kaggle

KaggleX- Showcase –Flight Price Prediction

By- Shruti Pandit

Flight Price Prediction System

Shruti Pandit

Background

- I am working as a Senior Director in a Corporate Travel Software Company
- I am also pursuing MSc. In AIML from Liverpool John Moores University
- I have done certifications in AIML,MLOps,GCP,AWS,Blockchain and Agile

Project Definition

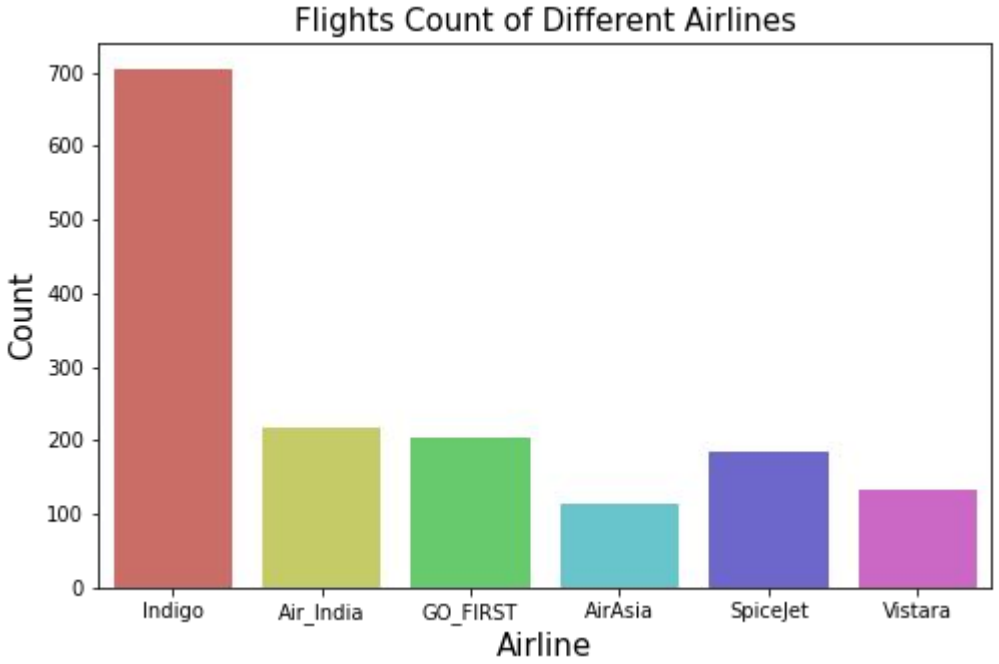
- An airline is a company that provides air transport services for traveling passengers and freight. Airlines use aircraft to supply these services and may form partnerships or alliances with other airlines for codeshare agreements, in which they both offer and operate the same flight. Generally, airline companies are recognized with an air operating certificate or license issued by a governmental aviation body. Airlines may be scheduled or charter operators.
- Airlines assign prices to their services in an attempt to maximize profitability. The pricing of airline tickets has become increasingly complicated over the years and is now largely determined by computerized yield management systems.
- The price of an Airline Ticket is affected by a number of factors, such as flight duration, days left for departure, arrival time and departure time etc. Airline organizations may diminish the cost at the time they need to build the market and at the time when the tickets are less accessible. They may maximize the costs. The price may rely upon different factors. Each factor has its own proprietary rules and algorithms to set the price accordingly. Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) makes it possible to infer such rules and model the price variation.

Project Definition

- Techniques Applied-
  - EDA
  - Linear Regression
  - Decision Tree
  - KNN
  - Extra Trees Regressor
  - XG Boost Regressor
  - Ridge Regressor
  - Lasso Regressor
  - Random Forest Regressor
  - Bagging Regressor
  - Gradient Boosting Regressor

