FINAL PROJECT

http://flightanalyst-env.us-east-1.elasticbeanstalk.com/

FLIGHT ANALYSIS FOR US

INFO 7390: Advances in Data Sciences/Architecture
Team 9

Bhavesh Patel
Ila Nigam
Ayushi Srivastava

Table of Contents

Introduction

1. *Background and Problem description*

Every year approximately 20% of airline flights are delayed or cancelled, costing travelers over 20 billion dollars in lost time and money. As we will see, some flights are more frequently delayed than others, and there is an interest in providing this information to travelers. As delays are a stochastic phenomenon, it is interesting to study their entire probability distributions, instead of looking for an average value. Many factors affect flight delays including air traffic control backups, equipment delays, and weather.

Our goal was to leverage the massive amount of data available on flight punctuality and reasons for delay and cancellation to forecast whether or not a flight will be delayed/cancelled. We also did analysis on prices to help customers to analyze the price changes for flights; they can analyze the delay and cancellation of flights and can compare various carriers in US under the various criteria through visualization

2. *Dataset Description*

The dataset is an airlines' data collection coming from the Bureau of Transportation Statistics (BTS) Airline On-time Performance dataset and it contains detailed facets of each air flight information between 2014-2015. It is huge information which includes 21 variables like Destination, Origin, Arrival time, and Departure time and so on. Here is an original list that shows the all variables. It is very important statistical records that any flight information could be tracked via special features. Our selected dataset is still having millions of observations which are definitely enough to obtain the satisfying outcomes.

The dataset contained csv files which had the monthly data. Since the number of records were huge in each month (around 0.5 million), we decided to limit the scope of the dataset to only 1 year of data i.e. 2014. Giving a brief idea about the data in each csv file, it had around 110 columns with a very detailed and granular data. But based on our business case scenario, we were focused more on building an application which our direct customers would be benefitted and can use on the go. For this we manually extracted 21 columns according to business context which would mostly help in making a decision about price, delay and cancellation of flights.

Here is a descriptive list of useful variables that we extracted out of the original dataset.

1. Day of Month - e.g. December 1st to December 31th.
2. Day of Week - 1 Refers to Monday and in a similar way, 7 refers to Sunday.
3. Departure Time Actual departure time
4. Arrival Time Actual arrival time
5. CRS Departure Time - Scheduled departure time
6. CRS Arrival Time - Actual arrival time
7. Unique Carrier - Unique carrier code
8. Flight Number- Flight number
9. Arrival Delay 15 – Binary information to determine whether the delay was more than fifteen minutes or not
10. Departure Delay 15 – Binary information to determine whether the delay was more than fifteen minutes or not
11. Year
13. Month
14. Day
15. Arrival Delay – Actual Arrival delay, in minutes
16. Departure Delay - Actual Departure delay, in minutes
17. Origin - Origin IATA airport code
18. Destination - Destination IATA airport code
19. Cancellation – Binary classifier whether a flight was cancelled or not
20. Cancellation Code – Reason for cancellation
21. Price – Cost of flight ticket

3. Business Case

➢ *Flight Price Prediction*

The first business case would be to predict the prices for the flights for their destination.

- *Business Problem: -*
  Most of the time we spend on comparing the prices of the flight ticket before we even book. This process takes a long time and is very tiring. Pricing is one of the most vital and highly demanded component. It helps consumers to have an image of the standards the firm has to offer through their prices and to have an exceptional reputation in the market. The Carrier's decision on the price of the ticket and the pricing strategy impacts the consumer's decision on whether or not to purchase the ticket.

- *Solution: -*
  So what if we provide the user with predictions on the flight tickets for their dates beforehand? This would help them make some important decision on whether to book the ticket or to wait for some time if they are flexible with their travel dates. We build an application which gives the consumers a common platform wherein they can compare future flight prices for multiple airlines in advance.

- *Outcome: -*
  This would give consumers the best deals as the application may provide them with additional prices for dates above and below their travel date and save their cost of travel which they can spend it wisely somewhere else.

- ***Arrival Delay Time Prediction***

The second business case would be to predict the arrival delay time for the flights.

- *Business Problem: -*

Flight delays are an inconvenience to passengers. A delayed flight can be costly to passengers by making them late to their personal scheduled events. A passenger who is delayed on a multi-plane trip could miss a connecting flight. Anger and frustration can occur in delayed passengers.

In the United States, passengers are not entitled to compensation when a delay occurs, not even a cut of fees airlines must pay federal authorities for long delays. Airlines are required to pay for lodging costs of passengers if the delay or a cancellation is through their own fault, but not if the cause is beyond their control, such as weather

- *Solution: -*
  Passengers are more concerned of their delay in arrival rather than if they are departing late. So what if we provide our passengers a real time application wherein they can get updates on how much delay would be there in their arrival while they are in flight.

- *Outcome: -*
  By getting their arrival delay time they can plan their personal schedules events or business meetings accordingly which would help them stay up to date with their future events. This would also help people to cancel their connecting flights if they get to know whether they would reach on time or not and accordingly can schedule another flight to arrive at their destination on time. Moreover, most important this would provide a better customer satisfaction and could help in minimizing the frustration of some of the customers.

- ***Flight Cancellation Prediction***

The third business case would be to predict whether a flight would be cancelled or not.

- *Business Problem: -*

Business trips are expensive and passengers cannot afford to miss those which cost them in millions of dollars in some cases. If the cancellation occurs due to airlines fault, the passengers would be refunded but what if passenger who has a connecting flight of different airline? This may cost them as the connecting flight cost might not be refunded. This may lead to dissatisfying the customer and reduce the possibility of travelling through that airline again.

- *Solution: -*

  So what if the passengers get a real time update of the chances of their flights being cancelled. We build an application which would help the consumers would know about the possibility of the cancellation.

- *Outcome: -*

  Business class passengers might have a good use of they cannot take a risk of flight getting cancelled and this can help them to reschedule a different flight or take necessary actions well in advance if they get to know that the flight might get cancelled. Airlines can also maintain their reputation if they let the passengers know of the chances of the flight getting cancelled.

4. *Data Visualization*

For visualization, we have used two years of data viz. 2014 and 2015. Total number of records exceeded more than 10 million.

The visualization was done using Power BI and have utilized the custom power BI visuals as well.

Flight Counts

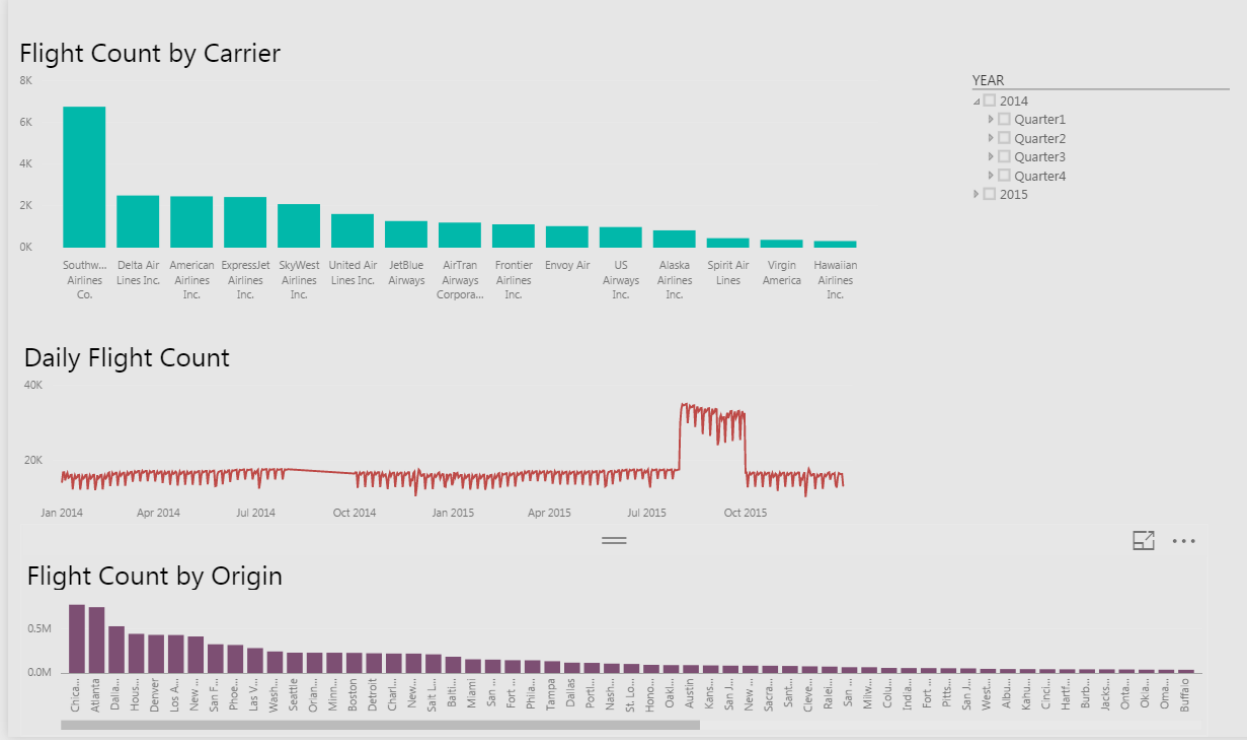


Fig 1.1

- *The first visualization describes the total unique flights that each carrier has throughout the year. Southern Airline Corporation has the most number of flights being owned and running*
- *The second visualization showed the total number of flights running altogether from all carrier for each month.*

- *The third visualization shows how many flights are originating from a location*
- *The visualization can be viewed for different year, quarters and months through the timeline slicer on the top right*

Cancellation and Delays

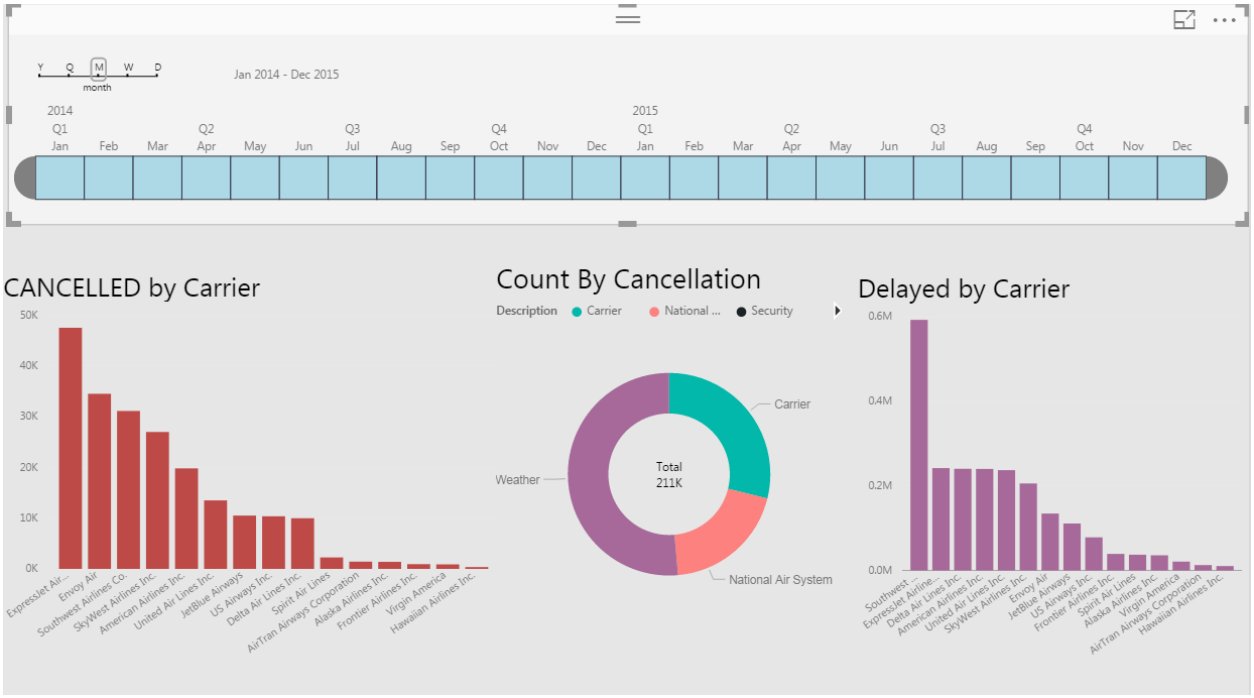


Fig 1.2

- *The first visualization shows total flights cancelled for each carrier which would let us know how good or bad an airline is.*
- *The second visualization categorized the flight cancellation to get a more insight on why the flight was cancelled.*
- *The third visualization shows how many flights were delayed per carrier which is also one of the factors deciding the reliability of an airline*
- *The visualization can be viewed for different year, quarters, months, week and day through the timeline at the top.*

