Weather Induced Airline Delays Prediction Using Machine Learning (Random Forests)

Aniruddha Humane

3321

STCL

Motivation

- Increase in volume of air travel from 450 billion passenger-miles in 1997 to 600 billion passenger-miles in 2014
- Challenging environment for National Airspace System (NAS).
- Scheduling parameters
- Average load factor of flight operations in 2012 was 83%
- Impact of Flight delays on passengers and government
- Why this area is important
- Work done in this area

Benefits

- Identify potential causes
- significant operational cost savings
- Better flight scheduling
- Better management of scheduled flights
- Improvement in quality of life
- Find ways to alleviate the impact

Why Machine learning

- Volume of historical flights and weather data are too large to analyse analytically.
- Correlation among factors are extremely complicated and highly non-linear.
- Machine learning is a clever method to analyse such data.

Analysis Process Flow

- Data Acquisition
- Data Exploration
- Predictive Models

Technologies and Libraries

- Python/ipython
- Numpy
- Matplotlib
- Scikit-Learn
- Scipy
- Pandas
- Imbalanced-Learn

Environment

• Processor	:	Intel i5-5200U 2.20GHz Dual core
• RAM	:	8GB
• Operating System	:	Windows 10 Pro
• Language platform	:	iPython

Data Acquisition

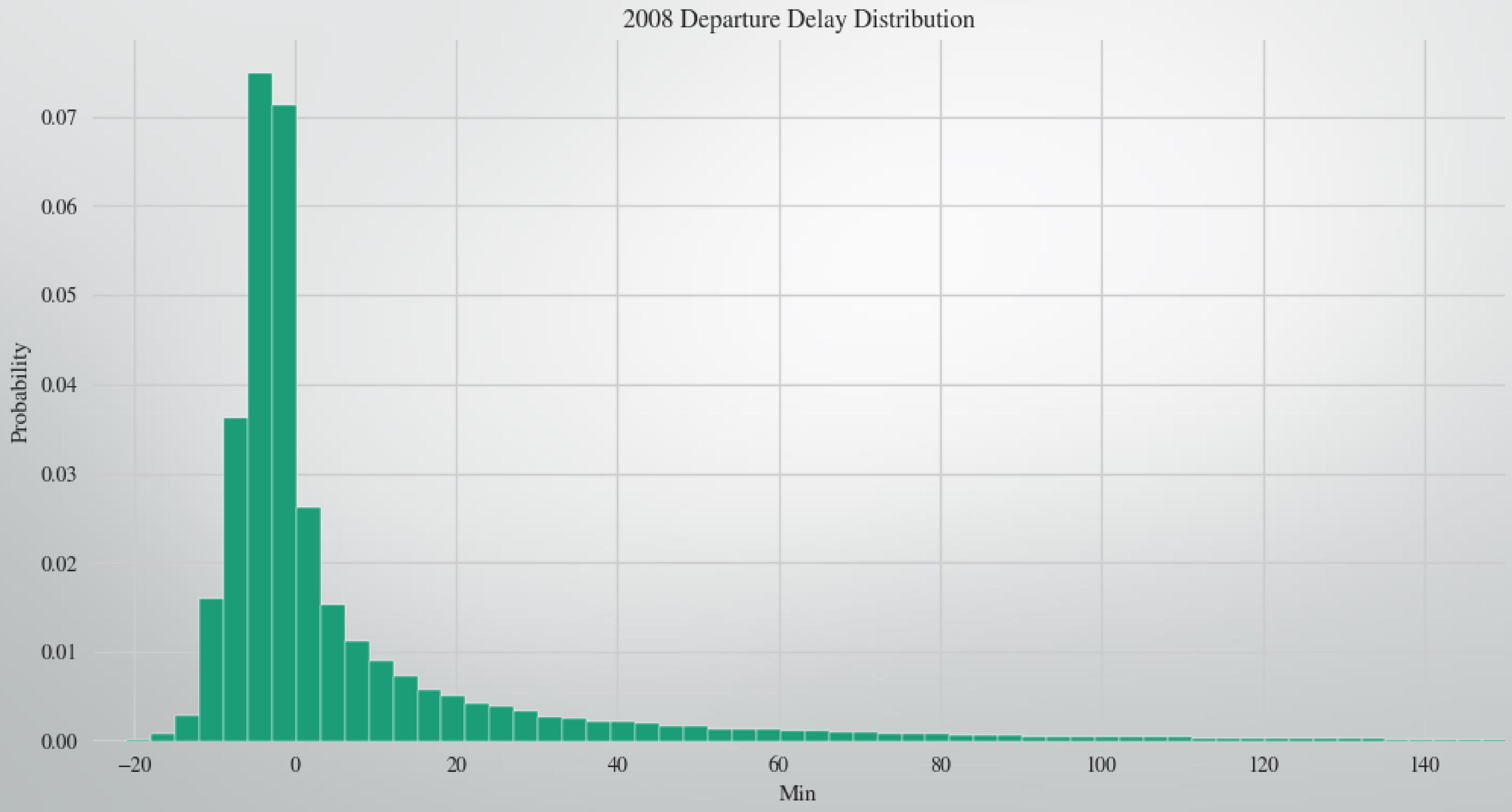
- The **flight data**, also known as on-time performance data can be downloaded from the American Statistical Association.
- **Historical weather data** and flight demand data for 2008 is from the FAA Aviation Systems Performance Metrics (ASPM).

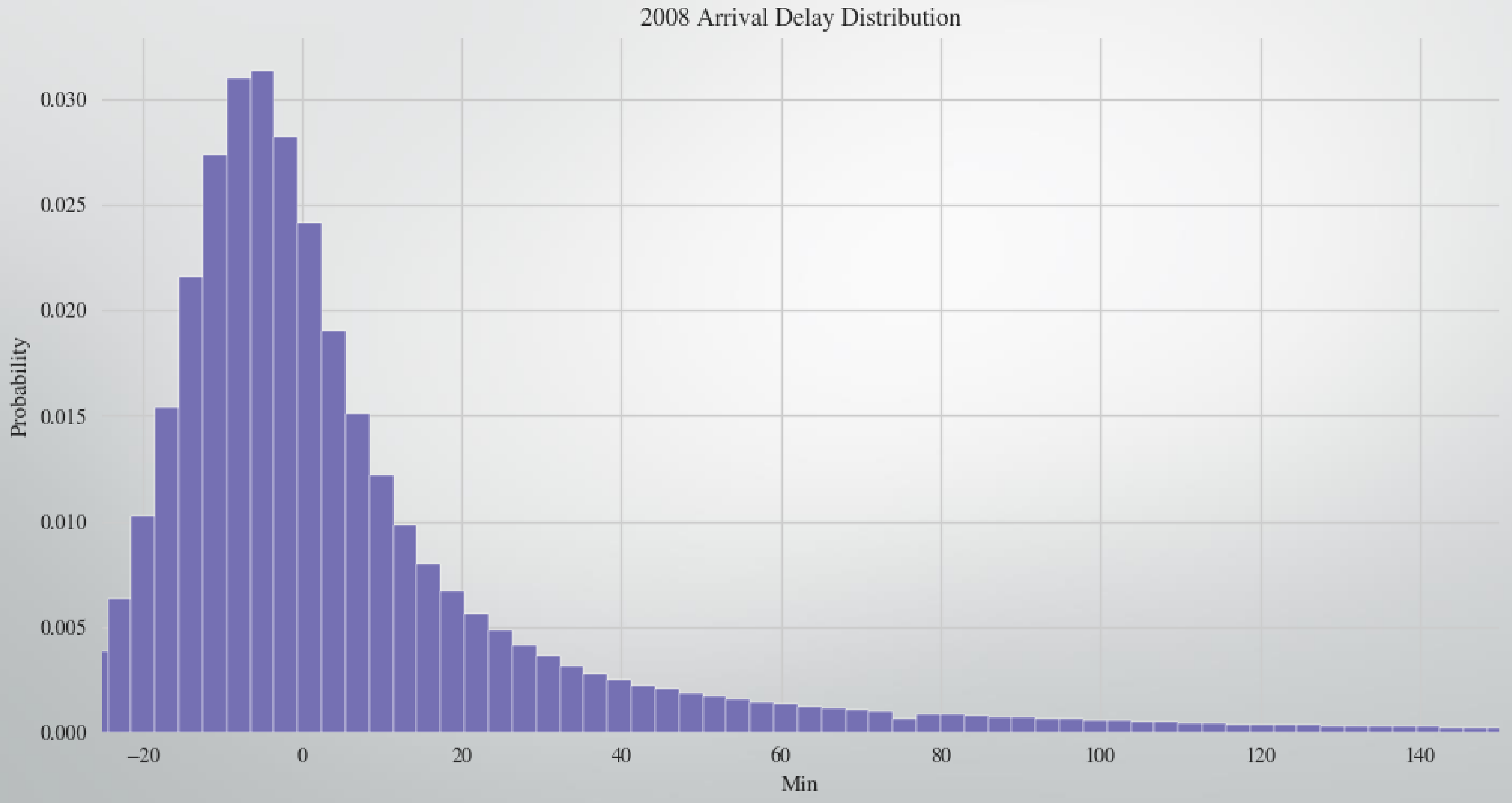
Features

Name	Description
Year	1987-2008
Month	1-12
DayofMonth	1-31
DayOfWeek	1 (Monday) - 7 (Sunday)
DepTime	actual departure time (local, hhmm)
CRSDepTime	scheduled departure time (local, hhmm)
ArrTime	actual arrival time (local, hhmm)
CRSArrTime	scheduled arrival time (local, hhmm)
UniqueCarrier	unique carrier code

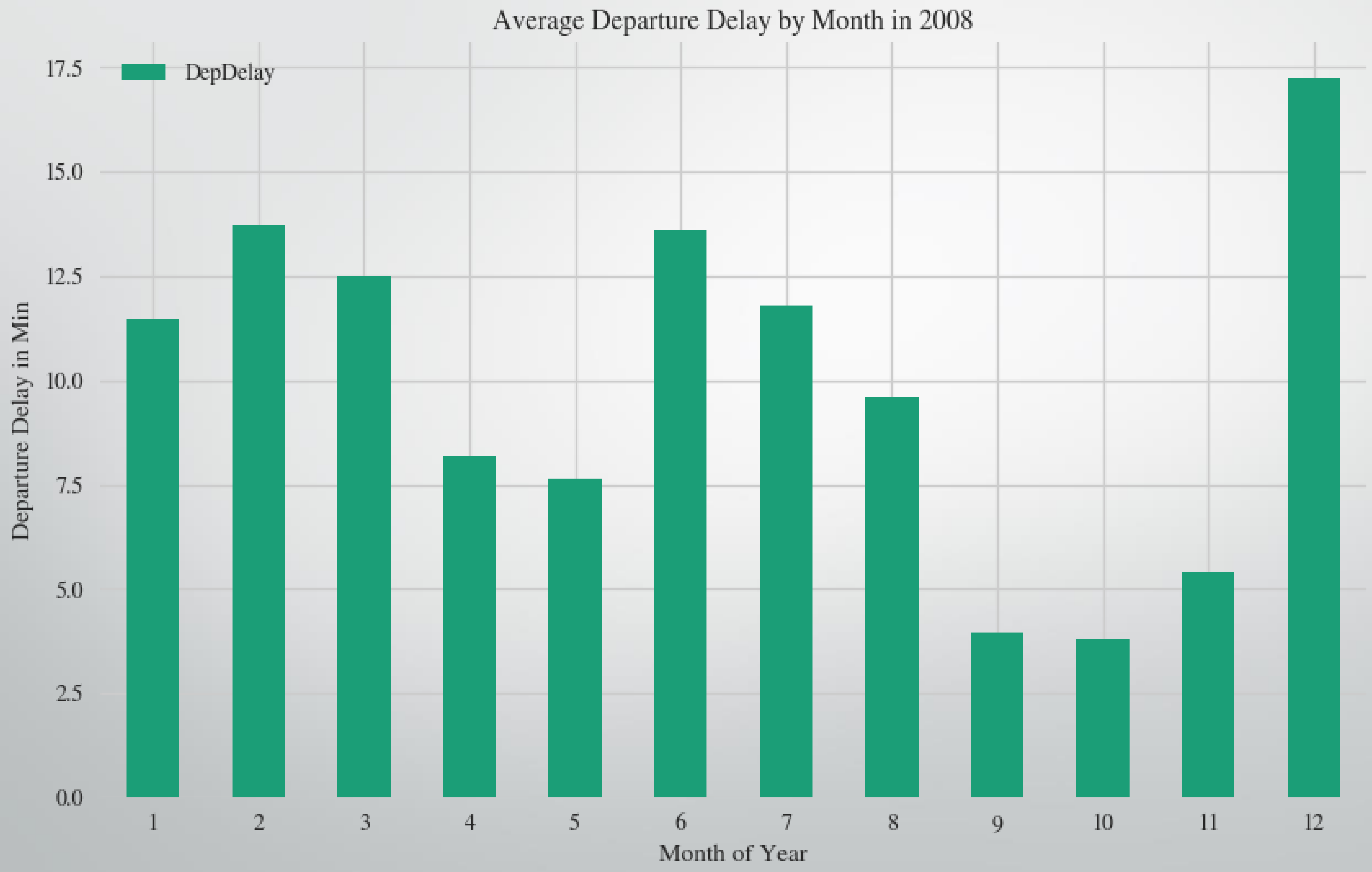
Name	Description
FlightNum	flight number
TailNum	plane tail number
ActualElapsedTime	in minutes
CRSElapsedTime	in minutes
AirTime	in minutes
ArrDelay	arrival delay, in minutes
DepDelay	departure delay, in minutes
Origin	origin IATA airport code
Dest	destination IATA airport code
Distance	in miles
Carrier	Carrier identifier code assigned by IATA

