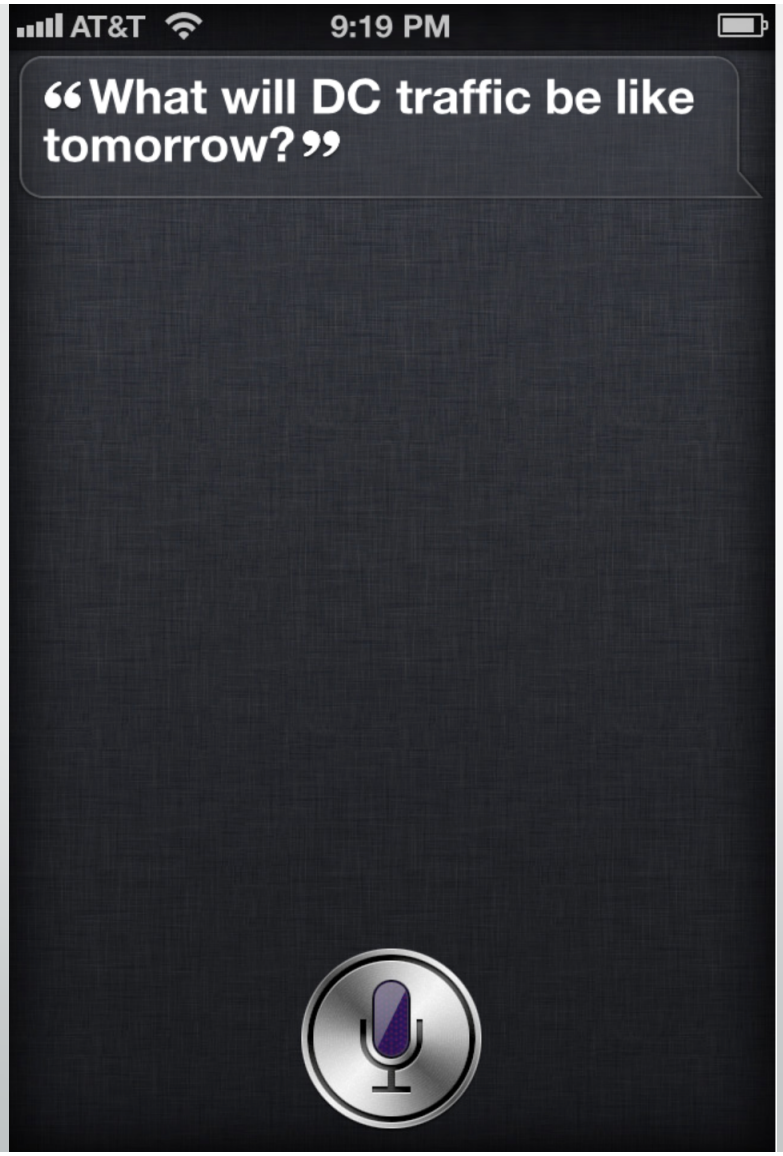
Predicting D.C. Traffic

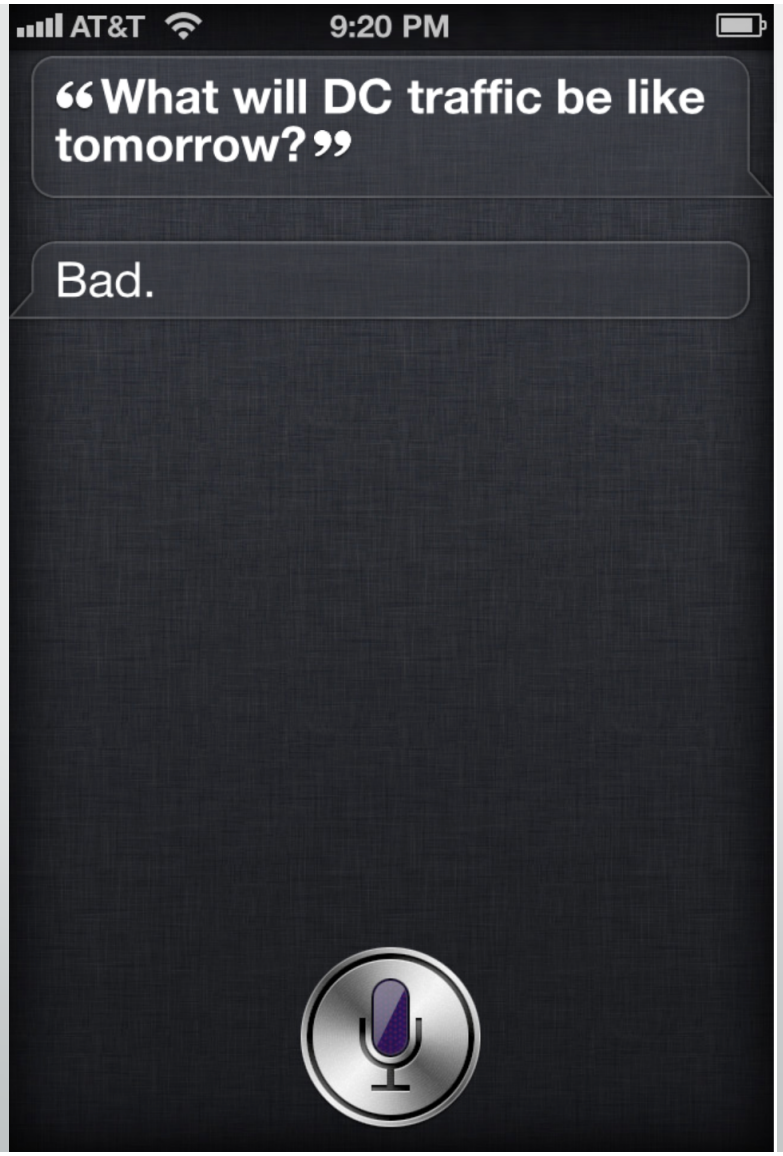
Michael Sanders
May 26, 2017

michaeljsanders.com

Agenda

- Problem statement
- Methodology
- EDA
- Modeling
- Results
- Project roadmap
- Q&A


AT&T
9:19 PM
“What will DC traffic be like tomorrow?”


AT&T
9:20 PM
“What will DC traffic be like tomorrow?”
Bad.


AT&T
9:21 PM
“What will DC traffic be like tomorrow?”
Bad.
“I know, I know. But HOW bad?”


AT&T
9:23 PM
“What will DC traffic be like tomorrow?”
Bad.
“I know, I know. But HOW bad?”
It's in the documentation... smh

The Usual Suspects are Only So Helpful

- Time of day
- Weather
- Accidents
- Construction
- “Safe”Track

What about Volume?

- How many people are commuting into the city on a given day?
- Why does that number change?
- Is it predictable?

The Hypothesis

- Political / governmental activities & events impact traffic
- More events ➔ more people coming into the city
- Every hearing, conference, legislative session, etc., involves an ecosystem of people
- Can we predict relative badness of traffic based on events?

