Flight Cost Analysis

Determining Optimal Times to Buy Airline Tickets for Domestic Flights within India

Team 1 Members

(names removed for ORIE 3120 submission)

Table of Contents

Figures and Tables

Abstract

Hundreds of millions of people take flights every year. These customers all pay for their trips, contributing greatly to the $470 billion airline industry. Domestic flights in India act as a good subset of data for this experiment, as the number of flights taken per year are a large enough sample size to be comparable to the worldscale population. Our team looks to provide individuals looking to fly domestically throughout India with the optimal times to purchase airline tickets to maximize their savings by answering three main questions, revolving around the variation of flight prices, optimal ticket prices for separate industries, and factors that increase airfare. To help answer these questions, we analyzed the vast array of data in our dataset, finding trends and patterns. After finding a commonality between airfare variation in subsequent days to flight departure, a linear regression prediction model confirmed that ticket price variation indeed had substantial correlation with what day the ticket was purchased before flight departure. We then found that ticket price in general was correlated with what day the ticket was purchased before flight departure.. To forecast ticket prices around the 20 to 50 day mark, we used the Holt Exponential Smoothing Forecast Model, which found all low range values. This recommendation report can conclude that the optimal time to buy tickets is between twenty and fifty days before the departure date, as ticket prices are significantly cheaper than they will be closer to the date, without much variation.

Introduction

With sanctions being imposed on the Russian economy, gas prices have soared. In the United States, these prices broke a new high since 2008. Soon, this rise in prices will reflect on the airline industry as airlines raise ticket prices to compensate for the new cost of fuel. The constant battle between airlines trying to make a profit and filling flight seats, causes ticket prices to drastically change. This complicates people's abilities to determine when flight tickets are the most affordable. Since ticket prices fluctuate significantly over the course of the booking period, our goal is to help customers of domestic flights within India find the optimal time to purchase airline tickets. Even though our data constrains us to Indian domestic flights, we hope to find price patterns that can be applied to all flights around the world. Ultimately, these passengers will be able to use this information to make the best financial decisions when purchasing airline tickets.

Our team found this dataset on Kaggle, consisting of 300,155 records in total. Each record contains information on a single ticket for domestic Indian flights across six different airlines operating in this country. For each ticket we have information pertaining to the flight departure city and time, the flight arrival city and time, the class of the ticket, the duration of the flight, the price of the ticket, and the number of days between purchase of the ticket and flight departure.

Research Questions

Descriptive Question

Our descriptive question is as follows: What factors have led to the increase in airfare in recent times? The world is still recovering from when Covid-19 ran amok in 2020. Indian airlines were one of the most damaged sectors in India's economy. Covid-19 considerably impaired the aviation sector's primary sources of income: tourism and global commerce. Additionally, profits decreased with the enforcement of travel restrictions, international travel bans, and periodic operation suspensions due to decreased passenger traffic: "[d]omestic air passenger traffic declined by 0.3 percent in 2019-20 and by 61.7 per cent in 2021 due to the pandemic".

The end of 2020 and the entirety of 2021 was a period of significant recovery in India's economy and aviation sector: Scientists developed the first viable Covid-19 vaccine in December 2020, and India started vaccine administration in January 2021.[2] [3] These critical events allowed Indian airlines to ease Covid-19 protocols and restrictions, leading to surges in passenger traffic as fewer people quarantined and more received vaccinations.[2] [3] However, more demand led to inflations in flight prices.

Recovery continued into 2022, accelerated by India's release of their own Covid-19 vaccine, Covaxin.[4] More restrictions were made as well, such India's withdrawal of its two-year international travel ban. Removing the ban improved passenger traffic and profit for Indian airlines but led to significantly high flight prices for travelers.[5]