Cancellation by Origin and Destination

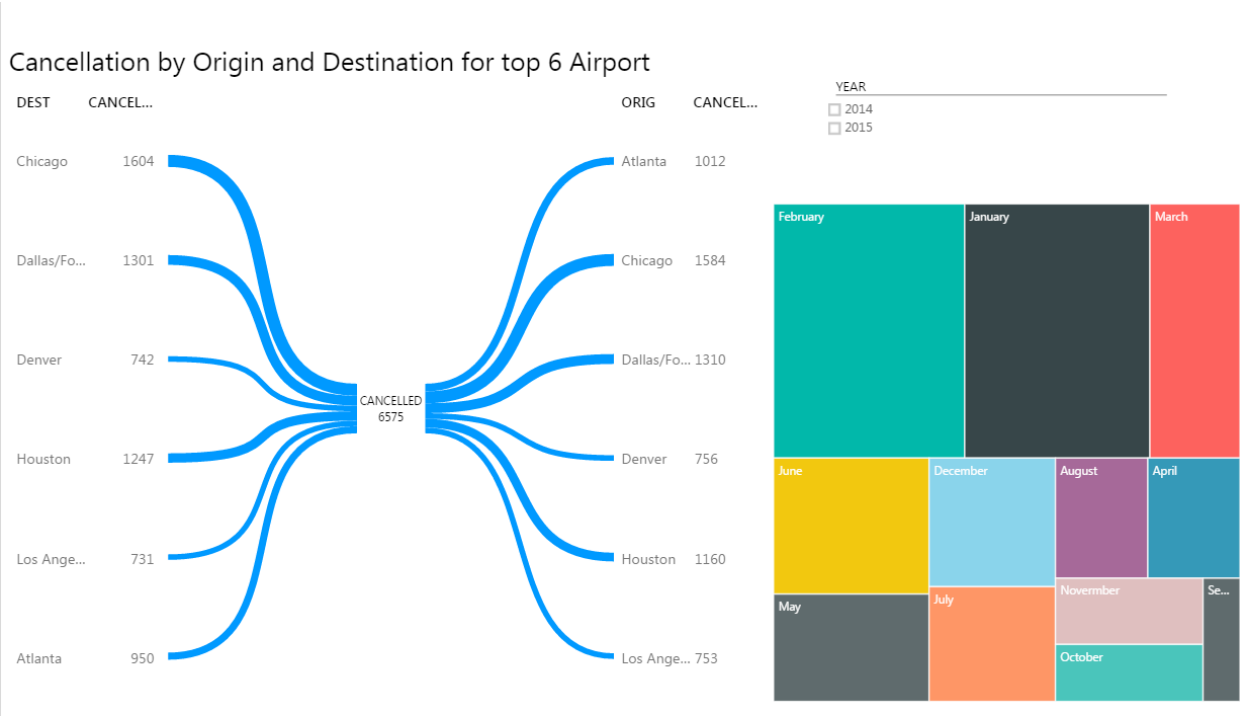


Fig 1.3

- *The first visualization is a custom visual demonstrating the number of flights being cancelled that were originating from a location and also the number of fights being cancelled which were destining at a particular location. E.g. 1012 flights heading from Atlanta were Cancelled and 1604 flights which were going to Chicago were cancelled. This visualization shows which are the places where most of the flights gets cancelled.*
- *The second visualization shows the number of flights getting cancelled at the origin for each month combining the year.*
- *Moreover, the visualization can be made for 2014 and 2014 using the timeline at the top right corner*

Busiest Airports based on number of flights Originating

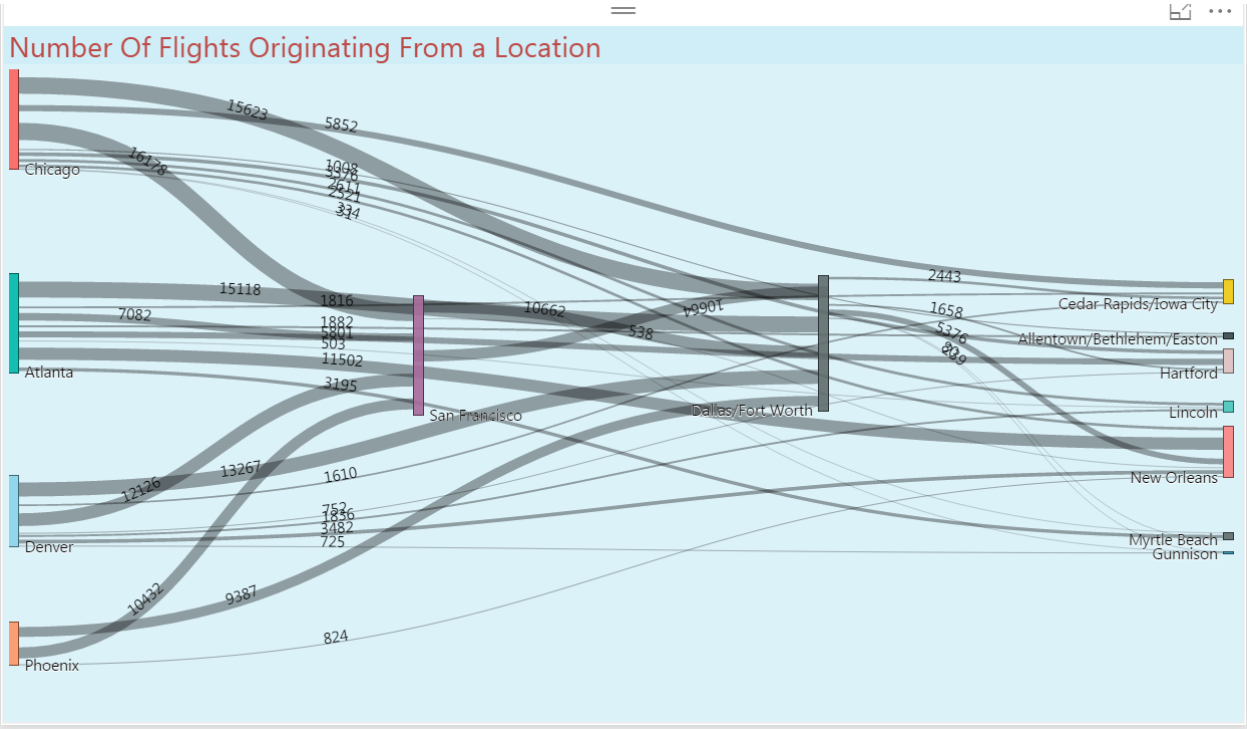


Fig 1.4

- *This visualization is a custom visual which shows the top Busiest Airports based on the frequency of flights originating from the location*
- *This also shows the destination where the most of the busiest airport flights head to.*
- *More thicker the lines from Origin t destination, busier the airport is.*
- *This visualization can help on deciding which places have much of the Air Traffic*

Hottest Places to Visit

Fig 1.5

- *The visualization is a custom power BI Globe map which shows how flights are travelling to a particular destination*
- *More the red area, more the number of flights travelling there.*
- *The visualization can also be interpreted by saying that most of the people travel to the coastal areas of the Country so may be those are the holiday destination*

5. Data Wrangling and Cleaning

1. The initial step was uploading the data in Azure from the source and creating a dataset with column names we require.
2. The major part of data transformation was done through R-script and as we worked on 12 different sheets of 12 months' data, we first combined all the data sheets into single sheet for further data processing and modelling.
3. We visualized each and every field in data set which helped us in understanding the nature of the data. Then we started excluding all the redundant, missing, junk data from the dataset.
4. Cancellation code column has missing data. But this missing data is actually not a missing data as the ones which has missing values are then ones wherein the flights are not cancelled. So we filled this data by filling the missing values as On-Time.
5. The departure delay new and arrival delay new fields had missing data but these were also not actually missing data. Another column named departure delay 15 and arrival delay 15 which specifies a binary classifier saying whether a flight was delayed or not. For all the field in departure delay new which were blank had 0 in the dep delay 15 column which implied the flight was not delayed so we filled the missing data in departure delay new and arrival delay new as 0 as there were 0 minutes of delay.
6. The real missing data were in the price column. Since we had around half a million records, the amount of missing data in price column were comparatively negligible which helped us in deciding to remove the rows with missing price values.
7. Then according to the nature of the data i.e. the datatypes, we used edit metadata block multiple times so to set the correct datatypes in azure and made some of the variables as categorical.
8. Finally, we trained the model on 90% of the data and tested in 10% of the data. The reason for choosing 90% train data was that we wanted to train the model which gives an output of high accuracy and because more the data more the model would be trained accurately to get the results as we cannot afford to make any wrong predictions on cancellation as these are very critical business problems.

Price Prediction Model

- ***Procedure and Approach***

1. Price amount for 2014 was given in the data set and we build the model to forecast and predict the price based on the model.
2. On evaluating the data we found that the main predictors were Origin, destination, start date, end date, carrier to predict the price.
3. Next we sampled the data into Test and Train to model the train data and then evaluate our model on based of Test data set.
4. Following Models were applied:
    - Neural Network Regression
    - Linear regression
    - Boosted Decision Tree Regression
    - Poisson Regression
    - Decision Forest Regression
    - Bayesian Linear Regression
5. The models were built and evaluated using varied combinations of factors in each algorithms so as to get the best results and then we calculated the performance matrix.
6. Graphs were plotted to evaluate the performance of the models.
7. In Azure the best model was hosted as web service and was integrated with the web application to predict the price for the given input.

- ***Prediction Analysis***

1. **Neural Network Regression**

Metrics

Mean Absolute Error	98.706655
Root Mean Squared Error	128.704774
Relative Absolute Error	0.409291
Relative Squared Error	0.148412
Coefficient of Determination	0.851588

2. Linear regression

Metrics

Mean Absolute Error	148.394157
Root Mean Squared Error	213.458665
Relative Absolute Error	0.615323
Relative Squared Error	0.408233
Coefficient of Determination	0.591767

3. Boosted Decision Tree Regression

Metrics

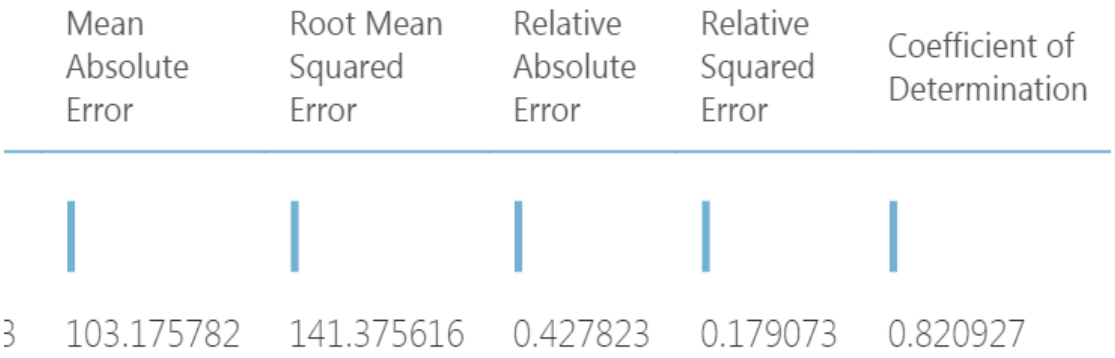
Mean Absolute Error	113.348325
Root Mean Squared Error	154.192362
Relative Absolute Error	0.470004
Relative Squared Error	0.213013
Coefficient of Determination	0.786987

4. Poisson Regression

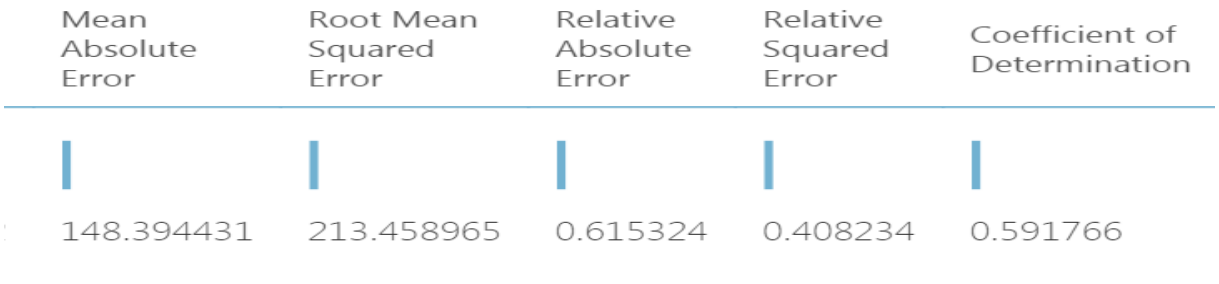
Metrics

Mean Absolute Error	139.313867
Root Mean Squared Error	204.273032
Relative Absolute Error	0.577671
Relative Squared Error	0.373854
Coefficient of Determination	0.626146

5. Decision Forest Regression



6. Bayesian Linear Regression



- ***Comparison of Models***

  On comparison we found that Neural Network Model worked best as RMSE was minimum for this Model.