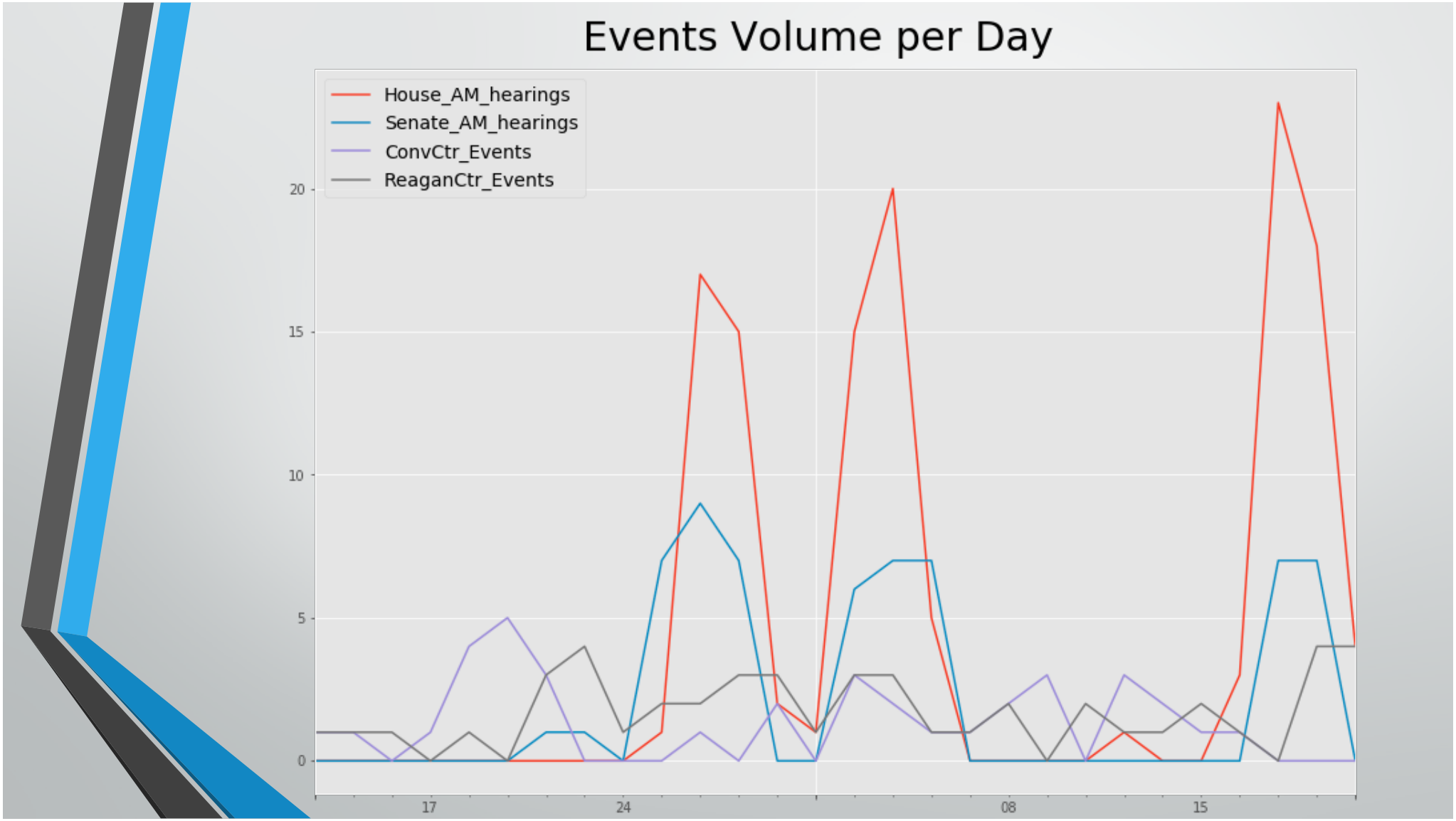
Approach

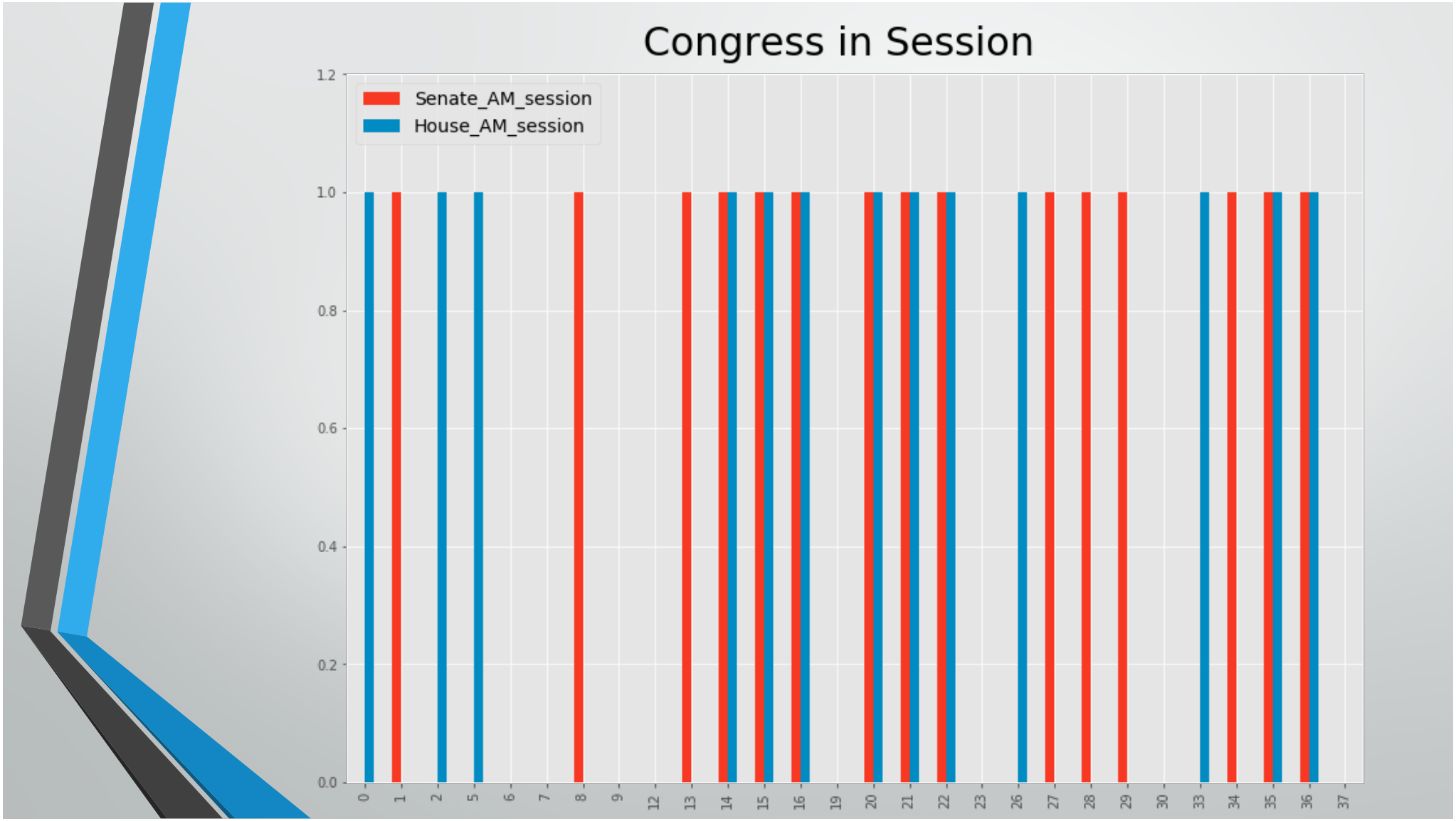
- Scope:
  - Morning commute
  - Upfront feature selection
- Collect events data
- Capture morning traffic data
- Build models
- Profit

