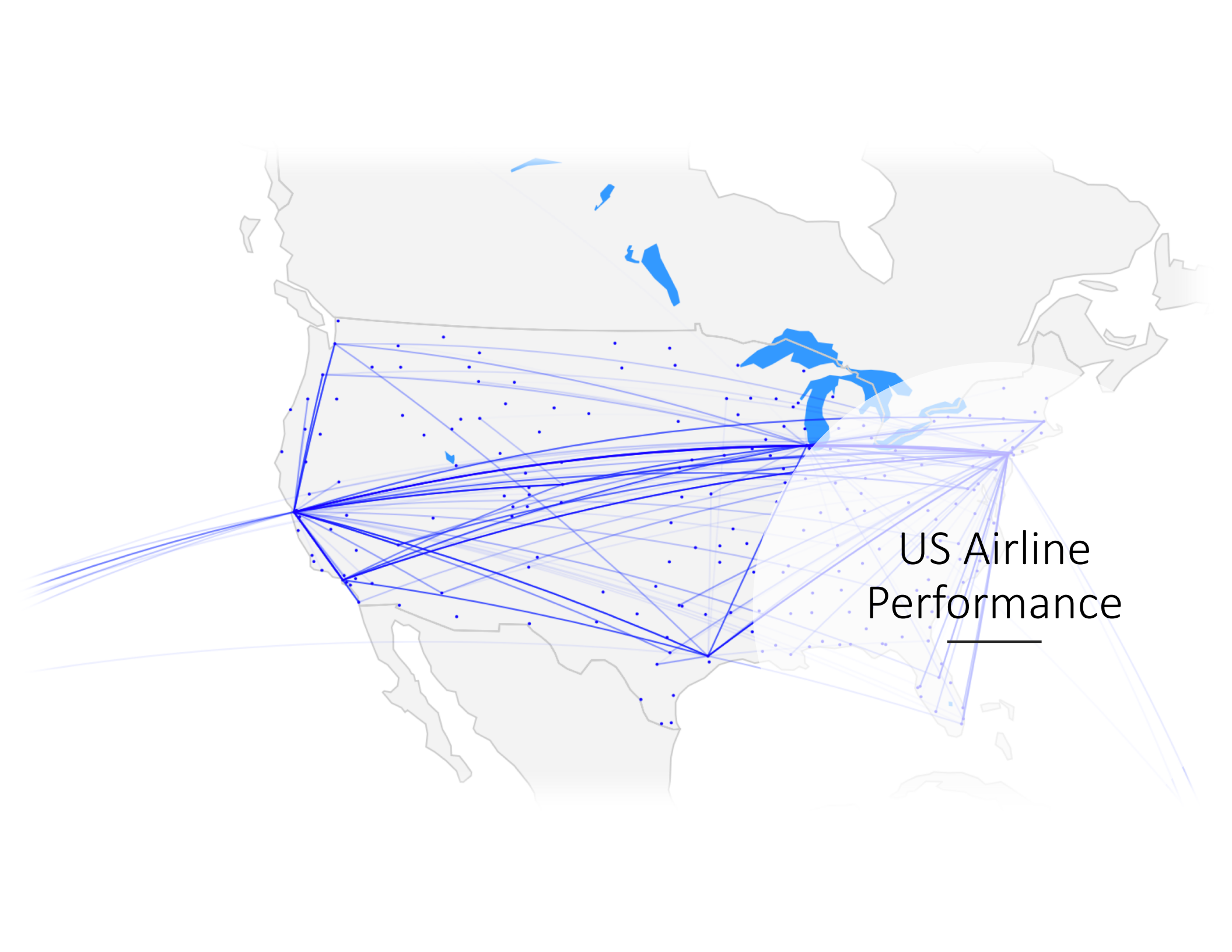

US Airline
Performance

Scope

PROBLEM

EXPLORE THE DATA

MODELING

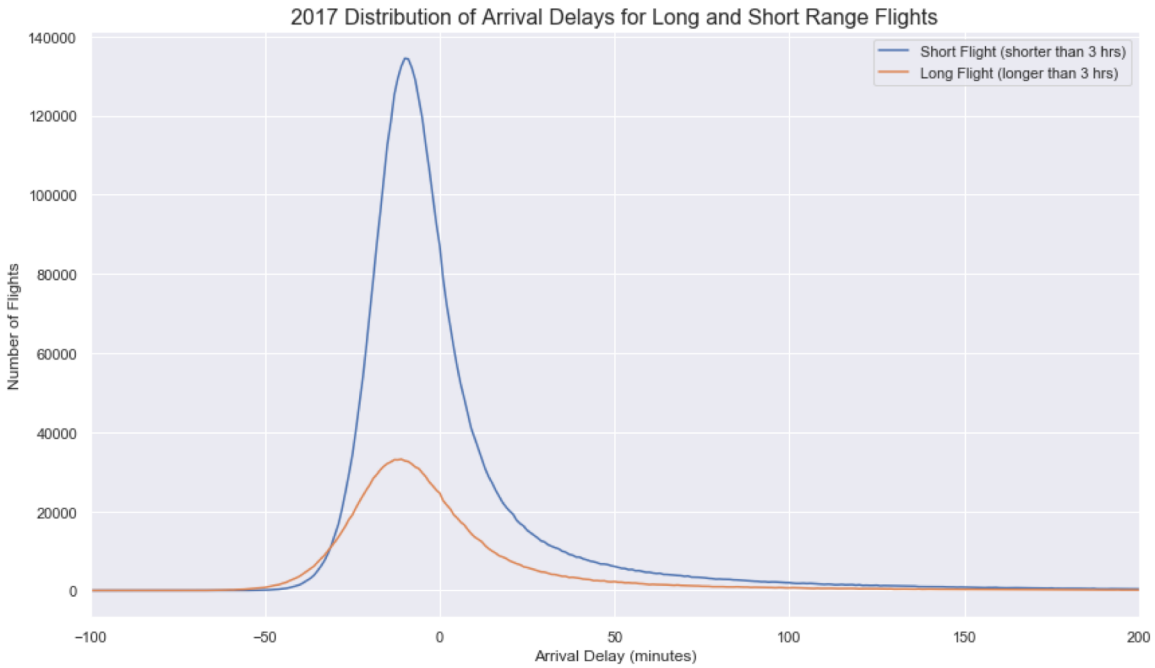
RESULTS

Problem

- Will the flight be delayed or on-time?
- Why are flights delayed?
- Are 80% of flights really on-time?
- Flight delays cost $32.9B ($16.7B cost to passengers alone!)