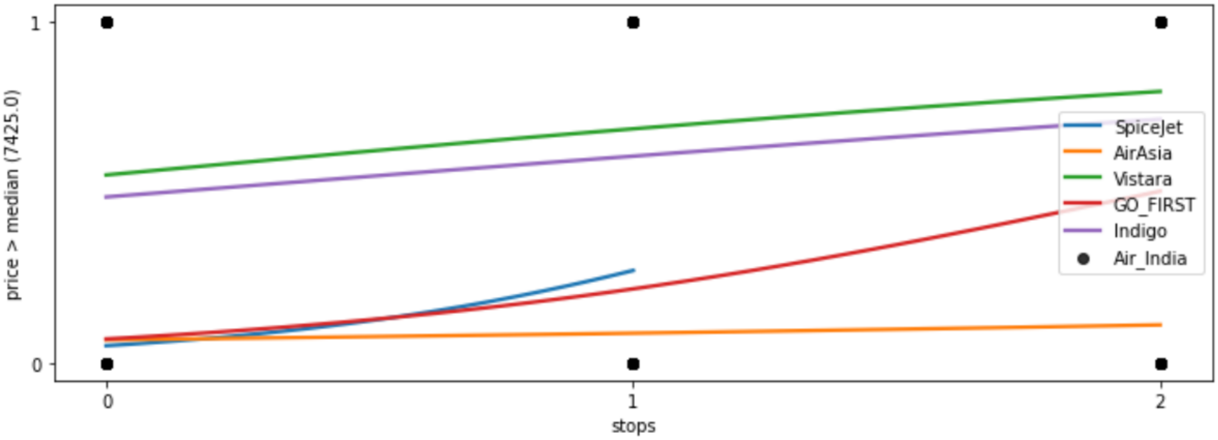
We provide financially conscious passengers with insights into optimal times to purchase tickets to address these hikes in cost. Our data is from February 11th to March 31st, 2022, capturing the flight price trends in airline fares during a recent recovery period in India's economy and the aviation sector.[6] Therefore, our analyses and predictions of flight price trends based on these data are still applicable today and can achieve our goal of giving airline passengers an advantage.

Logistic Regression of Price Model

Given these events, predicting price is extremely complex, especially when overlooked factors can impact price like the number of stops in a flight. Figure 1 shows a logistic regression model of price over median price and the number of stops. If the price were over the median price, it would be a one, otherwise zero. Additionally, stops were two if there were two or more. This gives another perspective as of what can impact price and helps us provide more comprehensive insights of optimal flight ticket purchase times.

Figure 1

Logistic Regression of Price vs. Number of Stops



Predictive Question

Our second research question is as follows: How much do flight prices vary within 1 week of departure? Some customers may feel as though price variations increase significantly as you get closer to the departure date as airlines attempt to maximize their profit while also trying to fill their planes.

From our findings, we discovered that the least variation in flight prices occurs one day before departure. We also discovered that within one week before departure, the price variance is largest two or three days before take-off. Referring to Figure 2b, the variance in pricing across all flights from our dataset is pretty stable between fifty and sixteen days before departure, with the median price also staying fairly similar around 100 USD. Once travelers get within 16 days of departure, the median price across all flights increases drastically and is fairly unstable. Referring to Figure 2a, you can see this instability in price variations and median prices. All these median prices are still much higher than those over two weeks out.

This means when booking domestic flights within India, you should not expect price variation to change significantly from day to day until you get within sixteen days of departure.

Once you get within that time period, you can expect prices to be significantly higher and for prices to vary a lot more than previously as airlines scramble to fill seats.

Our main takeaway from these visualizations is that customers should expect prices to change a lot the closer they get to the departure date with the average price of flights increasing overall.