	Neural Network	**Linear regression**	**Boosted Decision Tree**	**Poisson**	**Decision Forest**	**Bayesian Linear**
MAE	***98.70***	***148.39***	***113.34***	***139.31***	***103.17***	***148.39***
RMSE	***128.70***	***213.45***	***154.19***	***204.27***	***141.37***	***213.45***
Relative Absolute error	***0.40***	***0.61***	***0.47***	***0.57***	***0.42***	***0.61***
Relative squared error	***0.14***	***0.40***	***0.21***	***0.37***	***0.17***	***0.40***
Coefficient of	***0.85***	***0.59***	***0.78***	***0.62***	***0.82***	***0.59***

Determination						

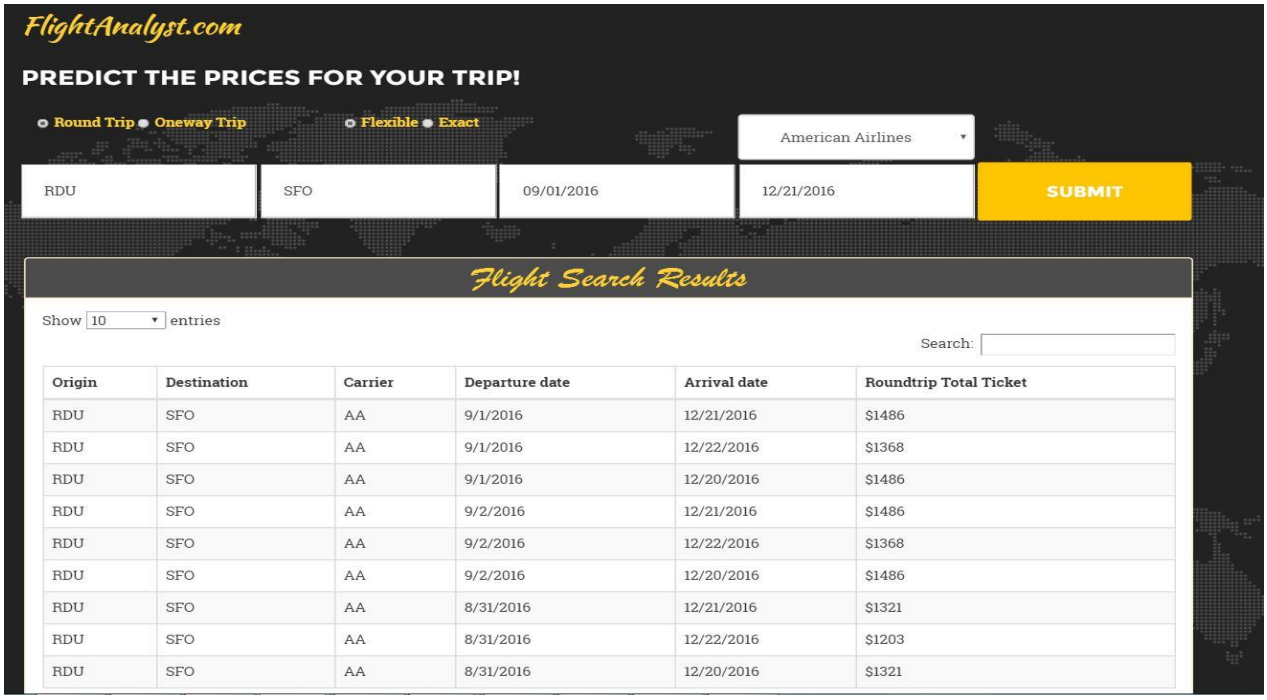
Conclusion: -

- ✓ **With the results derived above we can conclude that Neural Network has best Performance metrics, hold best results.**
- ✓ **Also, the RMSE supports neural network model the most accurate model on given dataset.**
- ✓ **In neural network, the prediction of price for next given date was very appropriate and could be verified against the trends of price range.**

➢ *Web Application Integrated and Use Cases*

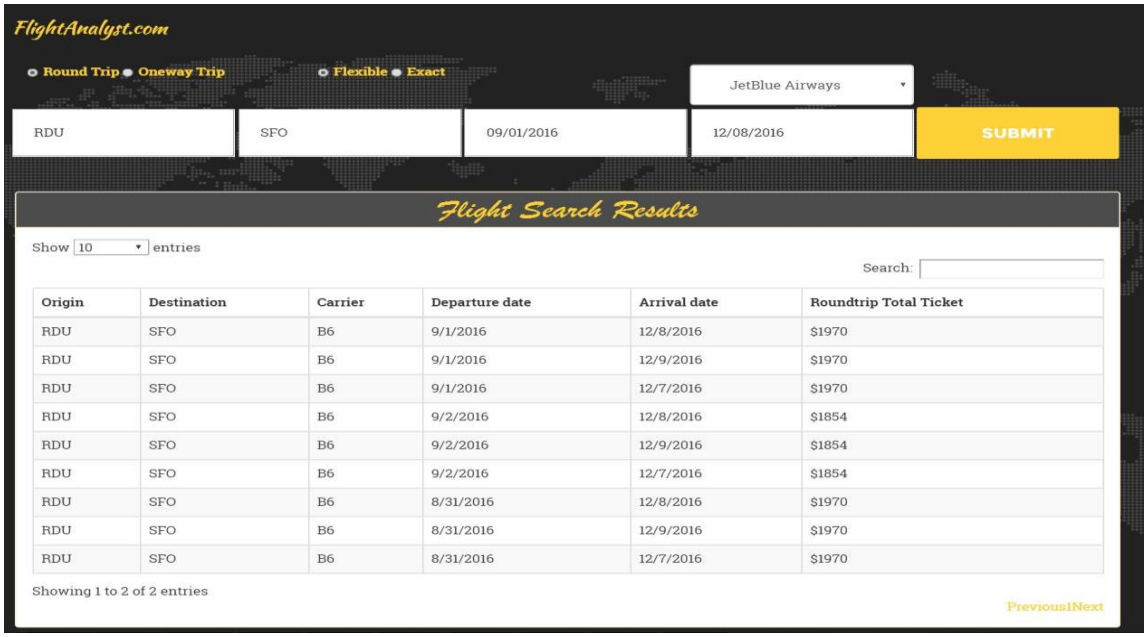
Use cases: -

1) **Price prediction for roundtrip journey for a date entered by the user along with giving them additional prices for other dates in the proximity**

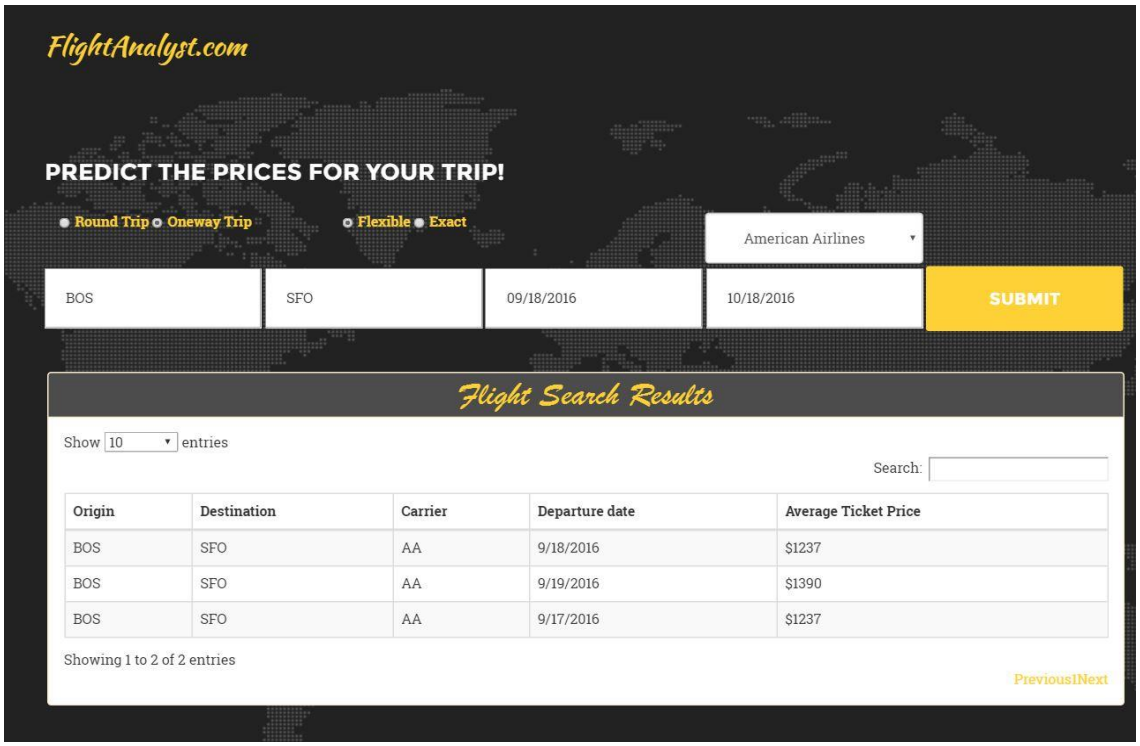


Origin	Destination	Carrier	Departure date	Arrival date	Roundtrip Total Ticket
RDU	SFO	AA	9/1/2016	12/21/2016	$1486
RDU	SFO	AA	9/1/2016	12/22/2016	$1368
RDU	SFO	AA	9/1/2016	12/20/2016	$1486
RDU	SFO	AA	9/2/2016	12/21/2016	$1486
RDU	SFO	AA	9/2/2016	12/22/2016	$1368
RDU	SFO	AA	9/2/2016	12/20/2016	$1486
RDU	SFO	AA	8/31/2016	12/21/2016	$1321
RDU	SFO	AA	8/31/2016	12/22/2016	$1203
RDU	SFO	AA	8/31/2016	12/20/2016	$1321

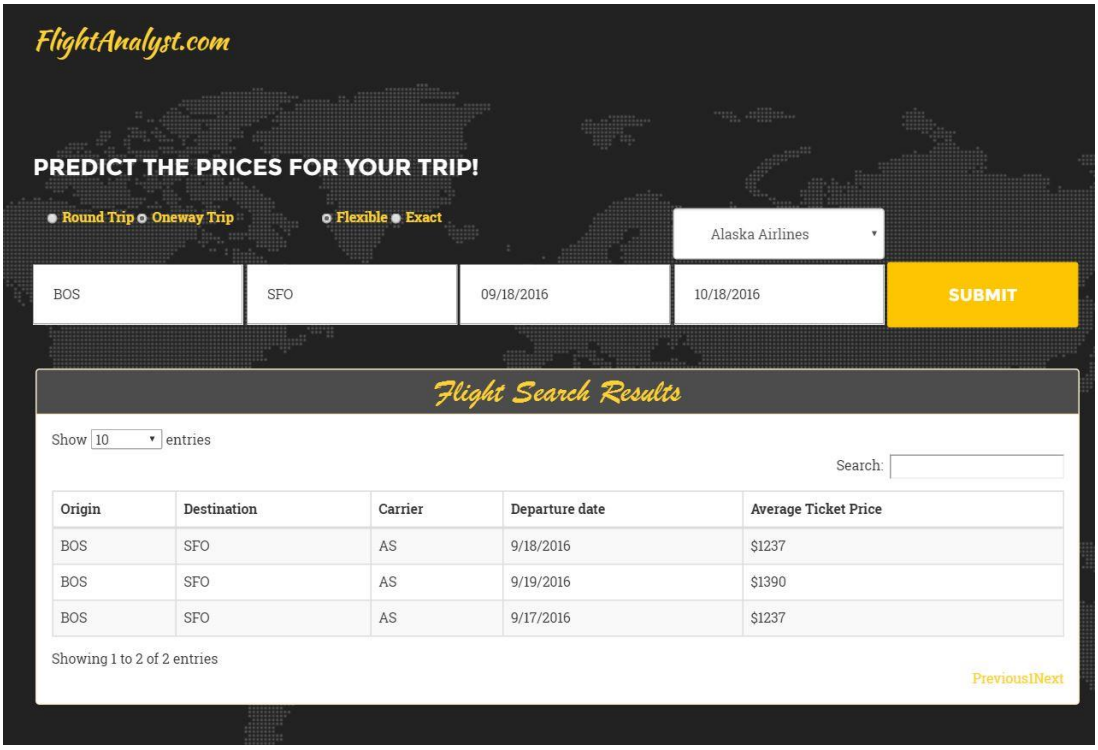
2) Price prediction for Different Airlines