Explore the Data – Delay Distribution

- Distribution of arrival delays is skewed for both long and short flights

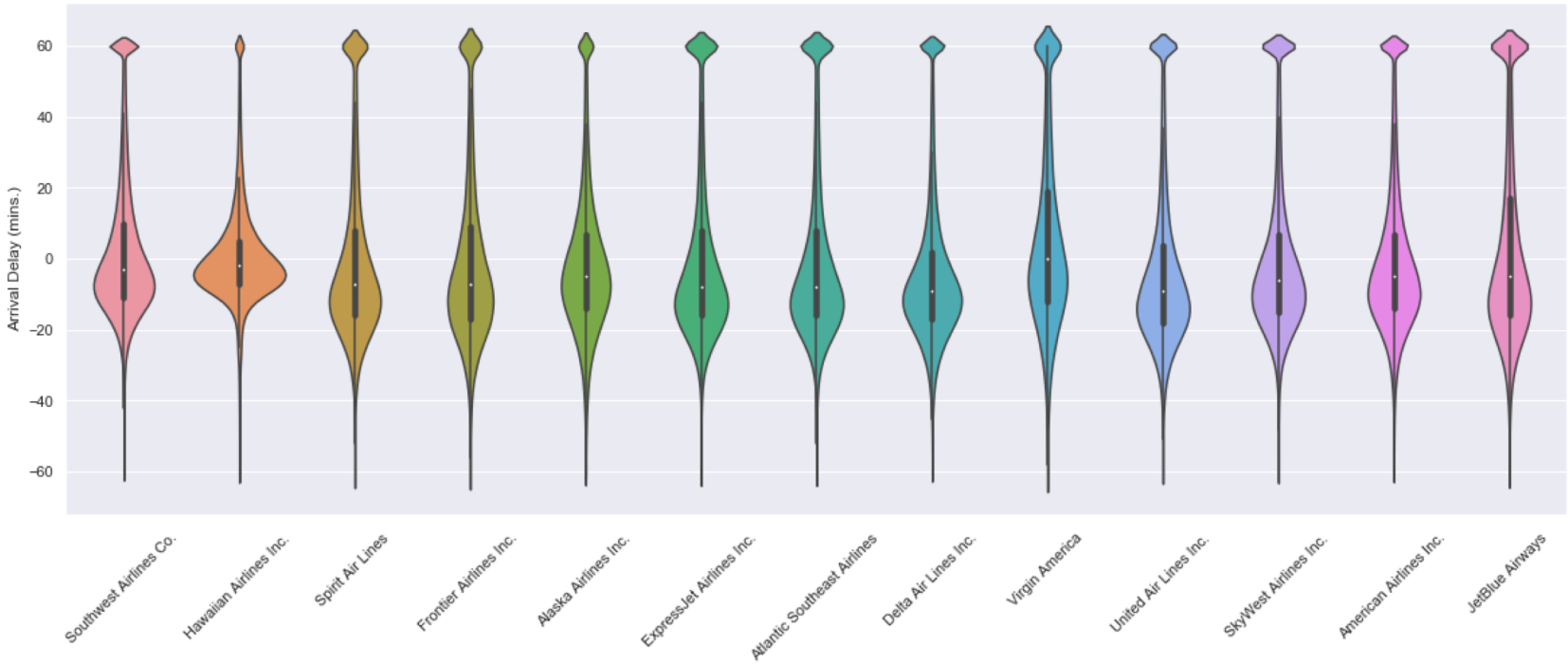