Figure 2a

Price Variation Of Airline Tickets Within a Week of Departure

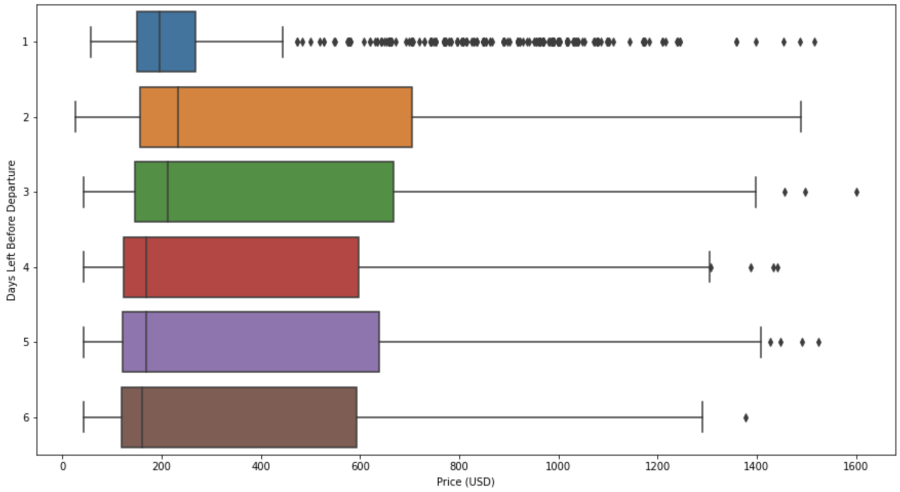
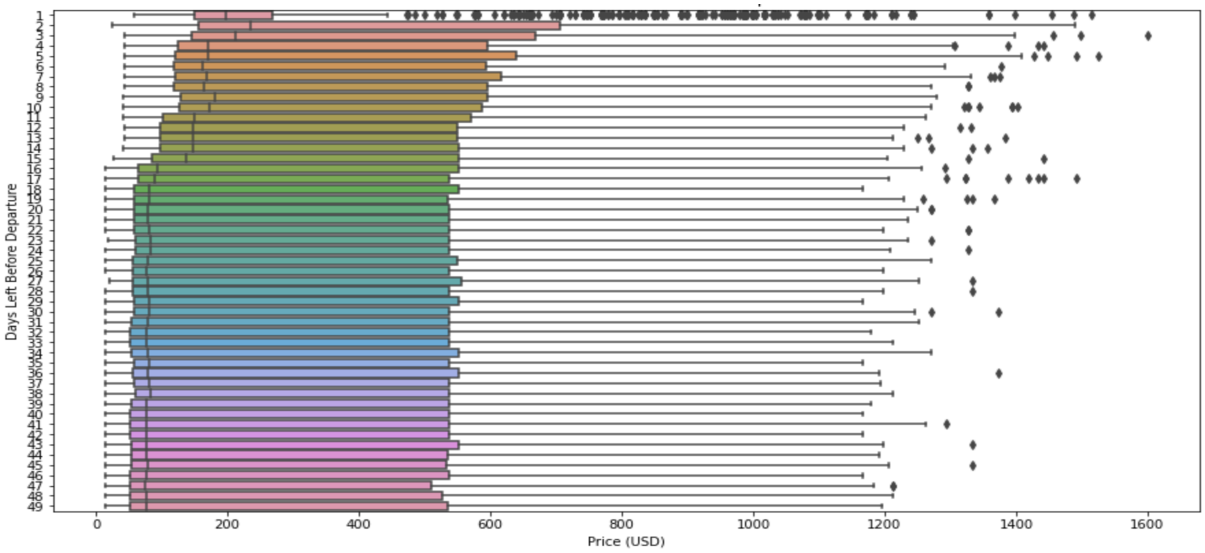


