Flip Robo Technologies

Flight Price Prediction Project

Submitted by: John Tojo

Data Science Intern at Flip Robo Technologies

ACKNOWLEDGMENT

It gives me immense pleasure to deliver this report. Working on this project was a great learning experience that helped me attain in-depth knowledge on data analysis process. Flip Robo Technologies (Bangalore) provided all of the necessary information and datasets, required for the completion of the project. I express my gratitude to my SME, **Gulshana Chaudhary**, for providing the dataset and directions for carrying out the case study procedure.

INTROUDCTION

Flight price prediction is a critical task for both airlines and consumers, as it helps airlines to optimize their pricing strategies, and consumers to make informed decisions when booking a flight. The airline industry is highly competitive, and airlines are always adjusting prices based on various factors such as time of purchase, flight occupancy, and duration. These adjustments can happen on a daily or even hourly basis, making it challenging for consumers to find the best deals. Therefore, to stay competitive and make informed pricing decisions, airlines use various techniques and models to predict flight prices.

These techniques and models take into account various factors such as route, departure and arrival times, number of stops, and duration of the flight, to make accurate predictions about the prices of flights. This allows airlines to optimize their pricing strategies by adjusting prices based on supply and demand, and by targeting specific groups of consumers. For example, airlines can offer lower prices to consumers who book flights well in advance, or offer higher prices to last-minute bookers.

Problem Statement

The problem of flight price prediction is a complex one, as the prices of flights are affected by various factors such as time of purchase, route distance, and flight occupancy. Airlines often adjust prices based on these factors in order to maximize revenue and fill seats on flights. For example, prices may be lower for early bookings, and higher for last-minute bookings or flights that are filling up quickly. Additionally, prices may vary depending on the route, with longer or more popular routes typically having higher prices. In order to stay competitive and make informed pricing decisions, airlines use various techniques and models to predict flight prices. The goal of this project is to develop a model that can accurately predict flight prices by analyzing historical data and other relevant factors. The project aims to:

- Collect and clean a dataset of flight fare data and other relevant features such as time of purchase, route distance, and airline carrier.
- Analyze the data to identify trends and patterns in flight prices.
- Implement and evaluate different machine learning models to predict flight prices.
- Optimize the model by fine-tuning hyperparameters and feature selection.

This project will provide valuable insights into the process of flight price prediction and the factors that influence it, which can be used by airlines to optimize their pricing strategies and by consumers to make informed decisions when booking flights.

.DATA SOURCES AND THEIR FORMATS

The data used in this project was scraped from the travel website MakeMyTrip.com, specifically for flights between the 8th and 11th day of the month. The scraped data contains the following columns: 'Airline', 'Travel_date', 'From', 'To', 'Departure_time', 'Arrival_time', 'Duration_minutes', 'Stops' and 'Price'. Each column provides a different piece of information that can be used to analyze and predict flight prices.

The 'Airline' column provides information about the airline company that is operating the flight. This information can be used to analyze the prices of different airlines and how they vary. For example, prices may be higher for flights operated by a well-known, premium airline compared to a budget airline.

The 'Travel_date' column provides information about the date of the flight, which can be used to analyze how prices vary over time. For example, prices may be higher during peak travel seasons and lower during off-peak seasons.

The 'From' and 'To' columns provide information about the departure and arrival cities or airports of the flight. This information can be used to analyze how prices vary depending on the route. For example, prices may be higher for flights between popular tourist destinations compared to less popular destinations.

The 'Departure_time' and 'Arrival_time' columns provide information about the departure and arrival times of the flight, and the duration of the flight. This information can be used to analyze how prices vary depending on the time of day or the duration of the flight.

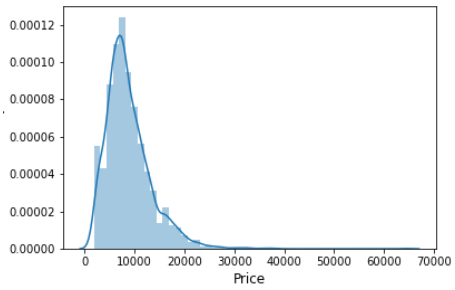
The 'Stops' column provides information about the number of stops on the flight. This information can be used to analyze how prices vary depending on the number of stops. For example, prices may be higher for flights with more stops.

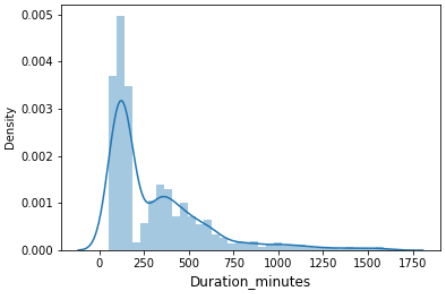
The 'Price' column is the target variable, which provides information about the price of the flight. This information is used to train and evaluate the model, and the ultimate goal of the project is to develop a model that can accurately predict the price of a flight based on the information provided in the other columns.

Exploratory Data Analysis (EDA)

1. Using df.info() to check for information about the data, such as the data types and number of non-null values for each column. This helped identify any potential issues or errors in the data.
2. Converting the 'travel_date' column to a datetime data type and extracting the day of the month. A new column called 'day' was added and the original 'travel_date' column was dropped. This helped analyze how prices vary over time.
3. Checking for and removing any duplicate records in the dataset. In this case, 6 records were duplicate.
4. Checking for unique data present in each column. This helped identify any errors or inconsistencies in the data.
5. Checking for null values and found 4, but since it existed on the same row so dropped it. This helped ensure that the data is clean and complete.
6. Creating new columns 'Arrival_time_hour' and 'Arrival_time_min' from the 'Arrival_time' column, and dropping the original 'Arrival_time' column. Similarly, did the same with the 'Departure_time' column as well. This helped analyze how prices vary depending on the time of day.

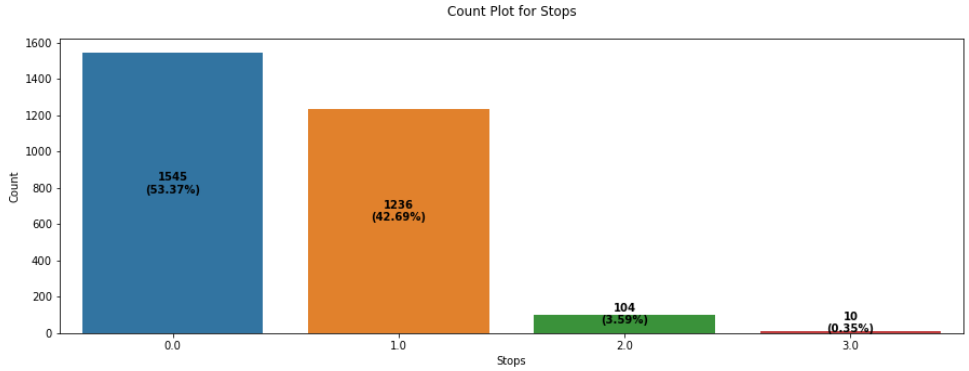
Visualization

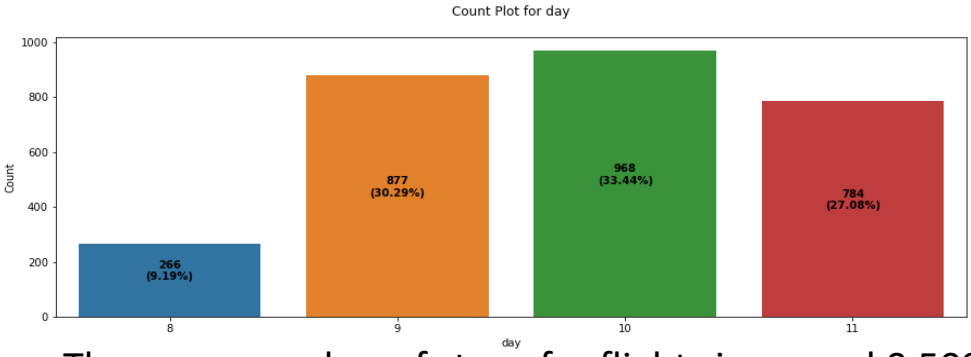




- The mean duration of the flights is around 313.8 minutes, with a standard deviation of 282.88 minutes
- The minimum duration of the flight is 55 minutes and the maximum duration of the flight is 1635 minutes

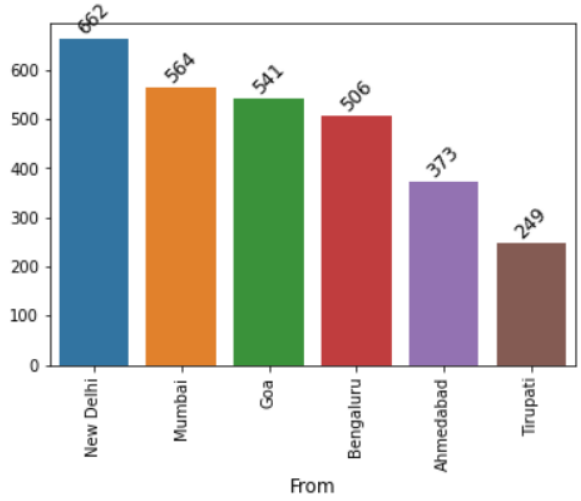
- The mean price of the flights is around 8825.76, with a standard deviation of 4651.56
- The minimum price of the flight is 1899 and the maximum price of the flight is 64163
- The duration of flight and prices of flight have wide range of variation
- presence of outliers

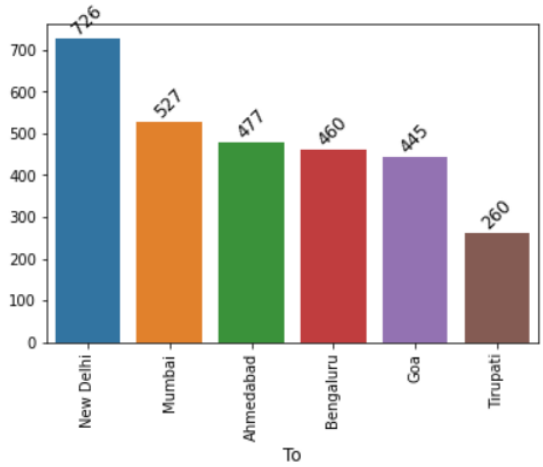




- The mean number of stops for flights is around 0.509154, with a standard deviation of 0.585327.
- The minimum number of stops for a flight is 0 and the maximum number of stops for a flight is 3.
- most of the flights are non stop followed by 1 stop
- The mean day of flight is around 9.784111, with a standard deviation of 0.946033.
- The minimum day of flight is 8 and the maximum day of flight is 11.

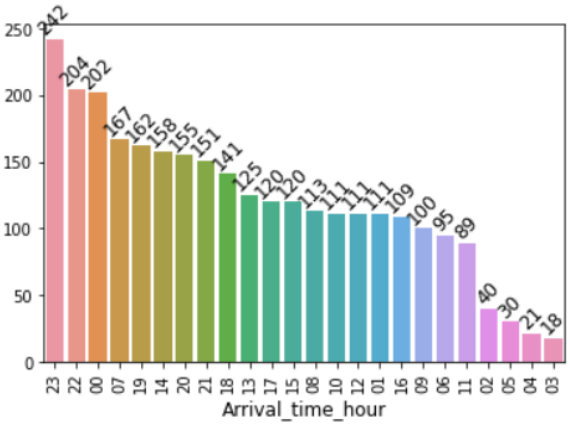
- most of the data is accumulated on day 10 followed by day 9

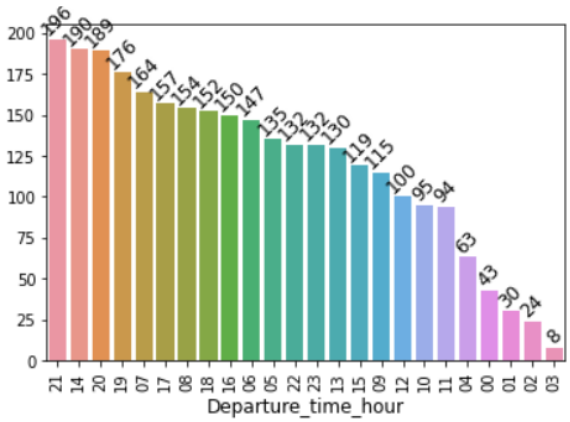




based on To and From inference

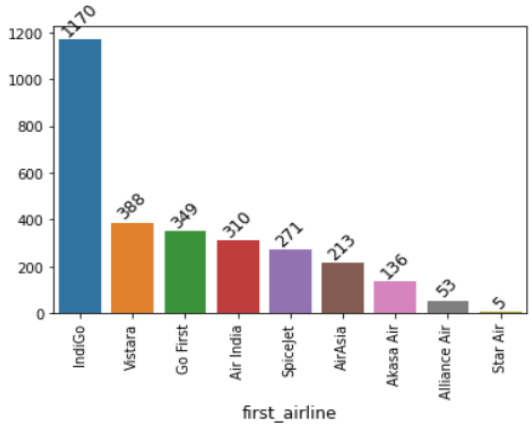
- New Delhi is a popular destination: The fact that most people are traveling to New Delhi suggests that it is a popular destination among travelers.
- Tirupati is not as popular a destination: The fact that least people want to travel to Tirupati indicates that it is not as popular a destination as New Delhi.
- New Delhi is a common starting point: The fact that most people are traveling from New Delhi suggests that it is a common starting point for many travelers.
- Tirupati may be a less-populated area: The fact that least people are traveling from Tirupati suggests that it may be a less-populated area compared to New Delhi.
- New Delhi may be a more accessible location: The fact that most people are traveling from and to New Delhi suggests that it may be more easily accessible to people.

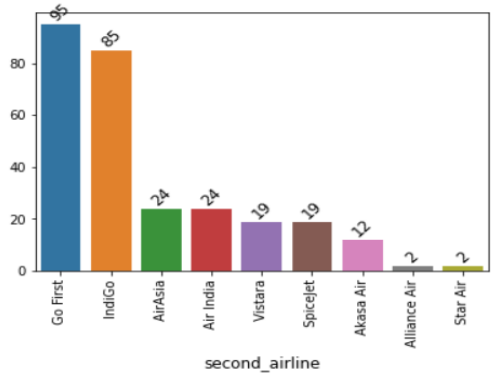




based on arrival and departure times

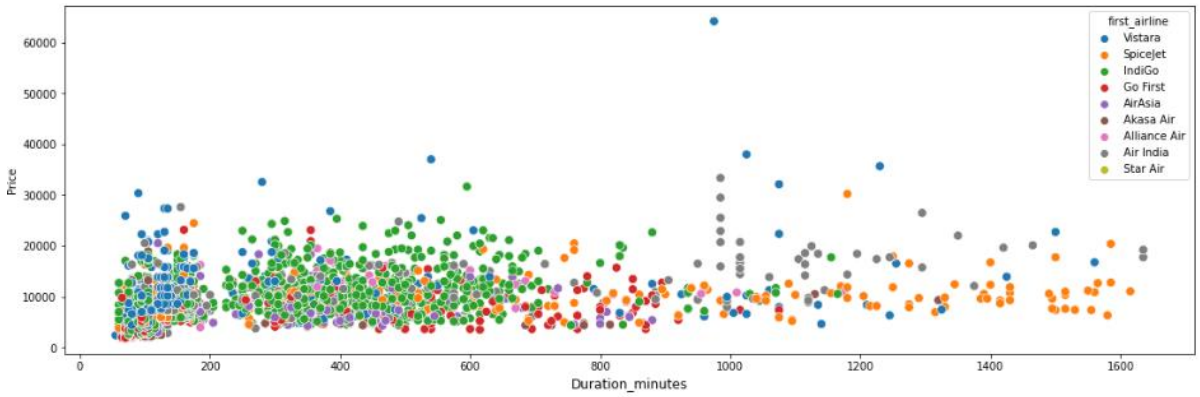
- There are more flights scheduled to arrive at 23 hr than at 03 hr.
- There may be a higher demand for flights at 23 hr than at 03 hr.
- The airport may have more capacity to handle flights at 23 hr than at 03 hr.
- There could be operational reasons such as less congestion or traffic at 23 hr than at 03 hr, so flights are scheduled to arrive at that time.
- there is higher arrival of flight at 21 hr than at 03 hr
- it can be seen that most of the flights depart and arrive at the latter half of the day
- The company has chosen that time as it may be cheaper to operate during those hours.