LONG FLIGHTS	
AVERAGE	2.8 min.
MEDIAN	-7 min.
STD	46 min.

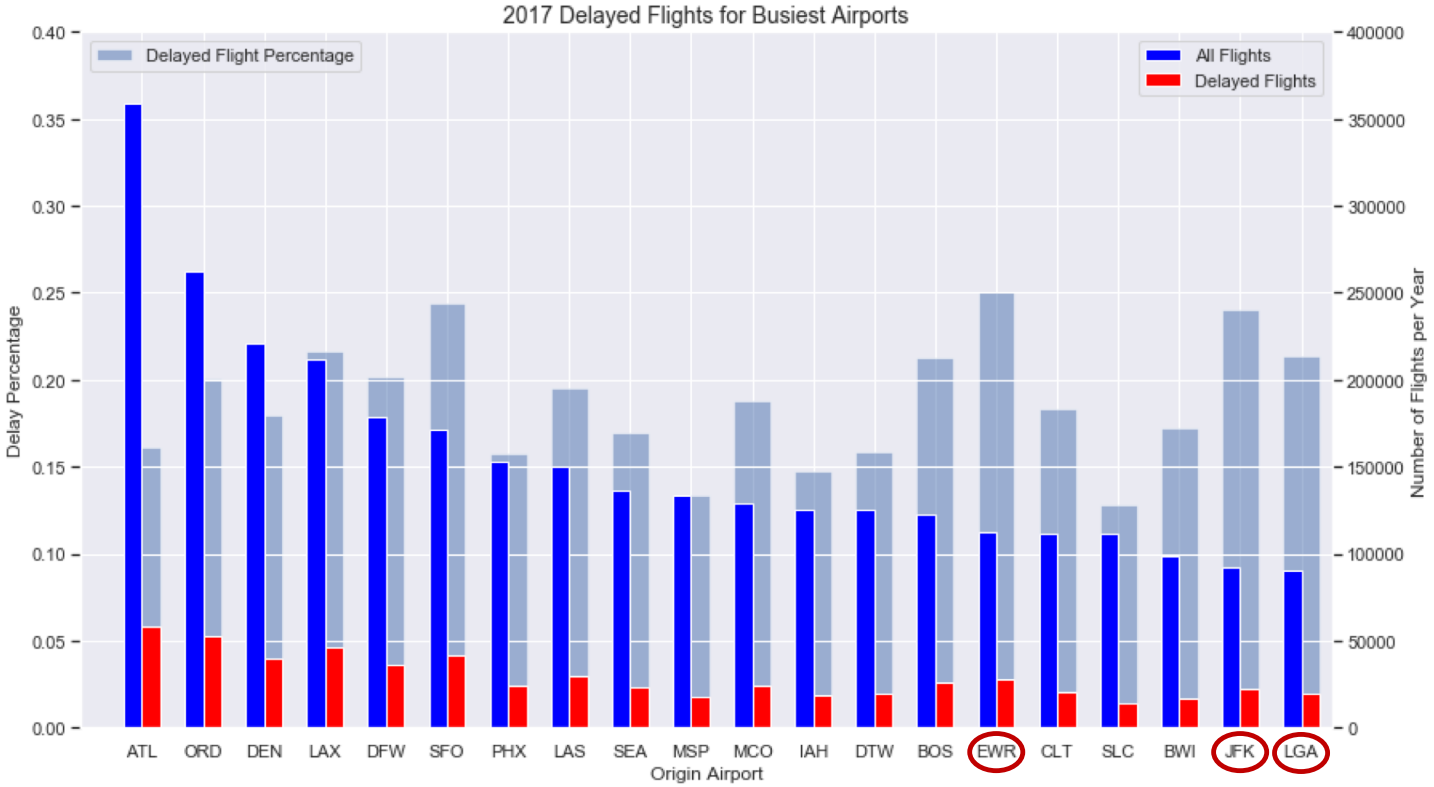
SHORT FLIGHTS	
AVERAGE	4.9 min.
MEDIAN	-6 min.
STD	45.8 min.