Figure 2b

Price Variation of Airline Tickets Within 50 Days of Departure



Linear Regression Prediction Model

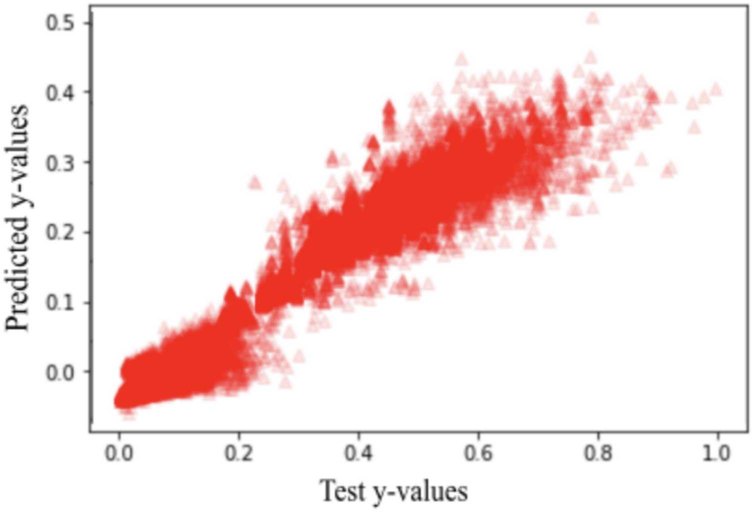
To better help these customers understand our predictive question, our team has built a linear regression model using recursive feature elimination to predict future flight prices (Appendix A). We first cleaned the data by adding dummy variables for flight arrival and departure information and dropping columns consisting of string objects, converting the remaining column values to floats. From there we were able to split the data into x and y training and test sets to build our predictive model.

The linear regression model was fitted on the x_train and y_train and predicted on the x_test. Using RFE, we eliminated columns in the training data that had little effect in predicting the target variable, giving us a more accurate x_train to fit the model on. We ended up with a mean squared error score of 0.193 and an r-squared value of 0.909 as you can see from the model summary in Appendix A. Using this model, customers can predict prices of future flights before their scheduled departure. Since this model uses the number of days before departure, flight duration, airline company, and departure and arrival cities as predictors, customers can enter this information into our model for a flight they wish to predict future prices on and receive an accurate estimate. In order to further estimate the importance of these different predictors, we ran many variations of the model, each time including a small subset of all these predictors. We discovered flight duration, number of days before flight departure, and the airline company were the most significant out of these predictors.

In Figure 2c, a clear linear trend is shown between the model's predicted values (y_pred plotted on the y-axis) and test values (y_test plotted on the x-axis). This is a good indication of how accurate our predictions are, with a majority of data points following along this linear trend with minimal outliers skewing our predictions. Since we are using a linear model, which is the least flexible model, and we are projecting a linear trend between test and predicted values, we can account for very little bias in the model as well.

Figure 2c

Correlation Between Test and Predicted Values (y_pred vs. y_test)



In summary, for customers wishing to purchase domestic flight tickets within India, one should only need to know the flight duration, number of days before departure, and airline company to get a pretty fair prediction of what they will need to pay for that flight ticket. Even though there can be a large variation in flight ticket prices before departure, as seen in Figures 2a and 2b, we can still get a fair prediction for these prices from the model built.

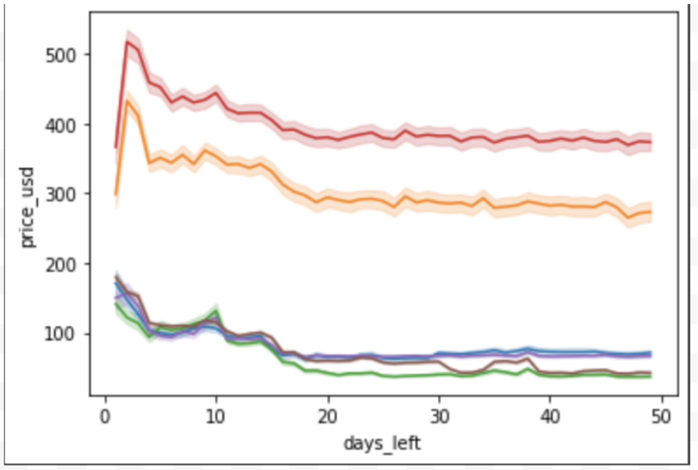
Prescriptive Question

Our third research question is as follows: What is the optimal time to buy airline tickets before their departure dates to maximize savings within each separate airline industry? This aims to help customers of domestic Indian flights determine when they should purchase the airline tickets before their departure date to maximize savings within each separate airline industry. Some customers may have the flexibility to choose the date of purchasing airline tickets any time they want, thus the goal of this research question is to make recommendations about what is the most ideal date before departure to purchase the ticket in each airline company. To see this clearly, we generated a graph from the days before departure vs. price (usd) for each airline company to try to find a pattern.