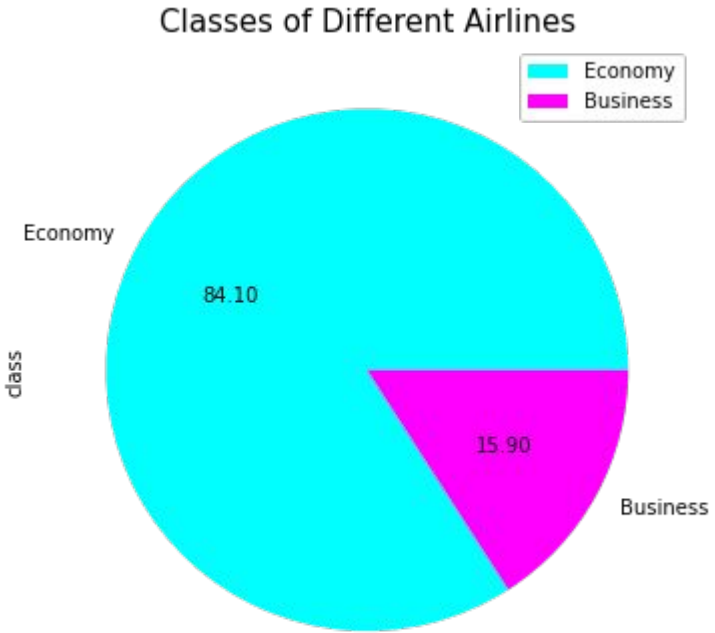
Exploratory Data Analysis

- Indigo appears to be the most popular Airline



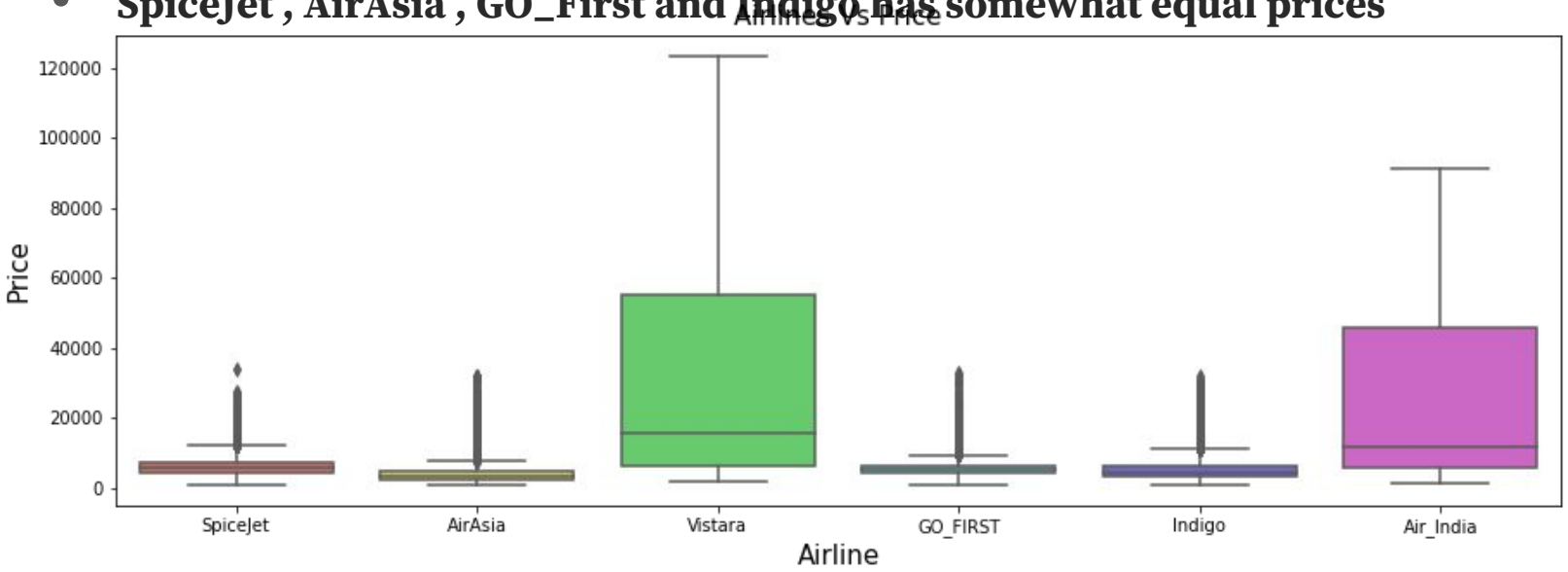
Exploratory Data Analysis

- Most Airlines have Economy Class as the Common Class



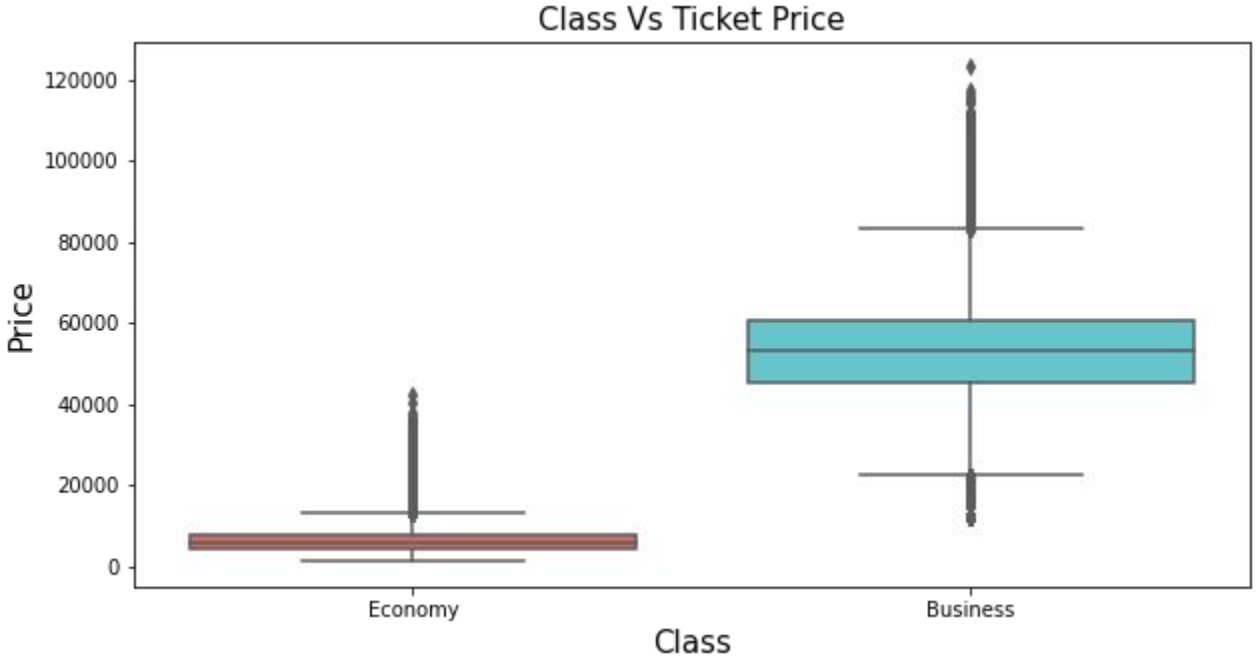
Exploratory Data Analysis

- **Does price vary with Airlines?**
- **As we can see Vistara has Maximum Price range**