Explore the Data – Airline Performance

- Distribution of arrival delays by airline



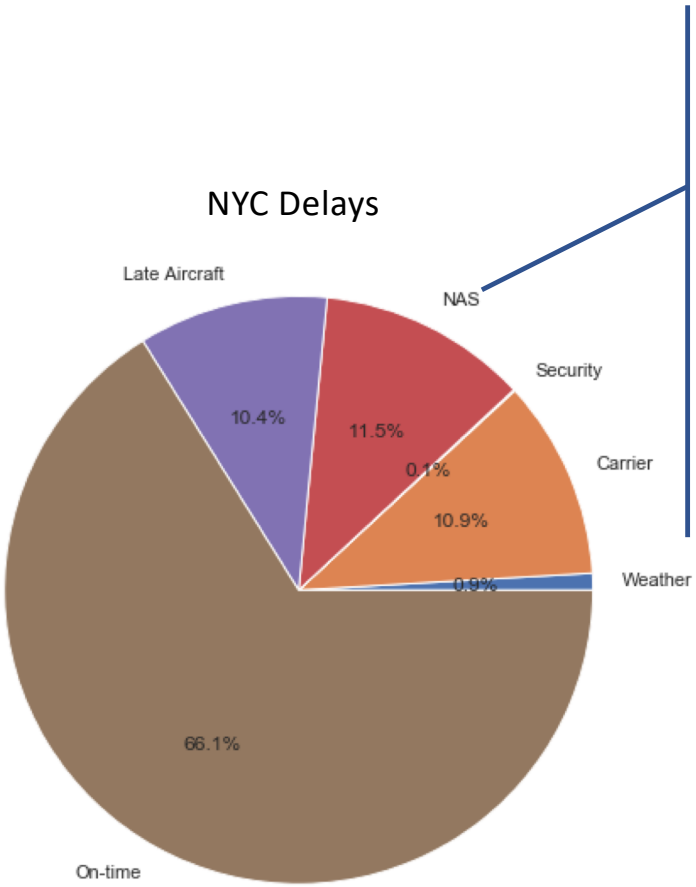
Explore the Data – Delays by Airport

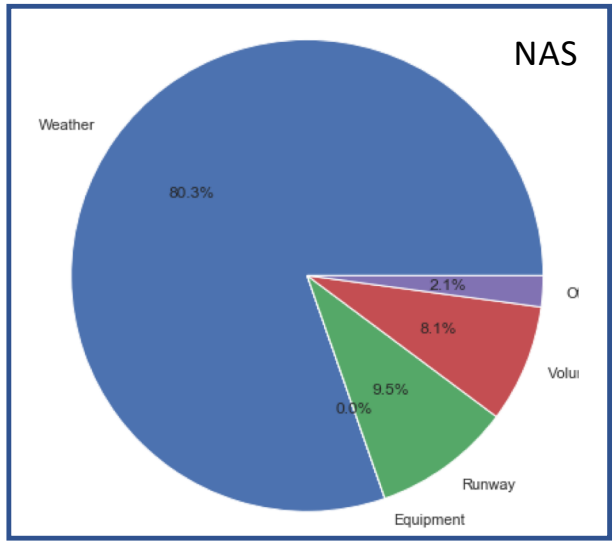


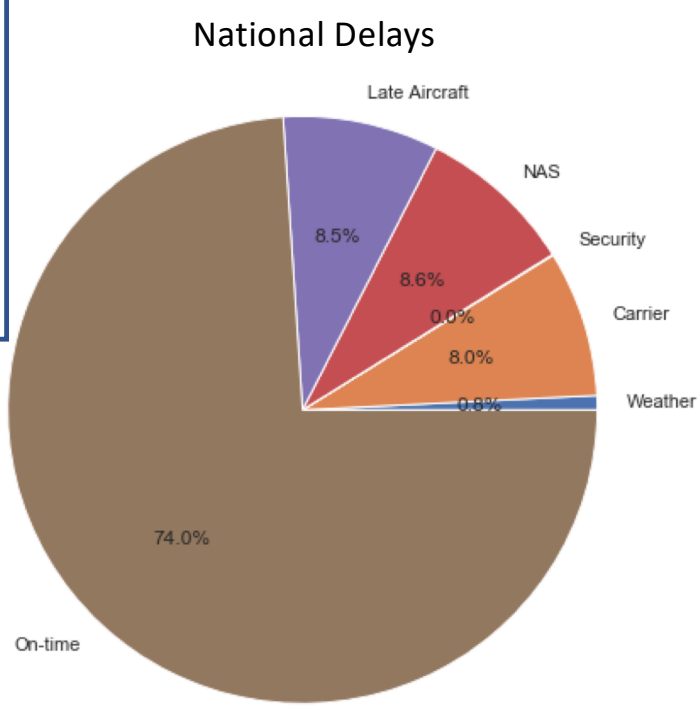
2017
TOP 10 MOST DELAYED AIRPORTS
(PERCENT OF DELAYED FLIGHTS)