Departure And Arrival Delay Distribution With Respect To Amount Of Delay

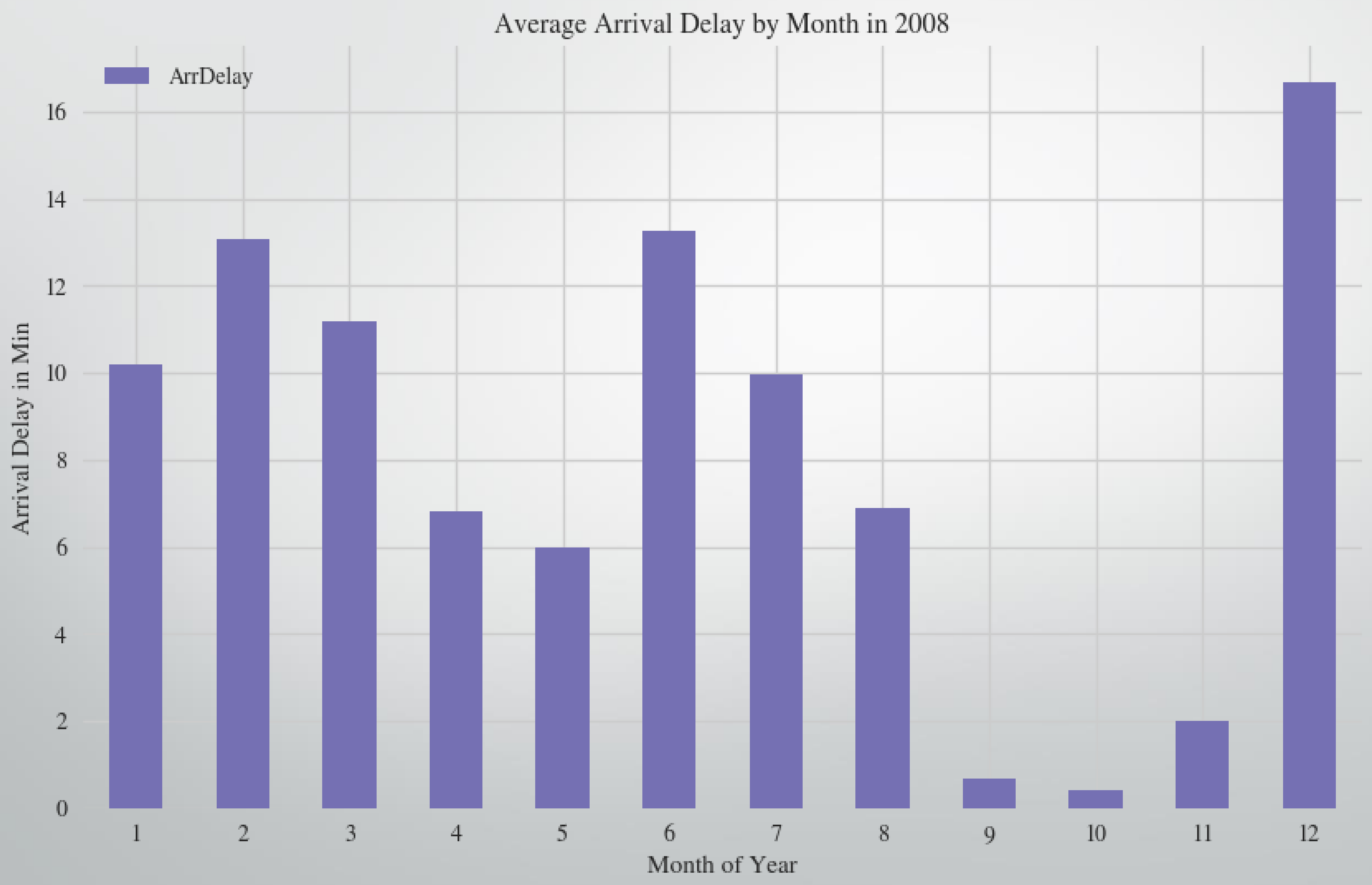

2008 Departure Delay Distribution
Probability
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
−20
0
20
40
60
80
100
120
140
Min


2008 Arrival Delay Distribution
Probability
0.030
0.025
0.020
0.015
0.010
0.005
0.000
−20
0
20
40
60
80
100
120
140
Min

Departure and Arrival Delay Distribution With respect to Months

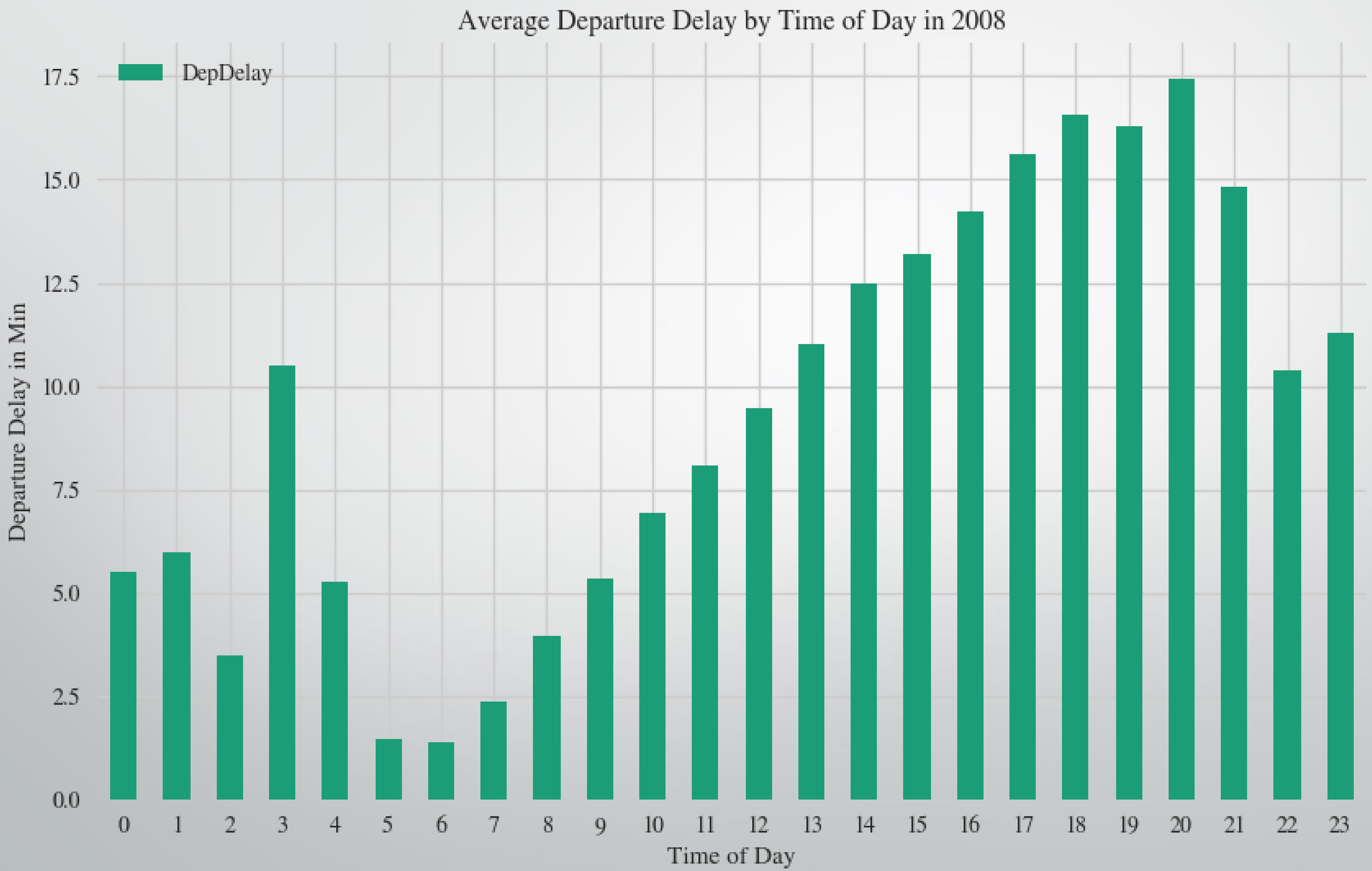


Average Departure Delay by Month in 2008
DepDelay
Departure Delay in Min
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
1
2
3
4
5
6
7
8
9
10
11
12
Month of Year

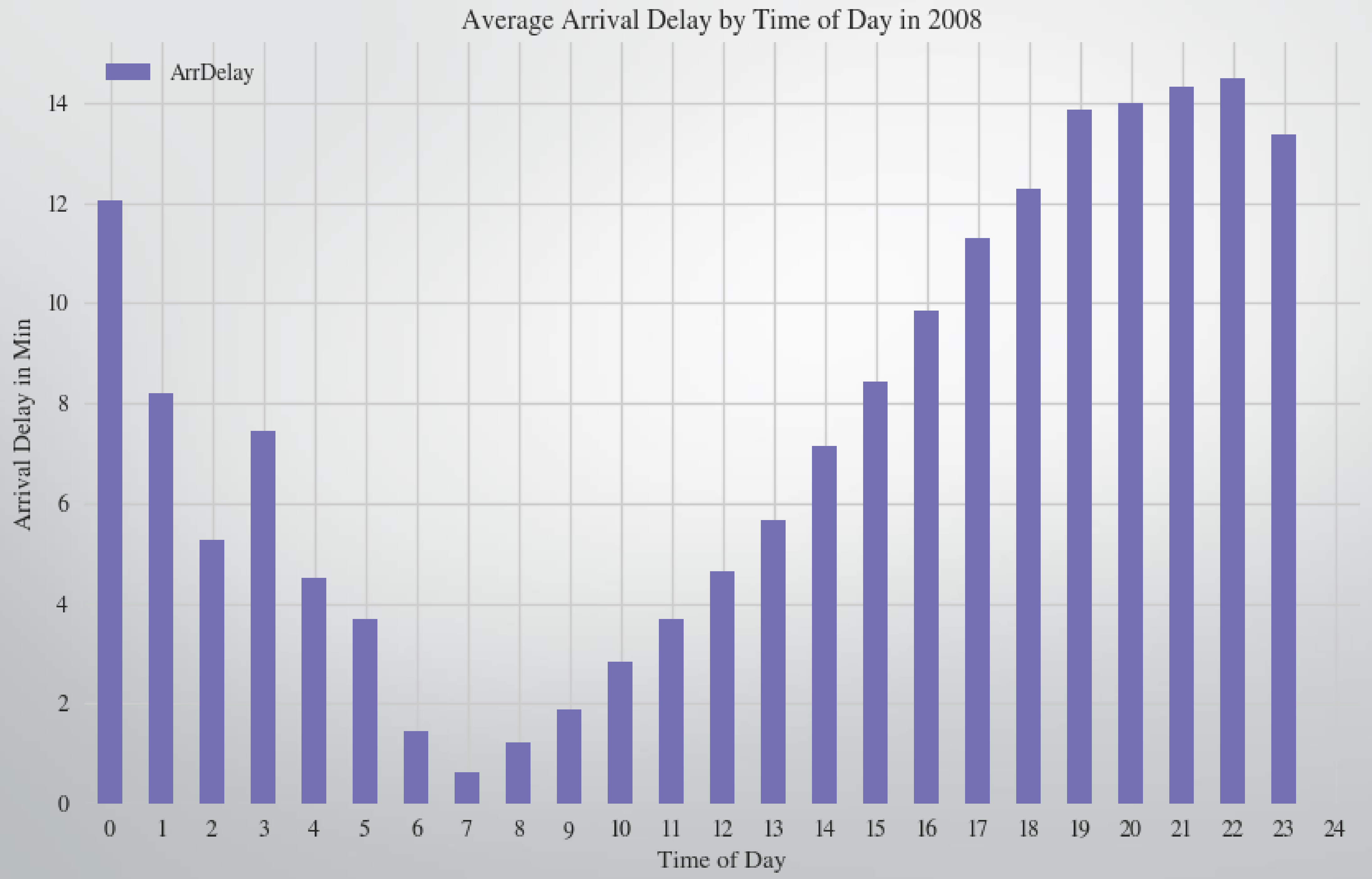


Average Arrival Delay by Month in 2008
ArrDelay
Arrival Delay in Min
0
2
4
6
8
10
12
14
16
1
2
3
4
5
6
7
8
9
10
11
12
Month of Year

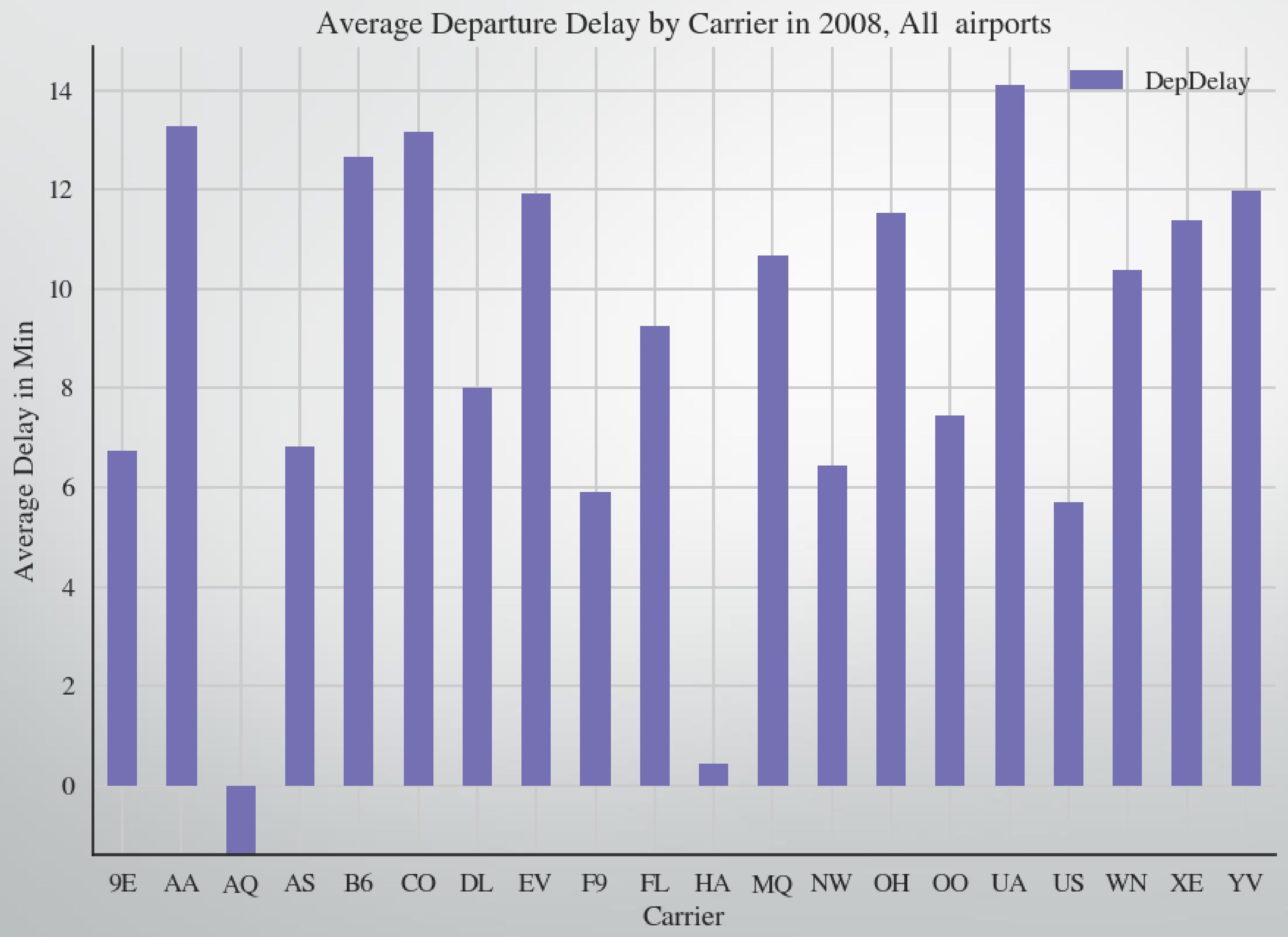
Departure and Arrival Delay Distribution With respect to Time of day