- **Vistara and Air_India Airlines Have Maximum Price when compared to Others**
- **SpiceJet , AirAsia , GO_First and Indigo has somewhat equal prices**



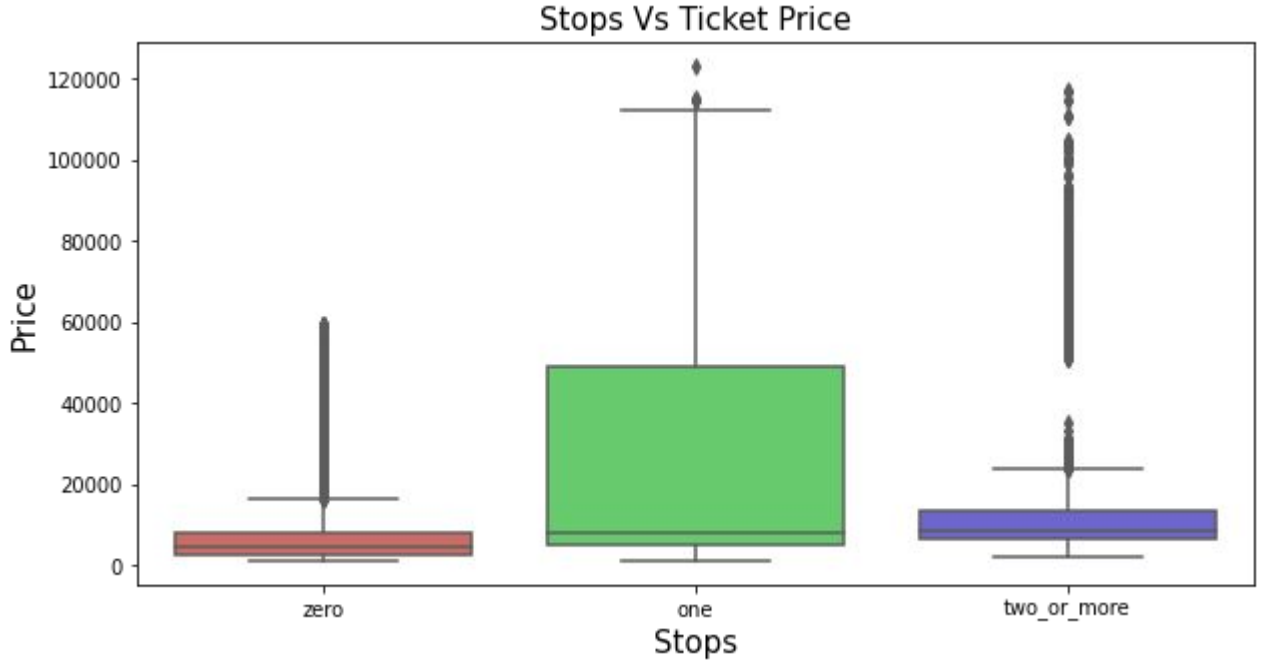
Exploratory Data Analysis

- **How Does the Ticket Price vary between Economy and Business Class?**
- **Ticket Price is Maximum for Business Class When compared to Economy Class**