EWR	**0.25**
SFO	0.24
JFK	**0.24**
LAX	0.22
LGA	**0.21**
BOS	0.21
DFW	0.20
ORD	0.20
LAS	0.19
MCO	0.19

Explore the Data – NYC vs National Avg.

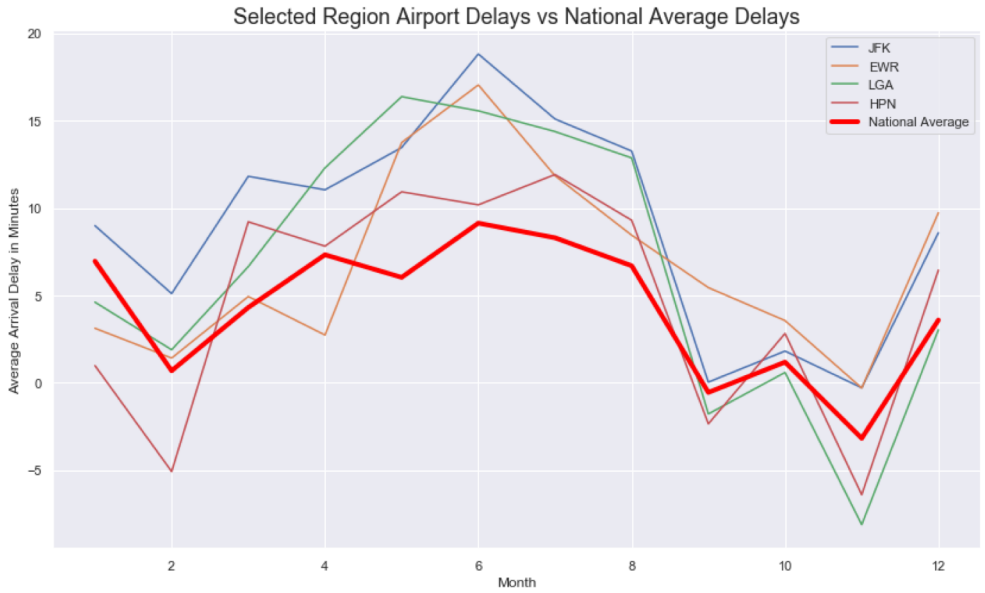


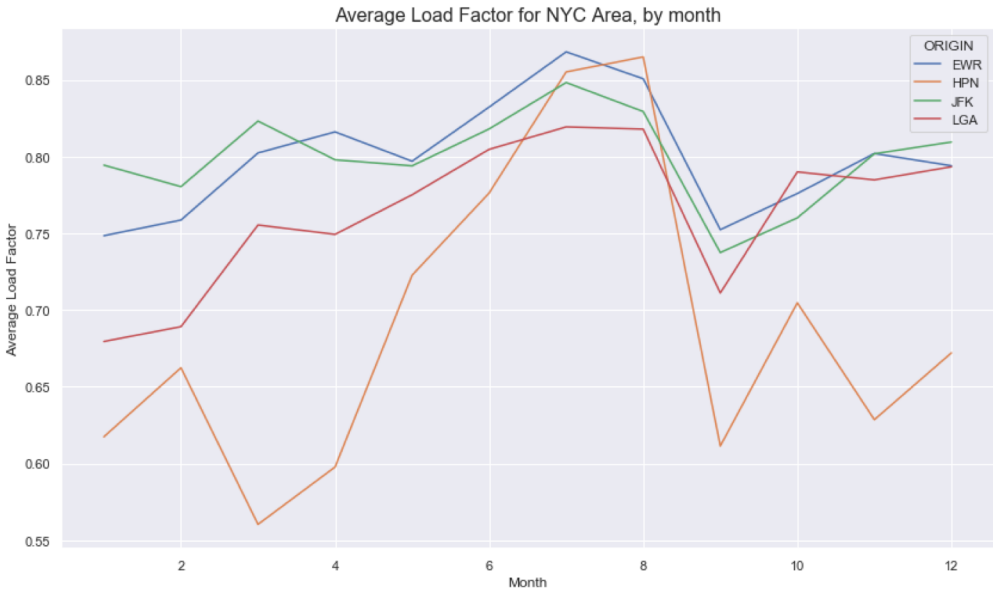