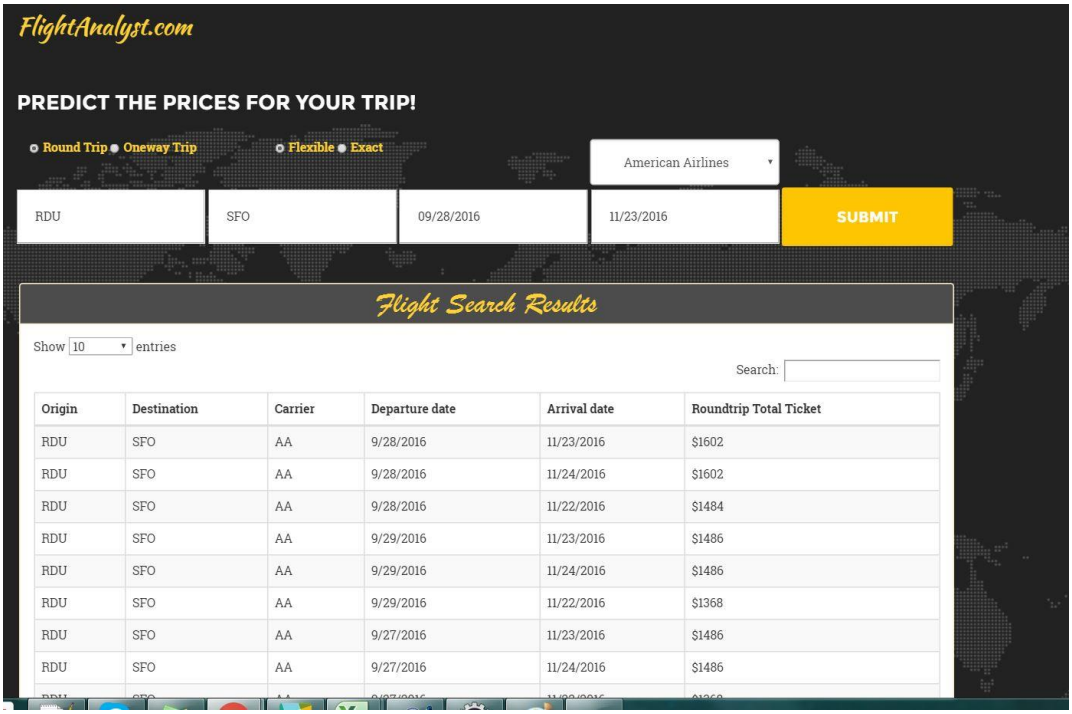
3) Price Prediction for one-way trip



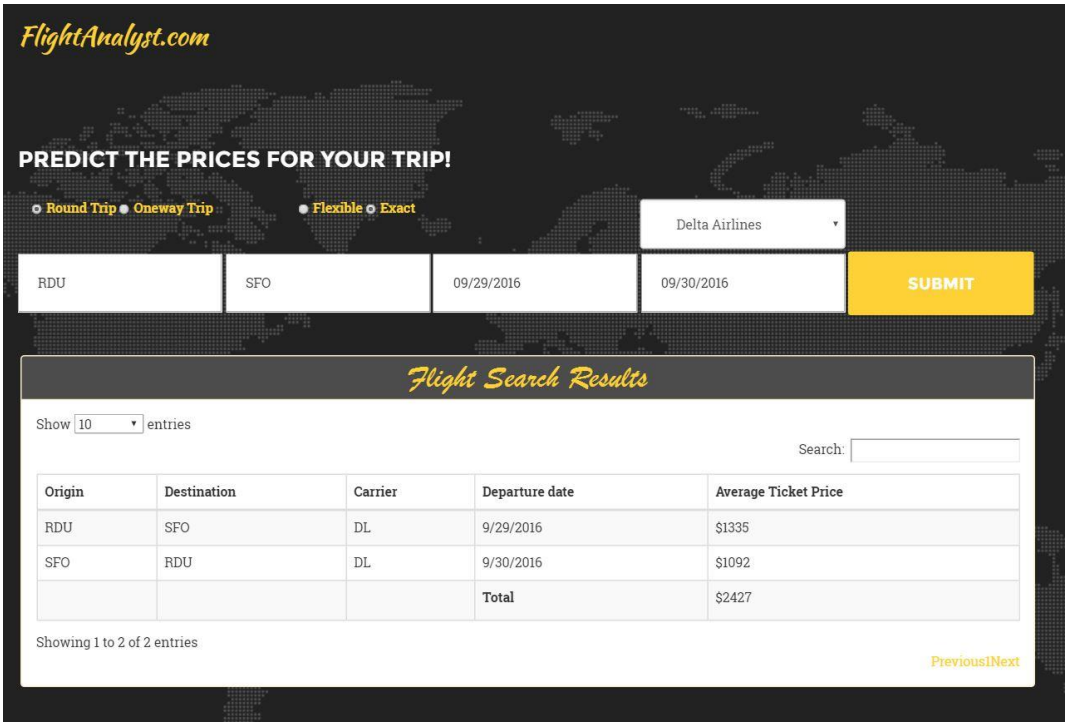
4) Price prediction for one way trip with flexible dates for different airlines



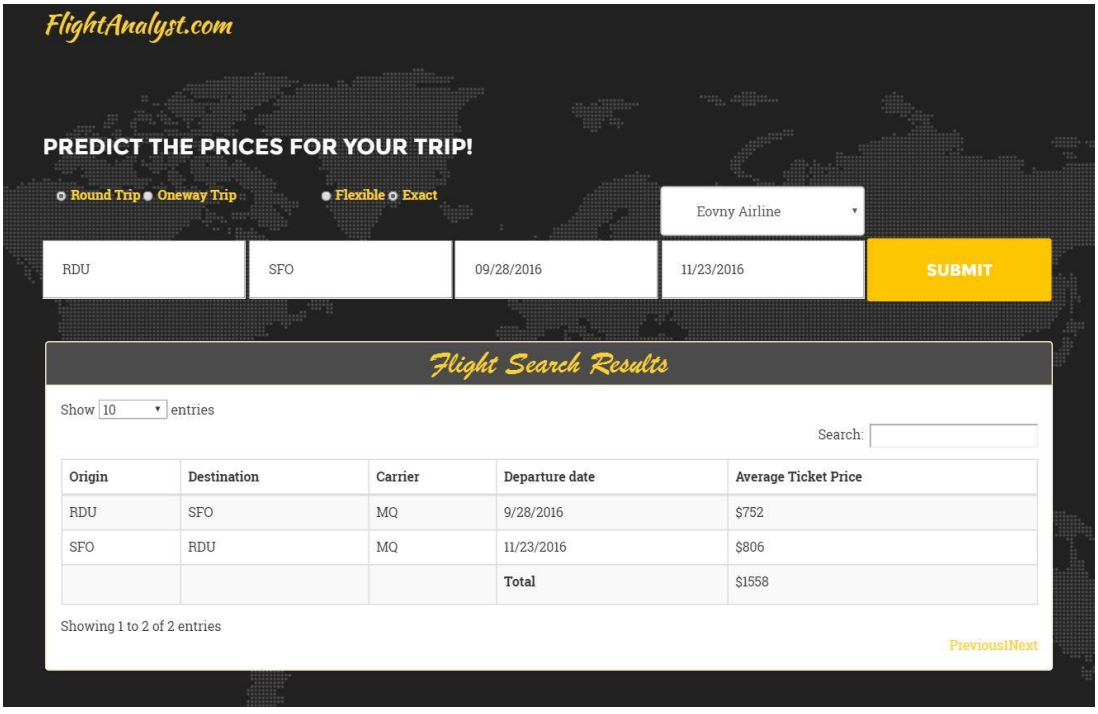
5) Price prediction with different origin and destination

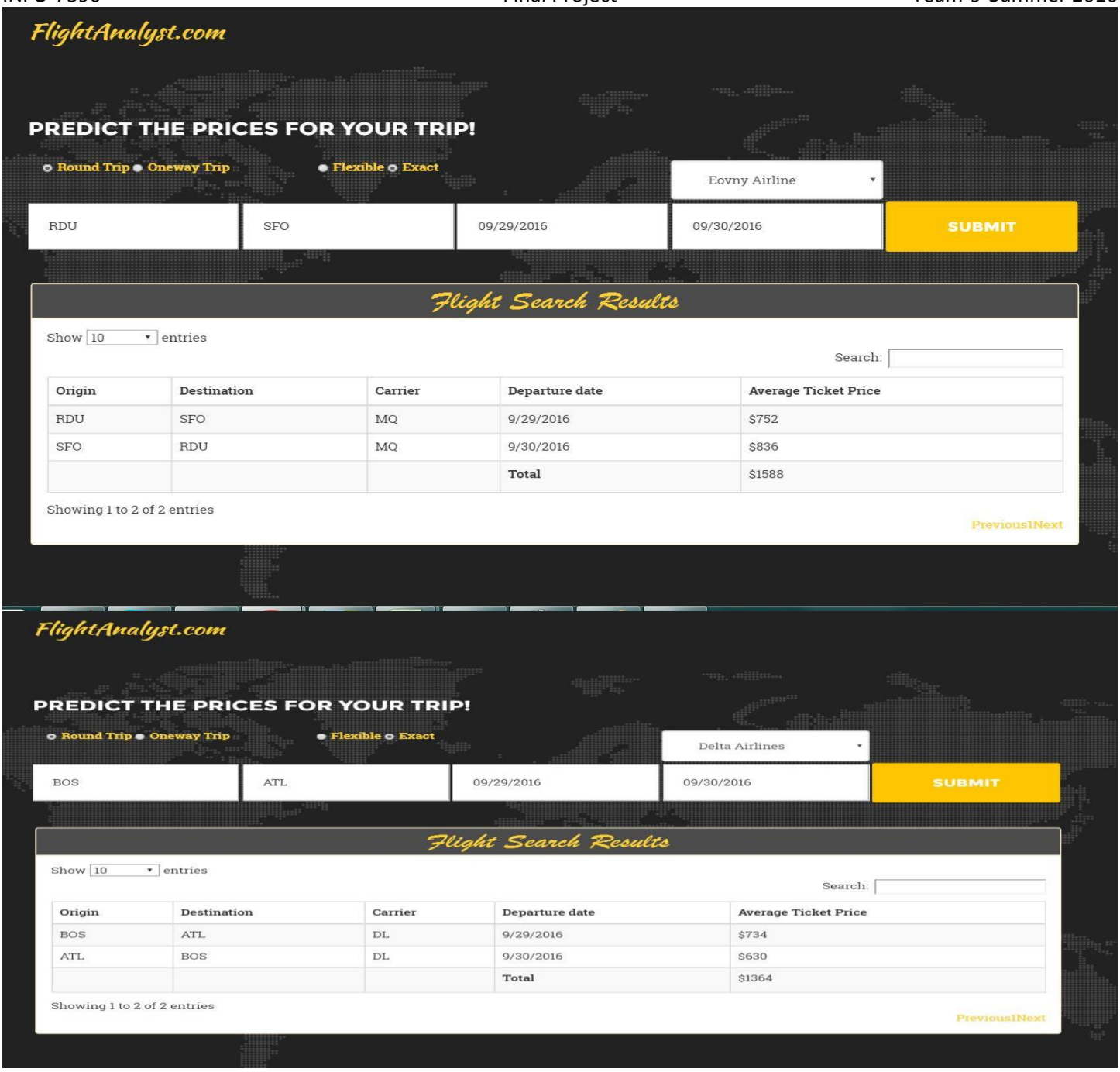


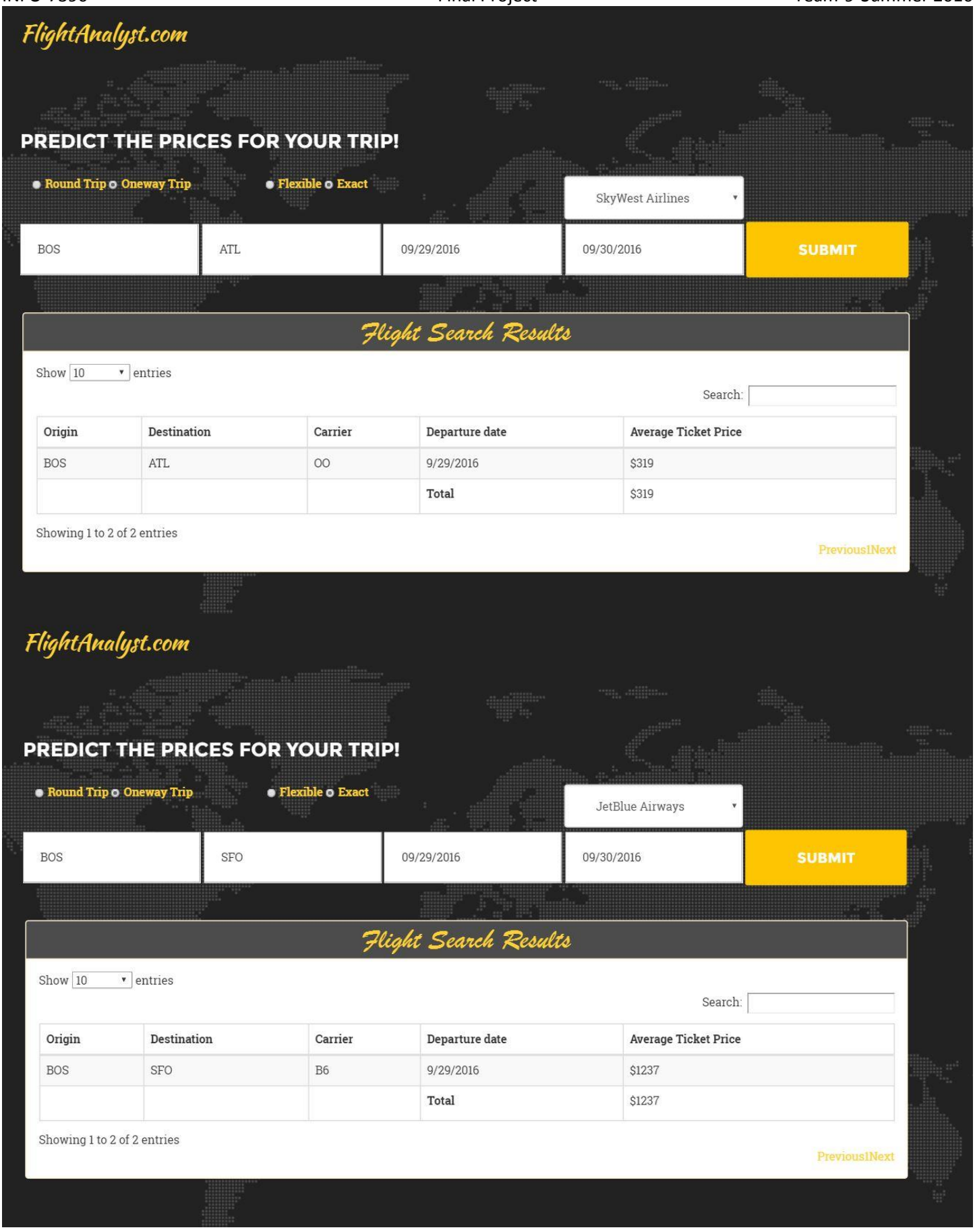
6) Price prediction for roundtrip with exact date