Event Data

Before 11 a.m.:

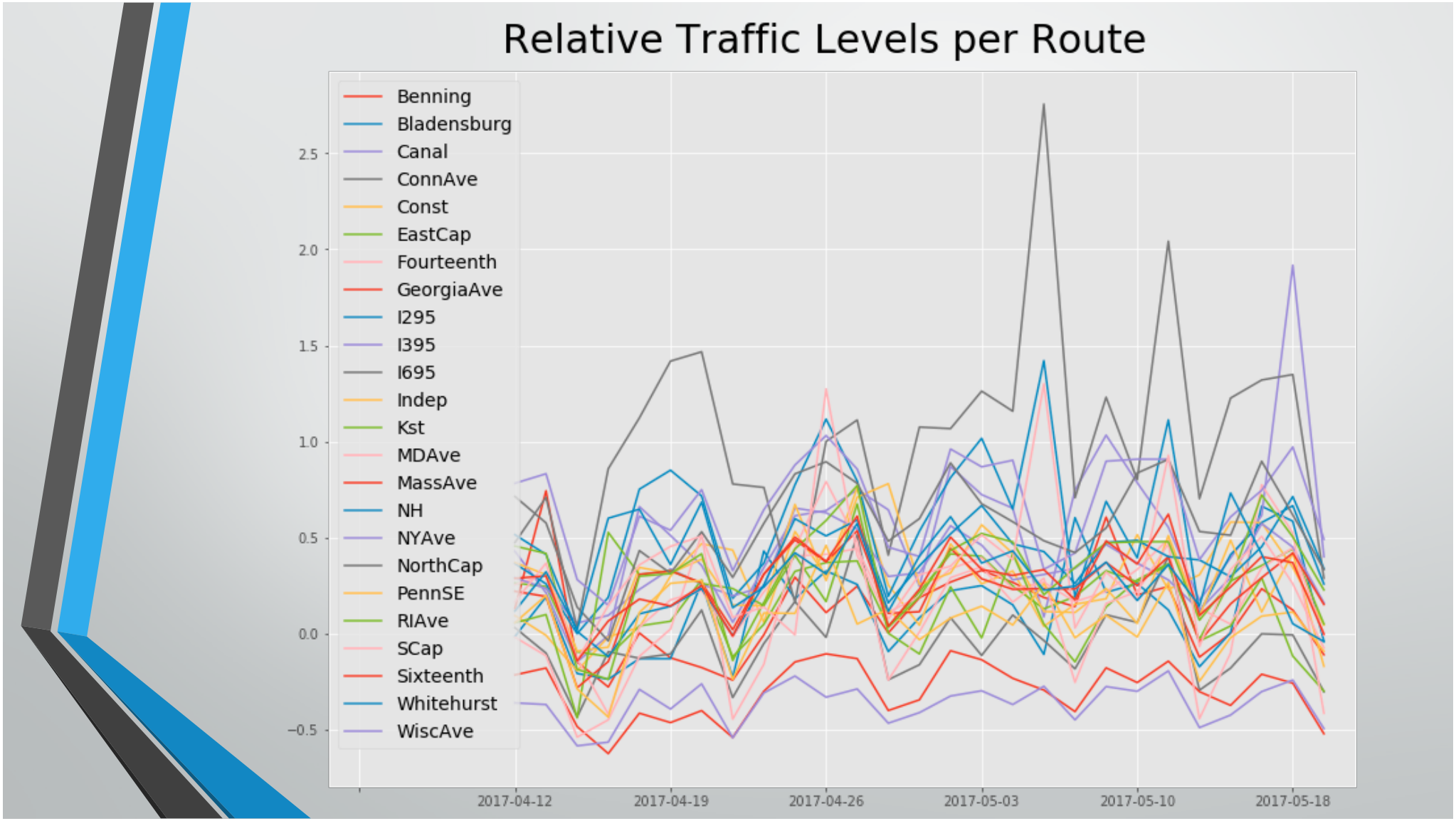
- Senate in session (binary)
- House in session (binary)
- Senate committee hearings
- House committee hearings
- Events at the Walter Washington Convention Center
- Events at the Ronald Reagan Building