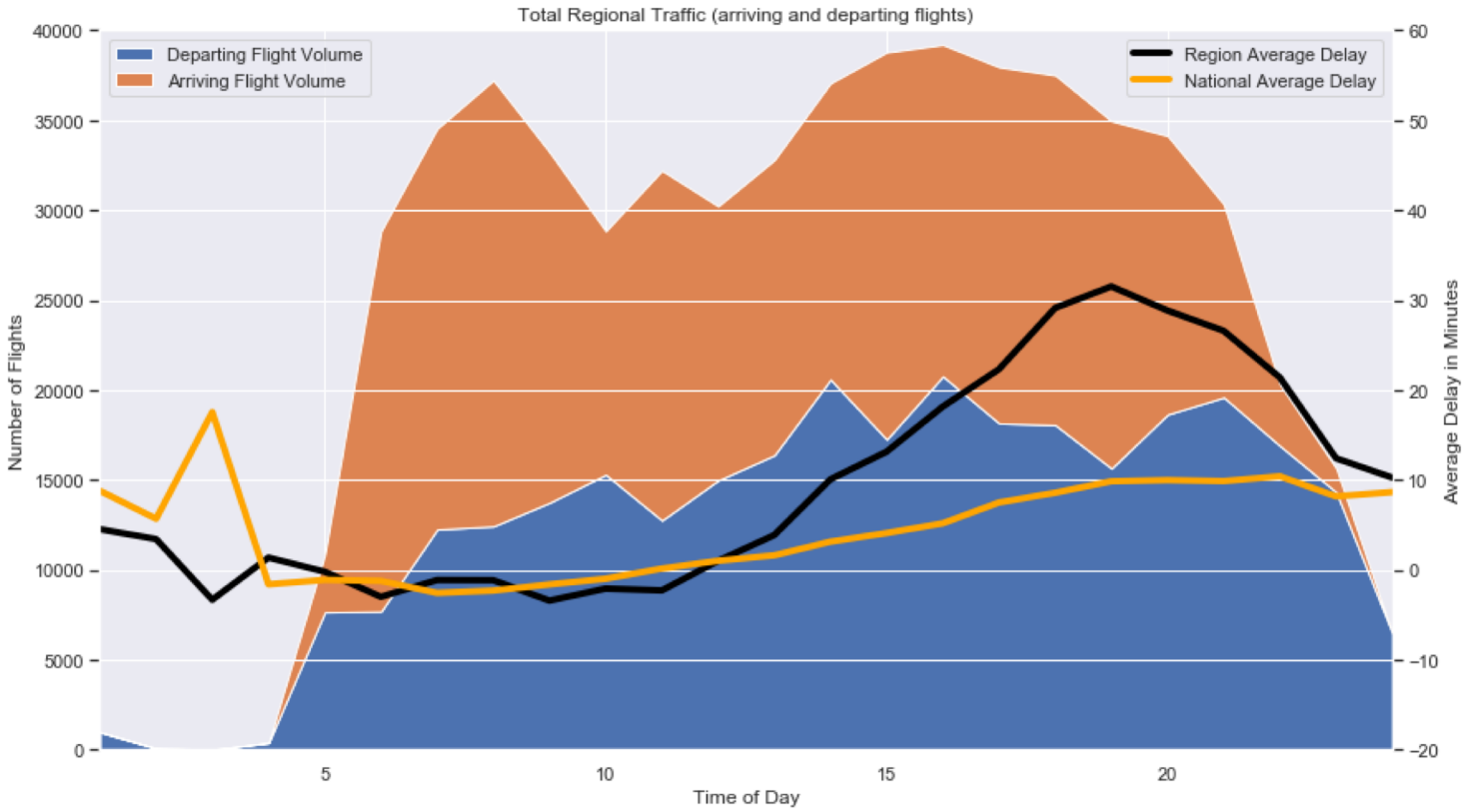
Explore the Data – NYC Performance

- Distribution of arrival delays by airport

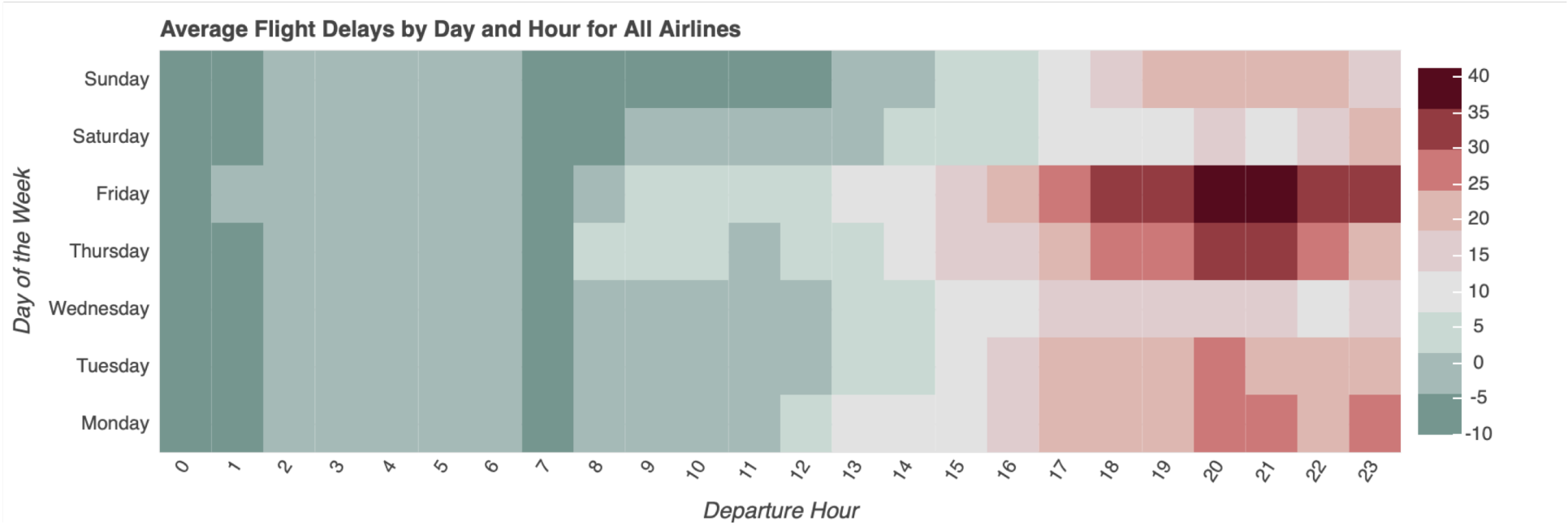




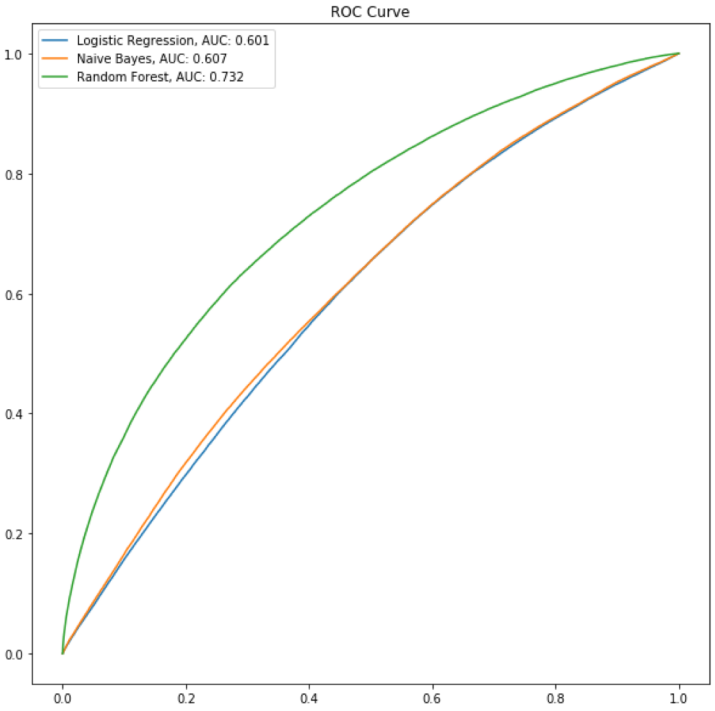
Explore the Data – NYC Flight Volume



Explore the Data – NYC Schedule



Modeling – Initial Models



Logistic Regression		Actual		
		On-time	Delayed	
Predicted	On-time	401,109	-	401,109
	Delayed	90,673	-	90,673
		491,782	-	