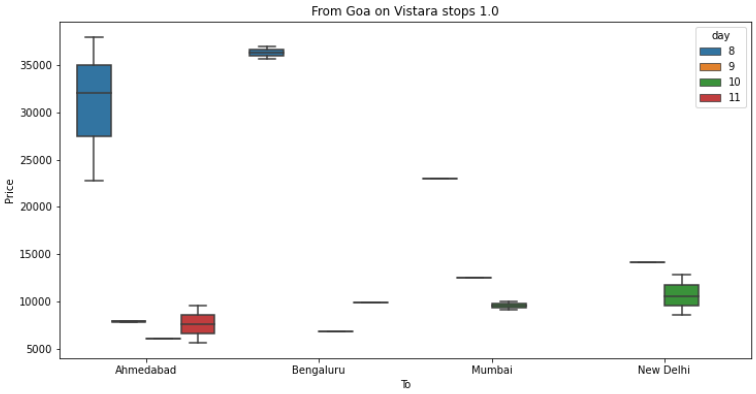
based on first_airline and second_airline(connection airline if there are stops)

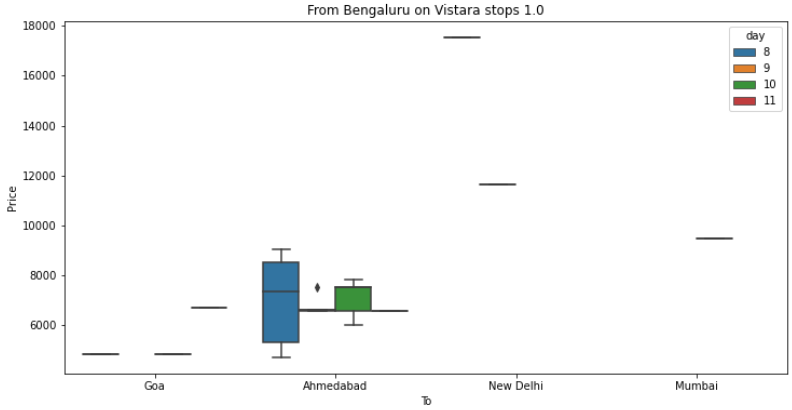
- Indigo has a better reputation and customer satisfaction than StarAir.
- People may find Indigo flights more affordable or convenient than StarAir flights.
- Gofirst may have a better reputation or provide better service than Star Air for connecting flights.
- People may find Gofirst flights more affordable or convenient than StarAir flights for connecting flights.

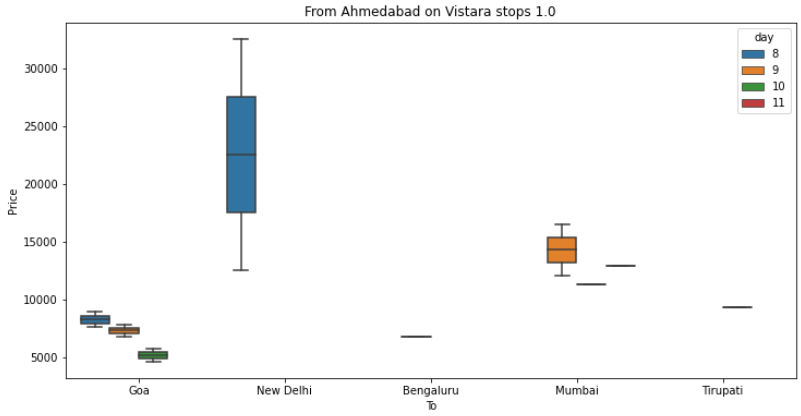


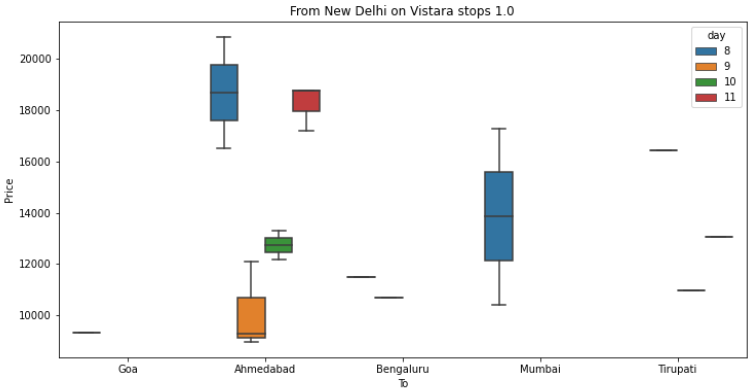
- Vistara is faster but more expensive than other airlines.
- Indigo offers more options for non-stop and connecting flights, and its connecting flights are comparatively cheaper.
- Air India flights are more expensive and take longer to reach the destination.
- Spice Jet has longer duration flights indicating multiple stops and is cheaper than the other airlines.

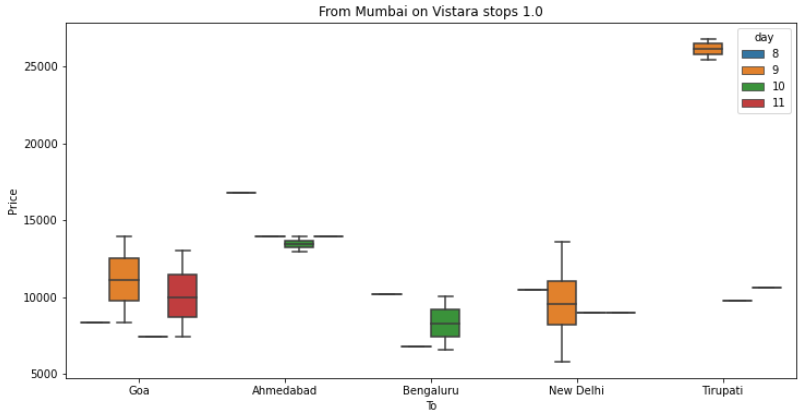
- Indigo may have a larger market share for domestic flights than other airlines.
- Vistara may be considered as a higher quality airline than the other options.
- Air India may be preferred for international flights rather than domestic flights.
- Spice jet may be used for low-budget or travel for time constraints, but it will take longer time to reach the destination.

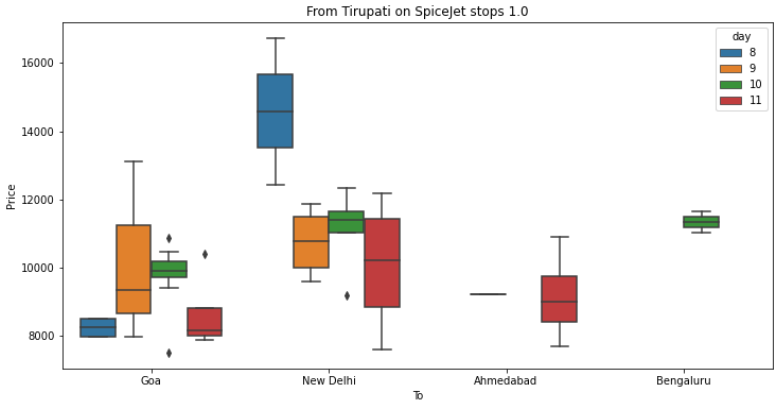

From Goa on Vistara stops 1.0
day
8
9
10
11
Price
35000
30000
25000
20000
15000
10000
5000
Ahmedabad
Bengaluru
Mumbai
New Delhi
To

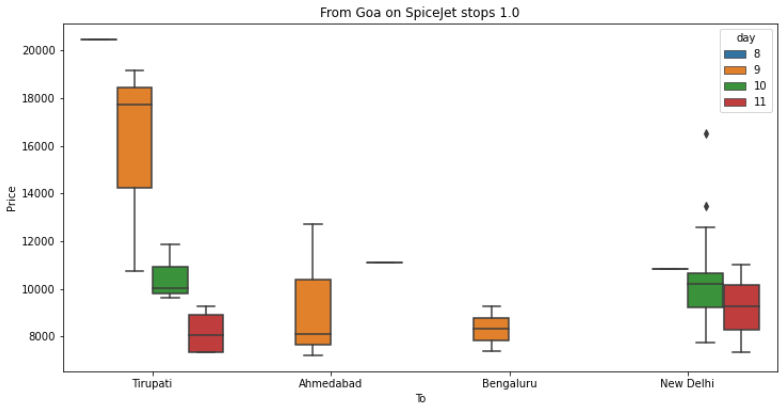

From Bengaluru on Vistara stops 1.0
day
8
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
Goa
Ahmedabad
New Delhi
Mumbai
To

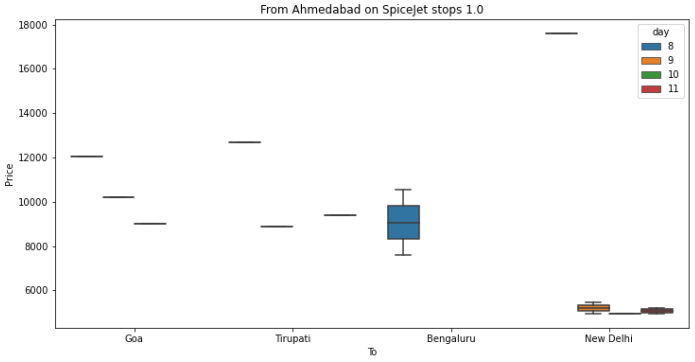

From Ahmedabad on Vistara stops 1.0
day
8
9
10
11
Price
30000
25000
20000
15000
10000
5000
Goa
New Delhi
Bengaluru
Mumbai
Tirupati
To

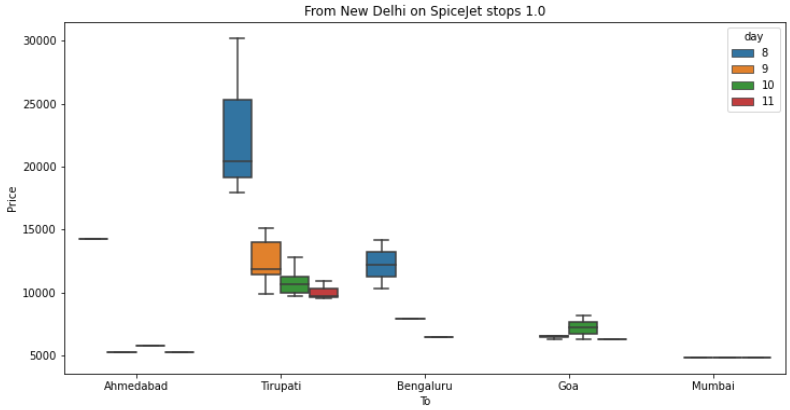

From New Delhi on Vistara stops 1.0
day
8
9
10
11
Price
20000
18000
16000
14000
12000
10000
Goa
Ahmedabad
Bengaluru
Mumbai
Tirupati
To


From Mumbai on Vistara stops 1.0
day
8
9
10
11
Price
25000
20000
15000
10000
5000
Goa
Ahmedabad
Bengaluru
New Delhi
Tirupati
To

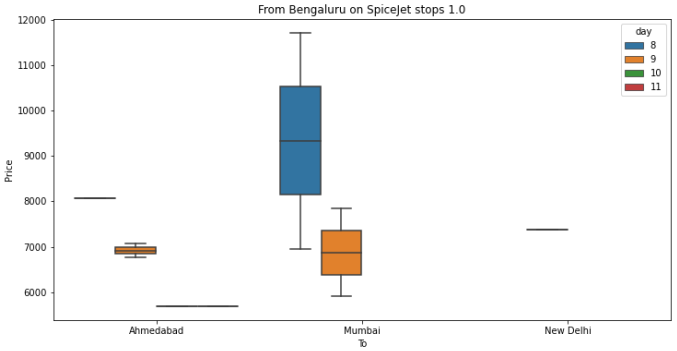

From Tirupati on SpiceJet stops 1.0
day
8
9
10
11
Price
16000
14000
12000
10000
8000
Goa
New Delhi
Ahmedabad
Bengaluru
To

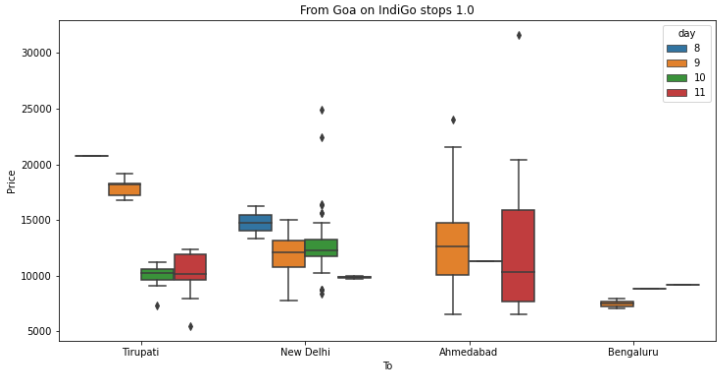

From Goa on SpiceJet stops 1.0
day
8
9
10
11
Price
20000
18000
16000
14000
12000
10000
8000
Tirupati
Ahmedabad
Bengaluru
New Delhi
To

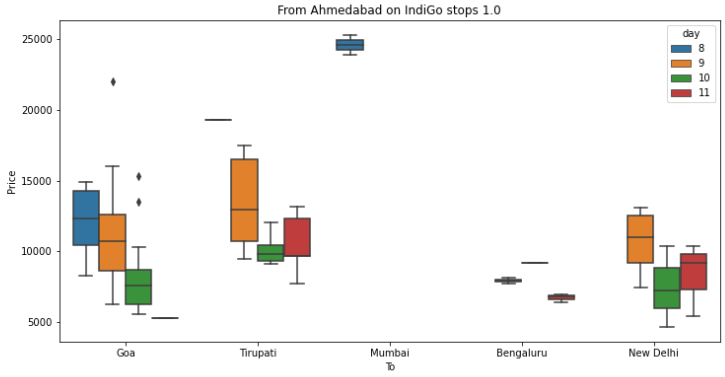

From Ahmedabad on SpiceJet stops 1.0
day
8
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
Goa
Tirupati
Bengaluru
New Delhi
To

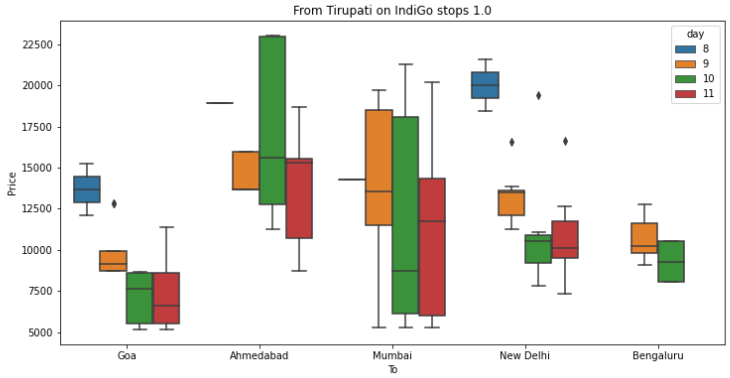

From New Delhi on SpiceJet stops 1.0
day
8
9
10
11
Price
30000
25000
20000
15000
10000
5000
Ahmedabad
Tirupati
Bengaluru
Goa
Mumbai
To

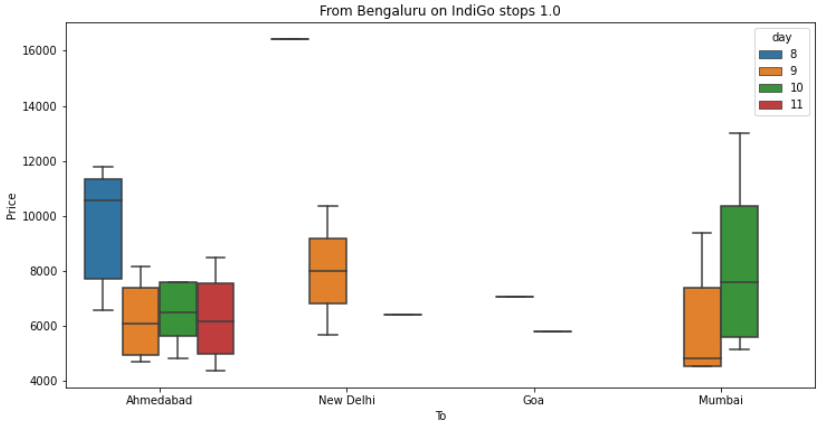

From Mumbai on SpiceJet stops 1.0
day
8
9
10
11
Price
7500
7000
6500
6000
5500
5000
4500
Bengaluru
New Delhi
To