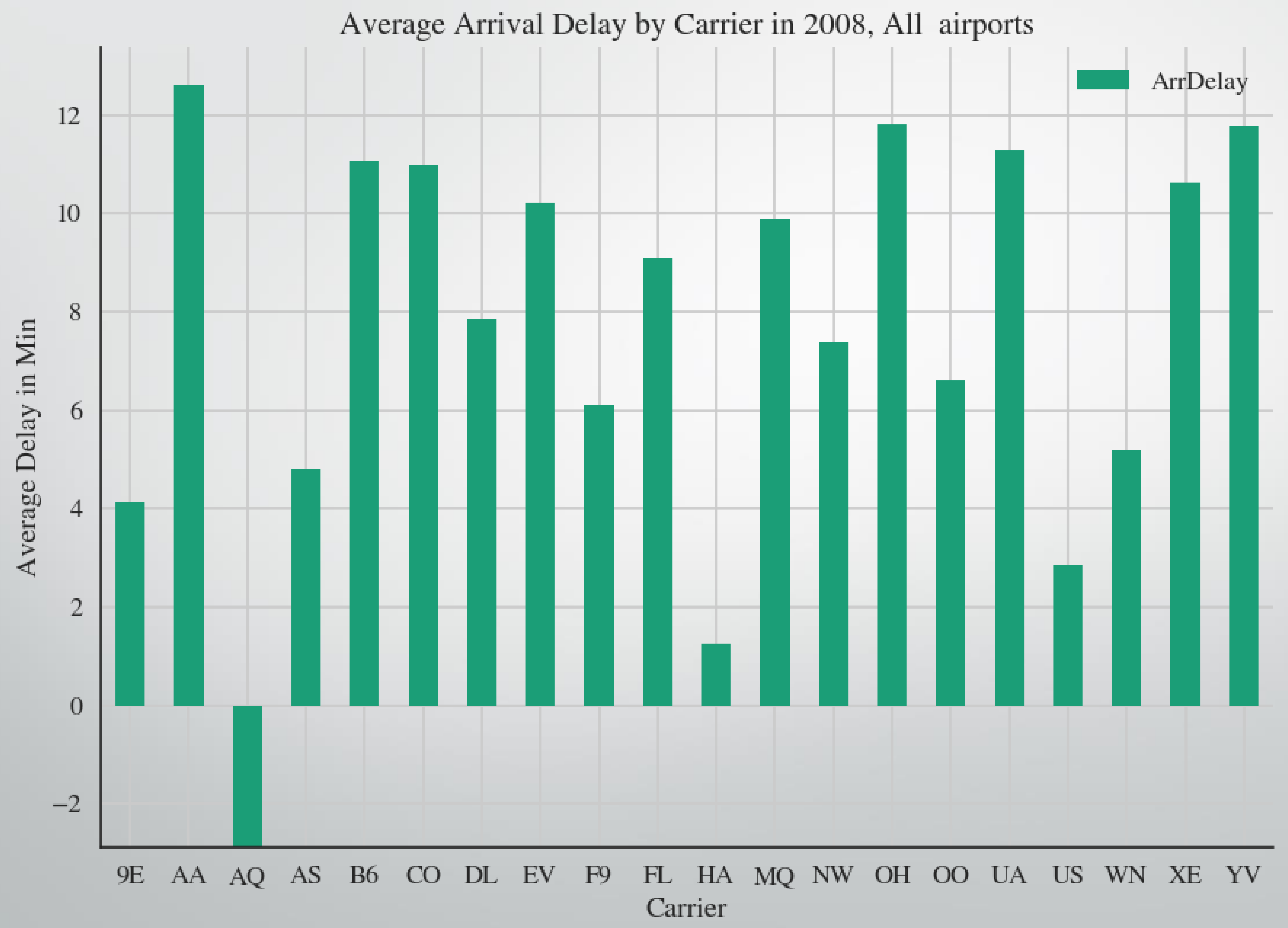
Average Departure Delay by Time of Day in 2008
DepDelay
Departure Delay in Min
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Time of Day


Average Arrival Delay by Time of Day in 2008
ArrDelay
Arrival Delay in Min
0
2
4
6
8
10
12
14
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Time of Day

Departure and Arrival Delay Distribution With respect to Carrier



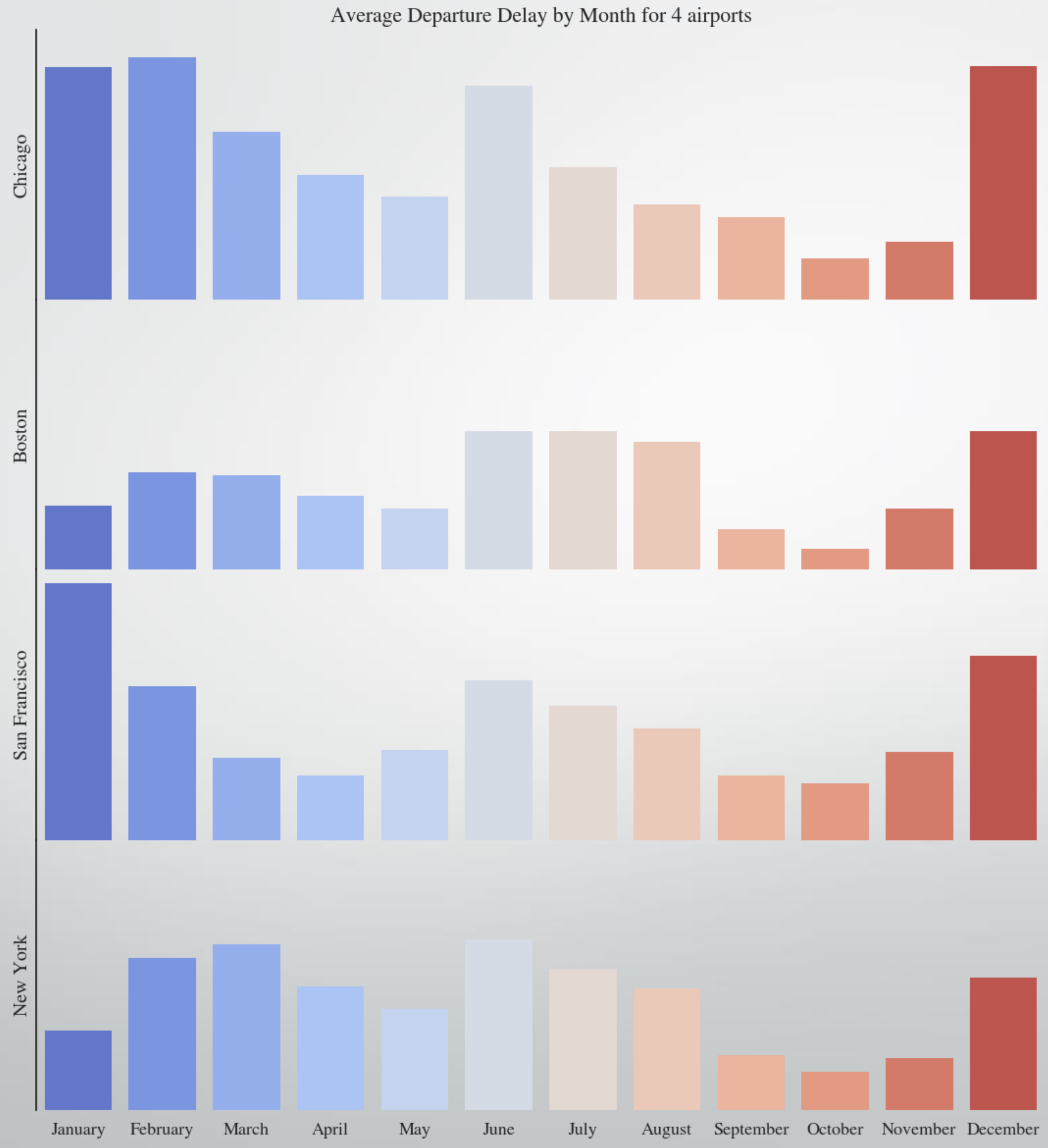
Average Departure Delay by Carrier in 2008, All airports
DepDelay
Average Delay in Min
14
12
10
8
6
4
2
0
9E
AA
AQ
AS
B6
CO
DL
EV
F9
FL
HA
MQ
NW
OH
OO
UA
US
WN
XE
YV
Carrier



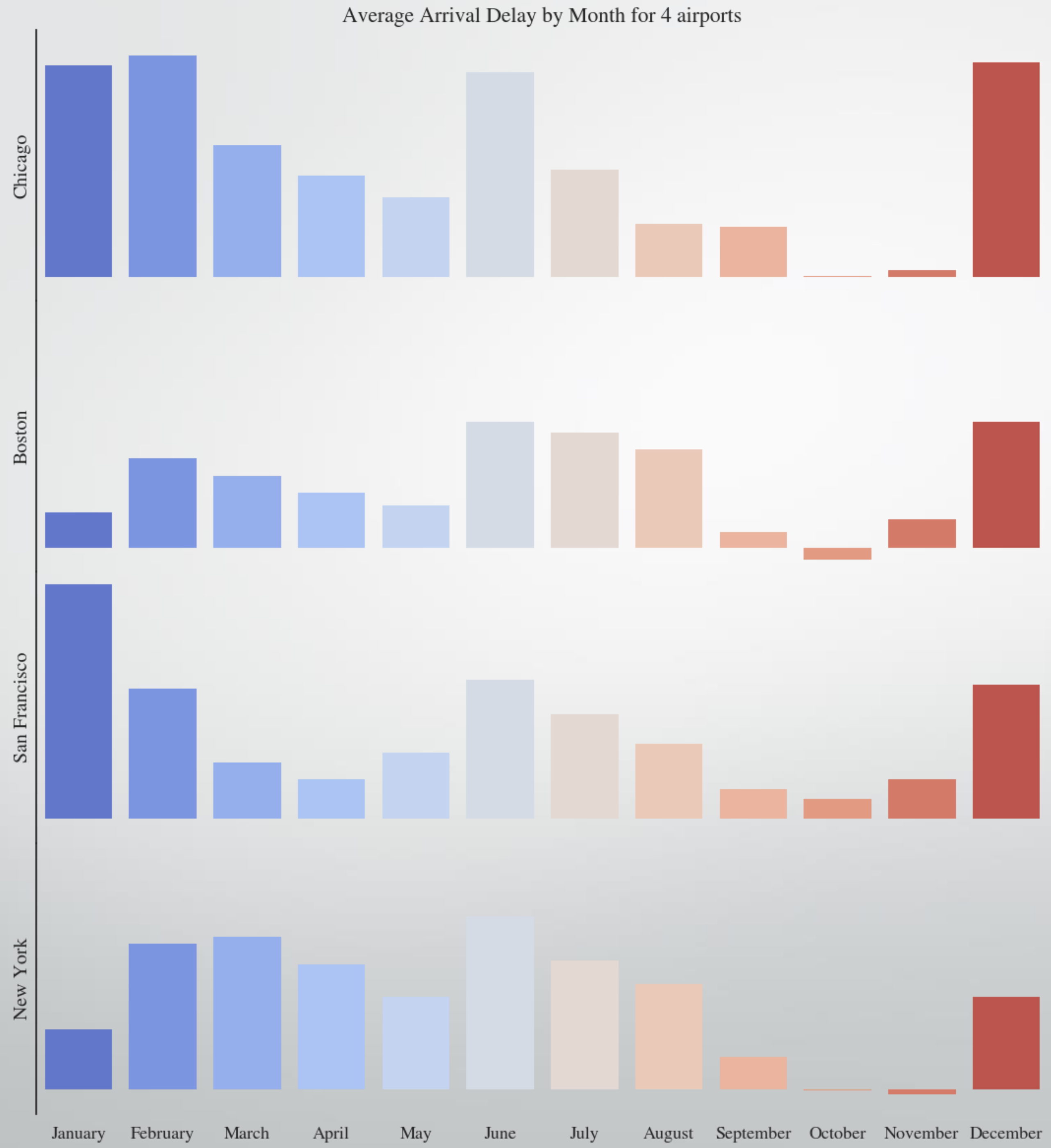
Average Arrival Delay by Carrier in 2008, All airports
ArrDelay
Average Delay in Min
12
10
8
6
4
2
0
−2
9E
AA
AQ
AS
B6
CO
DL
EV
F9
FL
HA
MQ
NW
OH
OO
UA
US
WN
XE
YV
Carrier

Departure and Arrival Delay Distribution With respect to Months for 4 major airports

- Chicago O'Hare (ORD)
- Boston Logan (BOS)
- San Francisco (SFO)
- New York LaGuardia(LGA)