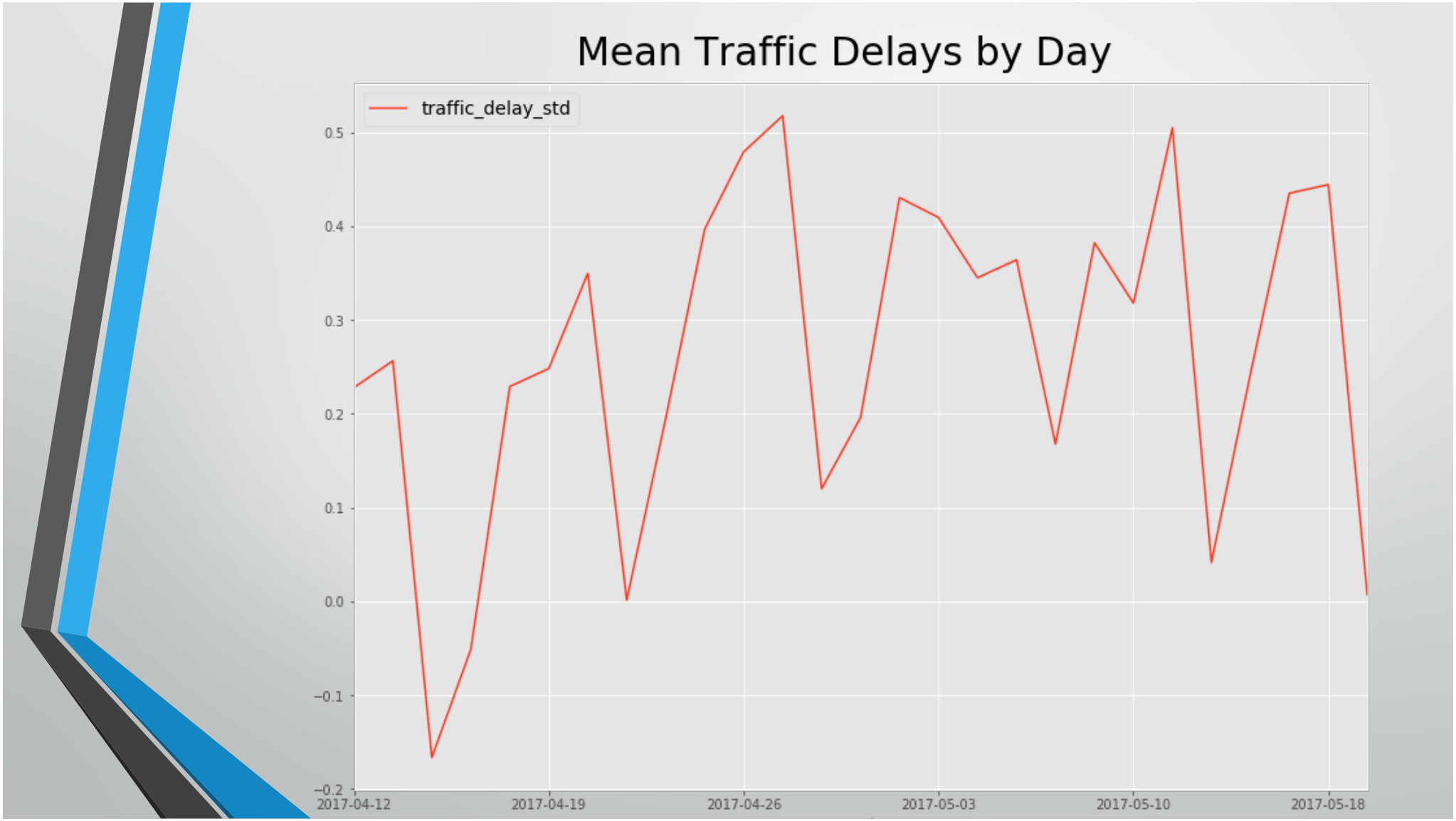

Events Volume per Day
House_AM_hearings
Senate_AM_hearings
ConvCtr_Events
ReaganCtr_Events
20
15
10
5
0
17
24
08
15

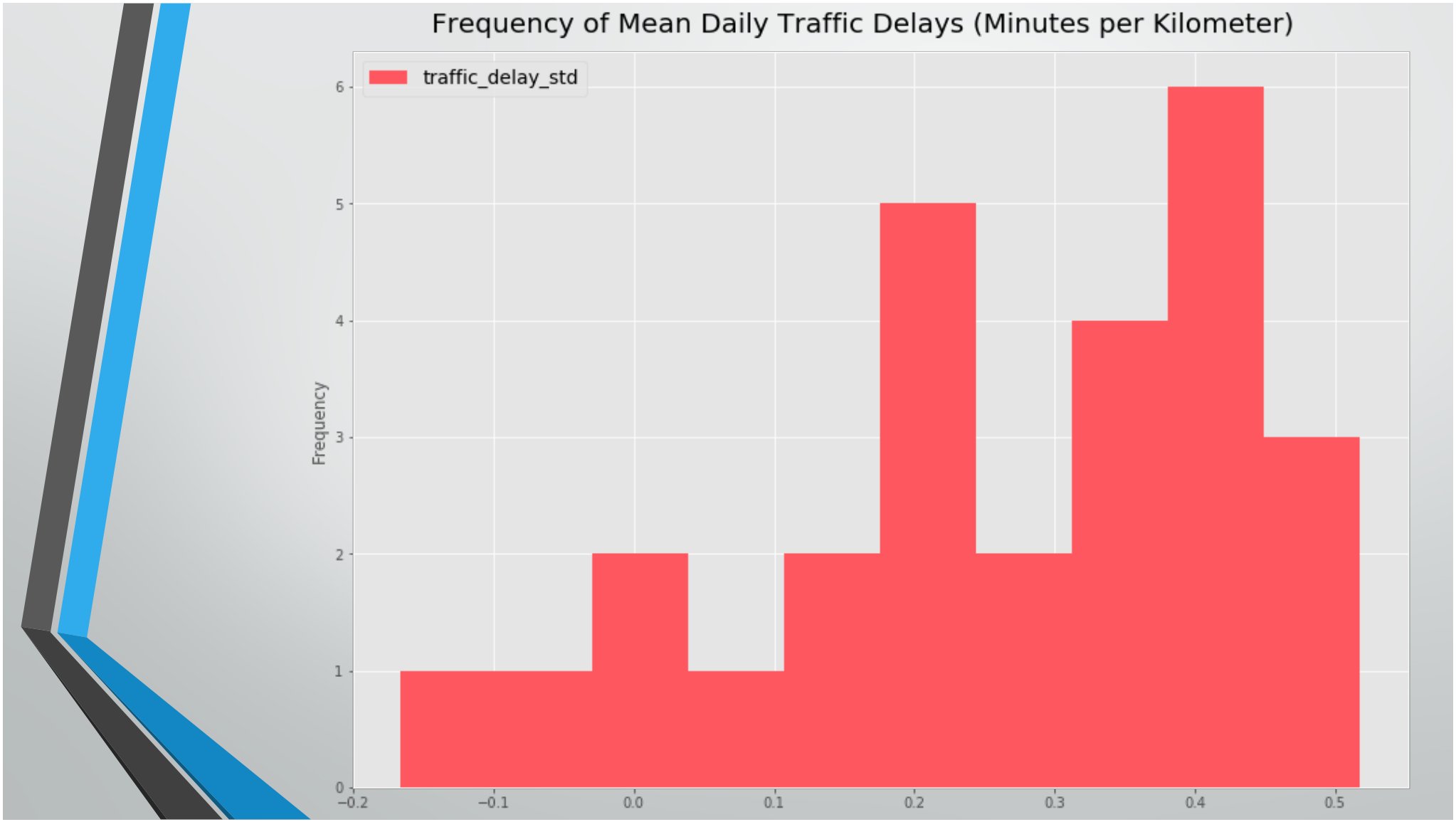

Congress in Session
Senate_AM_session
House_AM_session
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
5
6
7
8
9
12
13
14
15
16
19
20
21
22
23
26
27
28
29
30
33
34
35
36
37