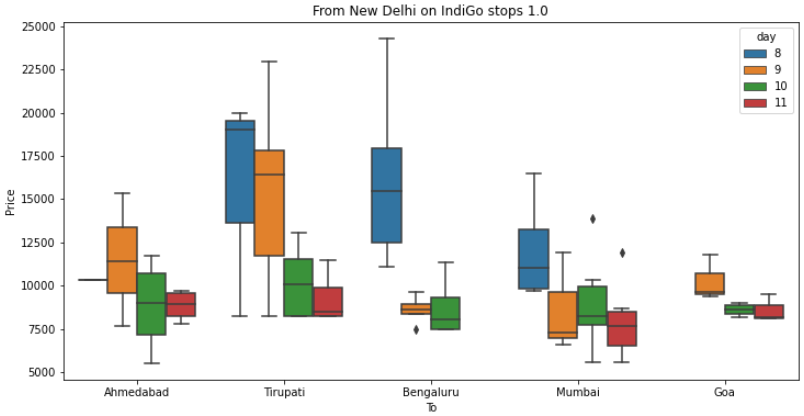

From Bengaluru on SpiceJet stops 1.0
day
8
9
10
11
Price
12000
11000
10000
9000
8000
7000
6000
Ahmedabad
Mumbai
New Delhi
To

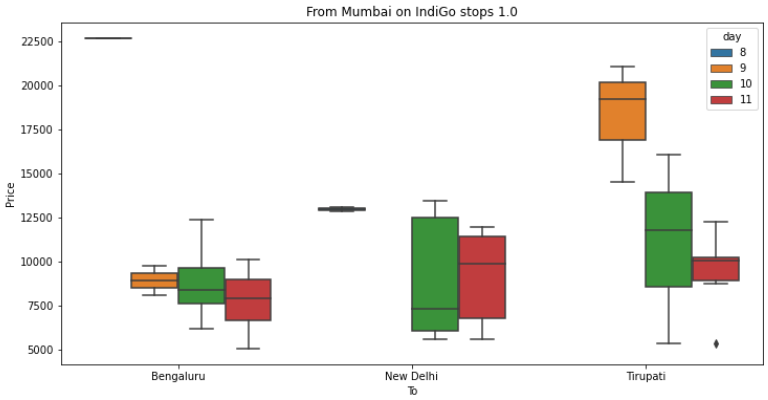

From Goa on IndiGo stops 1.0
day
8
9
10
11
Price
30000
25000
20000
15000
10000
5000
Tirupati
New Delhi
Ahmedabad
Bengaluru
To

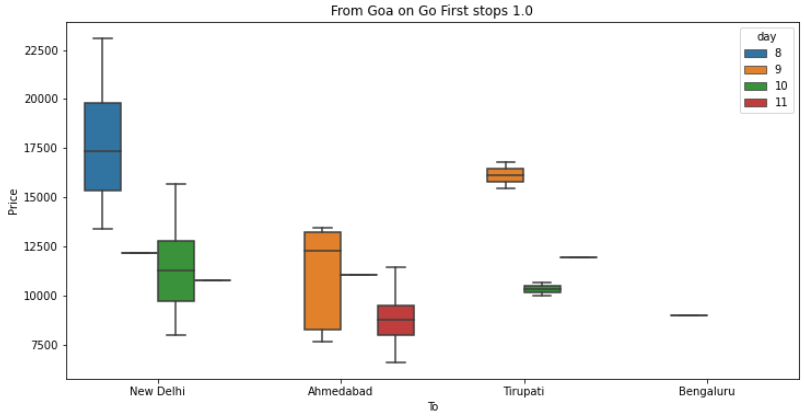

From Ahmedabad on IndiGo stops 1.0
day
8
9
10
11
Price
25000
20000
15000
10000
5000
Goa
Tirupati
Mumbai
Bengaluru
New Delhi
To

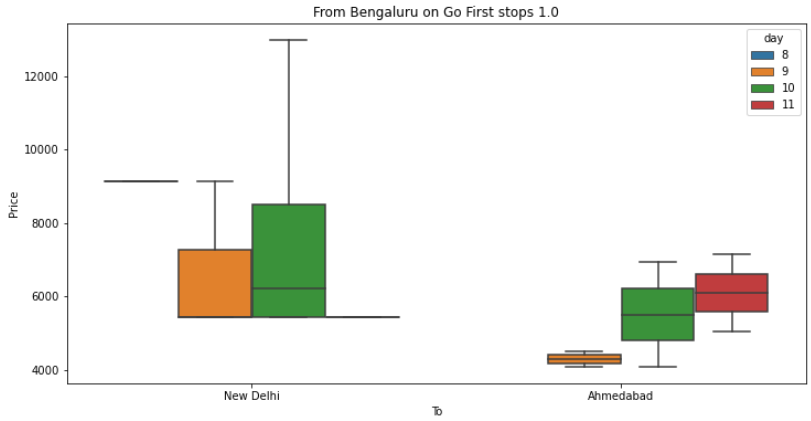

From Tirupati on IndiGo stops 1.0
day
8
9
10
11
Price
22500
20000
17500
15000
12500
10000
7500
5000
Goa
Ahmedabad
Mumbai
New Delhi
Bengaluru
To

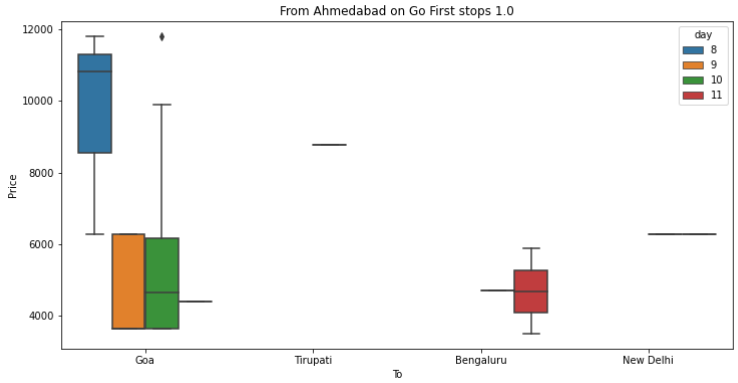

From Bengaluru on IndiGo stops 1.0
day
8
9
10
11
Price
16000
14000
12000
10000
8000
6000
4000
Ahmedabad
New Delhi
Goa
Mumbai
To

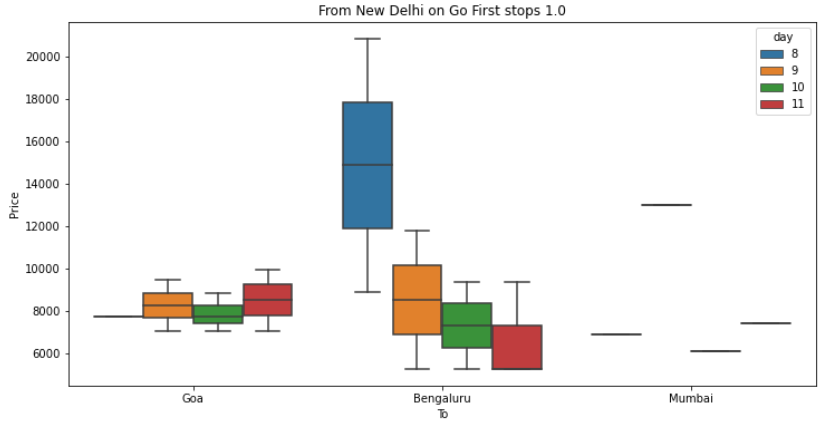

From New Delhi on IndiGo stops 1.0
day
8
9
10
11
Price
25000
22500
20000
17500
15000
12500
10000
7500
5000
Ahmedabad
Tirupati
Bengaluru
Mumbai
Goa
To

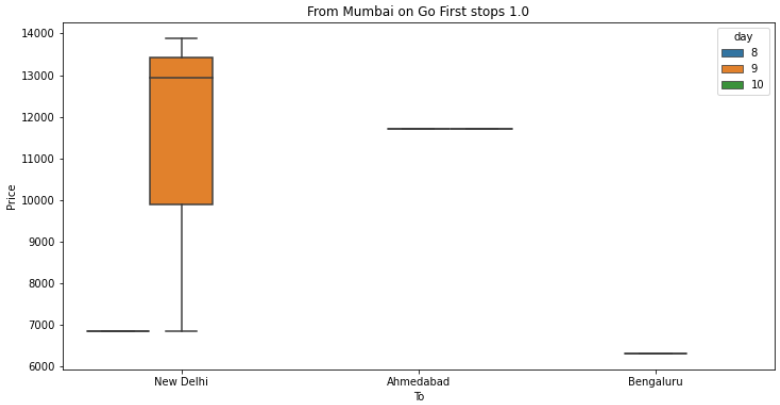

From Mumbai on IndiGo stops 1.0
day
8
9
10
11
Price
22500
20000
17500
15000
12500
10000
7500
5000
Bengaluru
New Delhi
Tirupati
To

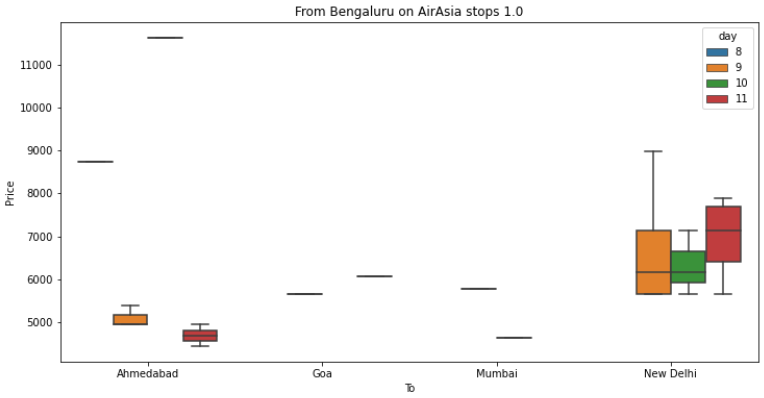

From Goa on Go First stops 1.0
day
8
9
10
11
Price
22500
20000
17500
15000
12500
10000
7500
New Delhi
Ahmedabad
Tirupati
Bengaluru
To

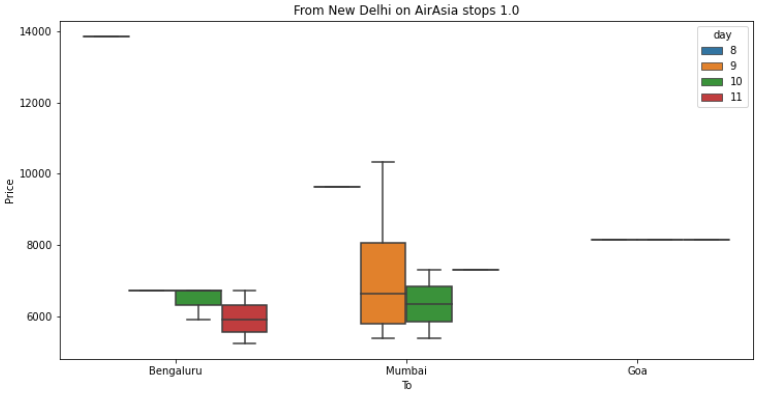

From Bengaluru on Go First stops 1.0
day
8
9
10
11
Price
12000
10000
8000
6000
4000
New Delhi
Ahmedabad
To

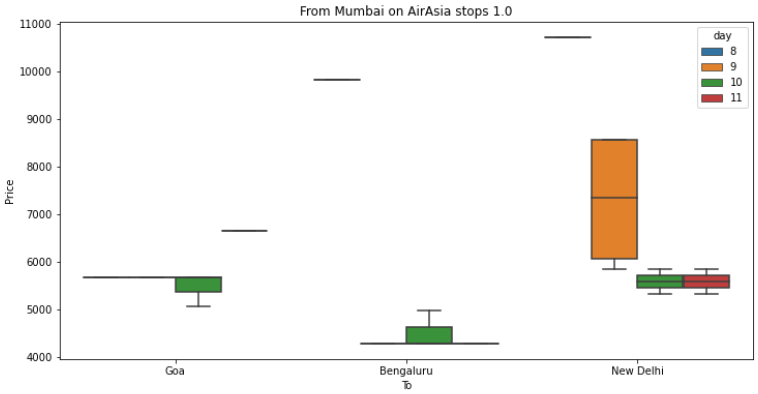

From Ahmedabad on Go First stops 1.0
day
8
9
10
11
Price
12000
10000
8000
6000
4000
Goa
Tirupati
Bengaluru
New Delhi
To

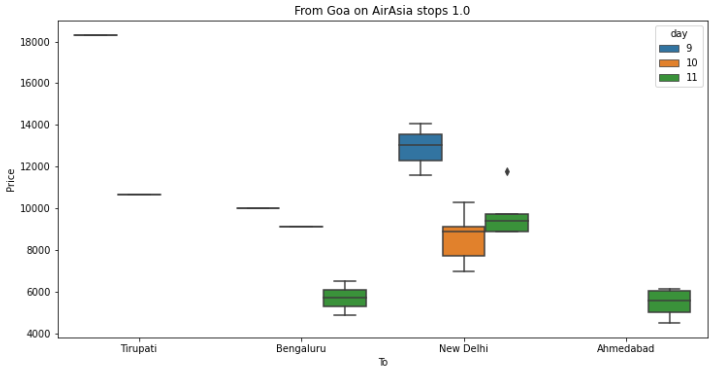

From New Delhi on Go First stops 1.0
day
8
9
10
11
Price
20000
18000
16000
14000
12000
10000
8000
6000
Goa
Bengaluru
Mumbai
To

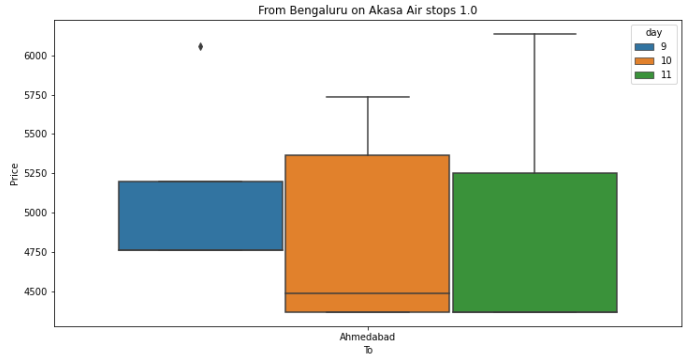

From Mumbai on Go First stops 1.0
day
8
9
10
Price
14000
13000
12000
11000
10000
9000
8000
7000
6000
New Delhi
Ahmedabad
Bengaluru
To

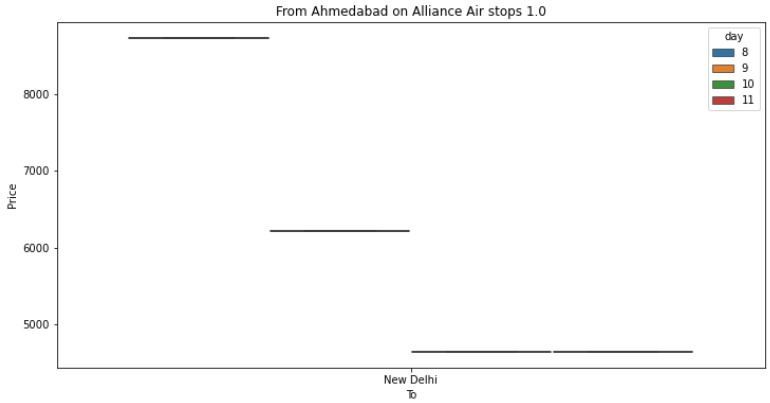

From Bengaluru on AirAsia stops 1.0
day
8
9
10
11
Price
11000
10000
9000
8000
7000
6000
5000
Ahmedabad
Goa
Mumbai
New Delhi
To

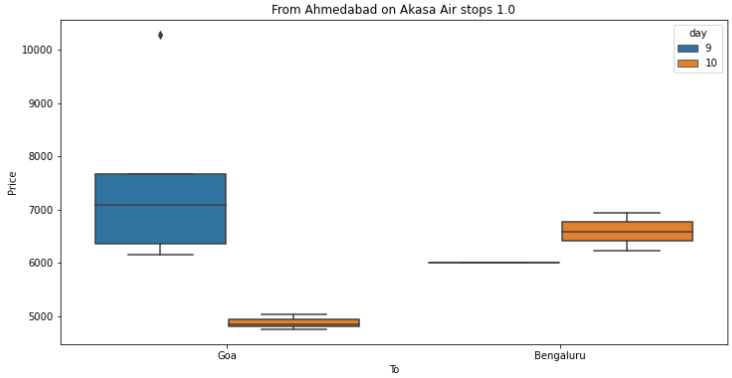

From New Delhi on AirAsia stops 1.0
day
8
9
10
11
Price
14000
12000
10000
8000
6000
Bengaluru
Mumbai
Goa
To