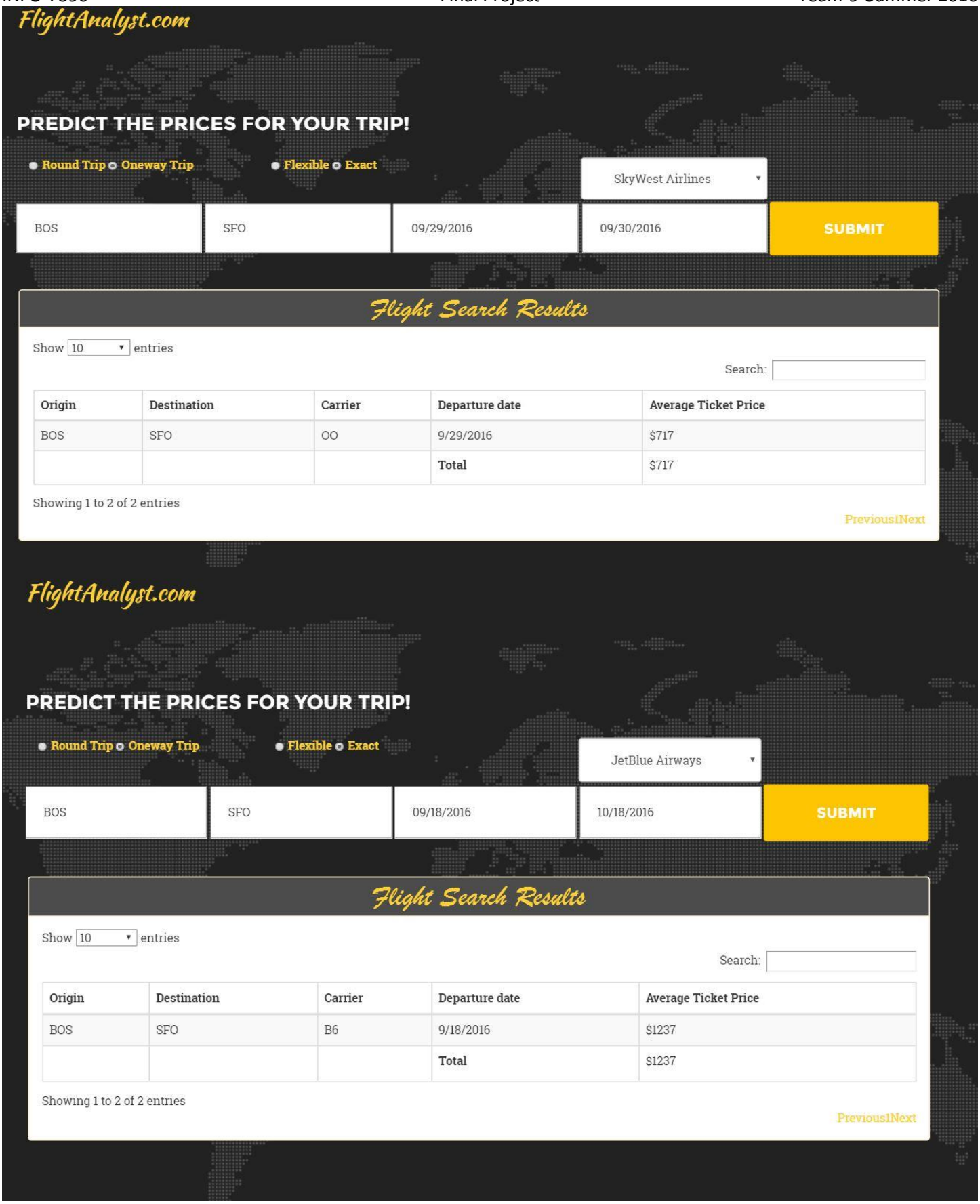


7) Price prediction for different dates




FlightAnalyst.com
PREDICT THE PRICES FOR YOUR TRIP!
Round Trip Oneway Trip
Flexible Exact
Eovny Airline
RDU
SFO
09/29/2016
09/30/2016
SUBMIT
Flight Search Results
Show 10 entries
Search:
Origin Destination Carrier Departure date Average Ticket Price
RDU SFO MQ 9/29/2016 $752
SFO RDU MQ 9/30/2016 $836
Total $1588
Showing 1 to 2 of 2 entries
Previous1Next
FlightAnalyst.com
PREDICT THE PRICES FOR YOUR TRIP!
Round Trip Oneway Trip
Flexible Exact
Delta Airlines
BOS
ATL
09/29/2016
09/30/2016
SUBMIT
Flight Search Results
Show 10 entries
Search:
Origin Destination Carrier Departure date Average Ticket Price
BOS ATL DL 9/29/2016 $734
ATL BOS DL 9/30/2016 $630
Total $1364
Showing 1 to 2 of 2 entries
Previous1Next


FlightAnalyst.com
PREDICT THE PRICES FOR YOUR TRIP!
Round Trip Oneway Trip
Flexible Exact
SkyWest Airlines
BOS
ATL
09/29/2016
09/30/2016
SUBMIT
Flight Search Results
Show 10 entries
Search:
Origin Destination Carrier Departure date Average Ticket Price
BOS ATL OO 9/29/2016 $319
Total $319
Showing 1 to 2 of 2 entries
Previous1Next
FlightAnalyst.com
PREDICT THE PRICES FOR YOUR TRIP!
Round Trip Oneway Trip
Flexible Exact
JetBlue Airways
BOS
SFO
09/29/2016
09/30/2016
SUBMIT
Flight Search Results
Show 10 entries
Search:
Origin Destination Carrier Departure date Average Ticket Price
BOS SFO B6 9/29/2016 $1237
Total $1237
Showing 1 to 2 of 2 entries
Previous1Next


FlightAnalyst.com
PREDICT THE PRICES FOR YOUR TRIP!
Round Trip Oneway Trip
Flexible Exact
SkyWest Airlines
BOS
SFO
09/29/2016
09/30/2016
SUBMIT
Flight Search Results
Show 10 entries
Search:
Origin Destination Carrier Departure date Average Ticket Price
BOS SFO OO 9/29/2016 $717
Total $717
Showing 1 to 2 of 2 entries
PreviousINext
FlightAnalyst.com
PREDICT THE PRICES FOR YOUR TRIP!
Round Trip Oneway Trip
Flexible Exact
JetBlue Airways
BOS
SFO
09/18/2016
10/18/2016
SUBMIT
Flight Search Results
Show 10 entries
Search:
Origin Destination Carrier Departure date Average Ticket Price
BOS SFO B6 9/18/2016 $1237
Total $1237
Showing 1 to 2 of 2 entries
PreviousINext

Arrival Delay Time Prediction Model

- ***Procedure and Approach***

1. Delay data of US flights for 2014 was given in the data set which we had saved in azure beforehand after cleaning in price prediction modeling and we build the model to forecast and predict the delay on the model.
2. On evaluating the data, we found that the main predictors were Origin, destination, Date, Day of Week, Quarter, Carrier, Departure Delay New, DepDelay15 and Flight Number to predict the arrival delay time.
3. Next we sampled the data into Test and Train to model the train data and then evaluate our model on based of Test data set.
4. Following Models were applied:
    a. Linear regression
    b. Boosted Decision Tree Regression
    c. Decision Forest Regression
    d. Bayesian Linear Regression
5. The models were built and evaluated using varied combinations of factors in each algorithms so as to get the best results and then we calculated the performance matrix.
6. Graphs were plotted to evaluate the performance of the models.
7. In Azure the best model was hosted as web service and was integrated with the web application to predict the delay of Flight for the given input.

- ***Prediction Analysis***

1. Linear regression

Metrics

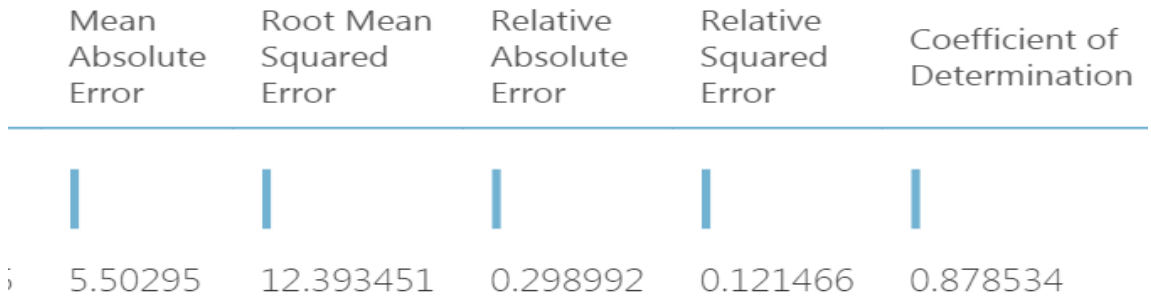
Mean Absolute Error	4.996491
Root Mean Squared Error	9.846345
Relative Absolute Error	0.271474
Relative Squared Error	0.076669
Coefficient of Determination	0.923331

2. Boosted Decision Tree Regression

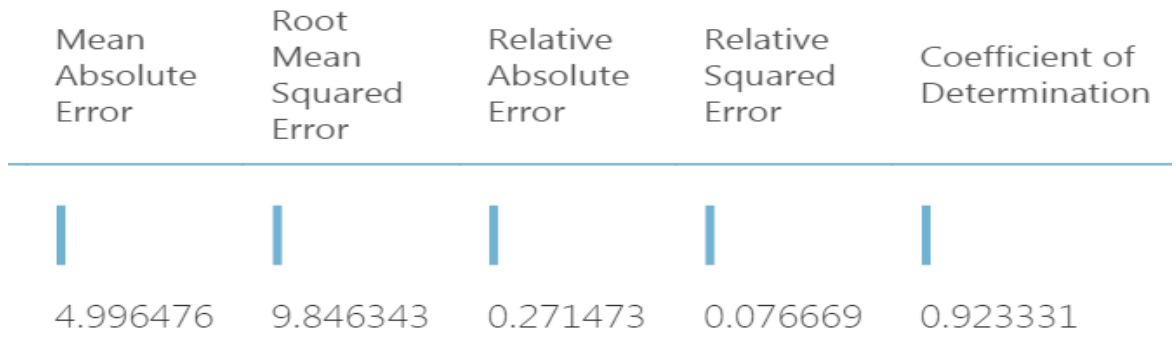
Metrics

Mean Absolute Error	4.963286
Root Mean Squared Error	9.915578
Relative Absolute Error	0.26967
Relative Squared Error	0.077751
Coefficient of Determination	0.922249

3. Decision Forest Regression



4. Bayesian Linear Regression



- ***Comparison of Models***

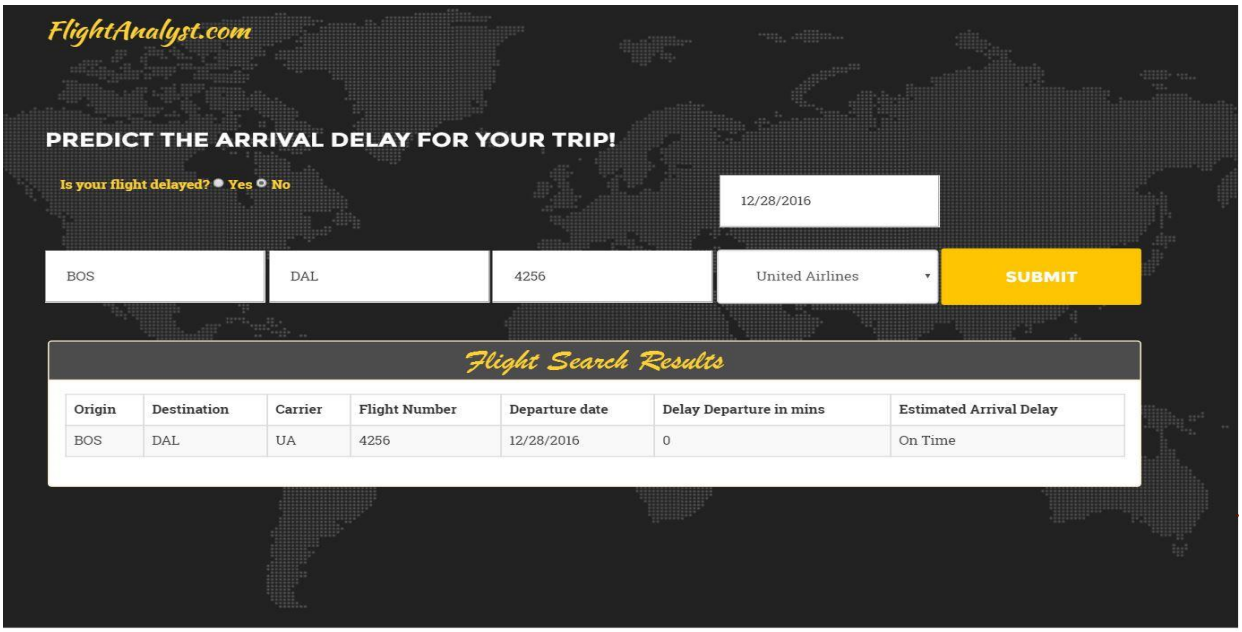
On comparison we found that Boosted Decision Tree Model worked best as RMSE was minimum for this Model.