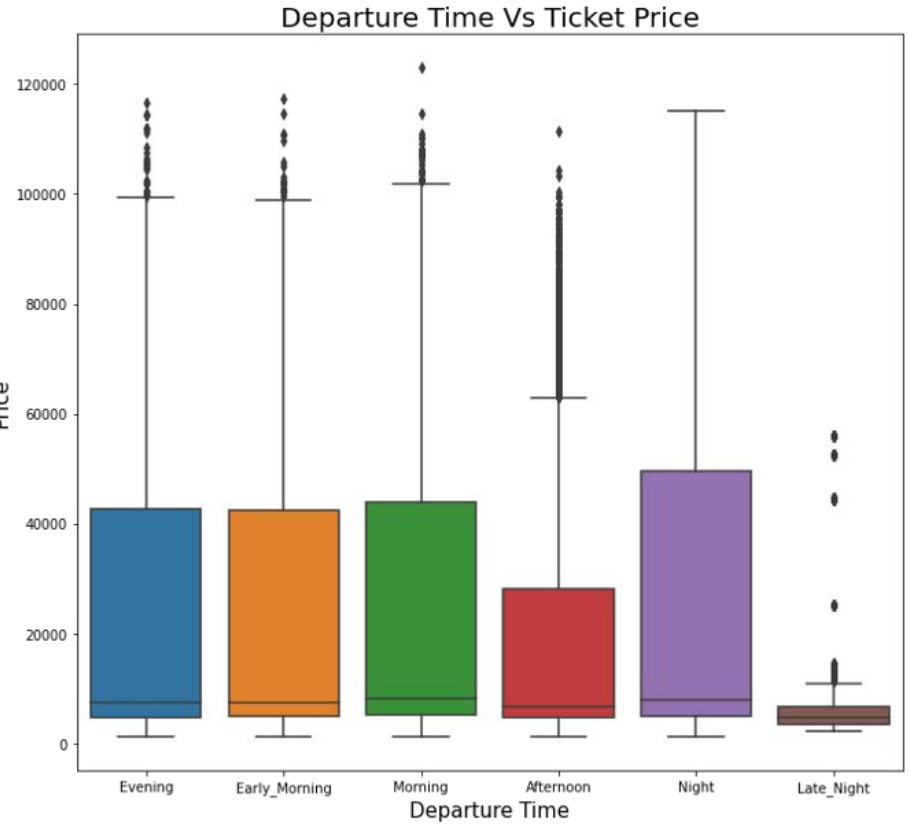


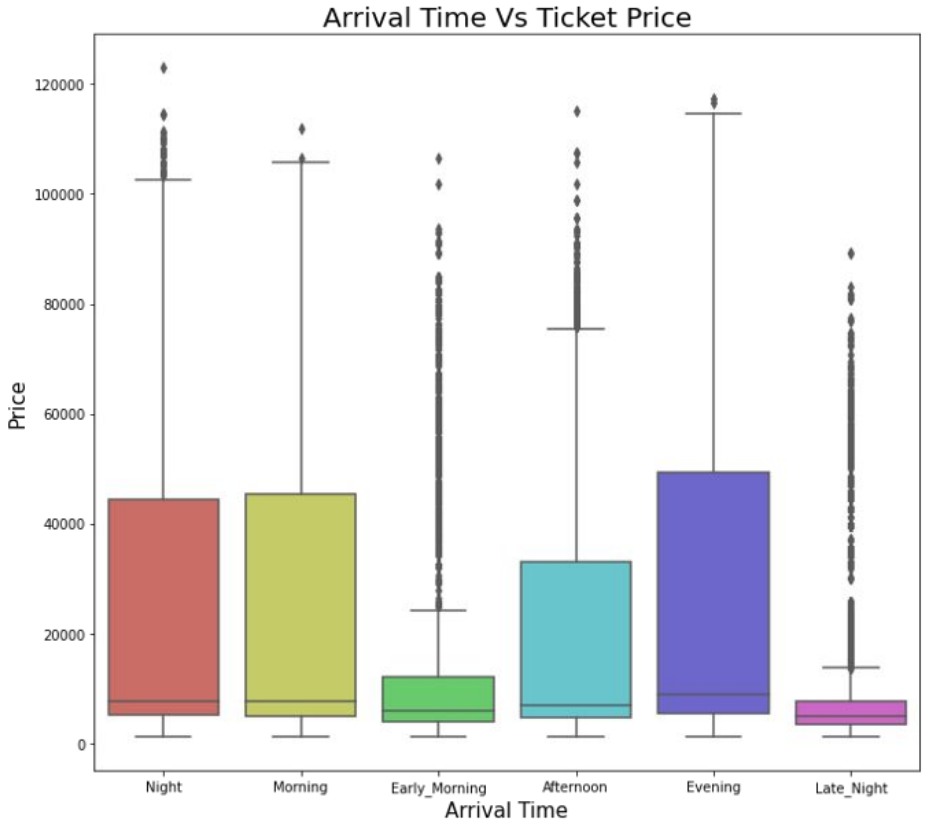
Exploratory Data Analysis

- **How Does the Ticket Price vary with the number of stops of a Flight?**
- **Flights having one stop has maximum ticket price**