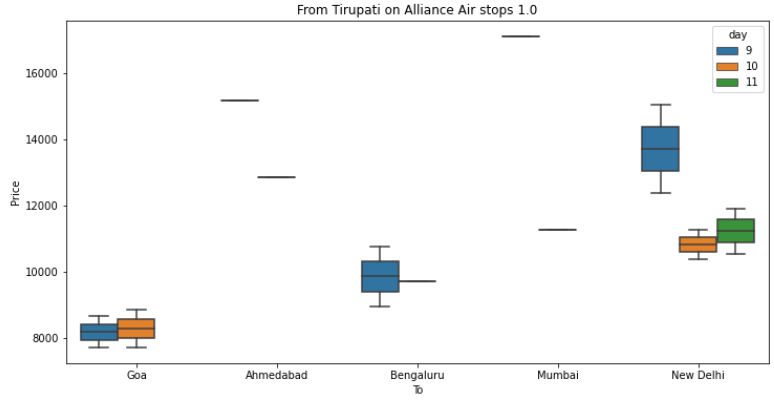

From Mumbai on AirAsia stops 1.0
day
8
9
10
11
Price
11000
10000
9000
8000
7000
6000
5000
4000
Goa
Bengaluru
New Delhi
To

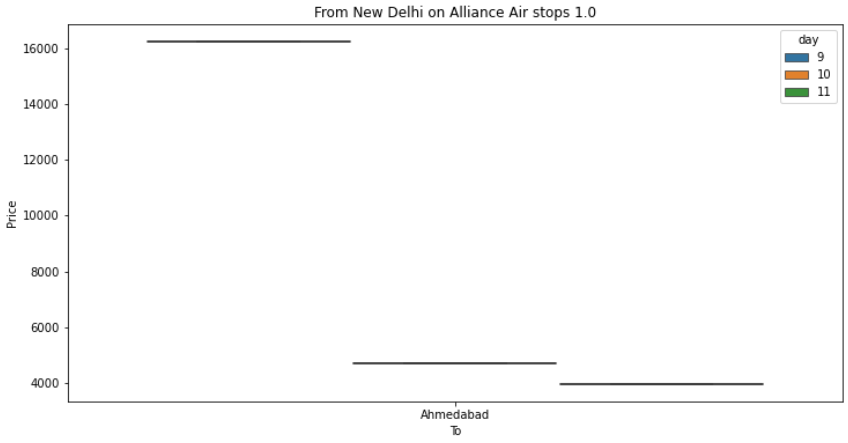

From Goa on AirAsia stops 1.0
day
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
4000
Tirupati
Bengaluru
New Delhi
Ahmedabad
To

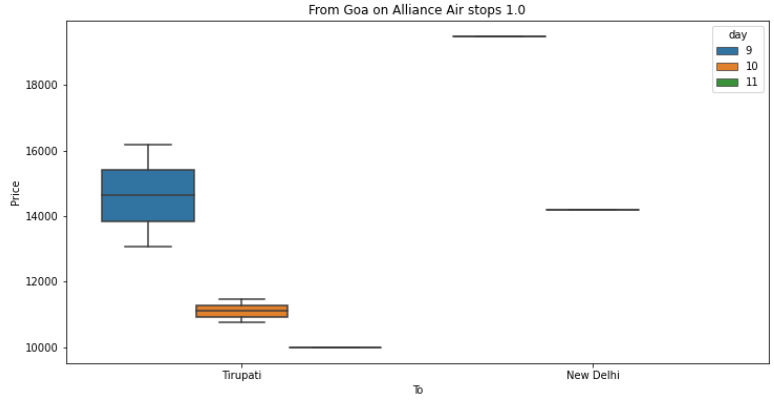

From Bengaluru on Akasa Air stops 1.0
day
9
10
11
Price
6000
5750
5500
5250
5000
4750
4500
Ahmedabad
To

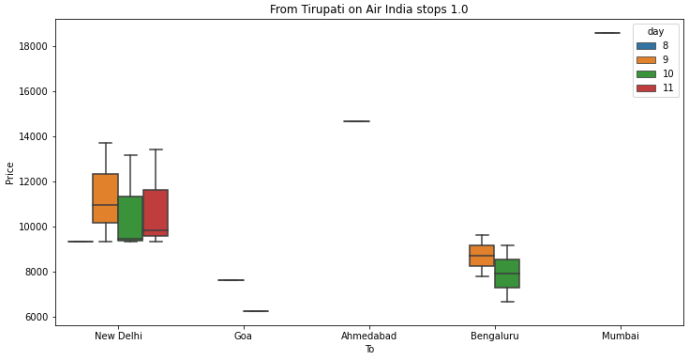

From Ahmedabad on Alliance Air stops 1.0
day
8
9
10
11
Price
8000
7000
6000
5000
New Delhi
To


From Ahmedabad on Akasa Air stops 1.0
day
9
10
Price
10000
9000
8000
7000
6000
5000
Goa
Bengaluru
To

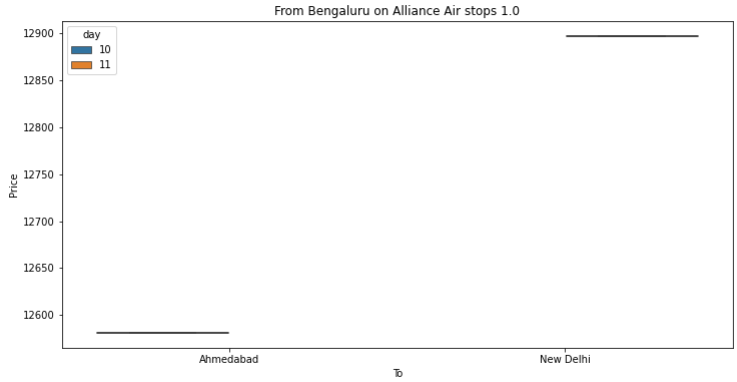

From Tirupati on Alliance Air stops 1.0
day
9
10
11
Price
16000
14000
12000
10000
8000
Goa
Ahmedabad
Bengaluru
Mumbai
New Delhi
To

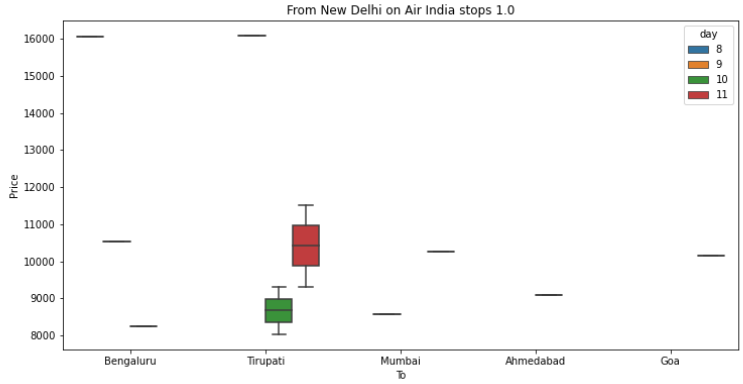

From New Delhi on Alliance Air stops 1.0
day
9
10
11
Price
16000
14000
12000
10000
8000
6000
4000
Ahmedabad
To

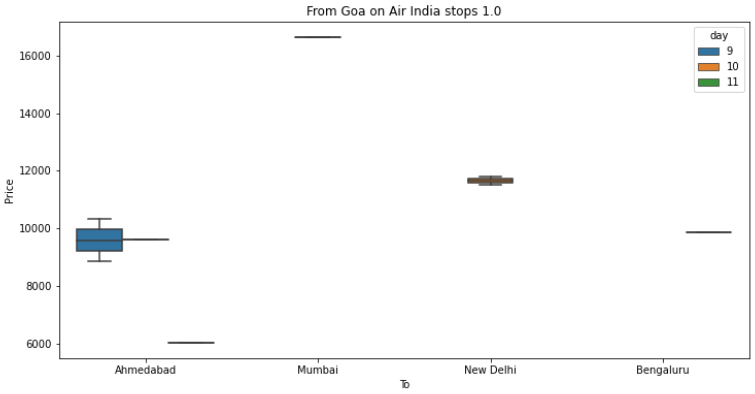

From Goa on Alliance Air stops 1.0
day
9
10
11
Price
18000
16000
14000
12000
10000
Tirupati
New Delhi
To

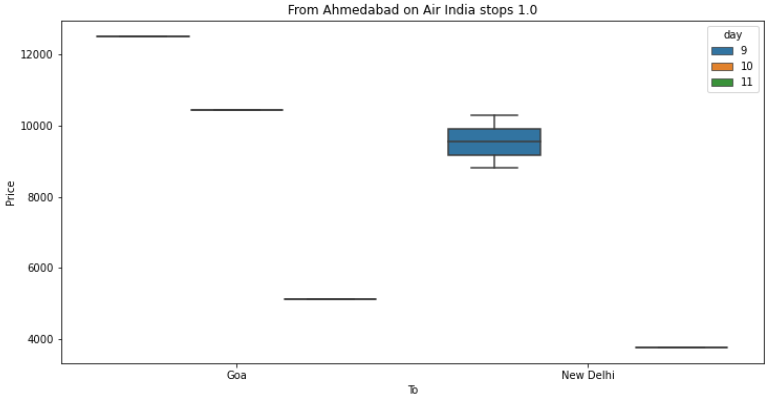

From Tirupati on Air India stops 1.0
day
8
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
New Delhi
Goa
Ahmedabad
Bengaluru
Mumbai
To

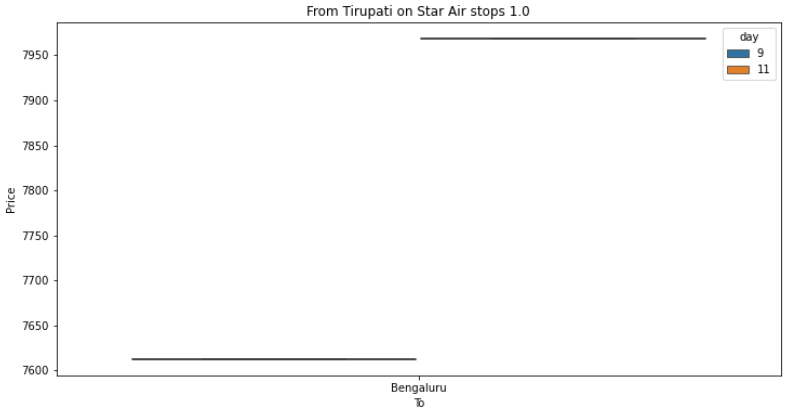

From Mumbai on Air India stops 1.0
day
8
9
10
11
Price
16000
14000
12000
10000
8000
6000
New Delhi
Ahmedabad
Bengaluru
Tirupati
To

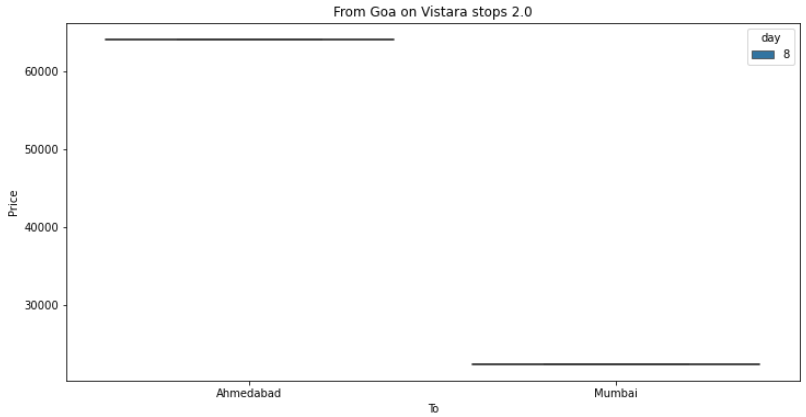

From Bengaluru on Alliance Air stops 1.0
day
10
11
Price
12900
12850
12800
12750
12700
12650
12600
Ahmedabad
New Delhi
To

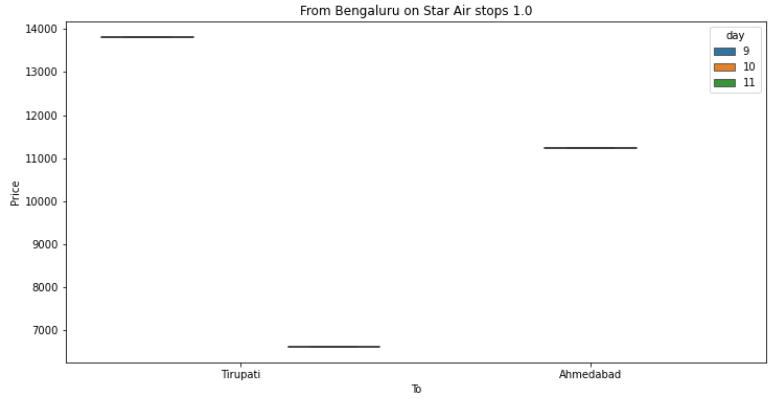

From New Delhi on Air India stops 1.0
day
8
9
10
11
Price
16000
15000
14000
13000
12000
11000
10000
9000
8000
Bengaluru
Tirupati
Mumbai
Ahmedabad
Goa
To

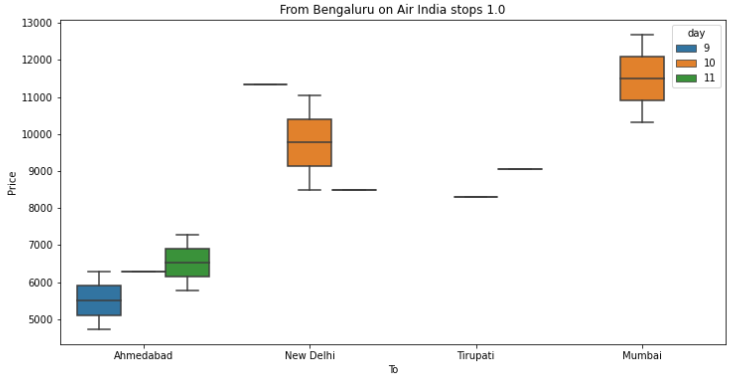

From Goa on Air India stops 1.0
day
9
10
11
Price
16000
14000
12000
10000
8000
6000
Ahmedabad
Mumbai
New Delhi
Bengaluru
To

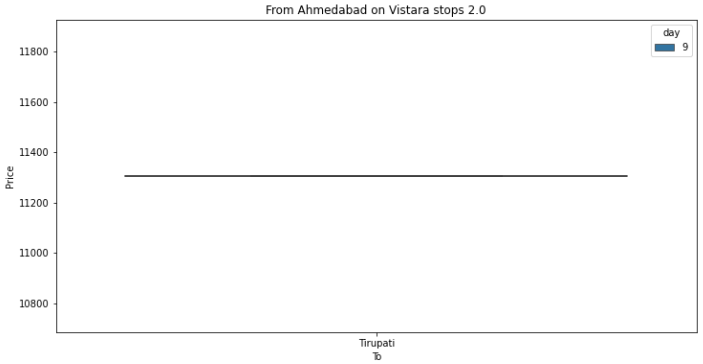

From Ahmedabad on Air India stops 1.0
day
9
10
11
Price
12000
10000
8000
6000
4000
Goa
New Delhi
To

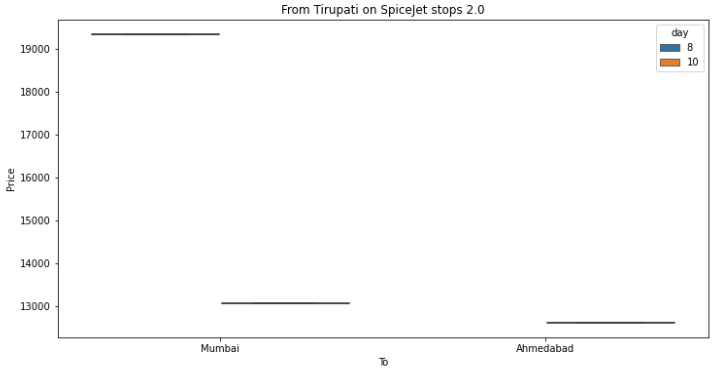

From Tirupati on Star Air stops 1.0
day
9
11
Price
7950
7900
7850
7800
7750
7700
7650
7600
Bengaluru
To

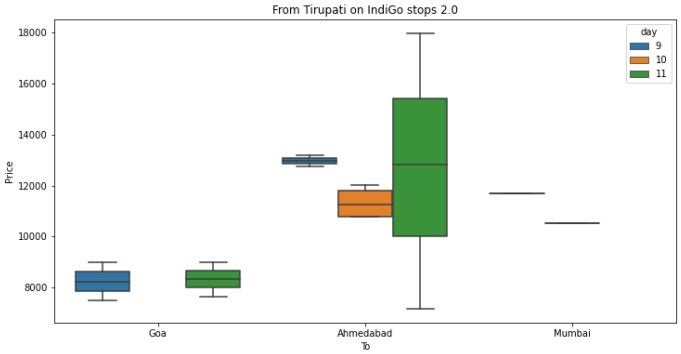

From Goa on Vistara stops 2.0
day
8
Price
60000
50000
40000
30000
Ahmedabad
Mumbai
To