From our findings, we discovered that in general, as the date grows further away from the departure date, the price of the airline tickets decreases accordingly. Specifically, at the point passing 20 days before the departure date, there exists a dramatic decrease in the price of the airline tickets. We compared the data for each airline company (each represented by a different color in Figure 3a), it turned out they all contain similar patterns. As we can see from each colored trend, during the time period between 20 to 50 days, the change in the airline ticket price stays small and lies in the range range. During the time period between 0 to 20 days, the slope of the price change becomes much steeper and decreases all the way from their highest points.

Figure 3a

Price of Six Different Airlines Days Before Departure



As analyzers, our suggestion for the customers is to avoid the period that the airline ticket prices change dramatically, but instead purchase them days before the price increase happens. The customers should attempt their best to avoid the price increase period and purchase tickets when their prices are still relatively low as much as possible. Our final recommendation given the data analysis we obtained is to suggest customers to buy tickets 20 to 50 days in advance to maximize the savings.

Holt Exponential Smoothing Forecast Model

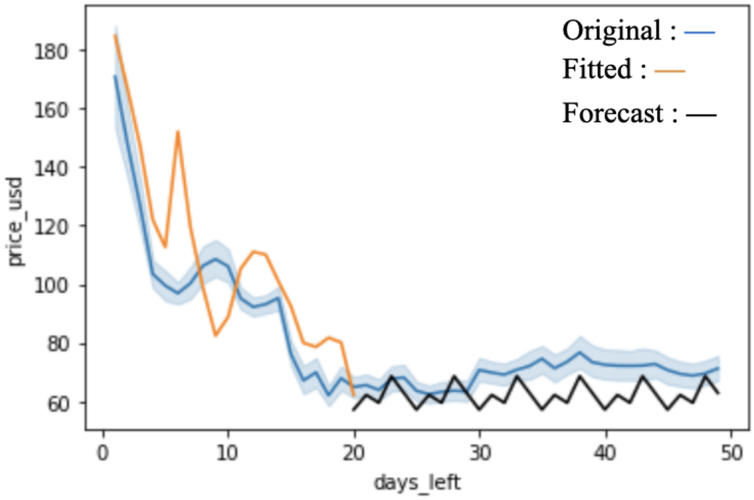
In order for our customers to better understand our prescriptive question, our team uses the method of Holt Exponential Smoothing to demonstrate the expected trend of the airline ticket prices as the day progresses further from the departure date.

According to Figure 3a, since all six airline companies contain a similar trend of prices, we will choose one of them (ex. Air SpiceJet) as the representative and fits the model on this specific dataset. To further prove our hypothesis of suggesting the price experiences its minimum value in the range from 20 to 50 days before the departure, we treat the range from 0 to 20 days before departure as our training data, and the range from 20 to 50 days before departure as our test data. We will first fit a model among training data that simulates the original data as closely as possible, then perform a forecast on the test data based on our fitting model. In the original data, there are several different prices corresponding to each day. This increases the difficulty in fitting an exponential model on the training data. Instead, we start by cleaning the data and take the average of prices on each day as our input for fitted values of the training data.

We then fitted a Holt Exponential Smoothing model to the training data which closely simulates the original data. The reason why we choose this model over other methods is because this model provides the most accurate forecasting compared to other methods. It produces the least sum of squared errors. The forecast of the price of tickets 20 to 50 days before departure is generated based on the fitting model, as shown in Figure 3b.

Figure 3b

Fitted, Forecasted, and Original Data of Prices 50 Days Before Departure



As we can see, based on the previous fitted trend, the forecast indicates that the price of airline tickets stays in the low value range from 20 to 50 days. The fitted trend model forecasts the test data to shape closely with our original data. This further strengthens our answer to the prescriptive answer, that 20 to 50 days before departure is the period that the ticket prices stay the lowest.

In conclusion, we use both our observations from data visualizations and analytics from Holt Exponential Smoothing model to give our customers information about the best time to purchase airline tickets before departure that maximize their savings, which is 20 to 50 days before the departure.