Exploratory Data Analysis





Exploratory Data Analysis

- **How the Ticket Price change based on the Departure Time and Arrival Time?**

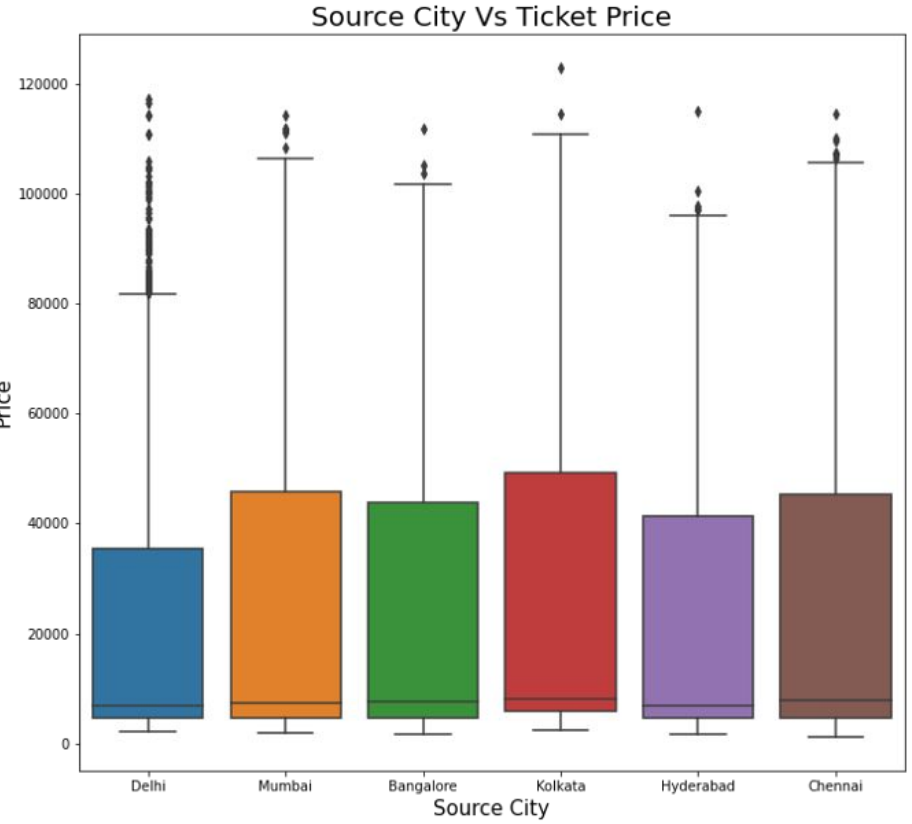
  1. Departure Time Vs Ticket Price

  - **Ticket Price is More for the Flights when the Departure Time is at Night**
  - **Ticket Price is almost equal for flights Having Departure time at Early_morning , Morning and Evening**
  - **Ticket Price is Low for the Flights Having Departure Time at Late_night**

- **2. Arrival Time Vs Ticket Price**

  - **Ticket Price is More for the Flights when the Arrival Time is at Evening**
  - **Ticket Price is almost equal for flights Having Arrival time is at Morning and Night**
  - **Ticket Price is Low for the Flights Having Arrival Time at Late_night as same as Departure Time**

Exploratory Data Analysis



Exploratory Data Analysis

- **How the price changes with change in Source city and Destination city?**

  1. Source City Vs Ticket Price

  - **Ticket Price is More for the Flights whose Source City is Kolkata**
  - **Ticket Price is almost equal for flights Having Source Cities as Mumbai and chennai , Hyderabad and Bangalore**
  - **Ticket Price is Low for the Flights Having Source City as Delhi**