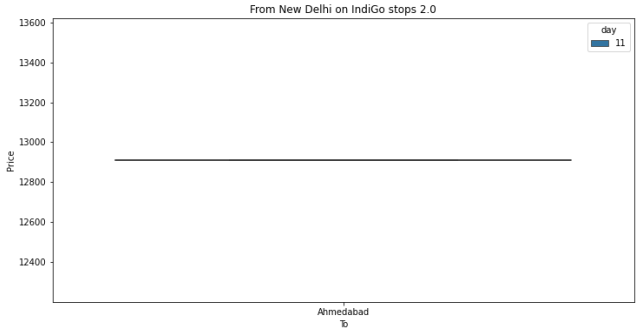

From Bengaluru on Star Air stops 1.0
day
9
10
11
Price
14000
13000
12000
11000
10000
9000
8000
7000
Tirupati
Ahmedabad
To

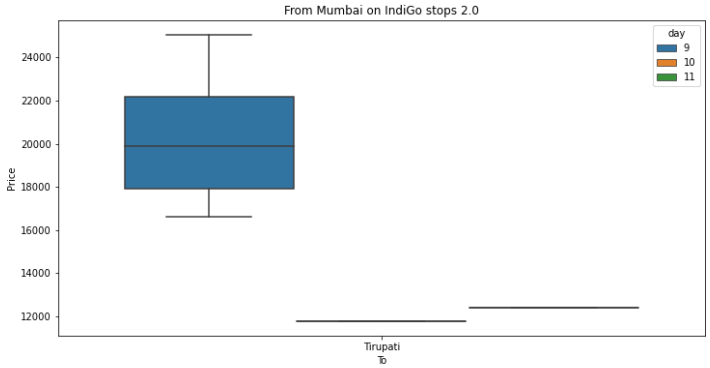

From Bengaluru on Air India stops 1.0
day
9
10
11
Price
13000
12000
11000
10000
9000
8000
7000
6000
5000
Ahmedabad
New Delhi
Tirupati
Mumbai
To

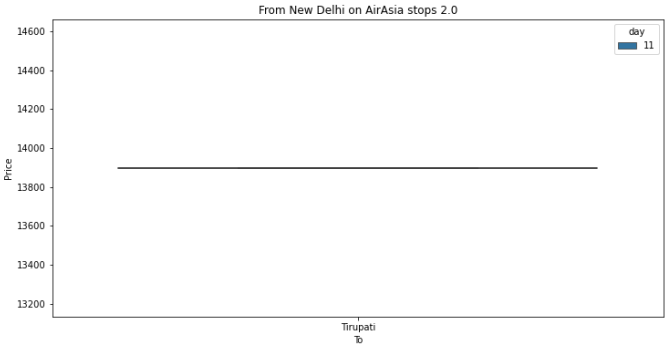

From Ahmedabad on Vistara stops 2.0
day
9
Price
11800
11600
11400
11200
11000
10800
Tirupati
To

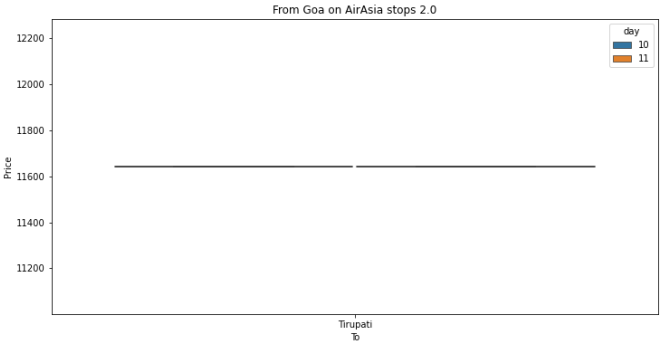

From Tirupati on SpiceJet stops 2.0
day
8
10
Price
19000
18000
17000
16000
15000
14000
13000
Mumbai
Ahmedabad
To

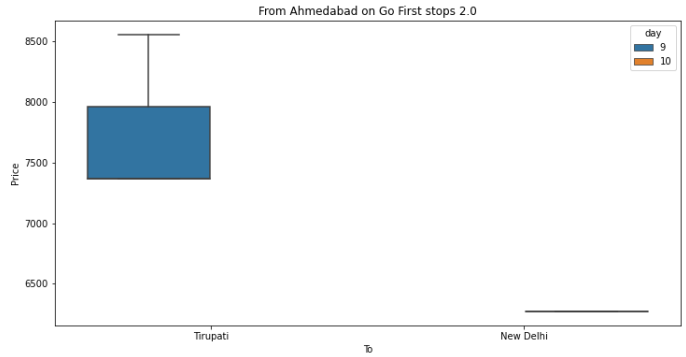

From Tirupati on IndiGo stops 2.0
day
9
10
11
Price
18000
16000
14000
12000
10000
8000
Goa
Ahmedabad
Mumbai
To

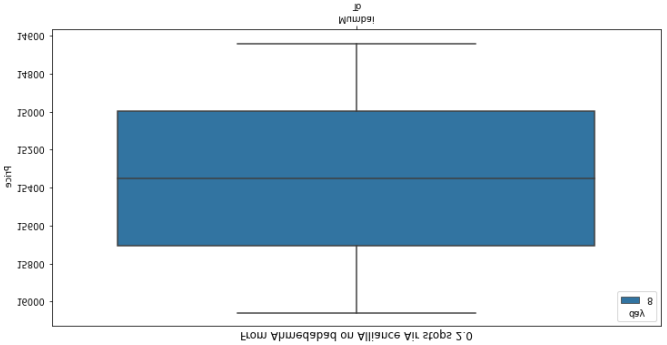

From New Delhi on IndiGo stops 2.0
day
11
Price
13600
13400
13200
13000
12800
12600
12400
Ahmedabad
To

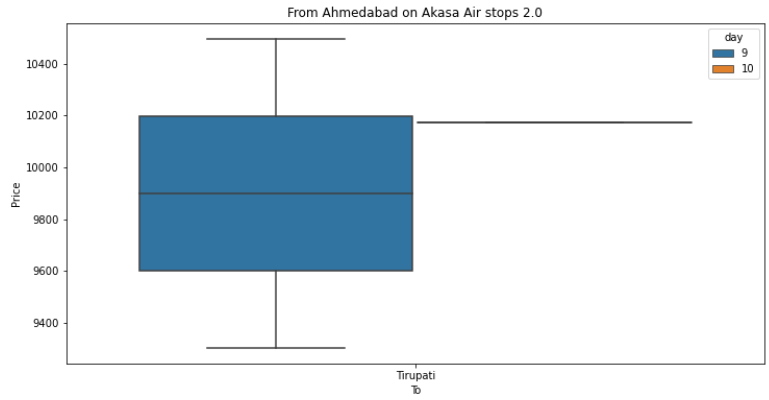

From Mumbai on IndiGo stops 2.0
day
9
10
11
Price
24000
22000
20000
18000
16000
14000
12000
Tirupati
To

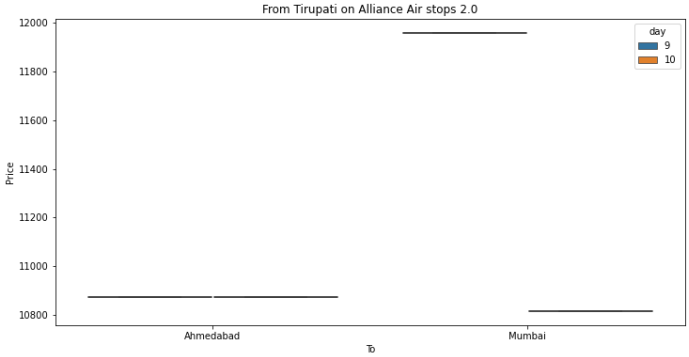

From New Delhi on AirAsia stops 2.0
day
11
Price
14600
14400
14200
14000
13800
13600
13400
13200
Tirupati
To

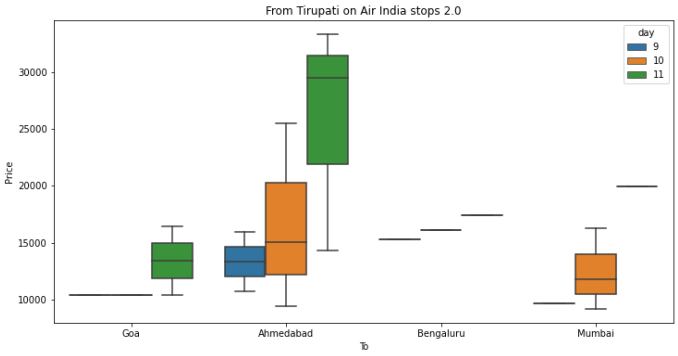

From Goa on AirAsia stops 2.0
day
10
11
Price
12200
12000
11800
11600
11400
11200
Tirupati
To

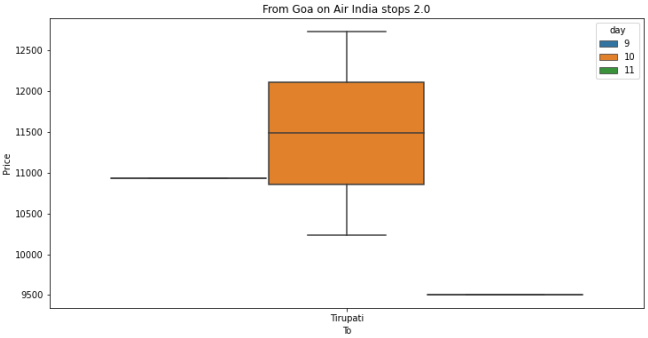

From Ahmedabad on Go First stops 2.0
day
9
10
Price
8500
8000
7500
7000
6500
Tirupati
New Delhi
To

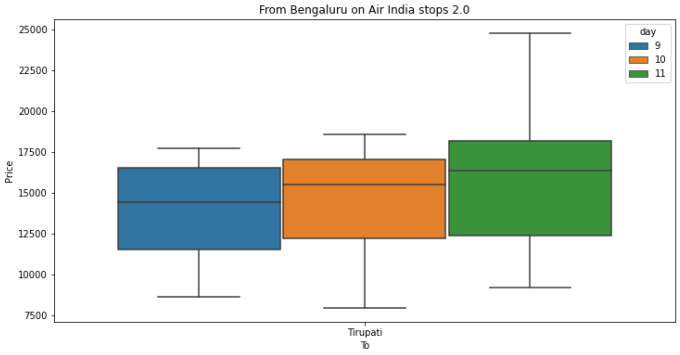


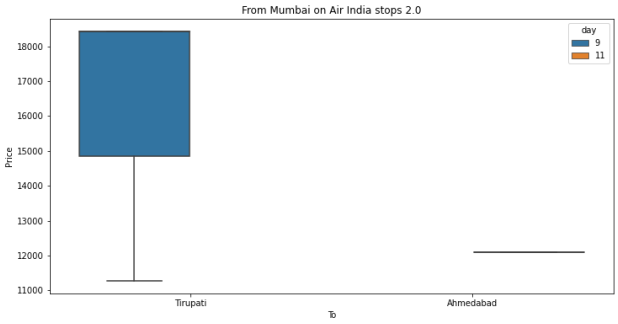

From Ahmedabad on Akasa Air stops 2.0
day
9
10
Price
10400
10200
10000
9800
9600
9400
Tirupati
To

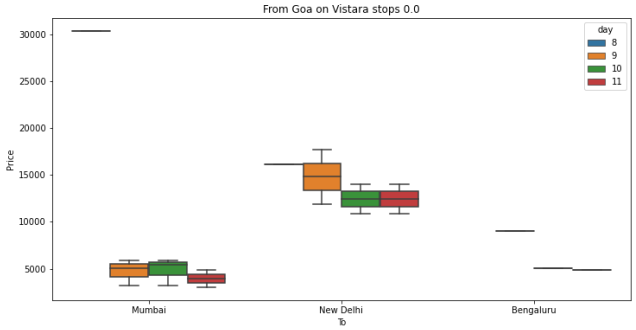

From Tirupati on Alliance Air stops 2.0
day
9
10
Price
12000
11800
11600
11400
11200
11000
10800
Ahmedabad
Mumbai
To

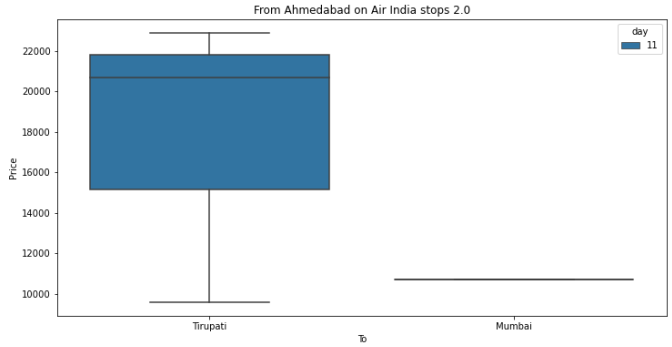

From Tirupati on Air India stops 2.0
day
9
10
11
Price
30000
25000
20000
15000
10000
Goa
Ahmedabad
Bengaluru
Mumbai
To

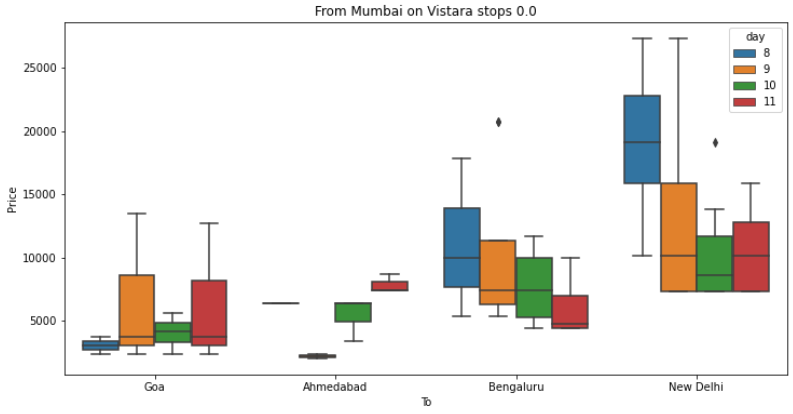

From Goa on Air India stops 2.0
day
9
10
11
Price
12500
12000
11500
11000
10500
10000
9500
Tirupati
To

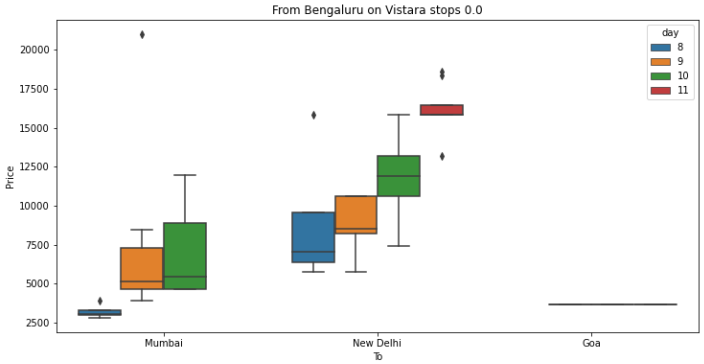

From Bengaluru on Air India stops 2.0
day
9
10
11
Price
25000
22500
20000
17500
15000
12500
10000
7500
Tirupati
To


From Mumbai on Air India stops 2.0
day
9
11
Price
18000
17000
16000
15000
14000
13000
12000
11000
Tirupati
Ahmedabad
To

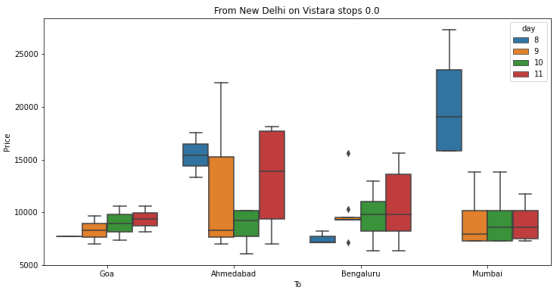

From Goa on Vistara stops 0.0
day
8
9
10
11
Price
30000
25000
20000
15000
10000
5000
Mumbai
New Delhi
Bengaluru
To