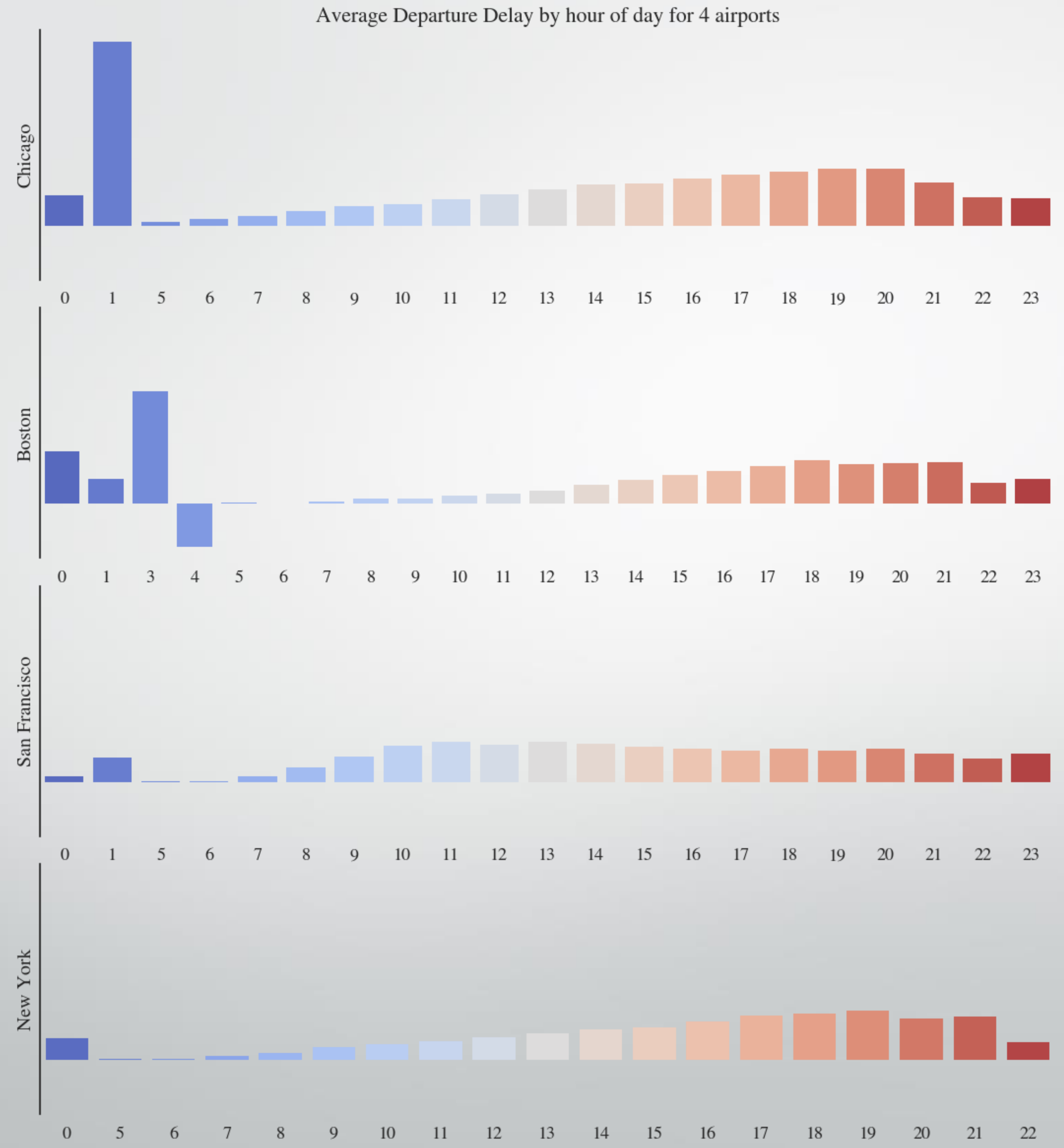


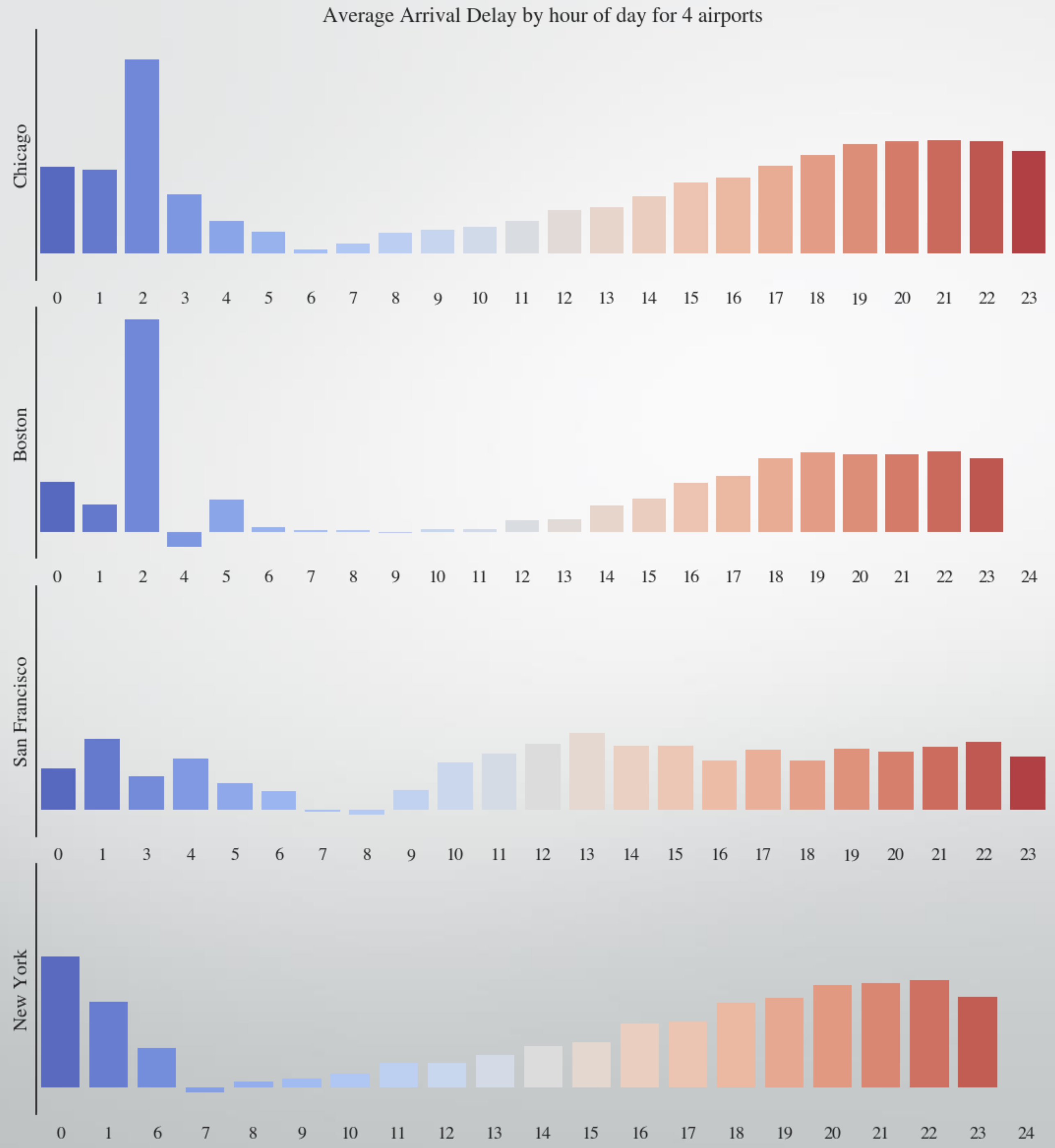
Average Departure Delay by Month for 4 airports
Chicago
Boston
San Francisco
New York
January
February
March
April
May
June
July
August
September
October
November
December



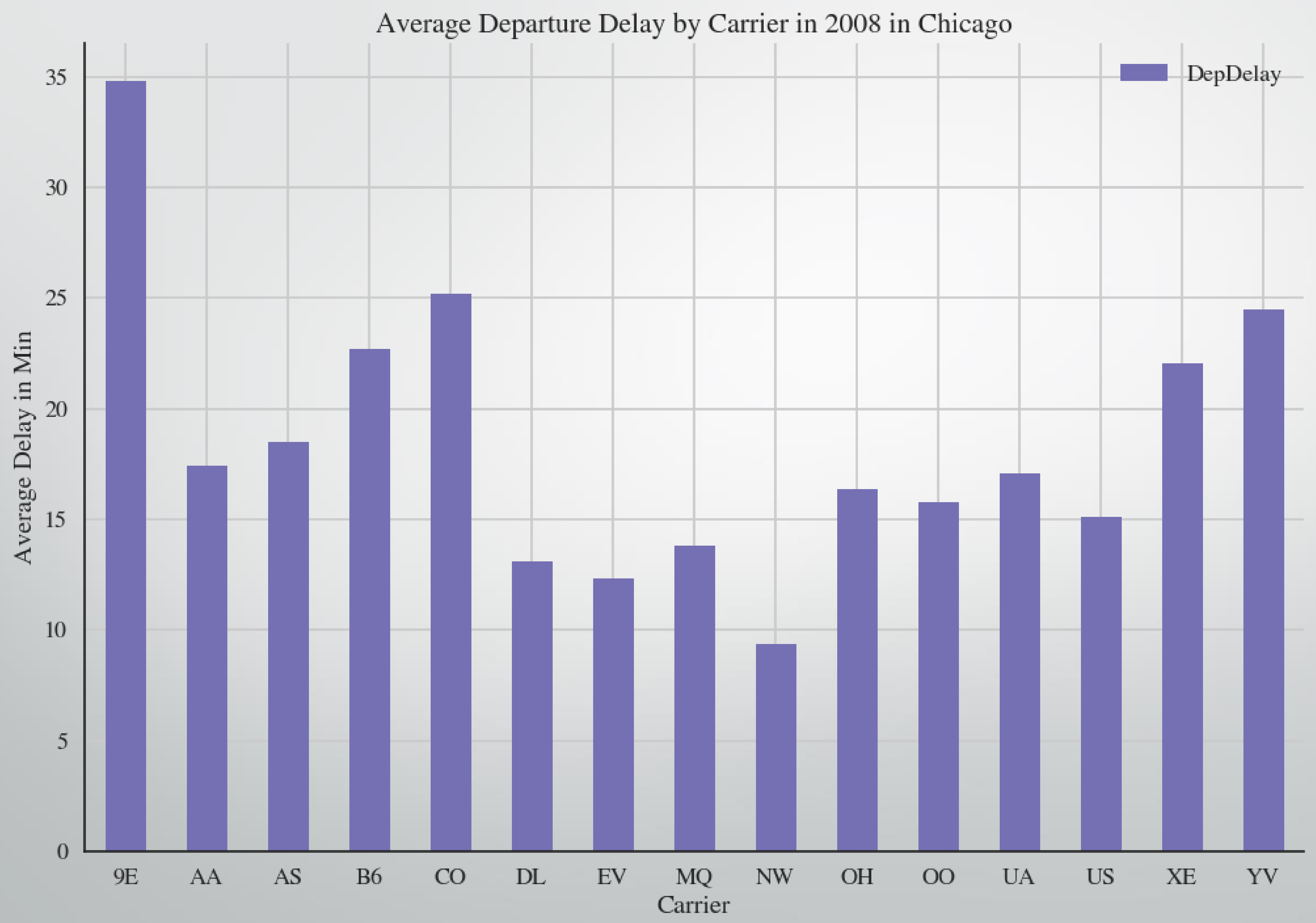
Average Arrival Delay by Month for 4 airports
Chicago
Boston
San Francisco
New York
January
February
March
April
May
June
July
August
September
October
November
December

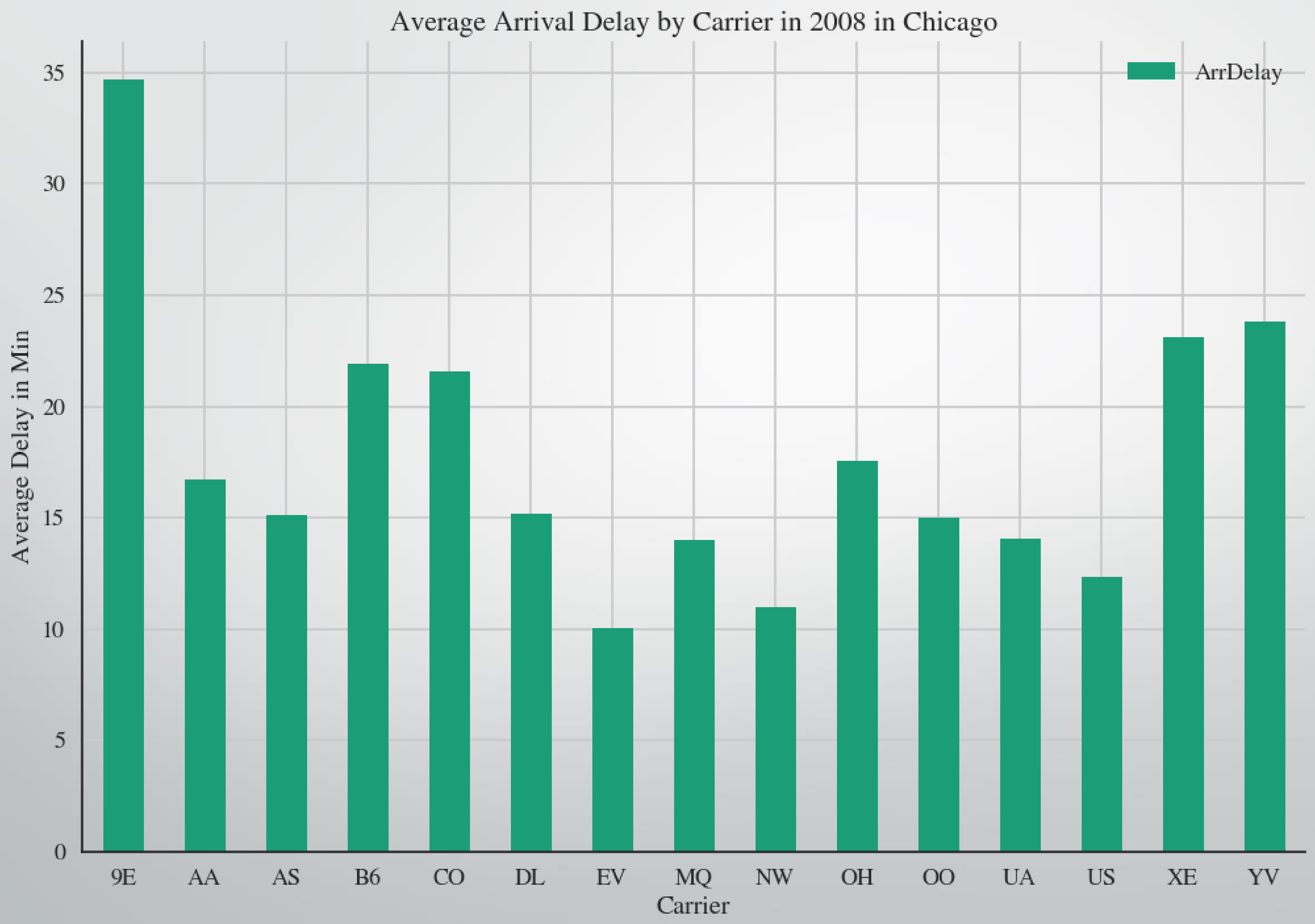
Departure and Arrival Delay Distribution With respect to Time of day for 4 major airports


Average Departure Delay by hour of day for 4 airports
Chicago
0 1 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Boston
0 1 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
San Francisco
0 1 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
New York
0 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22


Average Arrival Delay by hour of day for 4 airports
Chicago
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Boston
0 1 2 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
San Francisco
0 1 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
New York
0 1 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