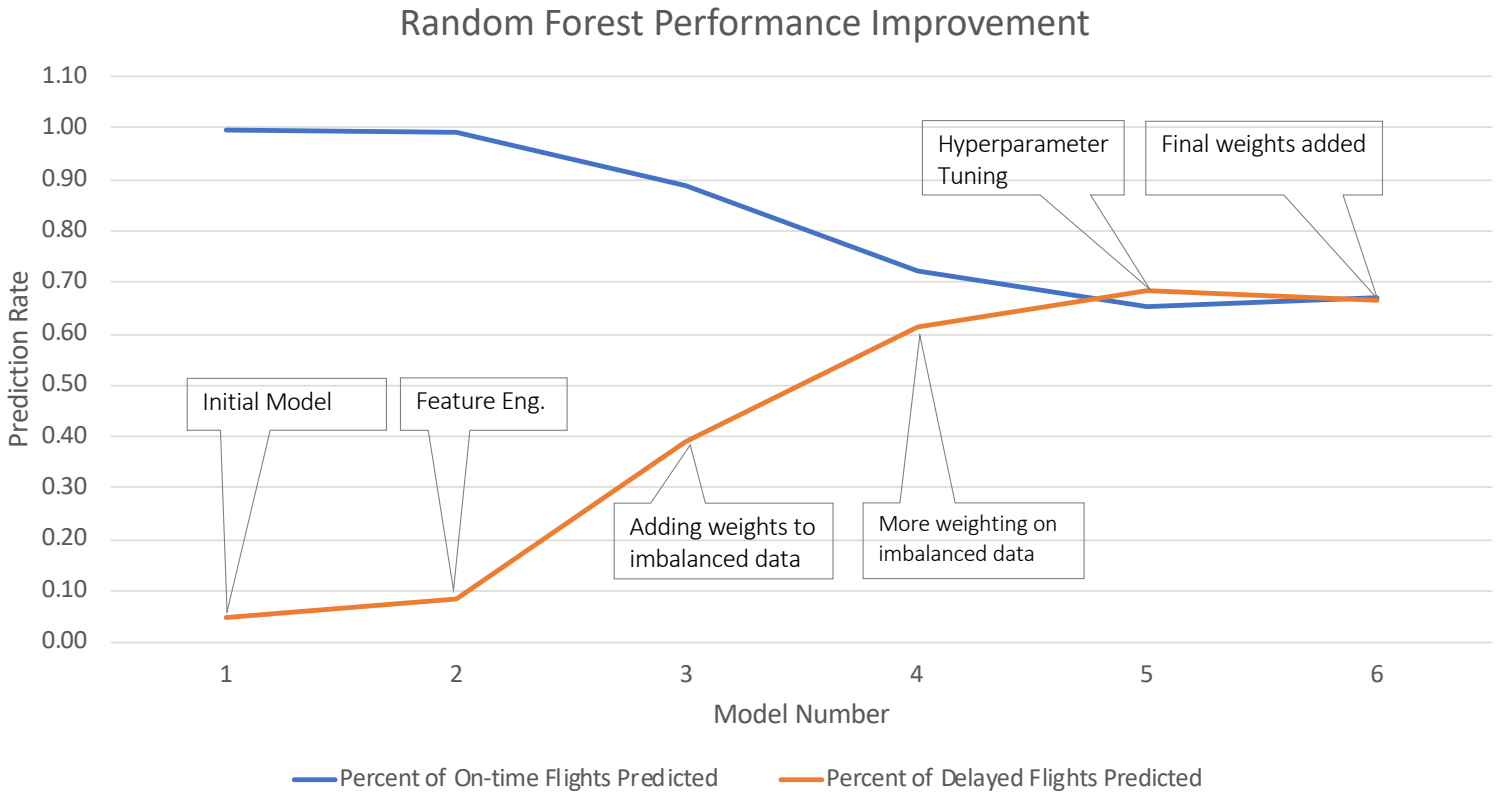
Naïve Bayes		Actual		
		On-time	Delayed	
Predicted	On-time	401,109	-	401,109
	Delayed	90,673	-	90,673
		491,782	-	

Random Forest		Actual		
		On-time	Delayed	
Predicted	On-time	399,344	1,765	401,109
	Delayed	86,270	4,403	90,673
		485,614	6,168	

Modeling – Feature Engineering

- Hourly buckets for departures and arrivals
- Direction of flight
- Weather (precipitation, visibility, wind, temperature)
- Origin flight volume during scheduled departure
- Monthly passenger loading data

Modeling – Improving Random Forest



Results

- Precision = 269,557 299,732 = 0.90
- False Positive Rate = 131,552 / 90,673 = 1.45
- True Negative Rate = 60,498 / 90,673 = 0.67 (Delayed Prediction Rate)
- Recall = 269,557 / 401,109 = 0.67 (On-time Prediction Rate)

Tuned Model and Final Weights Added 1:4.85		Actual		
		On-time	Delayed	
Predicted	On-time	**269,557**	131,552	401,109
	Delayed	30,175	**60,498**	90,673
		299,732	192,050	

Results – Next Steps

- Add Features – Plane information data form FAA
- Add Features – Plane maintenance history
- Add Features – Historical route performance
- Expand Flight Data – Add other years' data for training
- Expand Models – Introduce Neural Networks