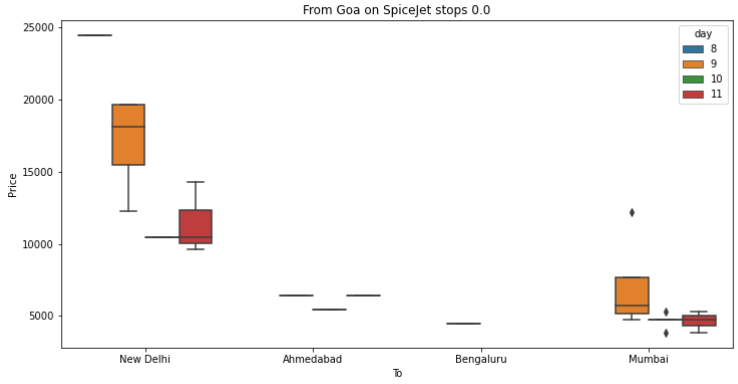

From Ahmedabad on Air India stops 2.0
day
11
Price
22000
20000
18000
16000
14000
12000
10000
Tirupati
Mumbai
To

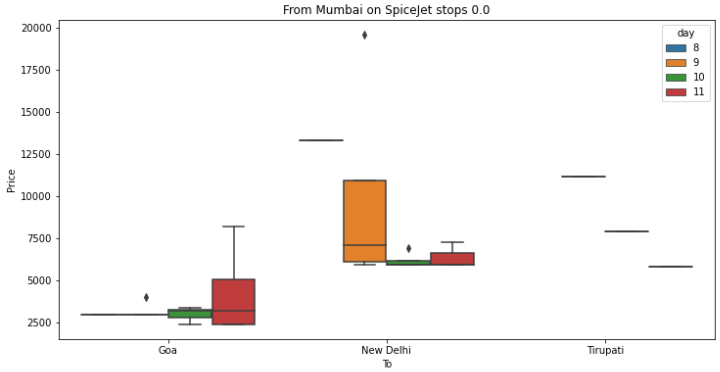

From Mumbai on Vistara stops 0.0
day
8
9
10
11
Price
25000
20000
15000
10000
5000
Goa
Ahmedabad
Bengaluru
New Delhi
To

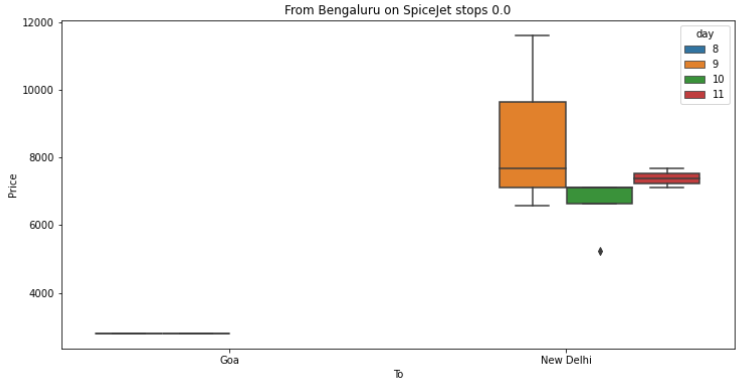

From Bengaluru on Vistara stops 0.0
day
8
9
10
11
Price
20000
17500
15000
12500
10000
7500
5000
2500
Mumbai
New Delhi
Goa
To

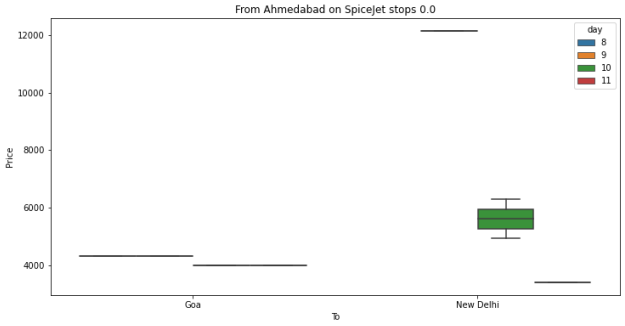

From Ahmedabad on Vistara stops 0.0
day
8
9
10
11
Price
25000
20000
15000
10000
5000
Mumbai
New Delhi
To

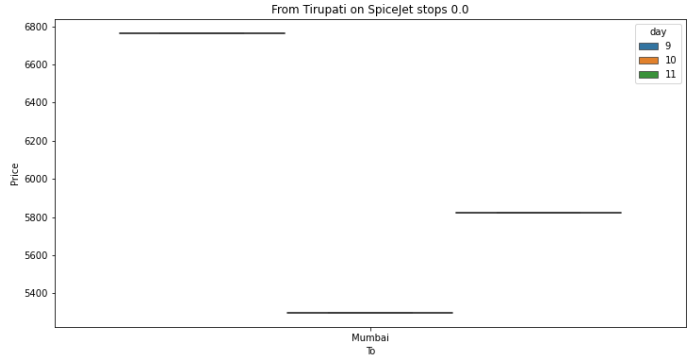

From New Delhi on Vistara stops 0.0
day
8
9
10
11
Price
25000
20000
15000
10000
5000
Goa
Ahmedabad
Bengaluru
Mumbai
To

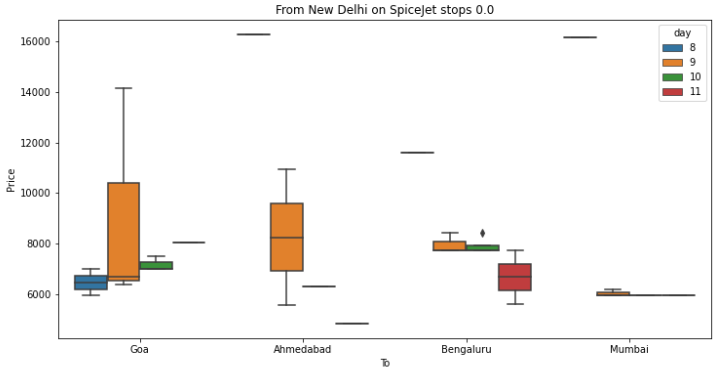

From Goa on SpiceJet stops 0.0
day
8
9
10
11
Price
25000
20000
15000
10000
5000
New Delhi
Ahmedabad
Bengaluru
Mumbai
To

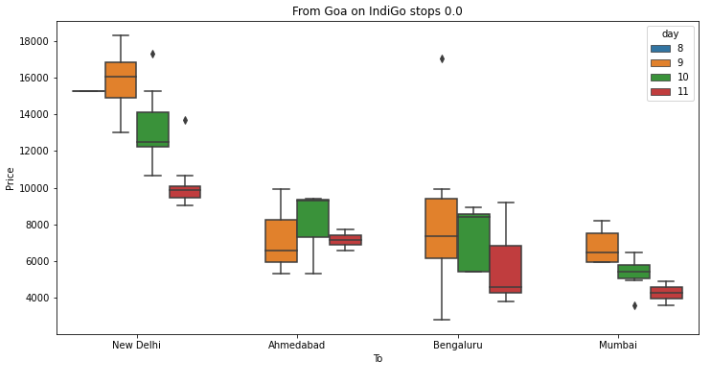

From Mumbai on SpiceJet stops 0.0
day
8
9
10
11
Price
20000
17500
15000
12500
10000
7500
5000
2500
Goa
New Delhi
Tirupati
To

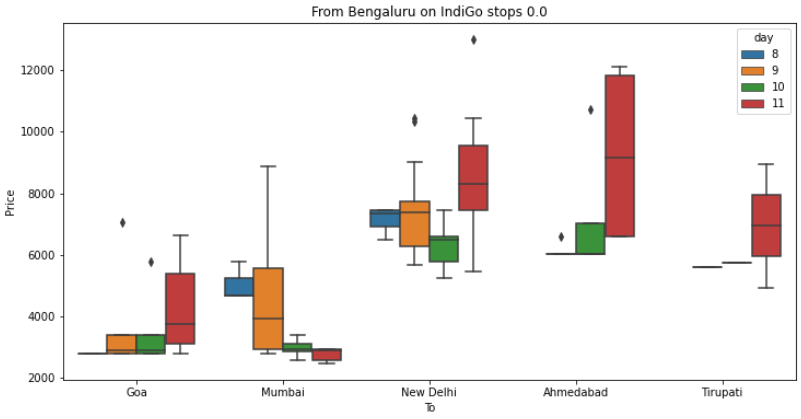

From Bengaluru on SpiceJet stops 0.0
day
8
9
10
11
Price
12000
10000
8000
6000
4000
Goa
New Delhi
To

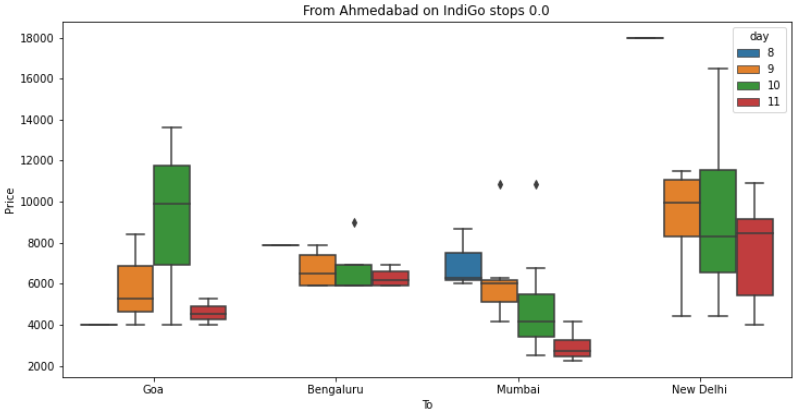

From Ahmedabad on SpiceJet stops 0.0
day
8
9
10
11
Price
12000
10000
8000
6000
4000
Goa
New Delhi
To

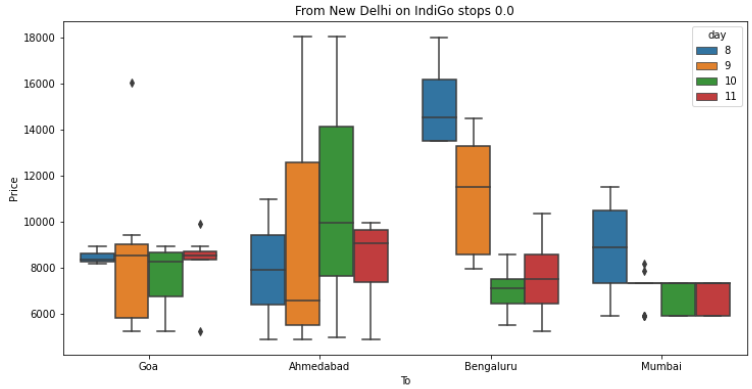

From Tirupati on SpiceJet stops 0.0
day
9
10
11
Price
6800
6600
6400
6200
6000
5800
5600
5400
Mumbai
To

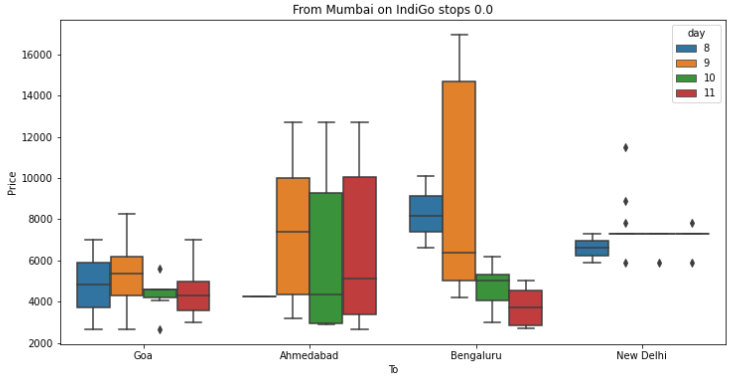

From New Delhi on SpiceJet stops 0.0
day
8
9
10
11
Price
16000
14000
12000
10000
8000
6000
Goa
Ahmedabad
Bengaluru
Mumbai
To

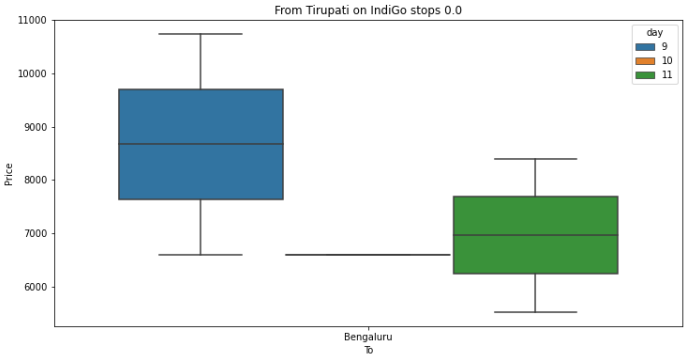

From Goa on IndiGo stops 0.0
day
8
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
4000
New Delhi
Ahmedabad
Bengaluru
Mumbai
To

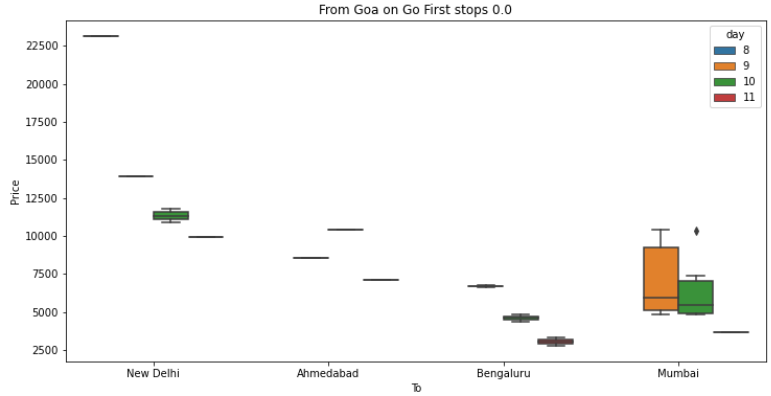

From Bengaluru on IndiGo stops 0.0
day
8
9
10
11
Price
12000
10000
8000
6000
4000
2000
Goa
Mumbai
New Delhi
Ahmedabad
Tirupati
To

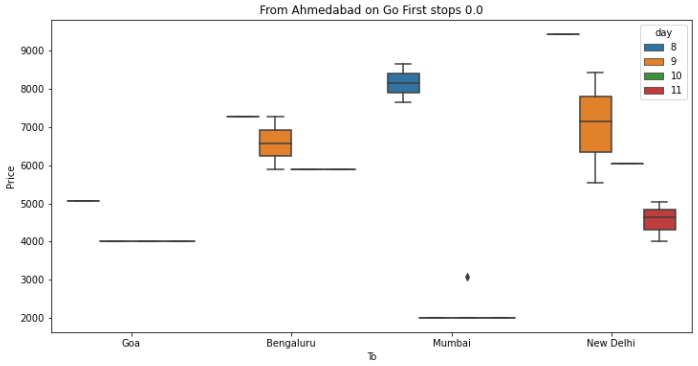

From Ahmedabad on IndiGo stops 0.0
day
8
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
4000
2000
Goa
Bengaluru
Mumbai
New Delhi
To

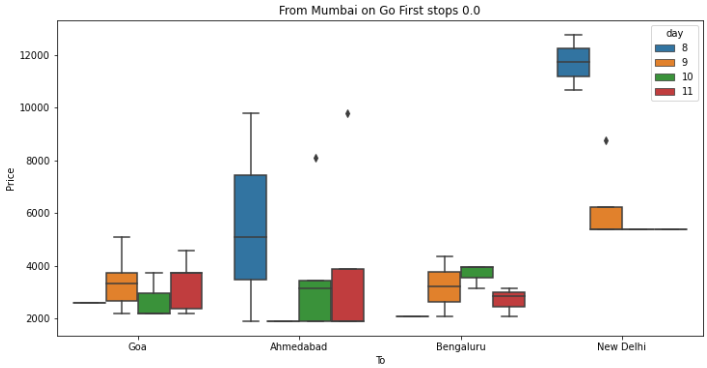

From New Delhi on IndiGo stops 0.0
day
8
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
Goa
Ahmedabad
Bengaluru
Mumbai
To

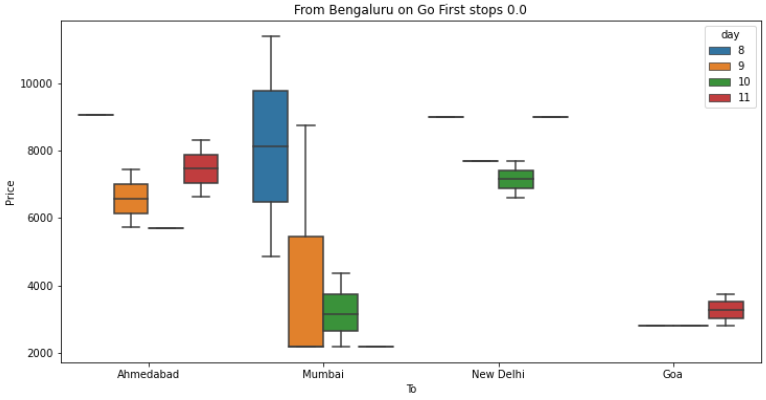

From Mumbai on IndiGo stops 0.0
day
8
9
10
11
Price
16000
14000
12000
10000
8000
6000
4000
2000
Goa
Ahmedabad
Bengaluru
New Delhi
To