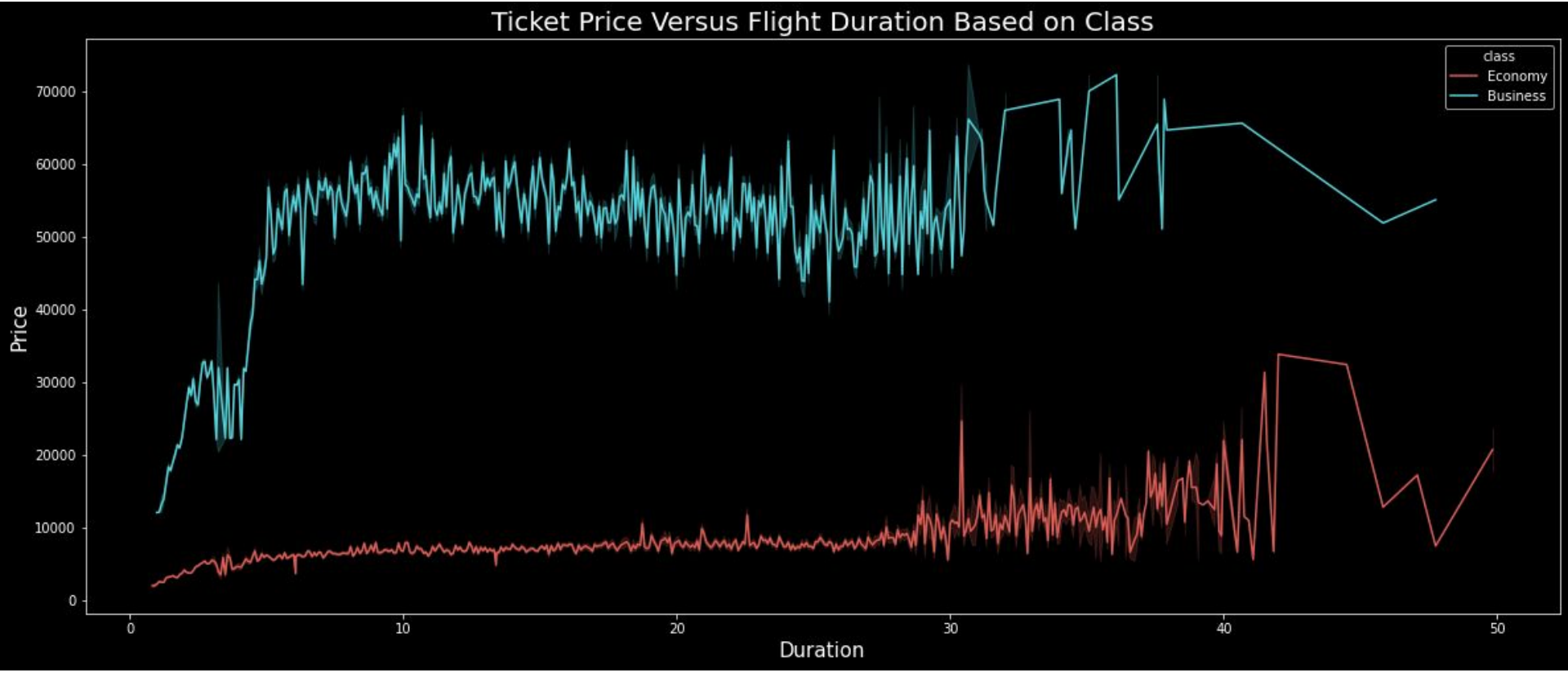
- **2. Destination City Vs Ticket Price**

  - **Ticket Price is More for the Flights whose Destination City is kolkata and Chennai**
  - **Ticket Price is almost equal for flights Having Destination Cities as Mumbai and Bangalore**
  - **Ticket Price is Low for the Flights Having Destination City as Delhi**

Exploratory Data Analysis

How Price Varies with the Flight Duration Based on Class?

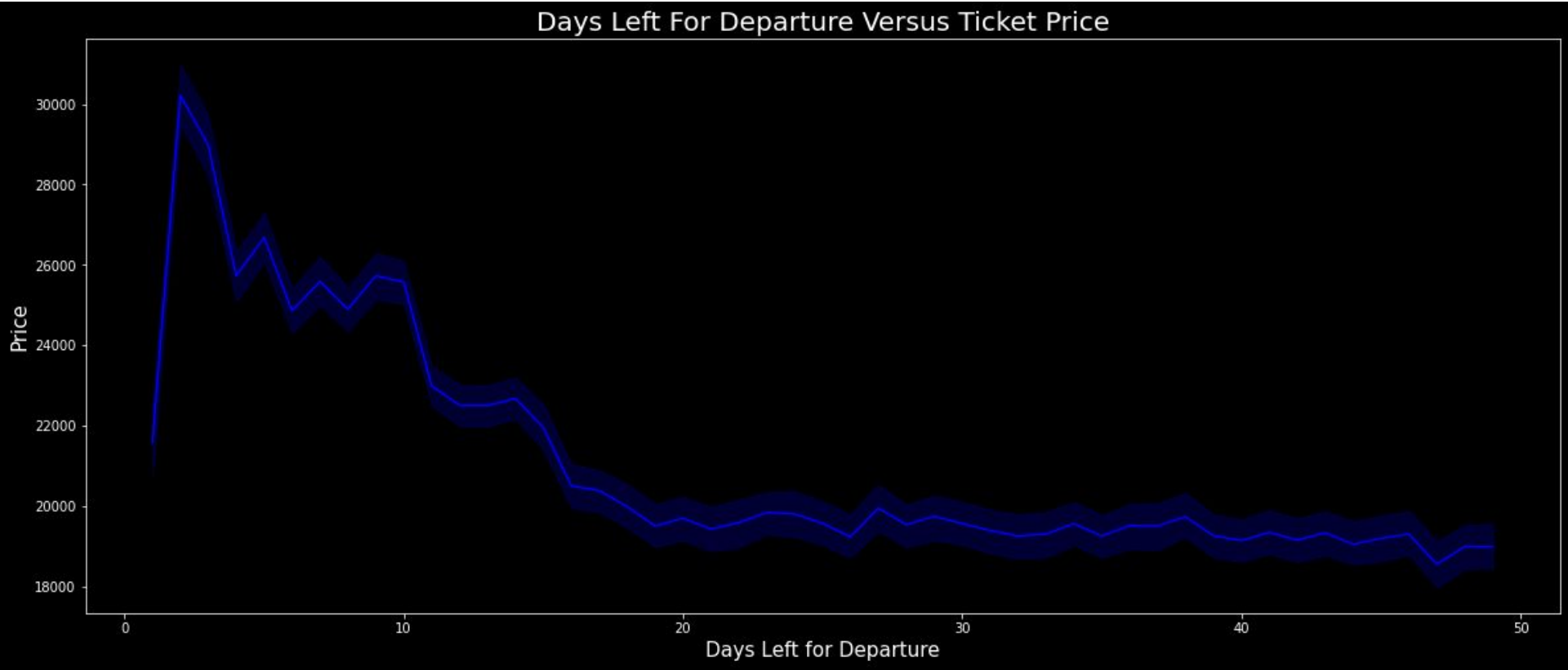
- **With increase in Duration, the Ticket Price is also Increases In both the Economy and Business classes**