Traffic Data

- Google Maps Distance Matrix API
- 24 routes
  - Within DC
  - Inbound direction
- 7 a.m. – 10 a.m., every 10 minutes
- 27 weekdays


Relative Traffic Levels per Route
Benning
Bladensburg
Canal
ConnAve
Const
EastCap
Fourteenth
GeorgiaAve
I295
I395
I695
Indep
Kst
MDAve
MassAve
NH
NYAve
NorthCap
PennSE
RIAve
SCap
Sixteenth
Whitehurst
WiscAve
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
2017-04-12
2017-04-19
2017-04-26
2017-05-03
2017-05-10
2017-05-18


Mean Traffic Delays by Day
traffic_delay_std
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
2017-04-12
2017-04-19
2017-04-26
2017-05-03
2017-05-10
2017-05-18


Frequency of Mean Daily Traffic Delays (Minutes per Kilometer)
traffic_delay_std
Frequency
0
1
2
3
4
5
6
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
0.5

Predictive Models

Predictions:

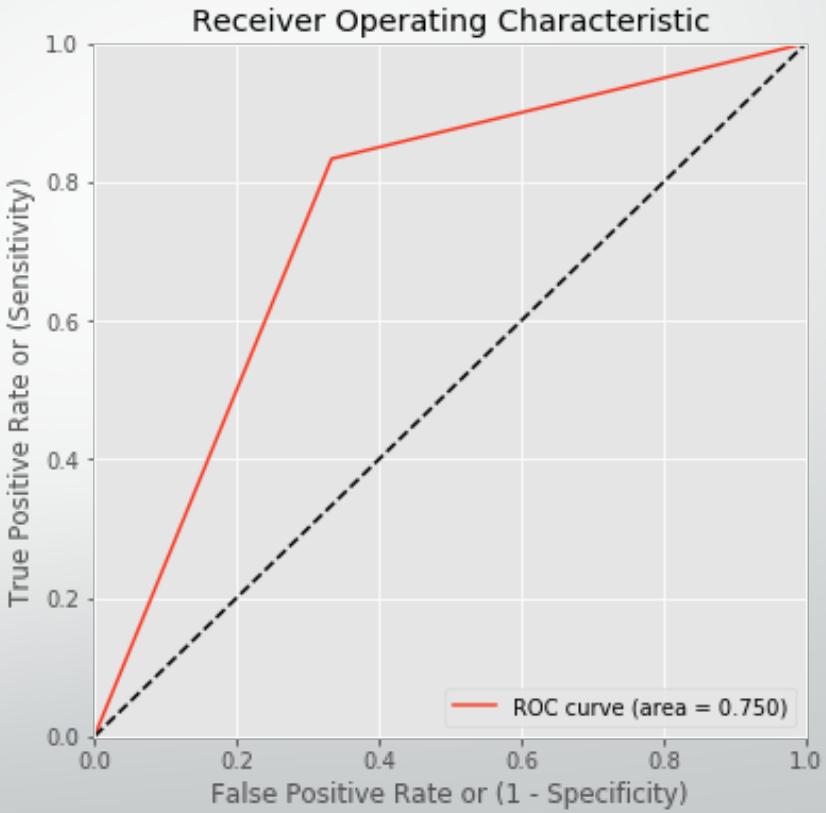
- Will traffic be above or below the median?
- Will traffic be very bad?
- Will traffic be less bad?

Model Evaluation

- Precision score is highest priority
- Consider the consequences of a false positive for “less bad” traffic