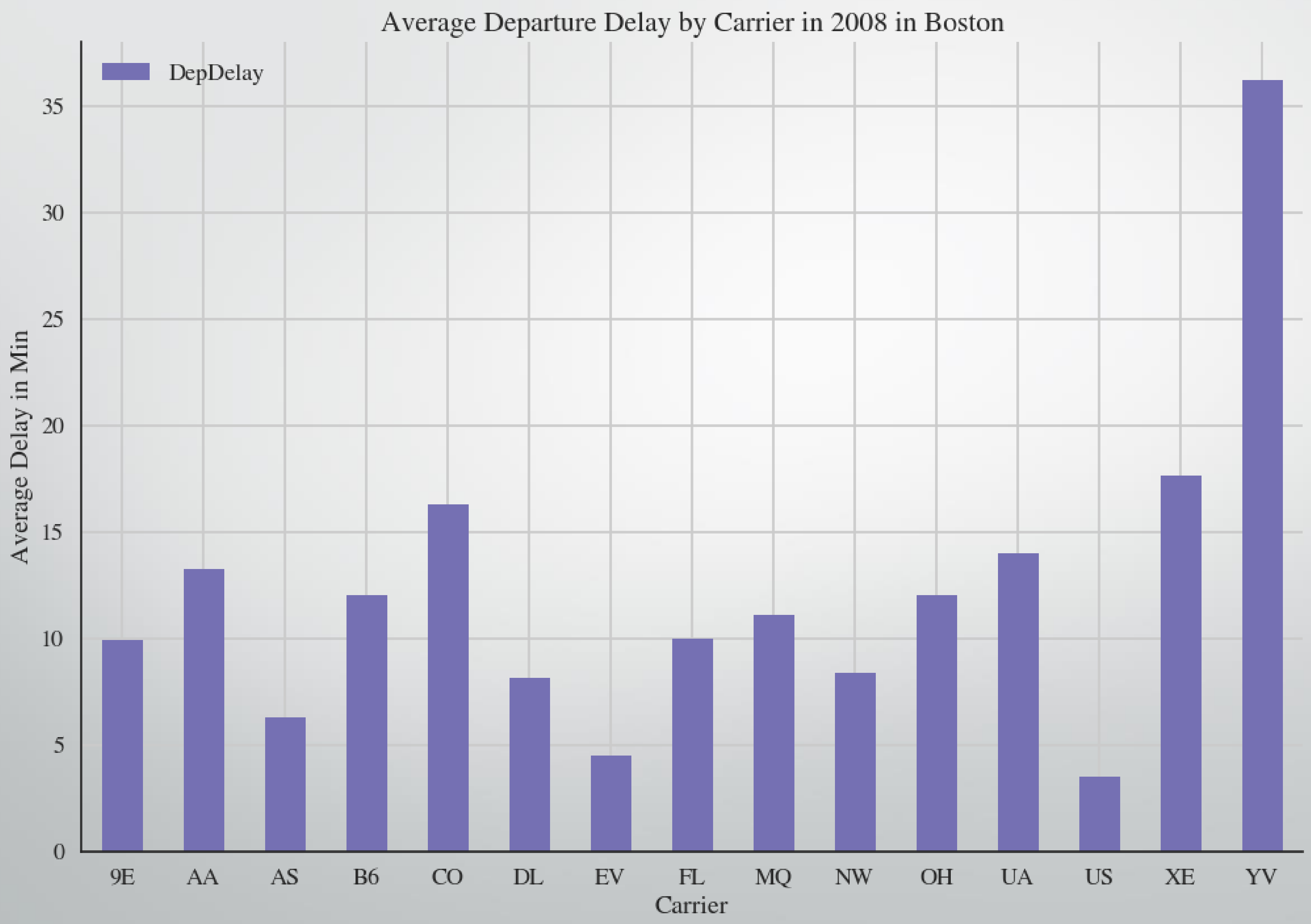
Departure and Arrival Delay Distribution With respect to Carrier for Chicago O'Hare (ORD)


Average Departure Delay by Carrier in 2008 in Chicago
DepDelay
Average Delay in Min
0
5
10
15
20
25
30
35
9E
AA
AS
B6
CO
DL
EV
MQ
NW
OH
OO
UA
US
XE
YV
Carrier

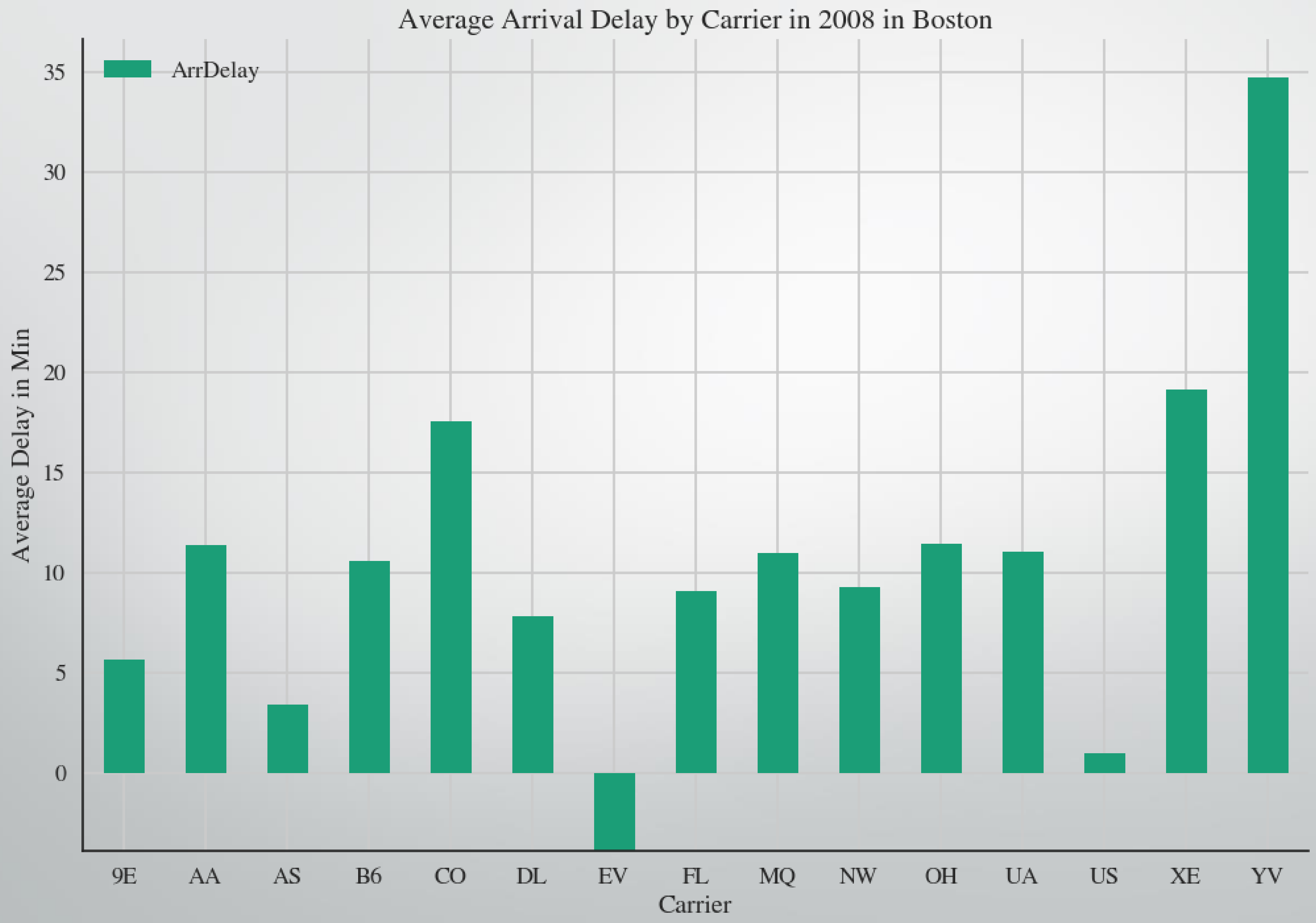


Average Arrival Delay by Carrier in 2008 in Chicago
ArrDelay
Average Delay in Min
0
5
10
15
20
25
30
35
9E
AA
AS
B6
CO
DL
EV
MQ
NW
OH
OO
UA
US
XE
YV
Carrier

Departure and Arrival Delay Distribution With respect to Carrier for Boston Logan (BOS)



Average Departure Delay by Carrier in 2008 in Boston
DepDelay
Average Delay in Min
0
5
10
15
20
25
30
35
9E
AA
AS
B6
CO
DL
EV
FL
MQ
NW
OH
UA
US
XE
YV
Carrier


Average Arrival Delay by Carrier in 2008 in Boston
ArrDelay
Average Delay in Min
35
30
25
20
15
10
5
0
9E
AA
AS
B6
CO
DL
EV
FL
MQ
NW
OH
UA
US
XE
YV
Carrier

Predictive Models for Flight Delays

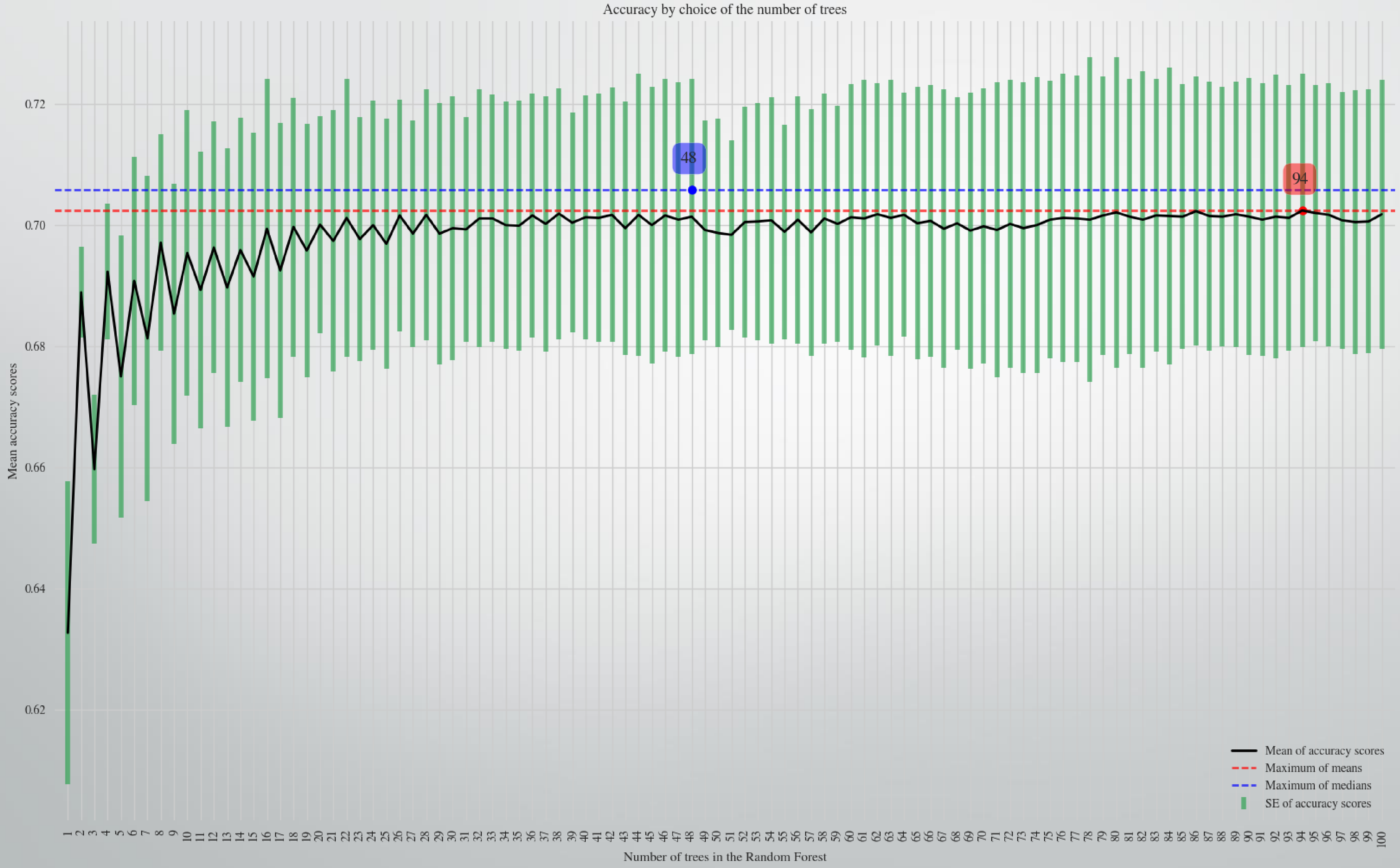
- Model to predict flight departure delays from Chicago O'Hare International Airport (ORD)
- Model to predict departure delays using subset of weather and flight data from Chicago O'Hare International Airport (ORD)
- Model to predict Departure delays using weather and flight data from Chicago O'Hare International Airport (ORD)
- Model to predict Departure delays from Chicago O'Hare International Airport (ORD) using weather and flight data with sampling.

Random Forest

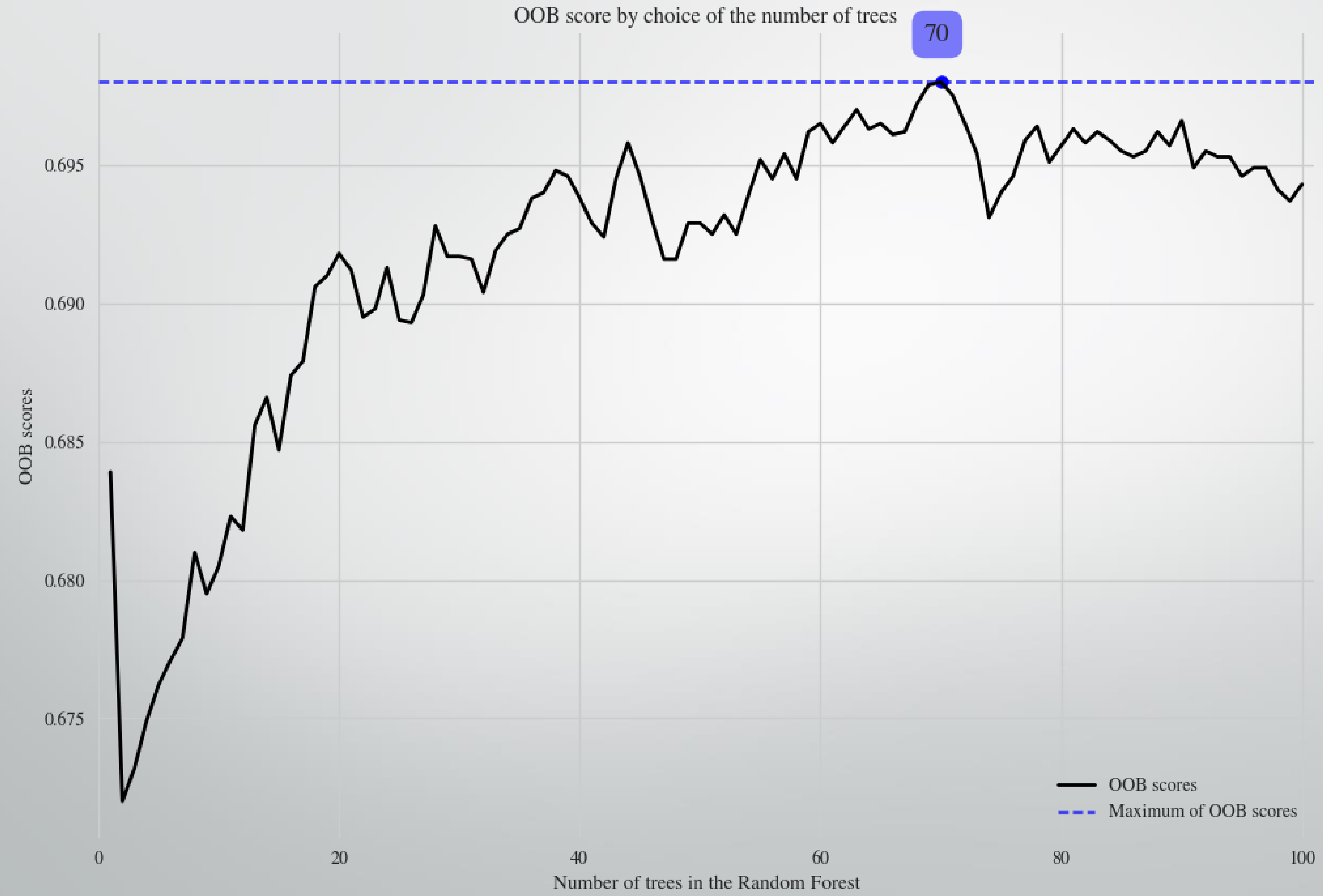
- Bagging i.e. bootstrap sampling
- Feature subsets
- Trains a number of decision trees
- Majority vote by decision trees decides the prediction
- Faster
- Better predictive performance
- Better Bias-Variance Trade-offs

