival\_timeEvening | 538.72 | destination\_cityKolkata:duration | 9.87 |
| source\_cityHyderabad:arrival\_timeEvening | -643.58 | destination\_cityMumbai:duration | 0.00 |
| source\_cityKolkata:arrival\_timeEvening | -748.32 | destination\_cityChennai:days\_left | 0.00 |
| source\_cityMumbai:arrival\_timeEvening | -0.18 | destination\_cityDelhi:days\_left | -13.03 |
| source\_cityChennai:arrival\_timeLate\_Night | 171.98 | destination\_cityHyderabad:days\_left | 12.35 |
| source\_cityDelhi:arrival\_timeLate\_Night | 825.89 | destination\_cityKolkata:days\_left | 0.00 |
| source\_cityHyderabad:arrival\_timeLate\_Night | 104.79 | destination\_cityMumbai:days\_left | -6.17 |
| source\_cityKolkata:arrival\_timeLate\_Night | 658.02 | classEconomy:duration | 87.35 |
| source\_cityMumbai:arrival\_timeLate\_Night | 333.81 | classEconomy:days\_left | -79.19 |
| source\_cityChennai:arrival\_timeMorning | -83.99 | duration:days\_left | -1.25 |
| source\_cityDelhi:arrival\_timeMorning | 0.00 |  |  |