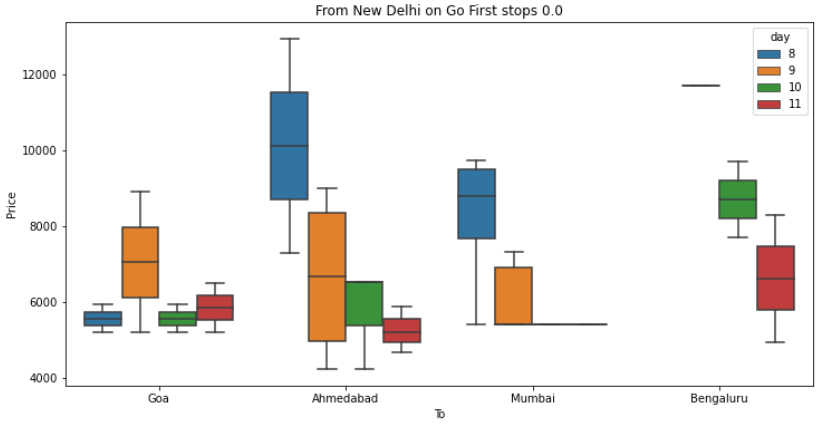

From Tirupati on IndiGo stops 0.0
day
9
10
11
Price
11000
10000
9000
8000
7000
6000
Bengaluru
To

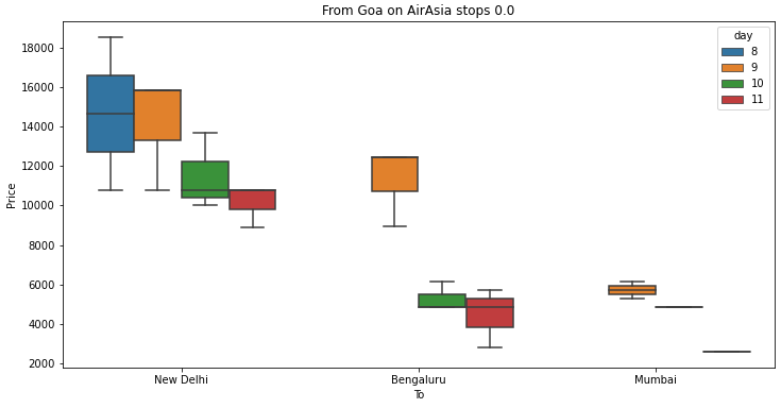

From Goa on Go First stops 0.0
day
8
9
10
11
Price
22500
20000
17500
15000
12500
10000
7500
5000
2500
New Delhi
Ahmedabad
Bengaluru
Mumbai
To

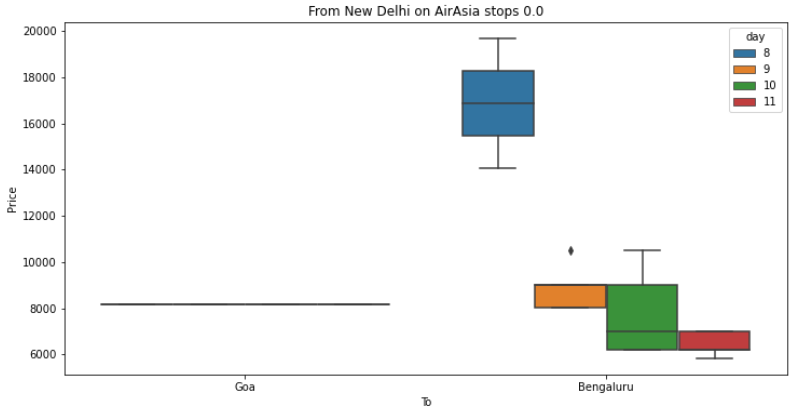

From Ahmedabad on Go First stops 0.0
day
8
9
10
11
Price
9000
8000
7000
6000
5000
4000
3000
2000
Goa
Bengaluru
Mumbai
New Delhi
To

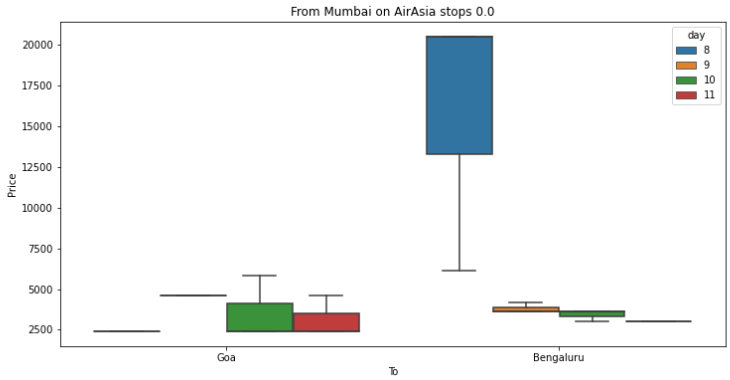

From Mumbai on Go First stops 0.0
day
8
9
10
11
Price
12000
10000
8000
6000
4000
2000
Goa
Ahmedabad
Bengaluru
New Delhi
To

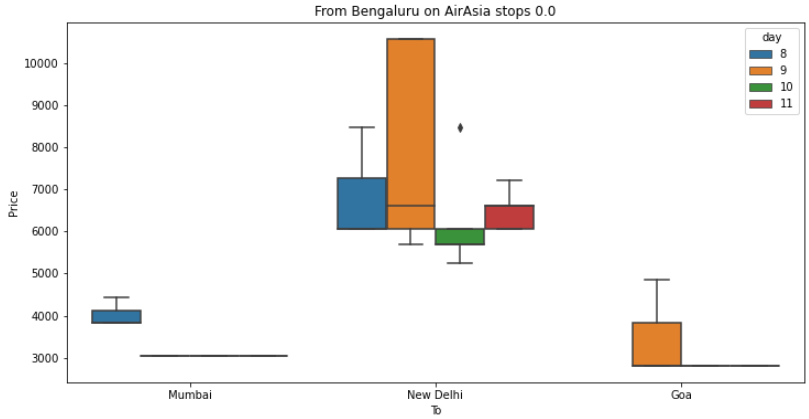

From Bengaluru on Go First stops 0.0
day
8
9
10
11
Price
10000
8000
6000
4000
2000
Ahmedabad
Mumbai
New Delhi
Goa
To

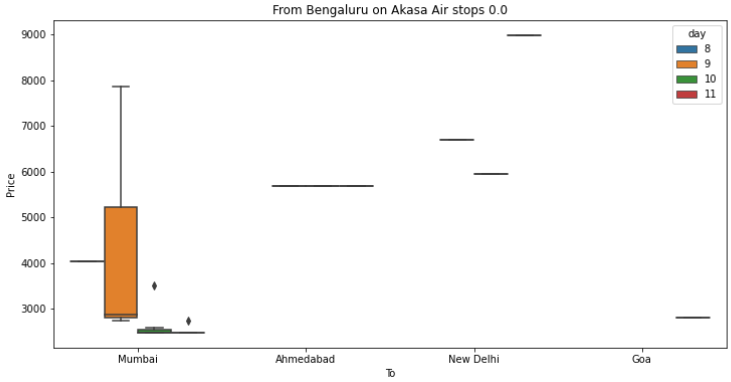

From New Delhi on Go First stops 0.0
day
8
9
10
11
Price
12000
10000
8000
6000
4000
Goa
Ahmedabad
Mumbai
Bengaluru
To

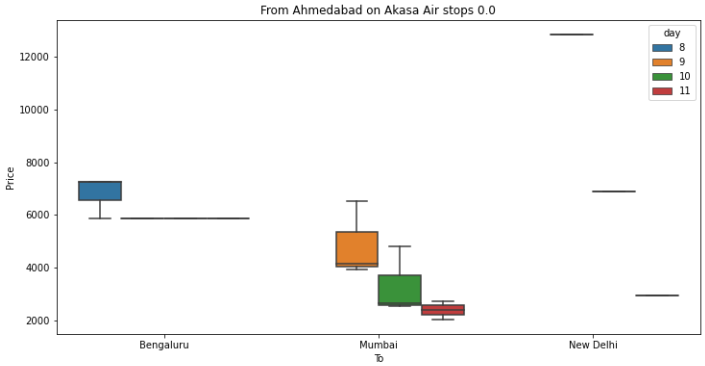

From Goa on AirAsia stops 0.0
day
8
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
4000
2000
New Delhi
Bengaluru
Mumbai
To

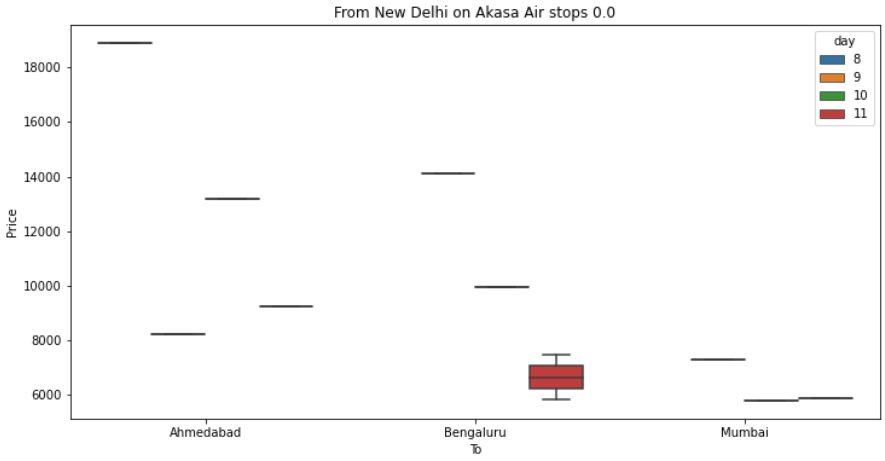

From New Delhi on AirAsia stops 0.0
day
8
9
10
11
Price
20000
18000
16000
14000
12000
10000
8000
6000
Goa
Bengaluru
To

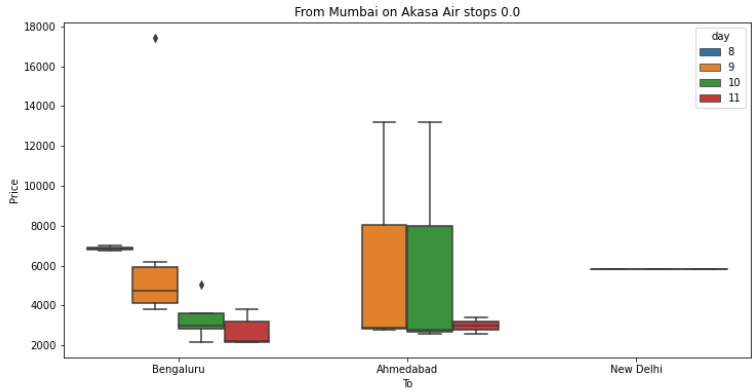

From Mumbai on AirAsia stops 0.0
day
8
9
10
11
Price
20000
17500
15000
12500
10000
7500
5000
2500
Goa
Bengaluru
To

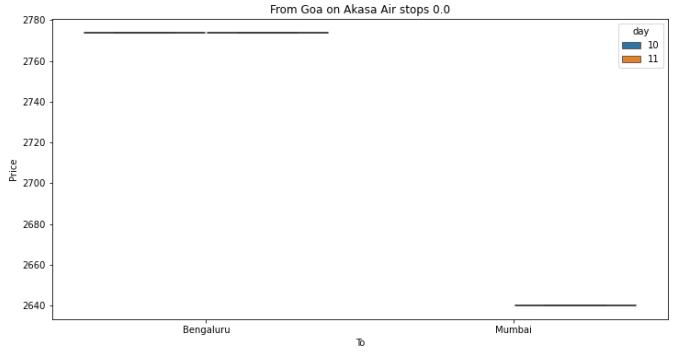

From Bengaluru on AirAsia stops 0.0
day
8
9
10
11
Price
10000
9000
8000
7000
6000
5000
4000
3000
Mumbai
New Delhi
Goa
To

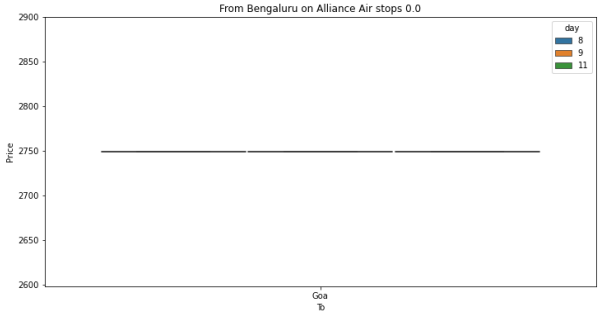

From Bengaluru on Akasa Air stops 0.0
day
8
9
10
11
Price
9000
8000
7000
6000
5000
4000
3000
Mumbai
Ahmedabad
New Delhi
Goa
To

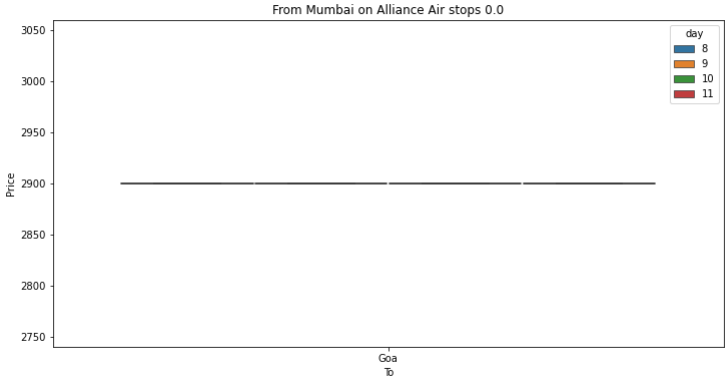

From Ahmedabad on Akasa Air stops 0.0
day
8
9
10
11
Price
12000
10000
8000
6000
4000
2000
Bengaluru
Mumbai
New Delhi
To

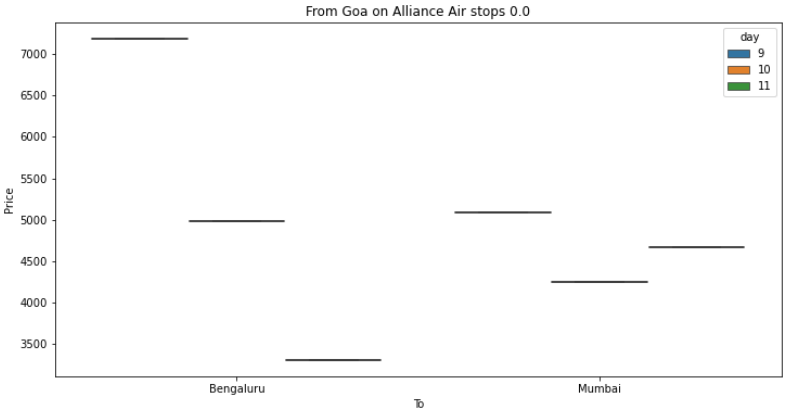

From New Delhi on Akasa Air stops 0.0
day
8
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
Ahmedabad
Bengaluru
Mumbai
To

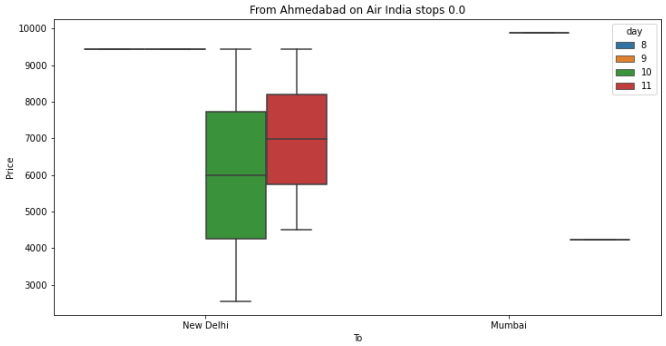

From Mumbai on Akasa Air stops 0.0
day
8
9
10
11
Price
18000
16000
14000
12000
10000
8000
6000
4000
2000
Bengaluru
Ahmedabad
New Delhi
To