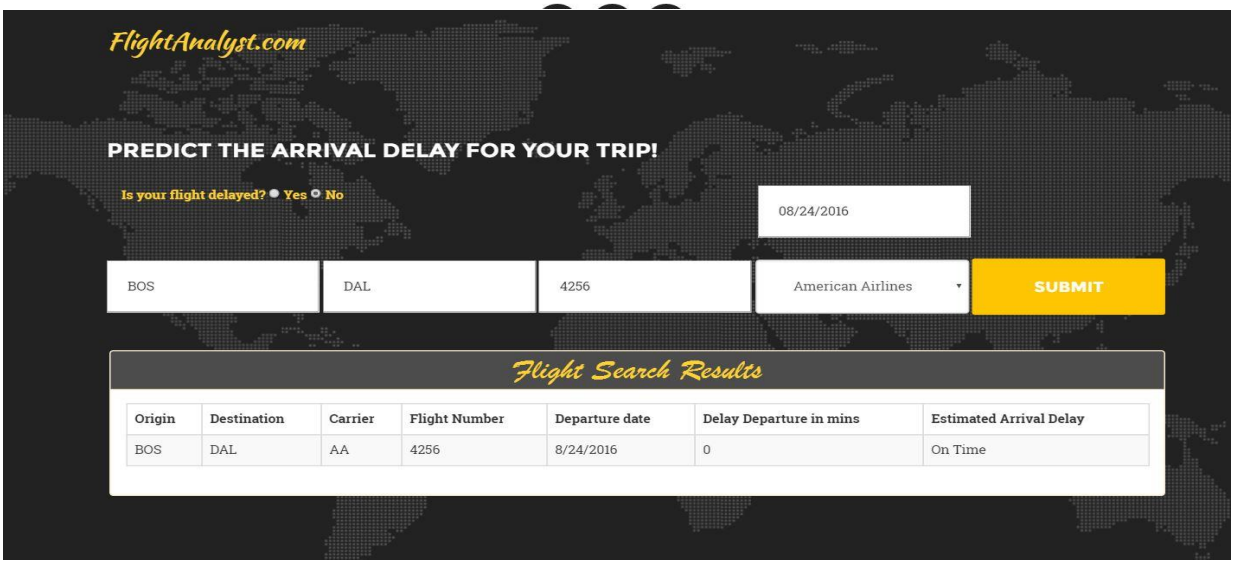
	Bayesian Linear	**Linear regression**	**Boosted Decision Tree**	**Decision Forest**
MAE	***4.99***	***4.99***	***4.96***	***5.5***
RMSE	***9.84***	***9.84***	***9.91***	***12.39***
Relative Absolute Error	***0.27***	***0.27***	***0.26***	***0.29***
Relative Squared error	***0.07***	***0.07***	***0.07***	***0.12***
Coefficient of Determination	***0.92***	***0.92***	***0.92***	***0.87***

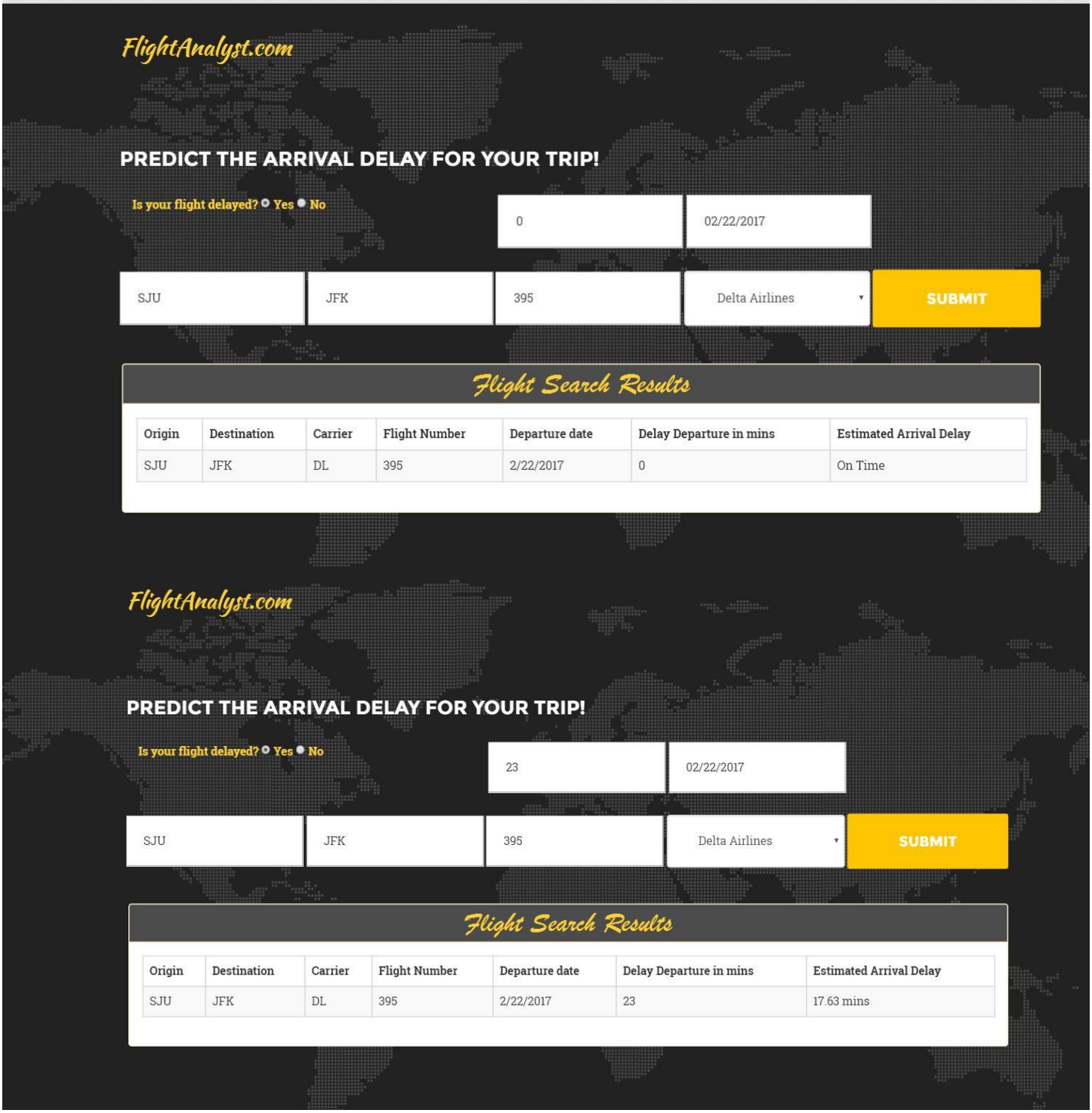
Conclusion: -

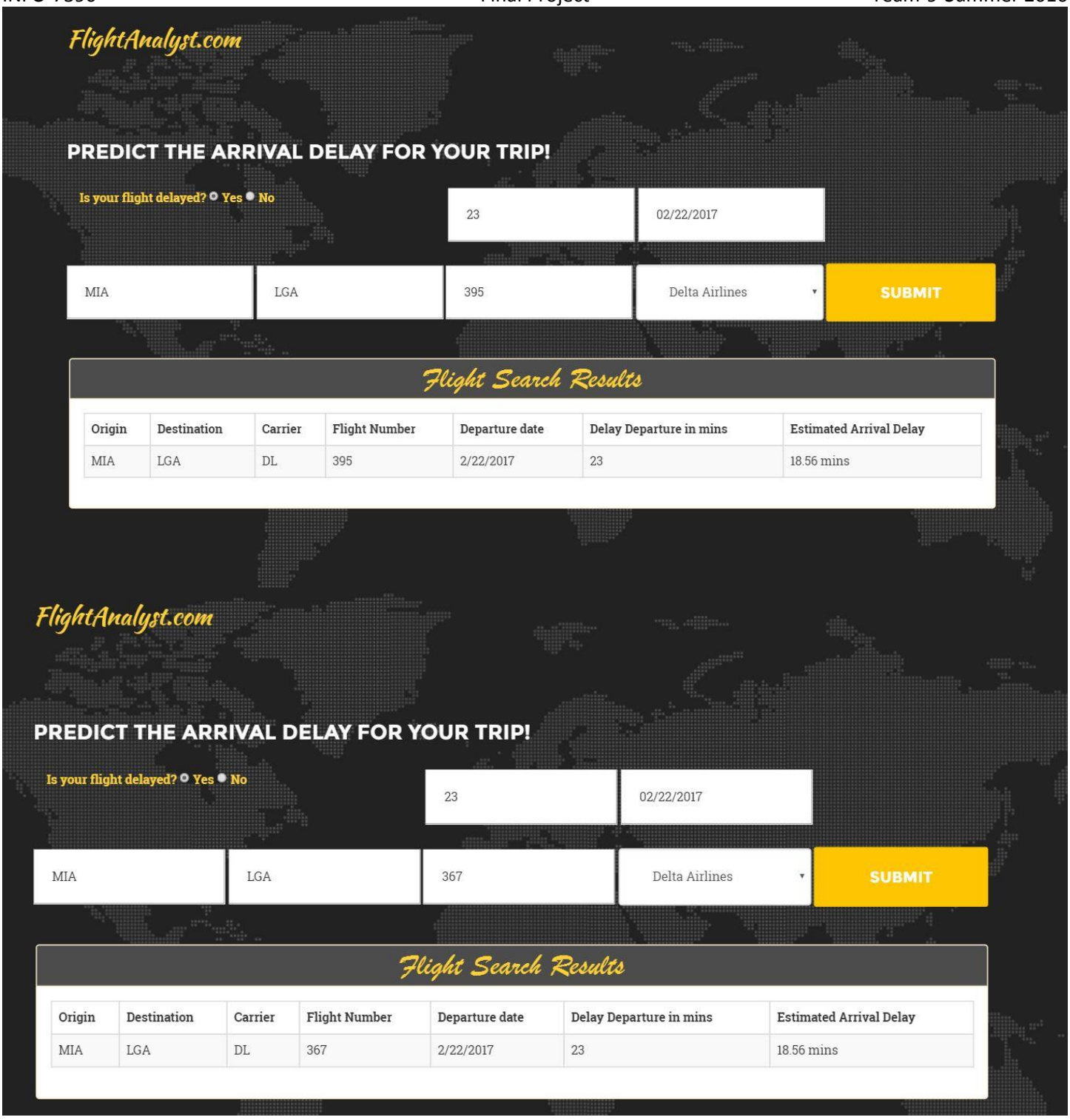
- ✓ **With the results derived above we can conclude that *Boosted Decision Tree Model* has best Performance metrics,**
- ✓ **Also, the performance metrics analysis supports *Boosted Decision Tree* model the most accurate model on given dataset.**
- ✓ **In prediction of arrival delay time were almost accurate as the errors were minimum in most of the cases**

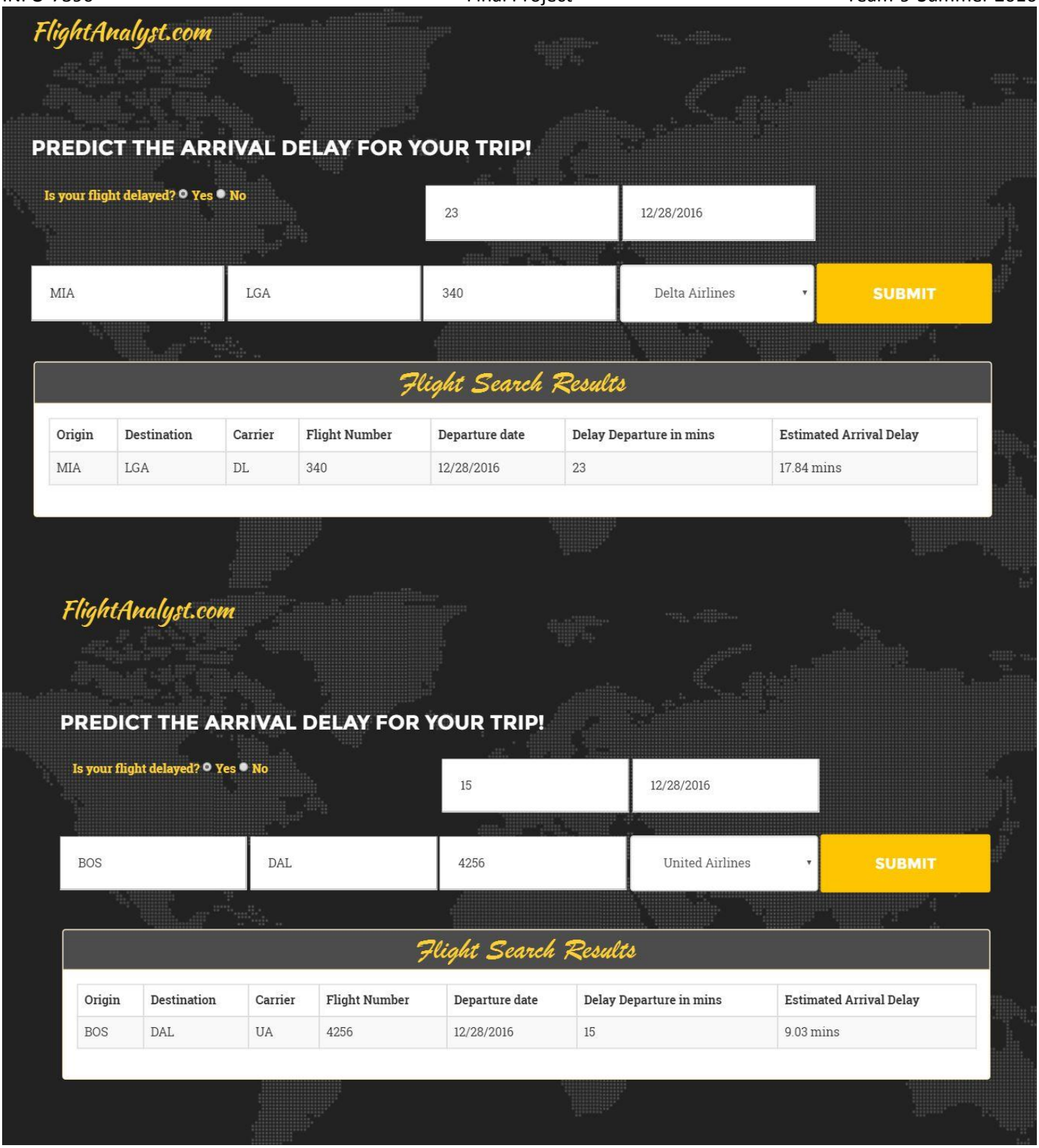
➢ ***Web Application Integrated and Use Cases***






FlightAnalyst.com
PREDICT THE ARRIVAL DELAY FOR YOUR TRIP!
Is your flight delayed? Yes No
0
02/22/2017
SJU
JFK
395
Delta Airlines
SUBMIT
Flight Search Results
Origin Destination Carrier Flight Number Departure date Delay Departure in mins Estimated Arrival Delay
SJU JFK DL 395 2/22/2017 0 On Time
FlightAnalyst.com
PREDICT THE ARRIVAL DELAY FOR YOUR TRIP!
Is your flight delayed? Yes No
23
02/22/2017
SJU
JFK
395
Delta Airlines
SUBMIT
Flight Search Results
Origin Destination Carrier Flight Number Departure date Delay Departure in mins Estimated Arrival Delay
SJU JFK DL 395 2/22/2017 23 17.63 mins