Conclusion

In this recommendation report, our team provided individuals looking to fly domestically throughout India with the optimal times to purchase airline tickets to maximize their savings. The three main questions mentioned throughout this report were key to finding a solution to this overarching query. From our predictive question, which was concerned with the variation of ticket prices, both graphs demonstrated that ticket prices typically change more dramatically the closer they get to the plane's departure date. Also, as this departure date draws nearer, ticket prices tend to increase on average. We confirmed this by using a linear regression analysis, which showed a high correlation between ticket price variation and how close the departure date is. Now that we understand the variation and overall trend of airline ticket prices as their departure date approaches, we can delve deeper to figure out the specifics. In our prescriptive question, covering the specific optimal times to purchase these airline tickets, we can see from the line graph that Indian airline ticket prices all seem to follow a relatively similar pattern. As we demonstrated in this graph, and as shown through the Holt Exponential Smoothing Forecast Model, the forecasted optimal time to buy tickets is between twenty and fifty days before the departure date, as ticket prices are significantly cheaper than they will be closer to the date, without much variation. Right after this twenty day mark, there tends to be a sharp increase, followed by variable increases up until the departure date, which is shown through the correlation graph, and confirmed by the linear regression model above. The combination of Covid-19 putting airline companies into debt as well as Russian sanctions raising gas prices has made airline ticket prices a serious issue for many people in India, but we hope that this recommendation report can help educate those on the optimal times to buy airline tickets, so they can save as much money as possible during this difficult time.

References

Covid-19 impact: India's airlines suffered loss of rs 19,564 cr in 20-21, says govt. Business Today. (2021, December 6). Retrieved May 8, 2022, from https://www.businesstoday.in/industry/aviation/story/covid-19-impact-indias-airlines-suffered-loss-of-rs-19564-cr-in-20-21-says-govt-314578-2021-12-06

FDA. (2021, August 23). *FDA approves first COVID-19 vaccine*. U.S. Food and Drug Administration. Retrieved May 8, 2022, from https://www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine#:~:text=Since%20Dec.%2011%2C%202020%2C,age%20on%20May%2010%2C%202021

Boye, B. A. (2021, January 28). *Covid-19 vaccine launch in India*. UNICEF India. Retrieved May 8, 2022, from https://www.unicef.org/india/stories/covid-19-vaccine-launch-india