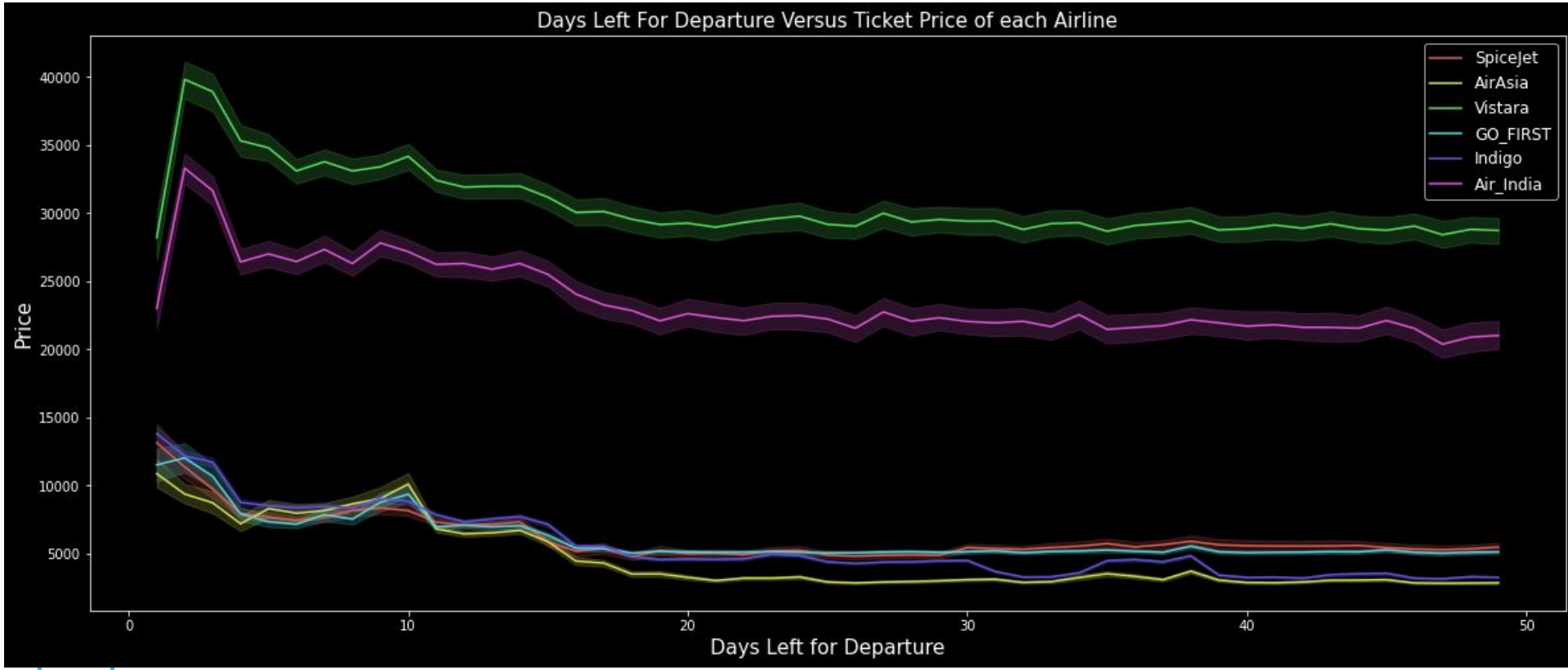
Exploratory Data Analysis

How does the price affected on the days left for Departure?