Predicting flight departure delays from Chicago O'Hare International Airport (ORD)

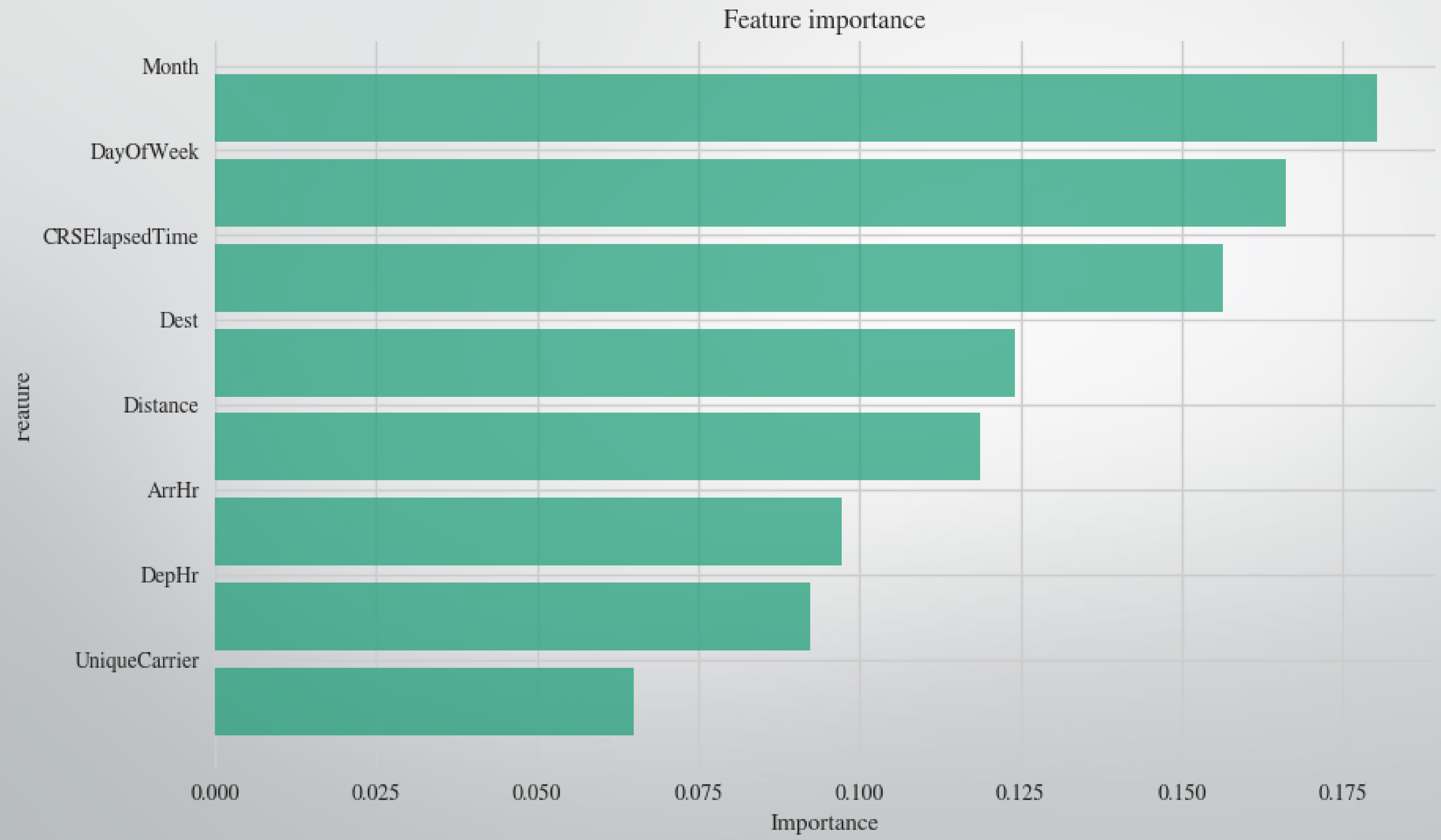
- Data Clean-up, considering random 20k samples
- No if trees = 1 to 100
- Max features for best split = 2
- Creation of classifier and calculating OOB score
- Estimating feature importance
- Calculating accuracy and Confusion Matrix



Accuracy by choice of the number of trees
Mean accuracy scores
Number of trees in the Random Forest
48
94
Mean of accuracy scores
Maximum of means
Maximum of medians
SE of accuracy scores



OOB score by choice of the number of trees
70
0.695
0.690
0.685
0.680
0.675
OOB scores
0
20
40
60
80
100
Number of trees in the Random Forest
OOB scores
Maximum of OOB scores

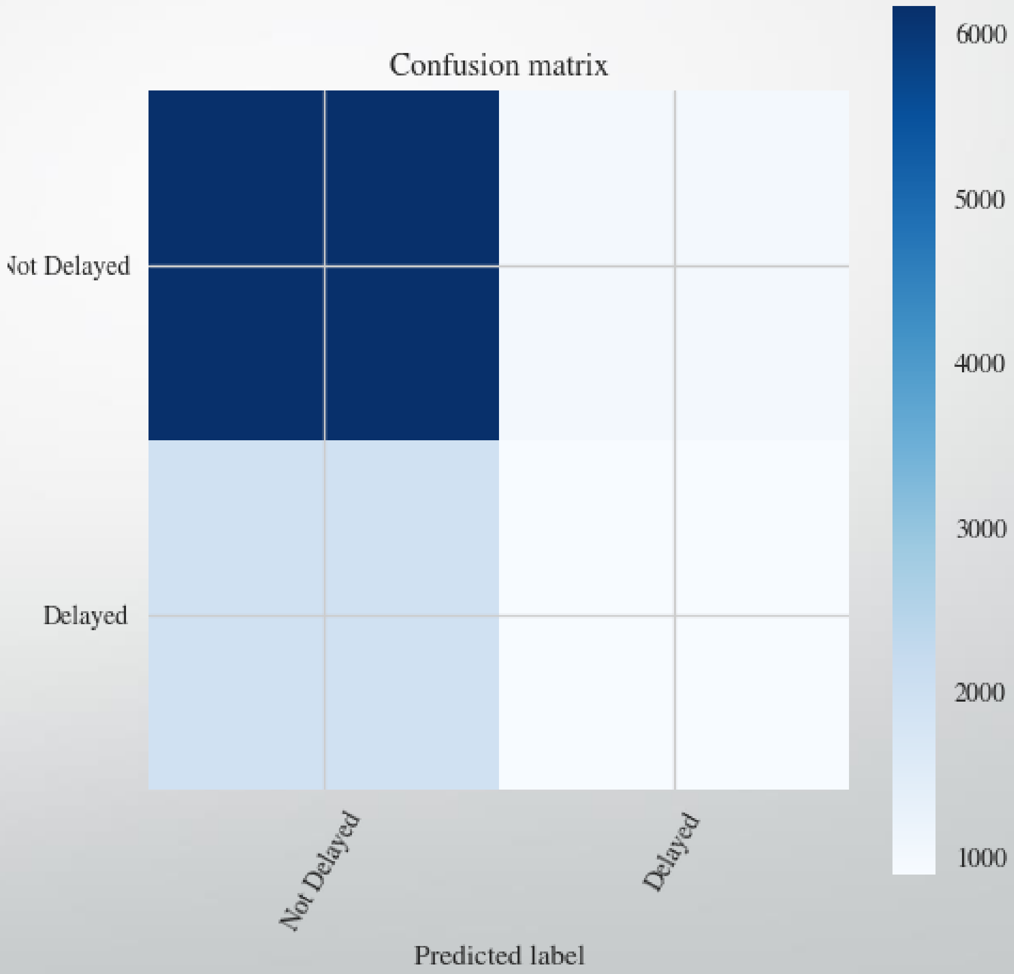

Feature importance
Month
DayOfWeek
CRSElapsedTime
Dest
Distance
ArrHr
DepHr
UniqueCarrier
feature
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
Importance

Confusion Matrix

	On time	Delayed
On time	6170	1000
Delayed	1934	896

	On time	Delayed
On time	0.860530	0.353357
Delayed	0.269735	0.316608

Precision	47%
Recall	32%
F1	38%
Accuracy	71%



Conclusion for first model

- Optimum number of trees : 70
- Cross validation score : { min= 69.97% , mean = 70.23%, max = 72.50%}
- Maximum Accuracy of model : 72%
- Classifier is guessing non-delayed flights more often than delayed
- Very less Precision and F1 score

Integrating 2008 weather data

- Total 84 features
- Important features are LOCID, Year, Month, DAYNUM, HR_LOCAL, ETMS_DEP, ETMS_ARR, MC, CEILING, VISIBLE, TEMP, WND_ANGL, WND_SPED
- Final dataset had 34 features with around 1.5 million records
- Exporting data

Predicting flight departure delays from Chicago airport with weather data

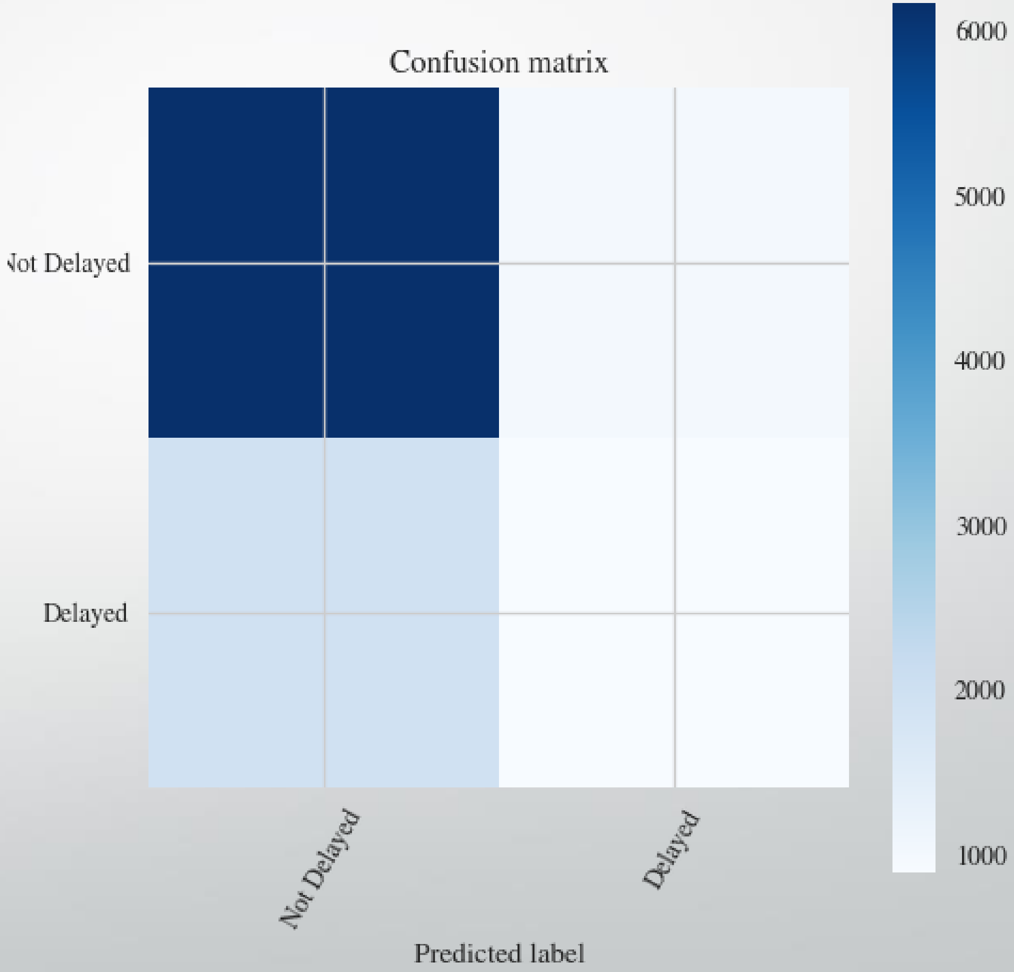
- Data Clean-up, considering random 20k samples
- No if trees = 70
- Max features for best split = 2
- Creation of classifier
- Calculating accuracy and confusion matrix
- Estimating feature importance
- Calculating Area Under Curve (AUC) for ROC curve