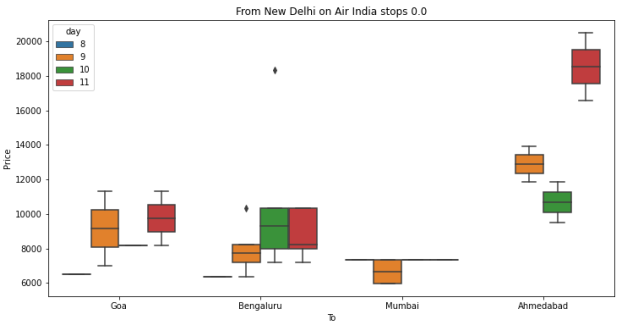

From Goa on Akasa Air stops 0.0
day
10
11
Price
2780
2760
2740
2720
2700
2680
2660
2640
Bengaluru
Mumbai
To

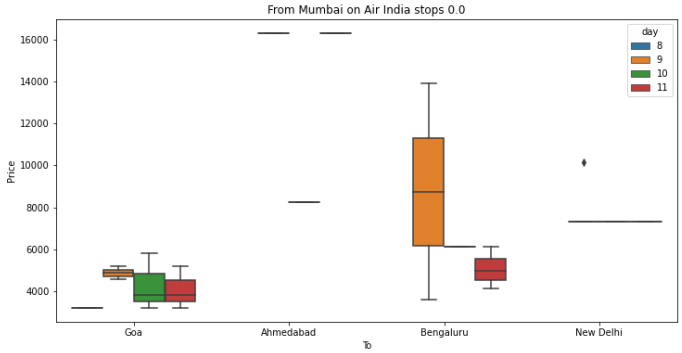

From Bengaluru on Alliance Air stops 0.0
day
8
9
11
Price
2900
2850
2800
2750
2700
2650
2600
Goa
To

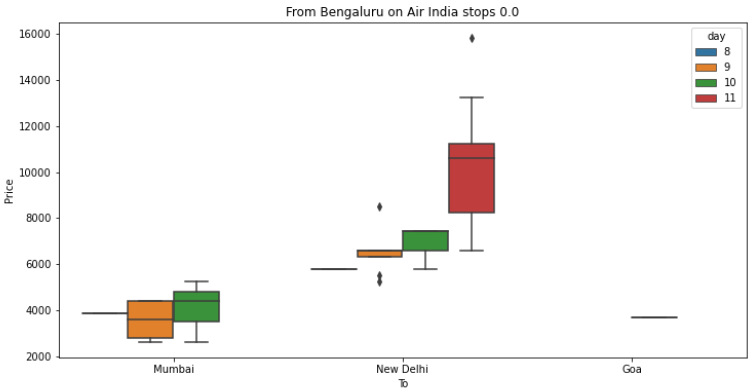

From Mumbai on Alliance Air stops 0.0
day
8
9
10
11
Price
3050
3000
2950
2900
2850
2800
2750
Goa
To

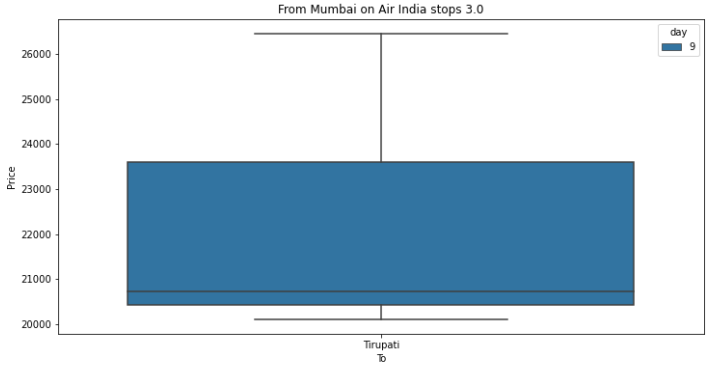

From Goa on Alliance Air stops 0.0
day
9
10
11
Price
7000
6500
6000
5500
5000
4500
4000
3500
Bengaluru
Mumbai
To

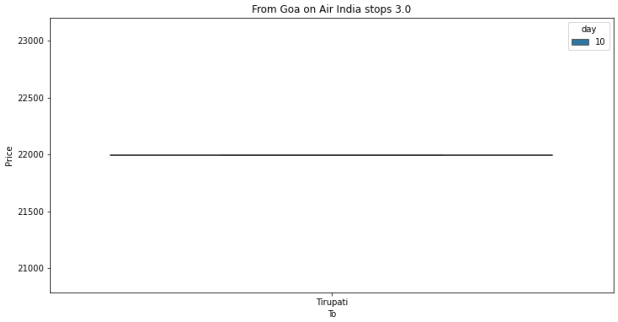

From Ahmedabad on Air India stops 0.0
day
8
9
10
11
Price
10000
9000
8000
7000
6000
5000
4000
3000
New Delhi
Mumbai
To

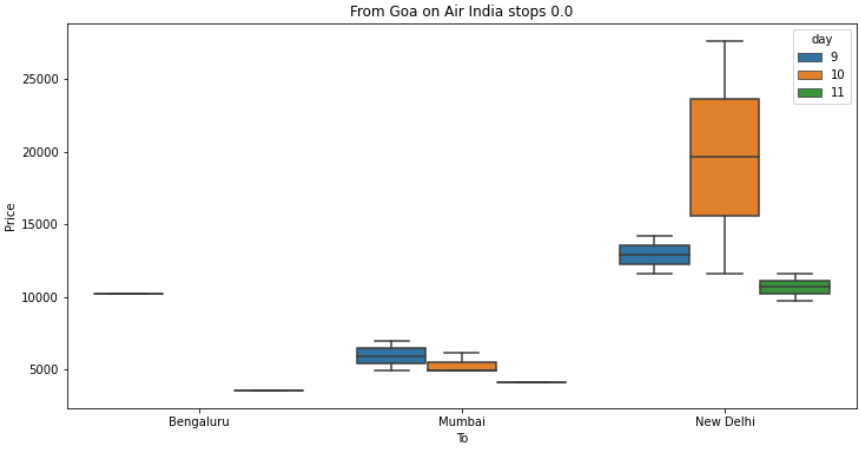

From New Delhi on Air India stops 0.0
day
8
9
10
11
Price
20000
18000
16000
14000
12000
10000
8000
6000
Goa
Bengaluru
Mumbai
Ahmedabad
To

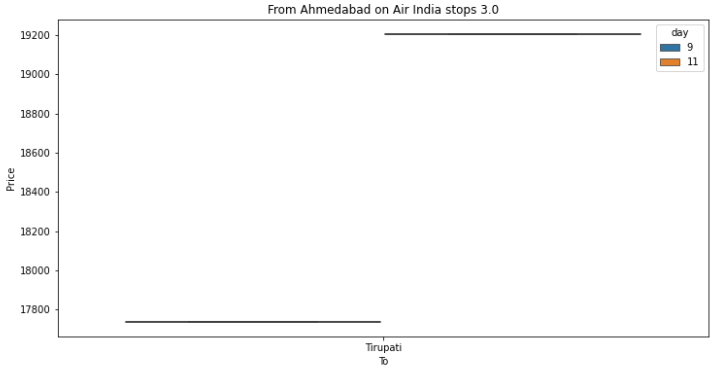

From Mumbai on Air India stops 0.0
day
8
9
10
11
Price
16000
14000
12000
10000
8000
6000
4000
Goa
Ahmedabad
Bengaluru
New Delhi
To

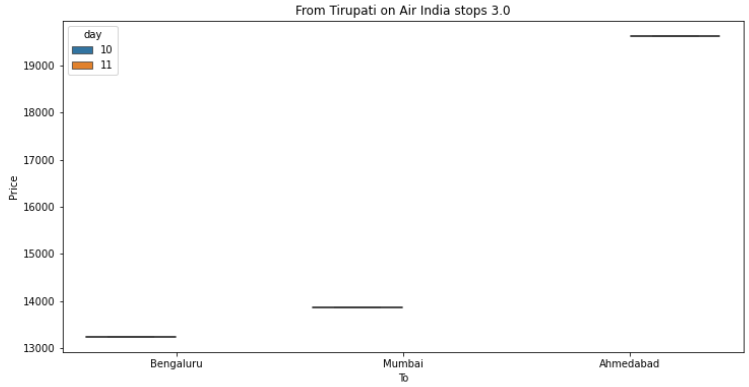

From Bengaluru on Air India stops 0.0
day
8
9
10
11
Price
16000
14000
12000
10000
8000
6000
4000
2000
Mumbai
New Delhi
Goa
To

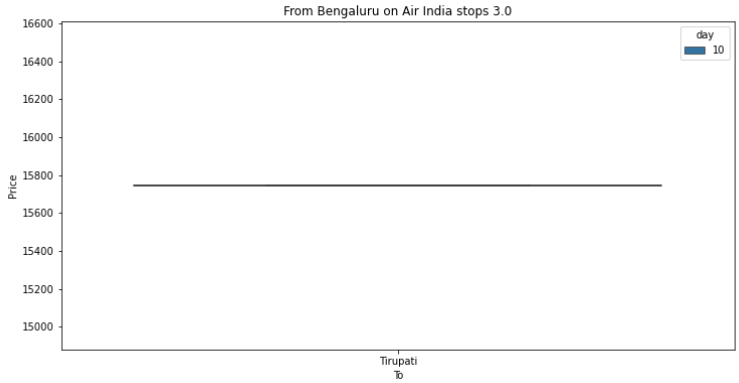

From Mumbai on Air India stops 3.0
day
9
Price
26000
25000
24000
23000
22000
21000
20000
Tirupati
To


From Goa on Air India stops 3.0
day
10
Price
23000
22500
22000
21500
21000
Tirupati
To


From Goa on Air India stops 0.0
day
9
10
11
Price
25000
20000
15000
10000
5000
Bengaluru
Mumbai
New Delhi
To


From Ahmedabad on Air India stops 3.0
day
9
11
Price
19200
19000
18800
18600
18400
18200
18000
17800
Tirupati
To





- Goa to Ahmedabad on Vistara with 1 stop, the median price is quite different depending on the day of the week. It's lowest on day 10 at 6101 and it's highest on day 8 at 32110. Similarly, for flights from Ahmedabad to Goa on Vistara with 1 stop, the median price is lowest on day 10 at 5214 and highest on day 8 at 8280.5.
- For flights between Goa and Tirupati on SpiceJet with 1 stop, the prices tend to be more expensive on days 8, 9, and 11. Additionally, it also appears that the prices for flights between Goa and Tirupati tend to be more expensive than flights between other destinations.
- For flights between Tirupati and Bengaluru on Star Air with 1 stop, the prices are consistent with Q1, Q3 and median fare being 7612 on day 9, and 7969 on day 11. This implies that the

fare prices do not vary much. Similarly, for the route Bengaluru to Tirupati, the prices are consistent with Q1, Q3 and median fare being 13822 on day 9, and 6612 on day 11.

- For flights between Goa to Ahmedabad on Vistara with 2 stops, the highest ticket price is observed on day 8 with 64163.0. Similarly, the highest median prices are observed on day 9 for Ahmedabad to Tirupati on Vistara with 2 stops, on day 8 for Tirupati to Mumbai on SpiceJet with 2 stops, on day 11 for Tirupati to Ahmedabad on IndiGo with 2 stops, on day 9 for Mumbai to Tirupati on IndiGo with 2 stops, on day 11 for New Delhi to Ahmedabad on IndiGo with 2 stops, on day 10 and 11 for Goa on AirAsia with 2 stops, and on day 11 for New Delhi on AirAsia with 2 stops.

Data Cleaning

1. Checking for skewness: The skewness of 'day','Stops','Duration_minutes' were checked which were not within limits. The data was transformed using power transformer to make it within limits. This helped ensure that the data is normally distributed and ready for modeling.
2. Encoding of object columns: The object columns were encoded using label encoding. This helped convert categorical variables into numerical variables, making it easier to analyze and model the data.
3. Correlation check: The correlation for numerical data was checked and it was found that 'stops' and 'duration_min' had a value of 0.9. However, checking VIF it was found to be within limits (less than 10) which means they are not correlated.
4. Outliers check: Outliers were checked using box plot and later using z-score method with no data loss. This helped identify

and remove any data points that may skew the analysis or modeling results.

Model Selection:

The target variable in this project is 'Price' which is a numerical continuous data type. This means that the goal of the project is to predict a numerical value, making it a regression problem.

Different models were used to predict the 'Price' variable:

1. DecisionTreeRegressor
2. RandomForestRegressor
3. ExtraTreesRegressor
4. GradientBoostingRegressor
5. AdaBoostRegressor

	classifier	r2	cv_score	diff	mse	mae
4	ETR	0.571828	0.610565	-0.038737	8.790282e+06	1874.266883
3	RF	0.525064	0.592563	-0.067499	9.750347e+06	1980.560832
1	GRAD	0.440001	0.550466	-0.110464	1.149666e+07	2174.638693
2	DT	-0.142031	0.323472	-0.465503	2.344567e+07	2591.167530
0	ada	-0.183906	0.038637	-0.222543	2.430535e+07	4307.625727

Based on the analysis of the different models used, it was found that ExtraTreesRegressor (ETR) performed the best. It had the highest r2_score, which is a measure of how well the model is able to explain the variability of the target variable. It also had the highest cross-validation score (cv_score), which is a measure of how well the model performs on unseen data.

Additionally, the difference between the r2_score and cv_score was the least for ETR, which indicates that the model is not overfitting and is able to generalize well to unseen data.

Furthermore, ETR had the least error value for mean squared error (MSE) and mean absolute error (MAE), which are measures of how well the model predicts the target variable.

All of these factors combined indicate that the ExtraTreesRegressor is the best model for predicting flight prices in this dataset.

Hyper Parameter tuning

Hyperparameter tuning is the process of fine-tuning the parameters of a model in order to improve its performance. In this project, different parameters were used for tuning the ExtraTreesRegressor (ETR) model. These parameters include:

1. **n_estimators**: The number of trees in the forest. A higher number of trees increases the model's accuracy but also increases the risk of overfitting.
2. **max_depth**: The maximum depth of the tree. A higher depth increases the model's accuracy but also increases the risk of overfitting.
3. **min_samples_split**: The minimum number of samples required to split an internal node. A higher value decreases the model's complexity and reduces the risk of overfitting.
4. **min_samples_leaf**: The minimum number of samples required to be at a leaf node. A higher value decreases the model's complexity and reduces the risk of overfitting.

The values of these parameters were set to different values and the model was trained and evaluated for each combination of

values. The combination of values that gave the best performance was then chosen as the final model.

```
param_grid = {'n_estimators': [150, 200,250, 300],
              'max_depth': [ 16, 17, 18, 19],
              'min_samples_split': [2, 3, 4, 5,],
              'min_samples_leaf': [1, 2, 3, 4],
              'random_state': [20]
             }
```

The parameter value obtained after running grid search

```
Best parameters:  {'max_depth': 17, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200,
'random_state': 20}
Best score:  0.6109264517441215
```

Conclusion

In conclusion, this project aimed to develop a model for predicting flight prices by analyzing historical flight fare data and other relevant factors. The data collected from MakeMyTrip.com showed that there is a lot of variation in flight prices depending on the day of the week, departure city, destination city, and airline. It was observed that prices tend to vary depending on the destination and day of travel, with some destinations having prices that are consistent across multiple days, indicating that prices do not vary much. Additionally, prices tend to be cheaper on certain days and more expensive on others. However, it was also noted that limited data is available for certain days, which could be due to flights not being operational on certain days, low demand on certain days, or data not being present in the dataset provided.

The analysis of the data also showed that prices are homogenous across days in certain cases, which could indicate that there is no significant change in prices across days. However, factors such as seasonality, holidays and special

events can also affect flight prices, which may not be captured in the data used in this project. Despite the limitations, the ExtraTreesRegressor (ETR) model was found to be the best model with the highest r2_score and cv_Score, least difference between cv_sccore and r2_score, and least error value for mse and mae. Overall, this project highlights the importance of understanding the factors that influence flight prices and how they vary over time, in order to make informed decisions when booking a flight.

Further Studies

Further studies on this project could include analyzing the impact of external factors such as holidays, festivals, and special events on flight prices. This could involve collecting data for a longer period of time, or for different months or seasons to see how prices change. Another area of focus could be to analyze the data by different segments such as travel class, as this could provide insight into how prices vary for different types of seats or classes. Additionally, studying the impact of airport taxes and fuel prices on flight prices could also be an interesting area to explore, as these factors can have a significant impact on the overall cost of a flight. A survey of consumers could also be conducted to understand their behavior and preferences while booking flights, which could provide valuable insights into how prices are perceived by consumers and how they make decisions about booking flights.