- **As we can see when compared to others when there are two days remaining for departure then the Ticket Price is very High for all airlines**



Exploratory Data Analysis

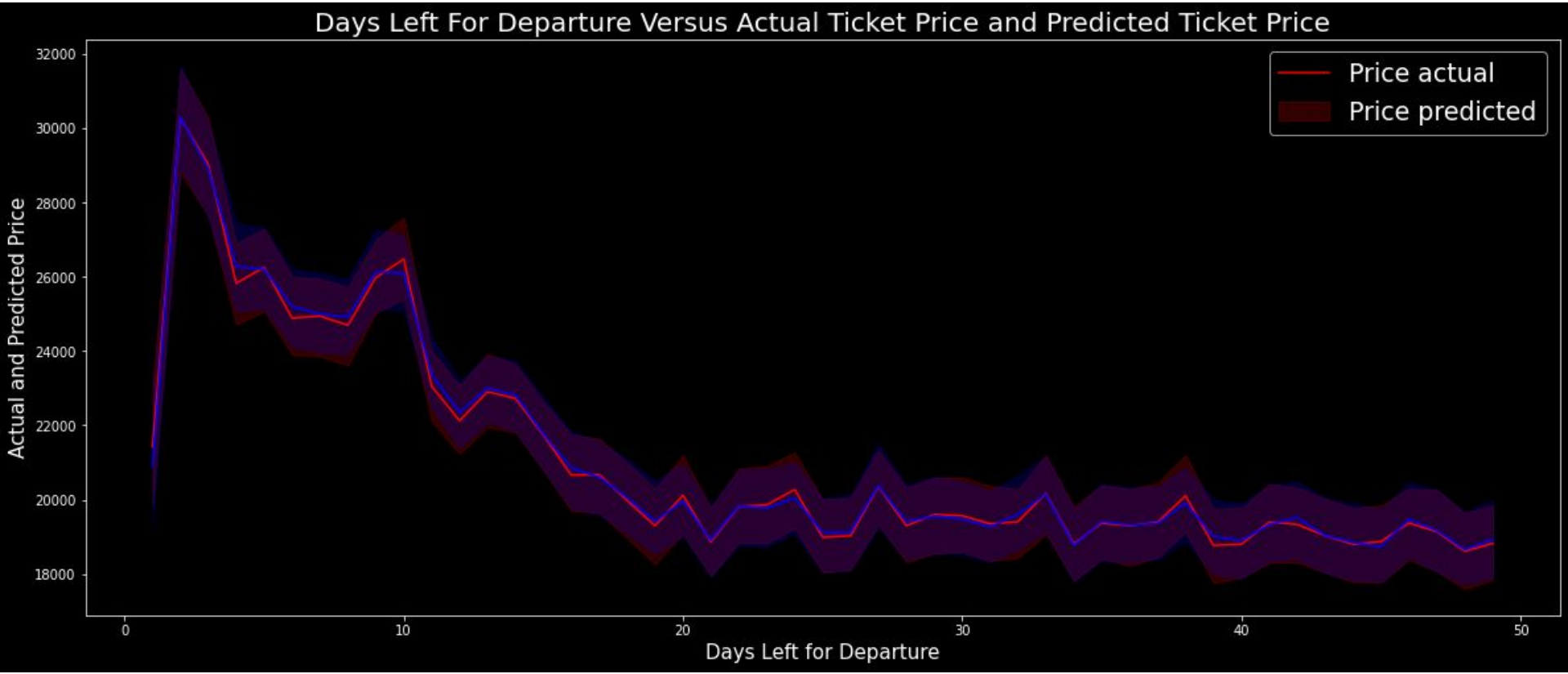


Results from various ML Models

From the Above Results, The Top 3 Models by comparing Errors , Adj_R_Square and R2_Score values are
1. ExtraTreesRegressor 2. RandomForestRegressor 3. Bagging Regressor

Sno	Model_Name	Adj_R_Square	Mean_Absolute_Error_MAE	Root_Mean_Squared_Error_RMSE	Mean_Absolute_Percentage_Error_MAPE	Mean_Squared_Error_MSE	Root_Mean_Squared_Log_Error_RMSLE	R2_score
0	ExtraTreesRegressor	0.984695	1148.84353	2806.484238	7.651255	7.88E+06	7.939688	0.984696
1	RandomForestRegressor	0.984127	1169.53345	2858.139321	7.925107	8.17E+06	7.957926	0.984128
2	BaggingRegressor	0.982889	1202.117007	2967.460291	8.284835	8.81E+06	7.995462	0.98289
3	XGBRegressor	0.977947	1866.030982	3368.903685	14.42194	1.13E+07	8.122343	0.977948
4	DecisionTreeRegressor	0.973256	1259.457598	3709.964181	8.587642	1.38E+07	8.218778	0.973257
5	KNeighborsRegressor	0.971657	1853.746812	3819.27082	11.109409	1.46E+07	8.247815	0.971658
6	GradientBoostingRegressor	0.956823	2808.067836	4713.929007	20.672946	2.22E+07	8.458277	0.956824
7	LinearRegression	0.904653	4630.295614	7005.015436	43.888567	4.91E+07	8.854382	0.904656
8	Ridge Regression	0.904653	4630.313301	7005.011583	43.888754	4.91E+07	8.854381	0.904656
9	Lasso Regression	0.904653	4630.179207	7005.006183	43.885459	4.91E+07	8.85438	0.904656

Making Predictions on the Test Data with Extra Regressor Algorithm



Making Predictions on the Test Data with Extra Regressor Algorithm


kaggle