FlightAnalyst.com
PREDICT THE ARRIVAL DELAY FOR YOUR TRIP!
Is your flight delayed? Yes No
23
02/22/2017
MIA
LGA
395
Delta Airlines
SUBMIT
Flight Search Results
Origin Destination Carrier Flight Number Departure date Delay Departure in mins Estimated Arrival Delay
MIA LGA DL 395 2/22/2017 23 18.56 mins
FlightAnalyst.com
PREDICT THE ARRIVAL DELAY FOR YOUR TRIP!
Is your flight delayed? Yes No
23
02/22/2017
MIA
LGA
367
Delta Airlines
SUBMIT
Flight Search Results
Origin Destination Carrier Flight Number Departure date Delay Departure in mins Estimated Arrival Delay
MIA LGA DL 367 2/22/2017 23 18.56 mins


FlightAnalyst.com
PREDICT THE ARRIVAL DELAY FOR YOUR TRIP!
Is your flight delayed? Yes No
23
12/28/2016
MIA
LGA
340
Delta Airlines
SUBMIT
Flight Search Results
Origin Destination Carrier Flight Number Departure date Delay Departure in mins Estimated Arrival Delay
MIA LGA DL 340 12/28/2016 23 17.84 mins
FlightAnalyst.com
PREDICT THE ARRIVAL DELAY FOR YOUR TRIP!
Is your flight delayed? Yes No
15
12/28/2016
BOS
DAL
4256
United Airlines
SUBMIT
Flight Search Results
Origin Destination Carrier Flight Number Departure date Delay Departure in mins Estimated Arrival Delay
BOS DAL UA 4256 12/28/2016 15 9.03 mins

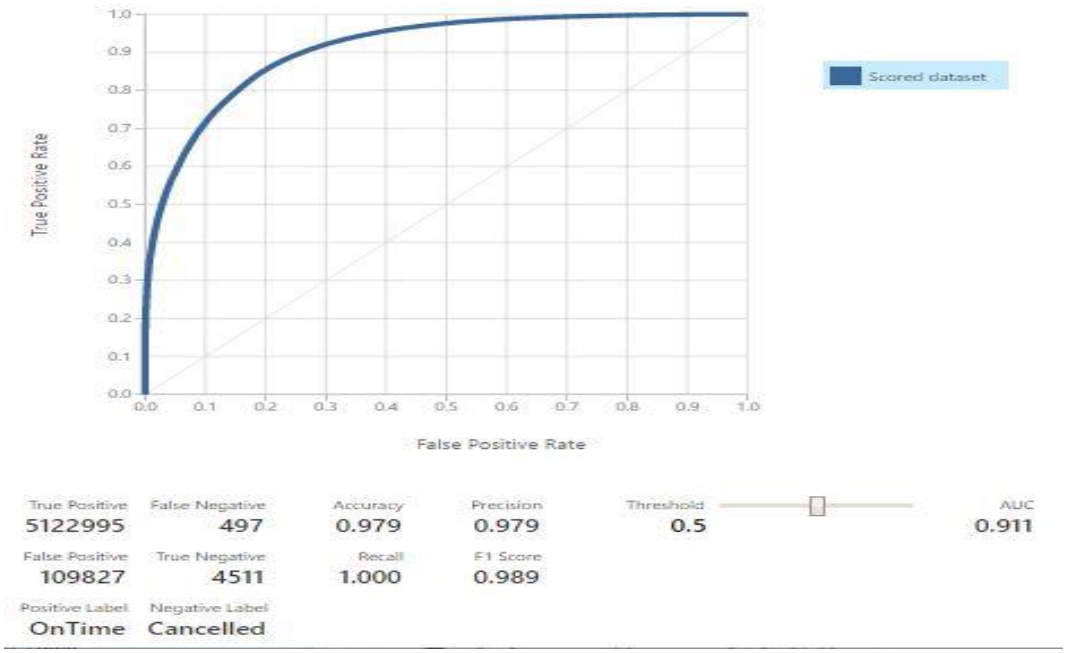
Flight Cancellation Classification Model

- ***Procedure and Approach***

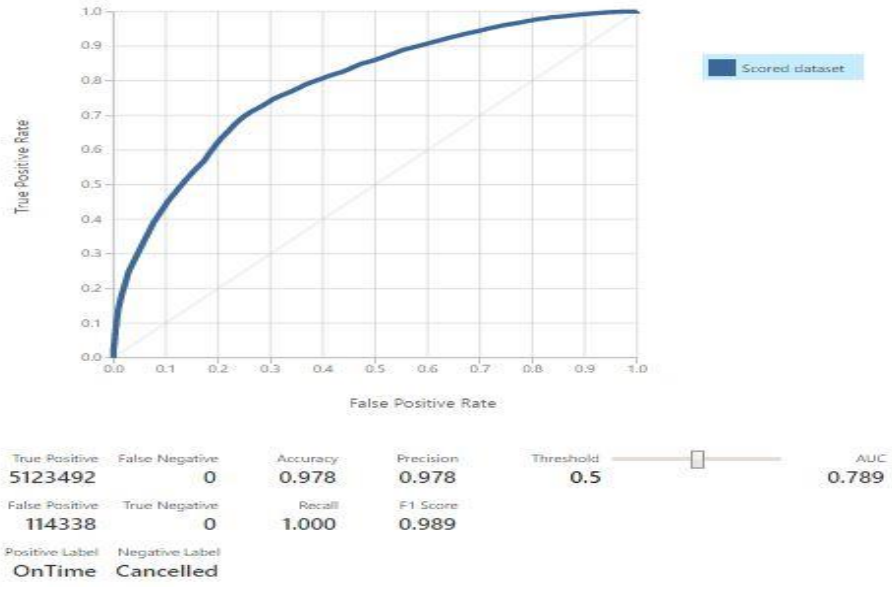
1. Cancellation code and cancellation description data of US flights for 2014 was given in the data set and we build the model to classify the cancellation on the model.
2. On evaluating the data, we found that the main predictors were Origin, destination, Date, Day of week, Quarter, carrier and cancelled to classify the cancellation.
3. Next we sampled the data into Test and Train to model the train data and then evaluate our model on based of Test data set.
4. Following Models were applied:
    a. Two class Decision Forest
    b. Two class Decision Jungle
    c. Two class Boosted Decision
    d. Two class Logistic Regression
5. The models were built and evaluated and the performance matrix was calculated.
6. Graphs were plotted to evaluate the performance of the models.
7. In Azure the best model was hosted as web service and was integrated with the front end to classify the cancellation of Flight for the given input.

- ***Classification Analysis***

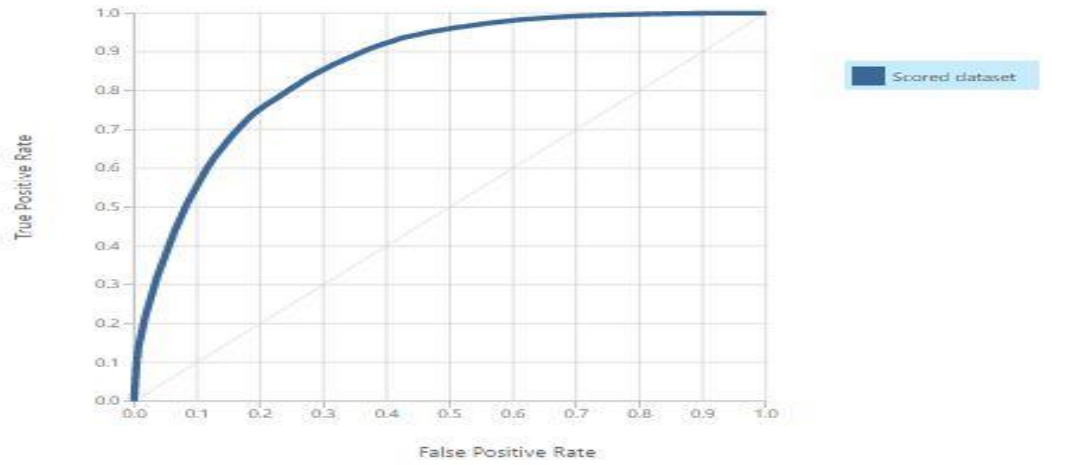
1. Two class Decision Forest



2. Two-class Decision Jungle

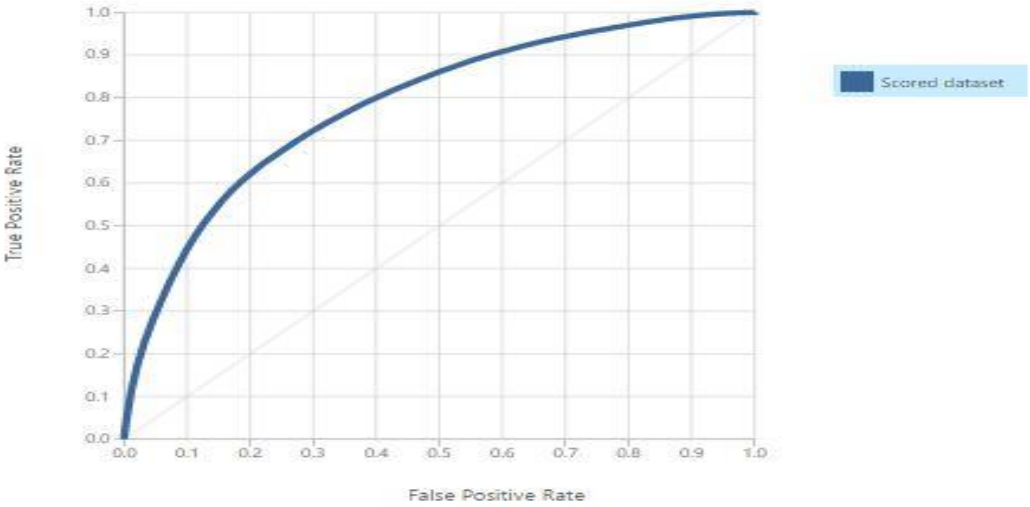


3. Two-class Boosted Decision



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
5114191	9301	0.980	0.982	0.5	0.862
False Positive	True Negative	Recall	F1 Score		
96017	18321	0.998	0.990		
Positive Label	Negative Label				
OnTime	Cancelled				

4. Two-class Logistic Regression



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
5123492	0	0.978	0.978	0.5	0.784
False Positive	True Negative	Recall	F1 Score		
114338	0	1.000	0.989		
Positive Label	Negative Label				
OnTime	Cancelled				