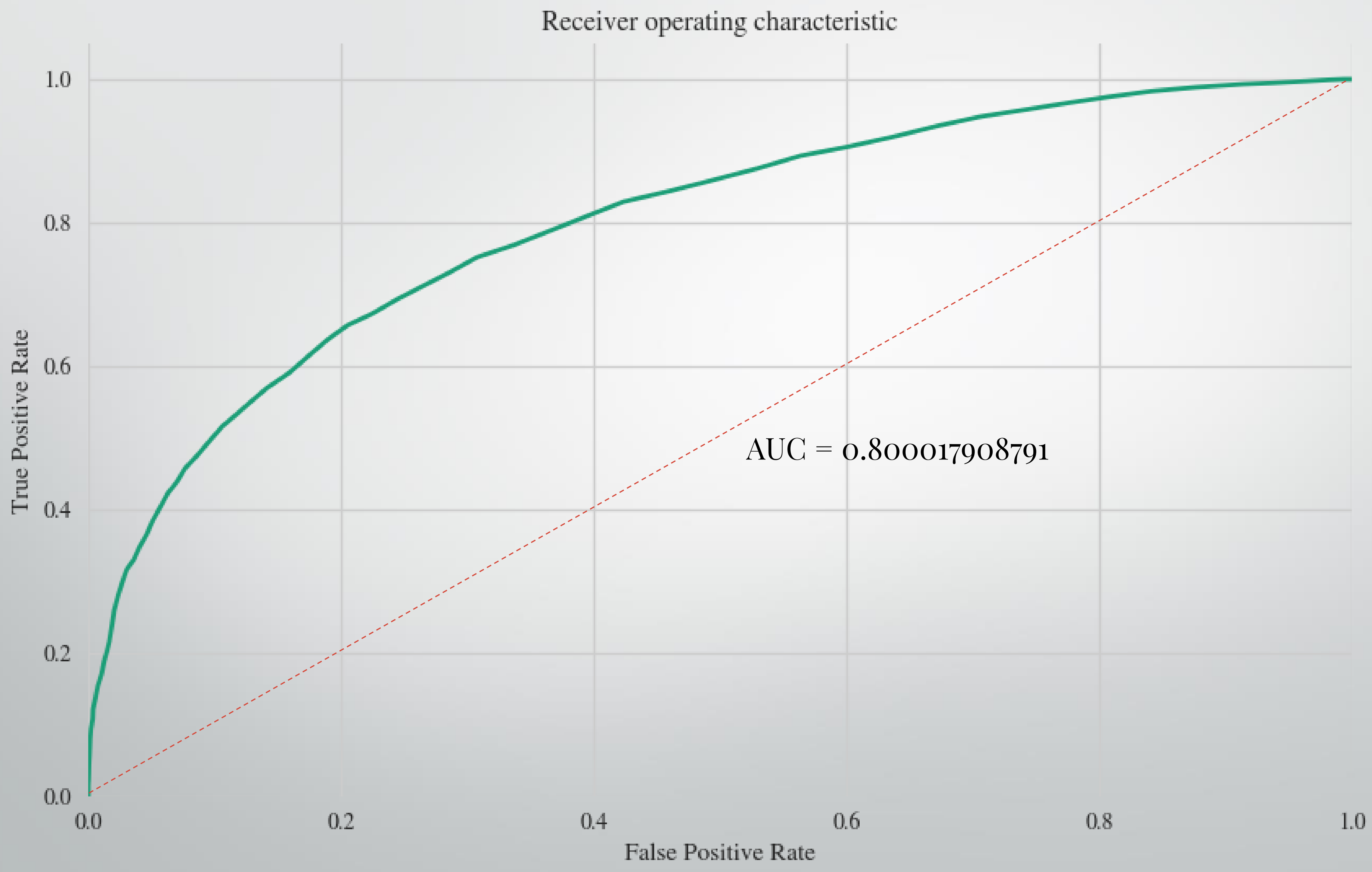
Confusion Matrix

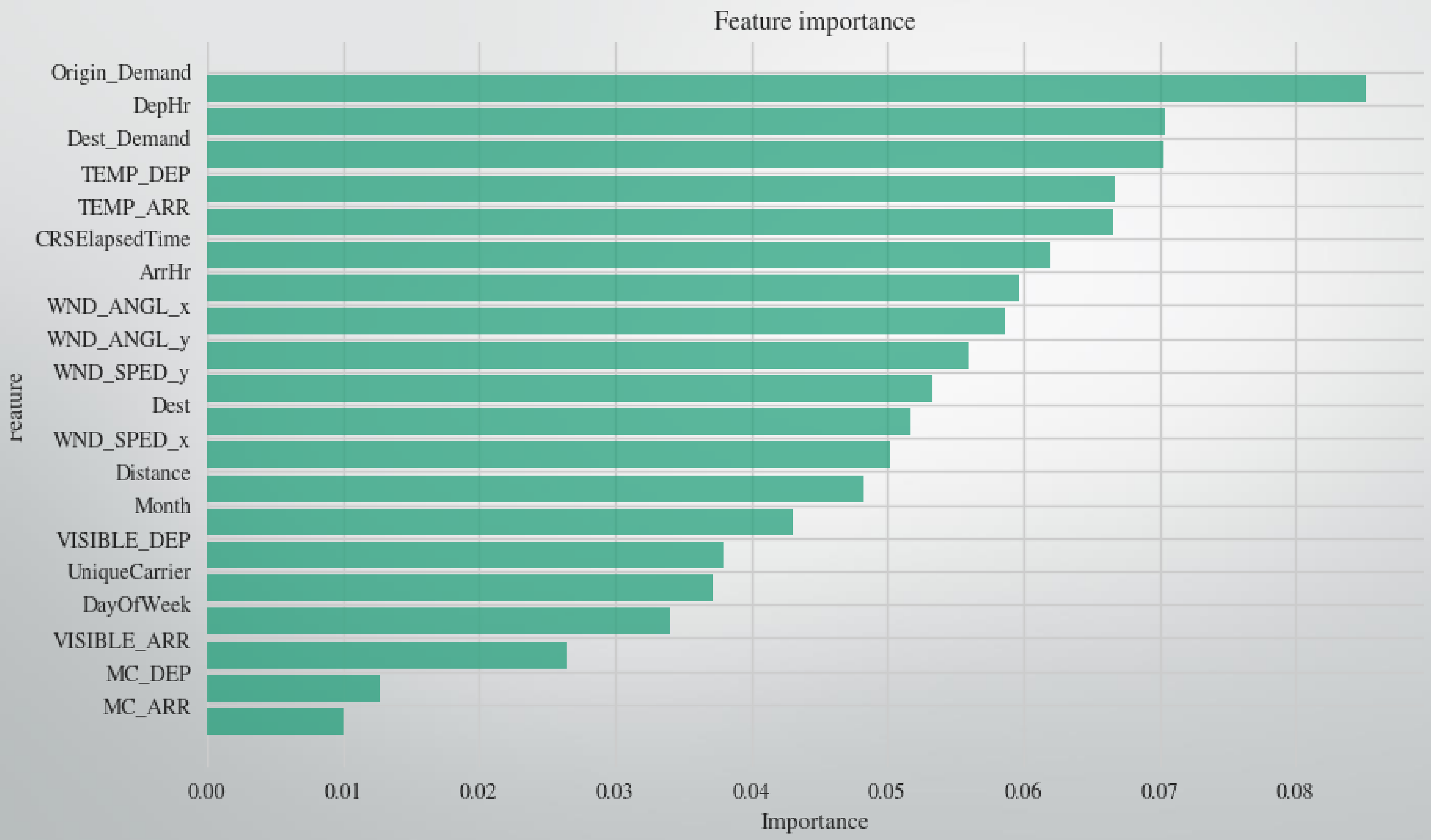
	On time	Delayed
On time	5929	663
Delayed	1706	1702

	On time	Delayed
On time	0.899424	0.194542
Delayed	0.258799	0.499413

Precision	72%
Recall	50%
F1	59%
Accuracy	76%




Receiver operating characteristic
True Positive Rate
False Positive Rate
AUC = 0.800017908791
0.0
0.2
0.4
0.6
0.8
1.0


Feature importance
Origin_Demand
DepHr
Dest_Demand
TEMP_DEP
TEMP_ARR
CRSElapsedTime
ArrHr
WND_ANGL_x
WND_ANGL_y
WND_SPED_y
Dest
WND_SPED_x
Distance
Month
VISIBLE_DEP
UniqueCarrier
DayOfWeek
VISIBLE_ARR
MC_DEP
MC_ARR
Feature
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Importance

Conclusion for second model

- Significant improvement in prediction performance
- Cross validation score : { min= 74.90% , mean = 76.15%, max = 79.92%}
- Maximum Accuracy of model : 79%
- Increase in delayed flights predictions by 18%
- 25% improvement in precision along with 21% improvement in in F1 score
- Better AUC under ROC curve

Predicting flight departure delays from Chicago airport with weather data

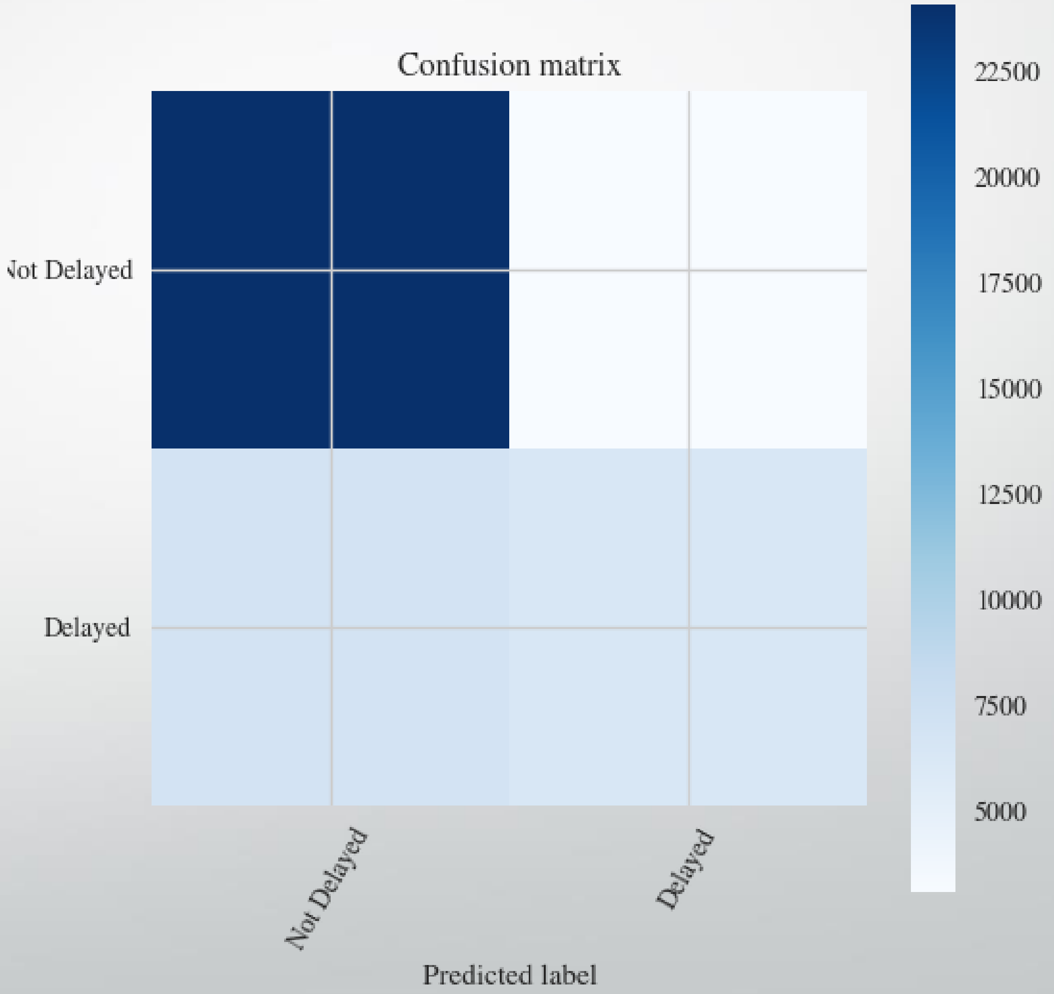
- Data Clean-up, considering <u>all samples</u>
- No if trees = 70
- Max features for best split = 2
- Creation of classifier
- Calculating accuracy and confusion matrix
- Estimating feature importance
- Calculating Area Under Curve (AUC) for ROC curve

Confusion Matrix

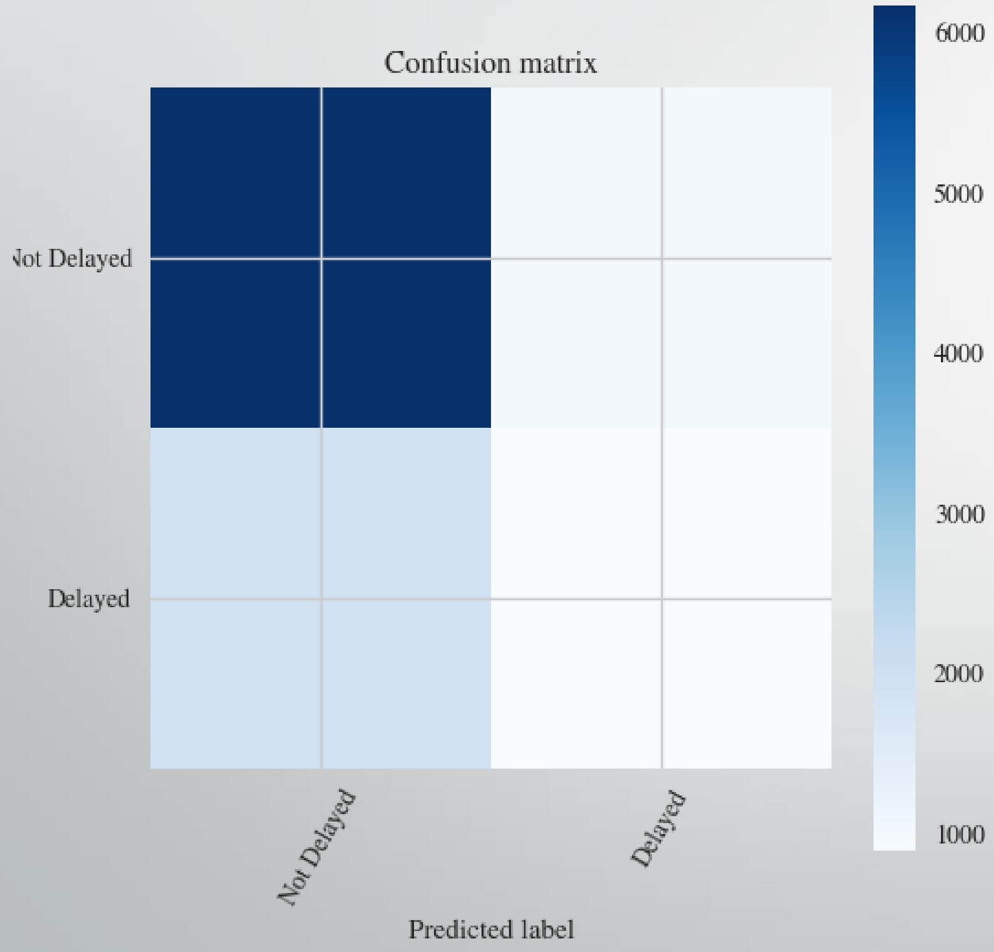
	On time	Delayed
On time	24085	3100
Delayed	6932	6360

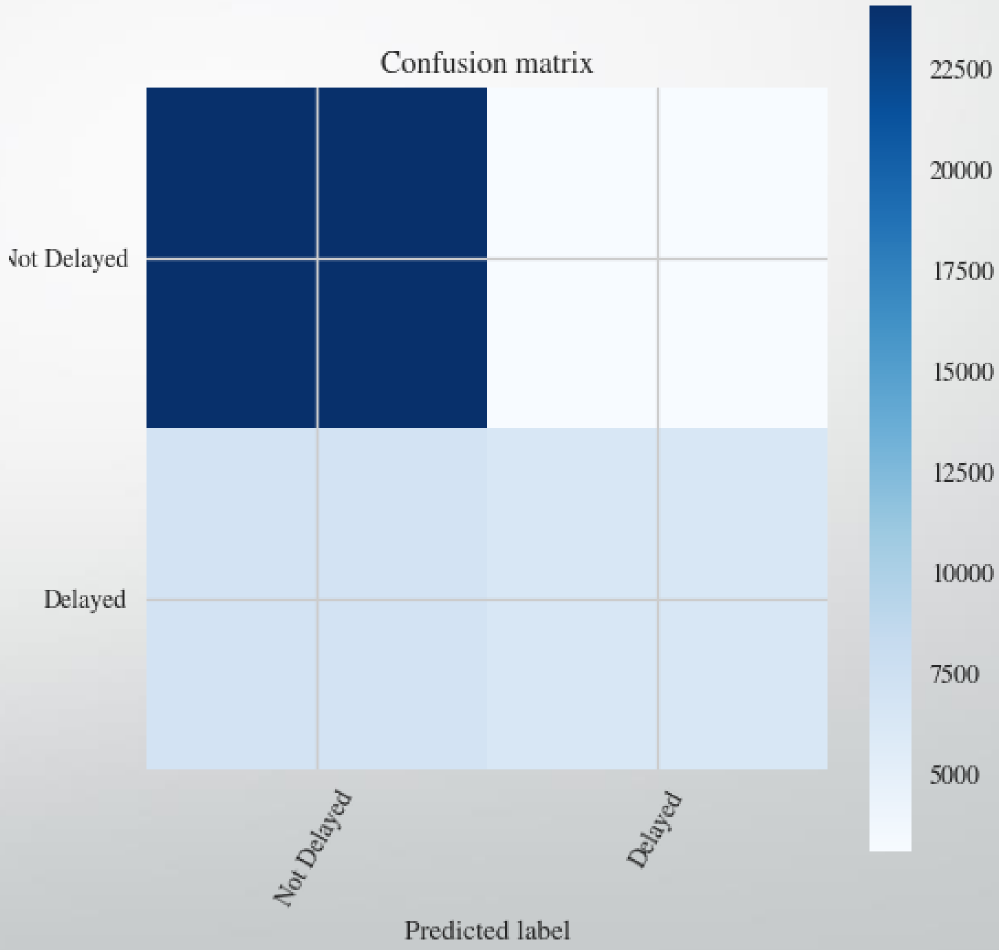
	On time	Delayed
On time	0.885967	0.233223
Delayed	0.254994	0.478483