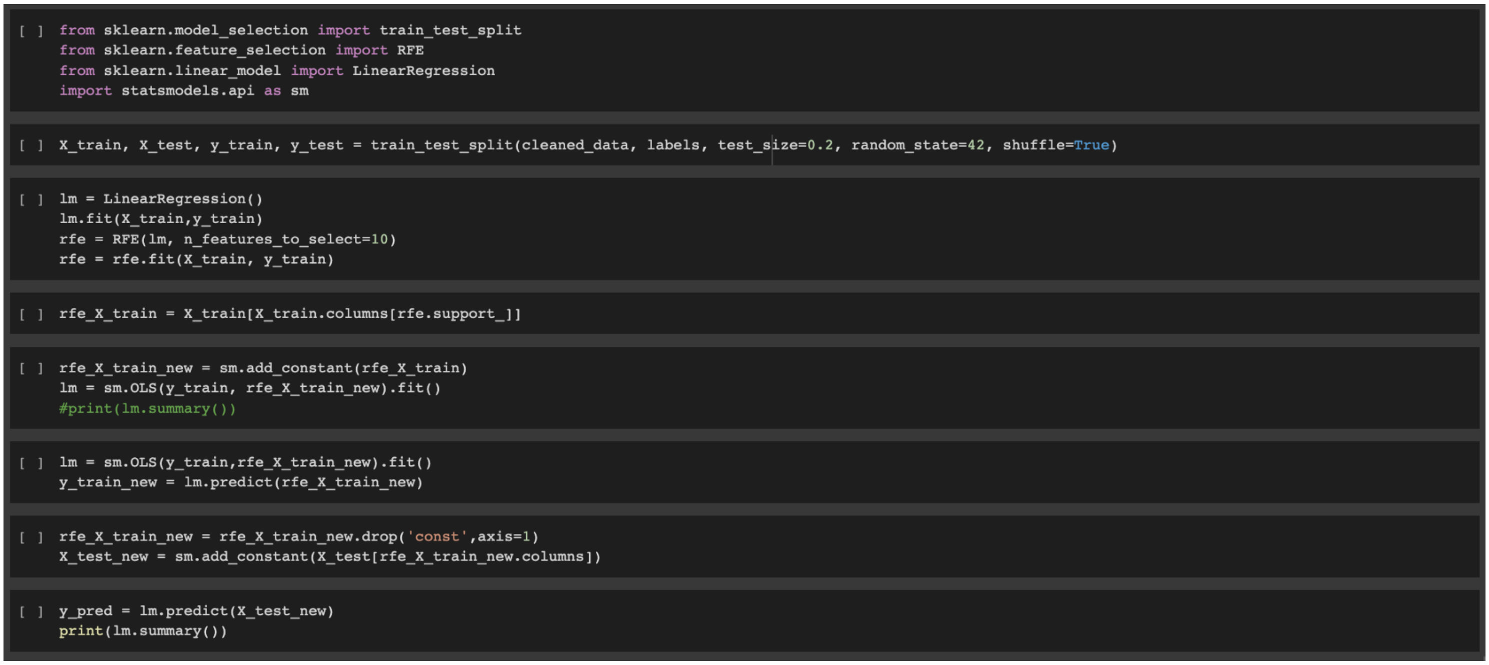
Livemint. (2021, August 29). *Covid-19 vaccine: 1st Batch of Covaxin released from Bharat Biotech's New Plant*. mint. Retrieved May 8, 2022, from https://www.livemint.com/news/india/covid19-vaccine-1st-batch-of-covaxin-released-from-bharat-biotech-s-new-plant-in-gujarat-11630222512270.html

Mukherji, B. (2022, March 28). *Ticket prices soar as India resumes international flights*. Fortune. Retrieved May 8, 2022, from https://fortune.com/2022/03/28/india-travel-restrictions-flight-ban-international-airfare-ticket-prices-cost-oil-jet-fuel/

Bathwal, S. (2022, February 25). *Flight price prediction*. Kaggle. Retrieved May 8, 2022, from https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction

Appendix A

Linear Regression Prediction Model Code



```
[ ] from sklearn.model_selection import train_test_split
    from sklearn.feature_selection import RFE
    from sklearn.linear_model import LinearRegression
    import statsmodels.api as sm

[ ] X_train, X_test, y_train, y_test = train_test_split(cleaned_data, labels, test_size=0.2, random_state=42, shuffle=True)

[ ] lm = LinearRegression()
    lm.fit(X_train,y_train)
    rfe = RFE(lm, n_features_to_select=10)
    rfe = rfe.fit(X_train, y_train)

[ ] rfe_X_train = X_train[X_train.columns[rfe.support_]]

[ ] rfe_X_train_new = sm.add_constant(rfe_X_train)
    lm = sm.OLS(y_train, rfe_X_train_new).fit()
    #print(lm.summary())

[ ] lm = sm.OLS(y_train,rfe_X_train_new).fit()
    y_train_new = lm.predict(rfe_X_train_new)

[ ] rfe_X_train_new = rfe_X_train_new.drop('const',axis=1)
    X_test_new = sm.add_constant(X_test[rfe_X_train_new.columns])

[ ] y_pred = lm.predict(X_test_new)
    print(lm.summary())
```

Model Summary

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.909
Model:                            OLS   Adj. R-squared:                  0.909
Method:                 Least Squares   F-statistic:                 2.402e+05
Date:                Tue, 15 Mar 2022   Prob (F-statistic):               0.00
Time:                        08:26:09   Log-Likelihood:             3.5108e+05
No. Observations:              240122   AIC:                        -7.021e+05
Df Residuals:                  240111   BIC:                        -7.020e+05
Df Model:                          10
Covariance Type:            nonrobust
------------------------------------------------------------------------------
```