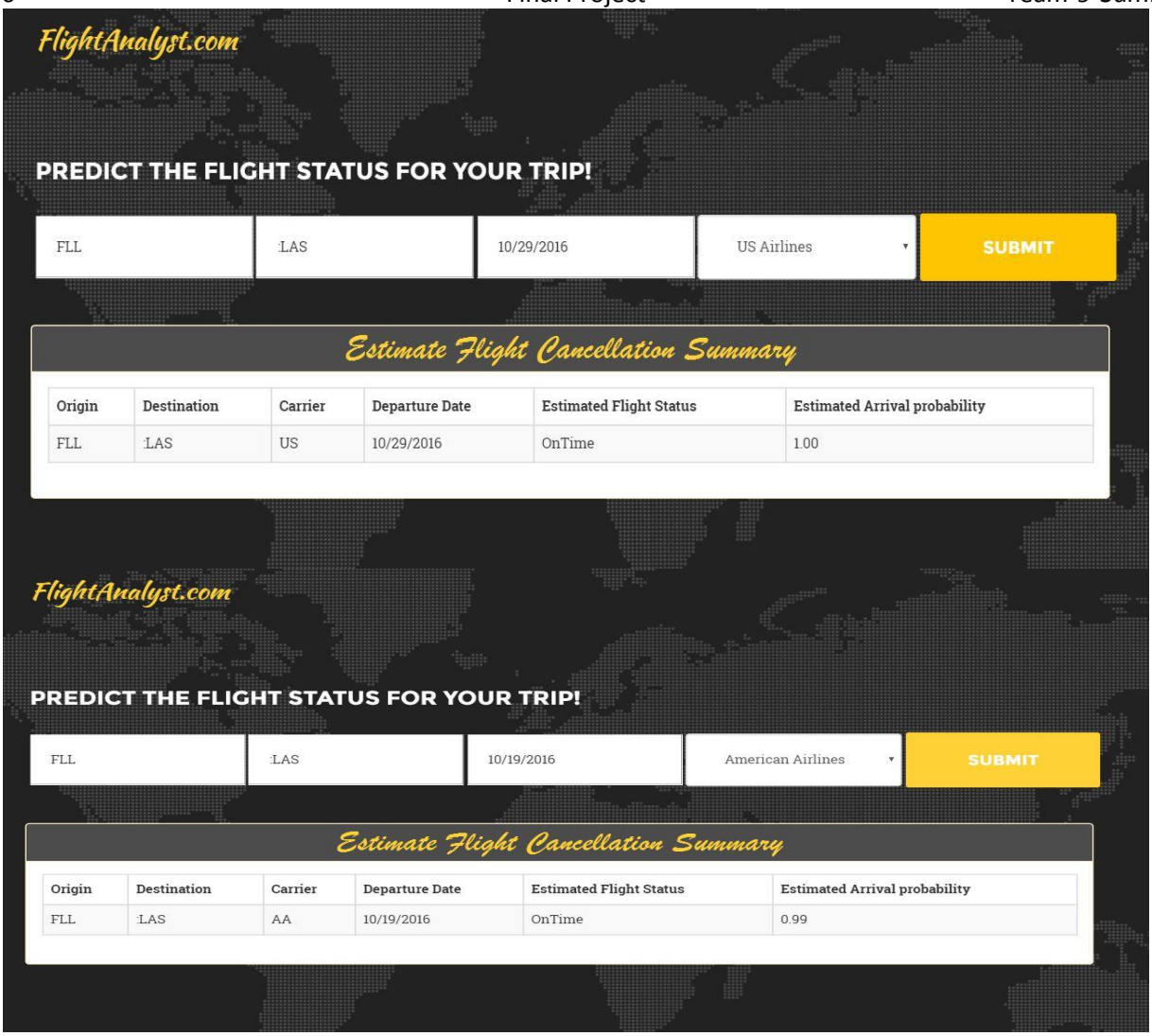
- ***Comparison of Models***

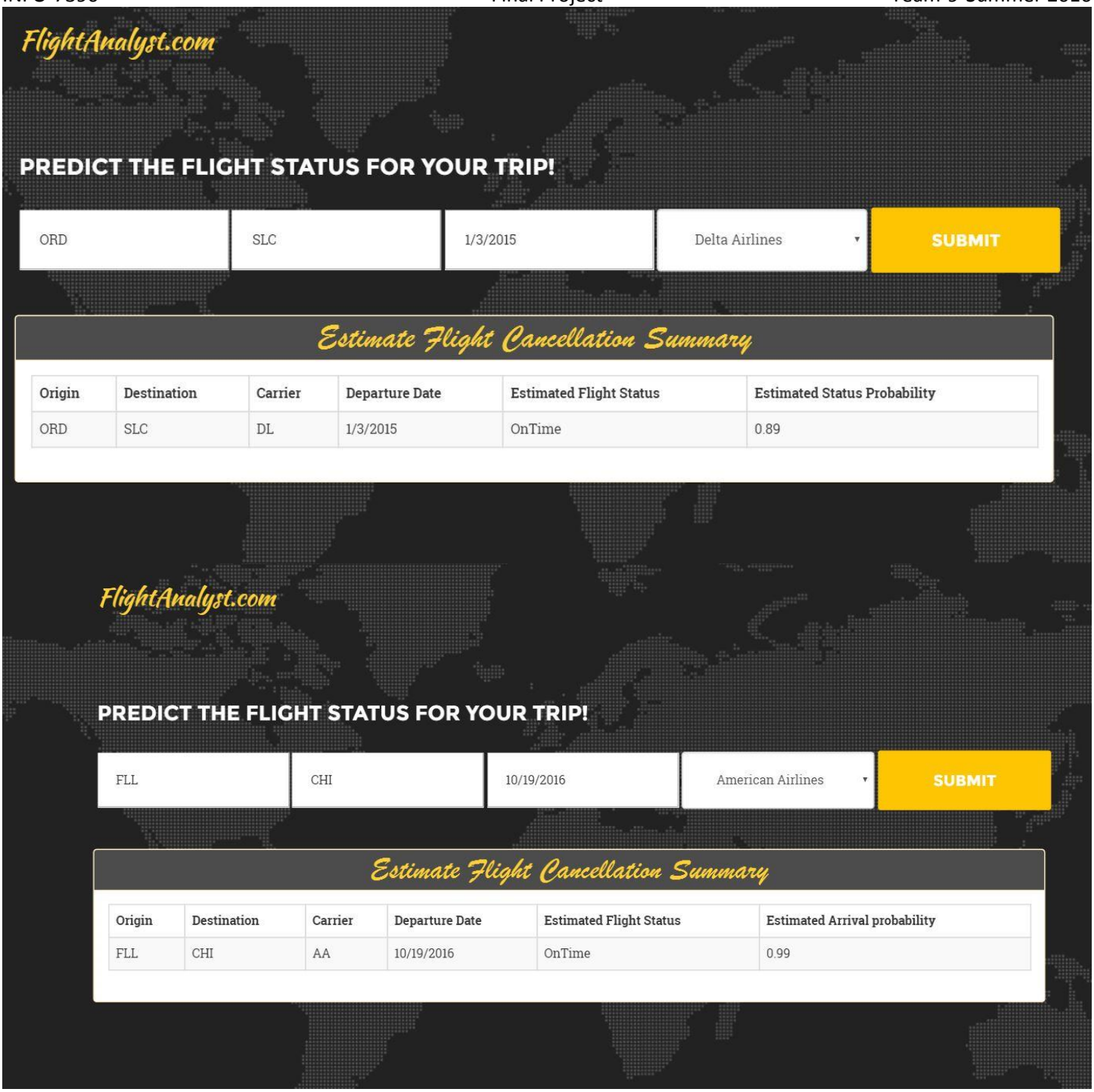
On comparison we found that Decision Forest worked best as the AUC is closest to 1 in this model and also the accuracy is good.

	Decision Forest	Decision Jungle	Boosted Decision	Logistic Regression
Accuracy	***0.979***	***0.978***	***0.98***	***0.978***
Precision	***0.979***	***0.978***	***0.982***	***0.978***
Recall	***1***	***1***	***0.998***	***1***
AUC	***0.991***	***0.789***	***0.862***	***0.784***

- ✓ **With the results derived above we can conclude that *Decision Forest* has best Performance metrics,**
- ✓ **Also, the area under ROC curve and lift curve analysis supports Decision Forest model the most accurate model on given dataset.**
- ✓ **Since it is critical in deciding whether a flight would be cancelled or not, we choose the model which more accurately predicts the cancellation rather than on time.**

- ***Web Application Integrated and Use Cases***


FlightAnalyst.com
PREDICT THE FLIGHT STATUS FOR YOUR TRIP!
FLL
LAS
10/29/2016
US Airlines
SUBMIT
Estimate Flight Cancellation Summary
Origin
Destination
Carrier
Departure Date
Estimated Flight Status
Estimated Arrival probability
FLL
LAS
US
10/29/2016
OnTime
1.00
FlightAnalyst.com
PREDICT THE FLIGHT STATUS FOR YOUR TRIP!
FLL
LAS
10/19/2016
American Airlines
SUBMIT
Estimate Flight Cancellation Summary
Origin
Destination
Carrier
Departure Date
Estimated Flight Status
Estimated Arrival probability
FLL
LAS
AA
10/19/2016
OnTime
0.99


FlightAnalyst.com
PREDICT THE FLIGHT STATUS FOR YOUR TRIP!
ORD
SLC
1/3/2015
Delta Airlines
SUBMIT
Estimate Flight Cancellation Summary
Origin | Destination | Carrier | Departure Date | Estimated Flight Status | Estimated Status Probability
ORD | SLC | DL | 1/3/2015 | OnTime | 0.89
FlightAnalyst.com
PREDICT THE FLIGHT STATUS FOR YOUR TRIP!
FLL
CHI
10/19/2016
American Airlines
SUBMIT
Estimate Flight Cancellation Summary
Origin | Destination | Carrier | Departure Date | Estimated Flight Status | Estimated Arrival probability
FLL | CHI | AA | 10/19/2016 | OnTime | 0.99

Twitter Sentiment Analysis For US Flights

Introduction

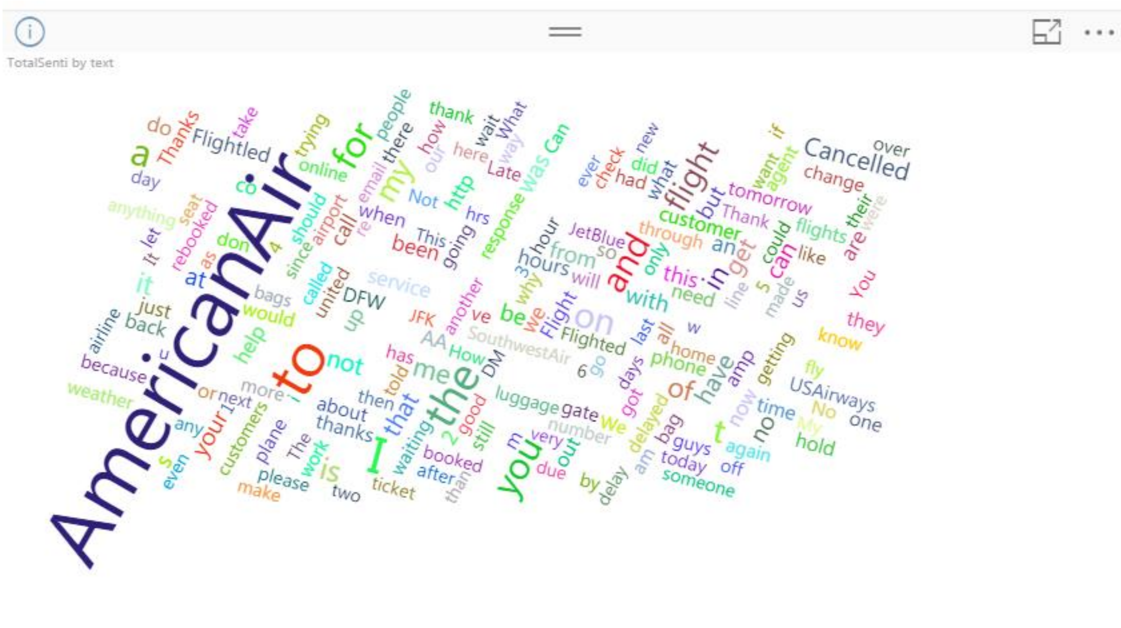
A sentiment analysis job about the problems of each major U.S. airline. Twitter data was scraped from February of 2015 and contributors were asked to first classify positive, negative, and neutral tweets, followed by categorizing negative reasons.

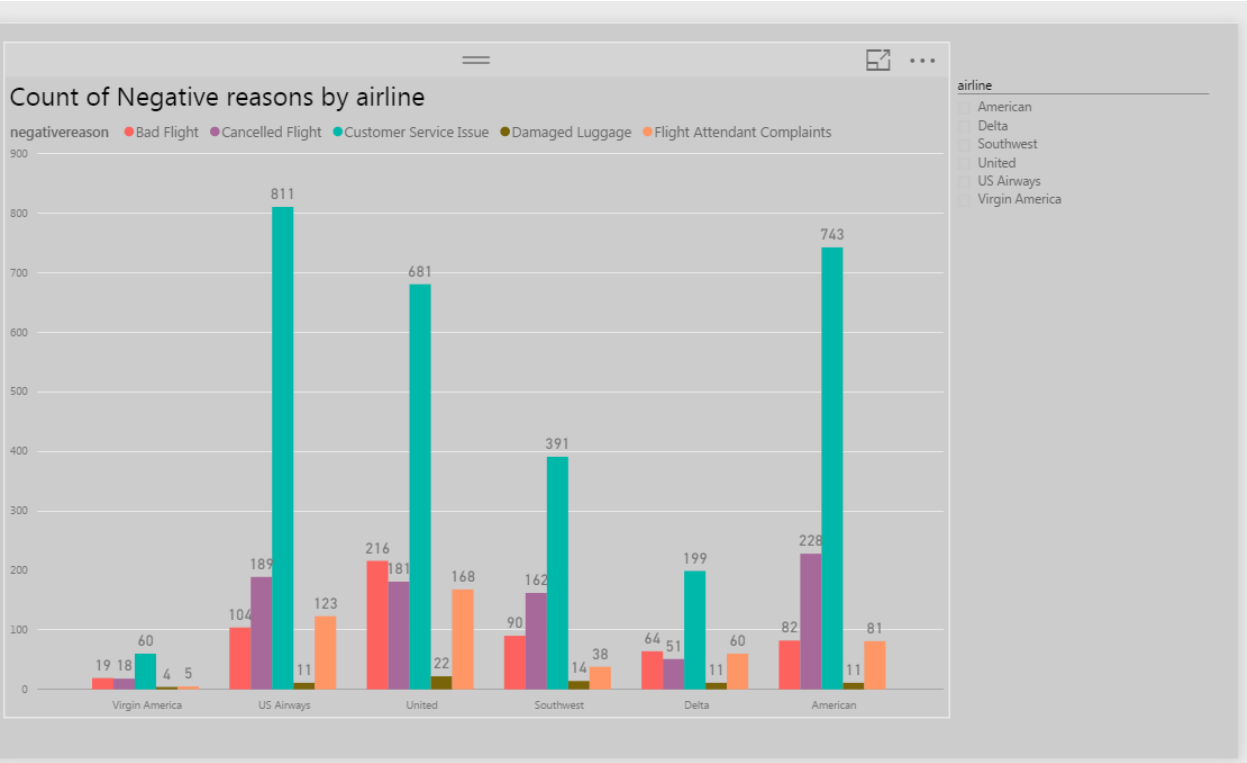
Dataset Description

The fields in the Tweets.csv file / Tweets database table are:

- tweet_id
- airline_sentiment
- airline_sentiment_confidence
- negativereason
- negativereason_confidence
- airline
- airline_sentiment_gold
- name
- negativereason_gold
- retweet_count
- text
- tweet_coord
- tweet_created
- tweet_location
- user_timezone


Count of Tweets By Sentiment
10K
8K
6K
4K
2K
0K
9K
3K
2K
negative
neutral
positive
airline_sentiment
negative
neutral
positive
airline
American
Delta
Southwest
United
US Airways
Virgin America
Sentiments by Carrier
airline_sentiment negative neutral positive
100%
80%
60%
40%
20%
0%
American
Delta
Southwest
United
US Airways
Virgin America


TotalSenti by text
AmericanAir
to
the
I
you
flight
and
on
for
a
my
Cancelled
Flightled
not
of
is
me
Thanks
thank
Late
customer
tomorrow
USAirways
JetBlue
SouthwestAir
service
luggage
gate
hold



Conclusion:

Web application link: **http://flightanalyst-env.us-east-1.elasticbeanstalk.com/**