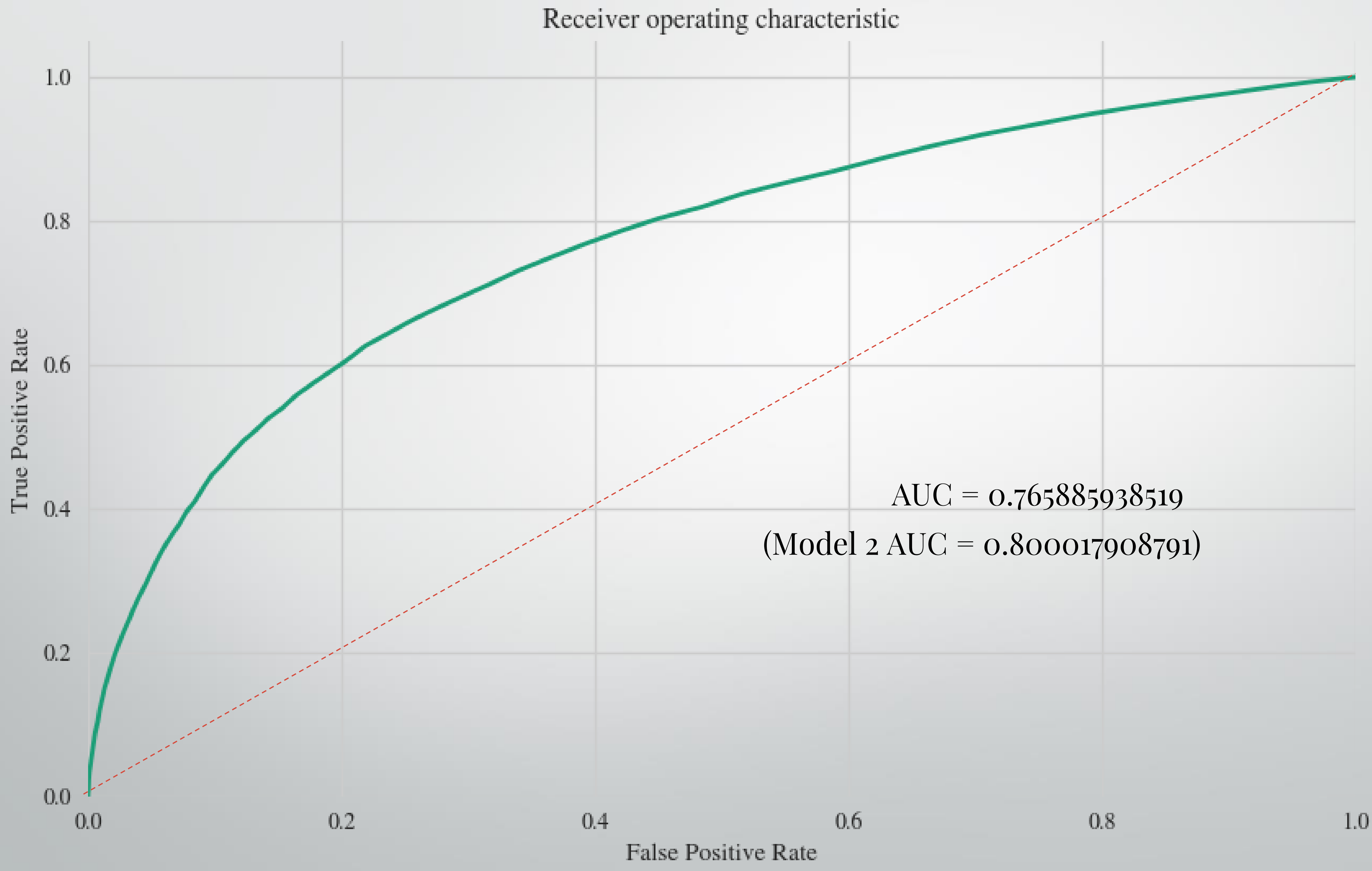
Precision	67%
Recall	48%
F1	56%
Accuracy	75%



Comparing Confusion Matrices






Receiver operating characteristic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.765885938519
(Model 2 AUC = 0.800017908791)

Conclusion for Third model

- Improved delay prediction performance
- Cross validation score : { min= 74.21% , mean = 74.86%, max = 75.52%}
- Maximum Accuracy of model : 75%
- Both accuracy and precision have decreased by 2%
- AUC have decreased
- Adding more number of samples did not help

Sampling (SMOTE)

- Need of sampling
  - Imbalanced data
  - Majority class has no of samples 3 to 4 times greater than minority class
  - Hence, 75% accuracy if every flight is classified as on time (Accuracy Paradox)
- Synthetic Minority Over-sampling Technique (SMOTE)
  - Oversampling as well as Undersampling
- But will sampling improve model performance ?

Predicting flight departure delays from Chicago airport with weather data and Sampling

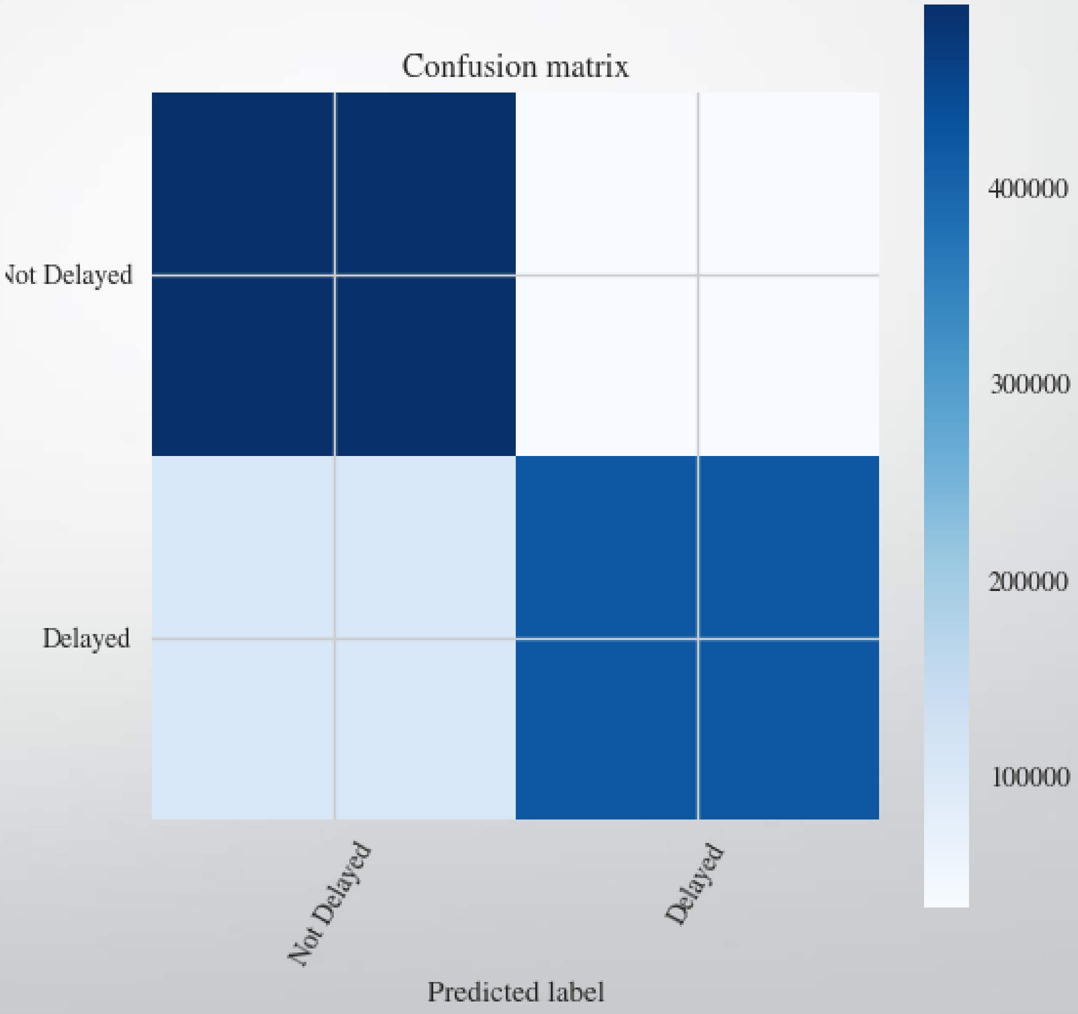
- Data Clean-up, considering all available samples
- No if trees = 70
- Max features for best split = 2
- Creation of classifier
- Calculating accuracy and confusion matrix
- Estimating feature importance
- Calculating Area Under Curve (AUC) for ROC curve

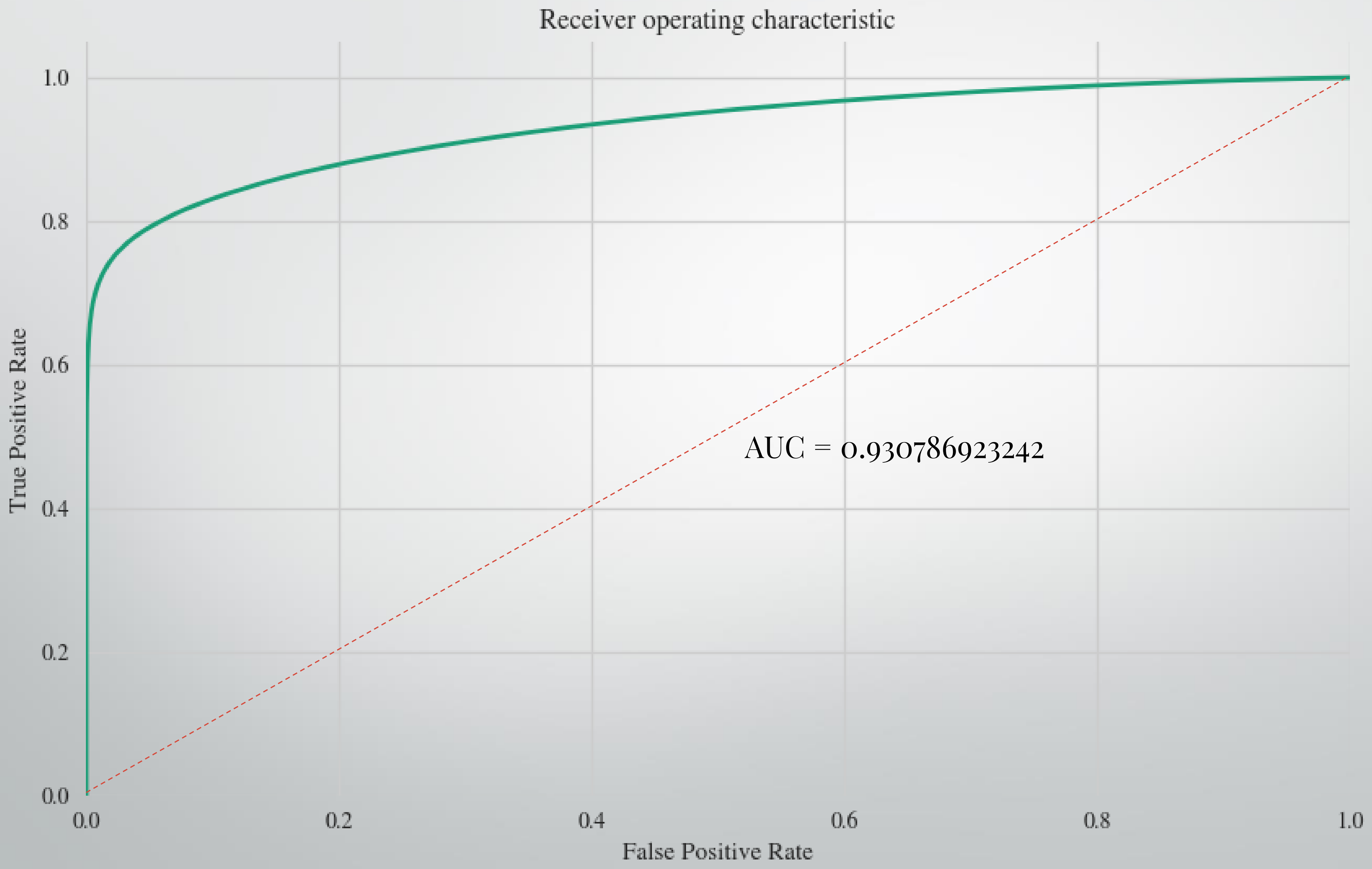
Confusion Matrix

	On time	Delayed
On time	493780	32630
Delayed	103239	422192

	On time	Delayed
On time	0.938014	0.062101
Delayed	0.196119	0.803516

Precision	93%
Recall	80%
F1	86%
Accuracy	87%




Receiver operating characteristic
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
AUC = 0.930786923242

Final results

- 80% of the time our model is predicting delayed flights
- Cross validation score : { min= 86.73% , mean = 86.87%, max = 87.02%}
- Maximum Accuracy of model : 87%
- Increase in delayed flights predictions by 33%
- 26% improvement in precision along with 30% improvement in in F1 score
- Much better AUC under ROC curve
- Sampling improved performance of model

How to further increase performance

- Number of trees
  - More the better
  - Diminishing results (once you go above a certain number of trees the predictive power to algorithm doesn’t improve much)
  - Slower to construct with more no of trees
- Number of features
  - More features reduce bias
  - But increases the correlation of trees
- More samples

References

- BTS, US Passenger Miles Table
- Sun Choi, Young Jin Kim, Simon Briceno and Dimitri Mavris - Prediction of Weather-induced Airline Delays Based on Machine Learning Algorithms Georgia 30332–0250
- L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001
- N. V. Chawla, “C4.5 and imbalanced data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure,” in *Proceedings of the ICML, Workshop on Learning from Imbalanced Datasets II*, Washington DC, 2003.
- S. S. P. Reshma C. Bhagat, “Enhanced smote algorithm for classification of imbalanced big-data using random forest,” in Proceedings of the Advance Computing Conference (IACC), 2015.