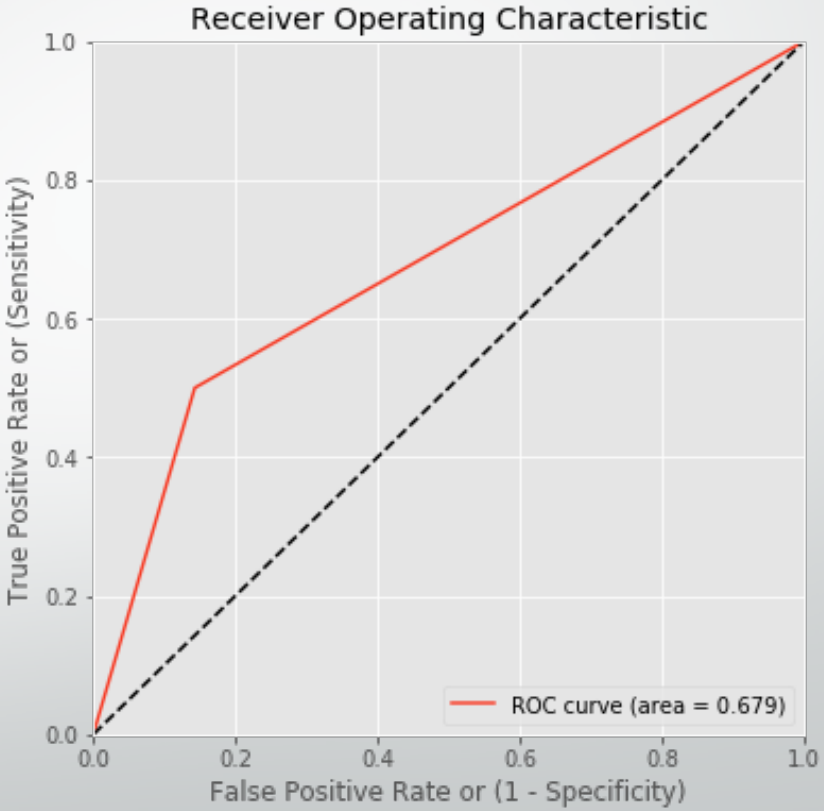
Will Traffic Be Above or Below the Median?

- Bernoulli Naïve Bayes
- 1 = above; 0 = below
- Scores:
  - Precision for 1: 0.67
  - Precision for 0: 0.83
  - ROC-AUC: 0.75
  - Accuracy: 0.778



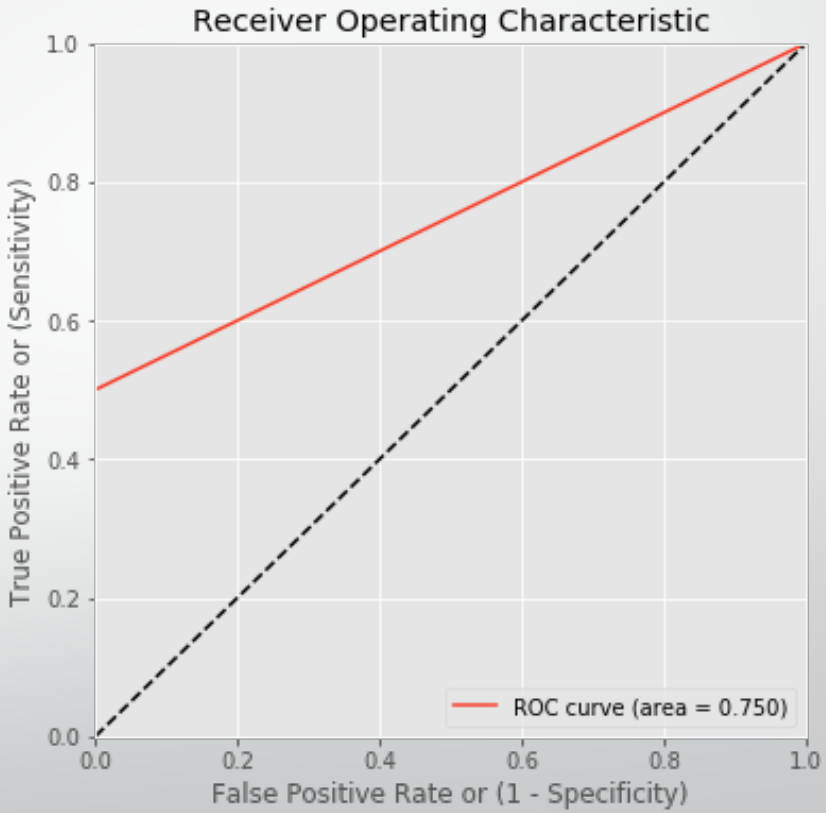
Will Traffic Be Very Bad?

- Logistic Regression
- 1 = yes; 0 = no
- Scores:
  - Precision for 1: 0.86
  - Precision for 0: 0.50
  - ROC-AUC: 0.679
  - Accuracy: 0.778



Will Traffic Be Less Bad?

- Logistic Regression (C=2)
- 1 = yes; 0 = no
- Scores:
  - Precision for 1: 0.88
  - Precision for 0: 1.0*
  - ROC-AUC: 0.75
  - Accuracy: 0.889

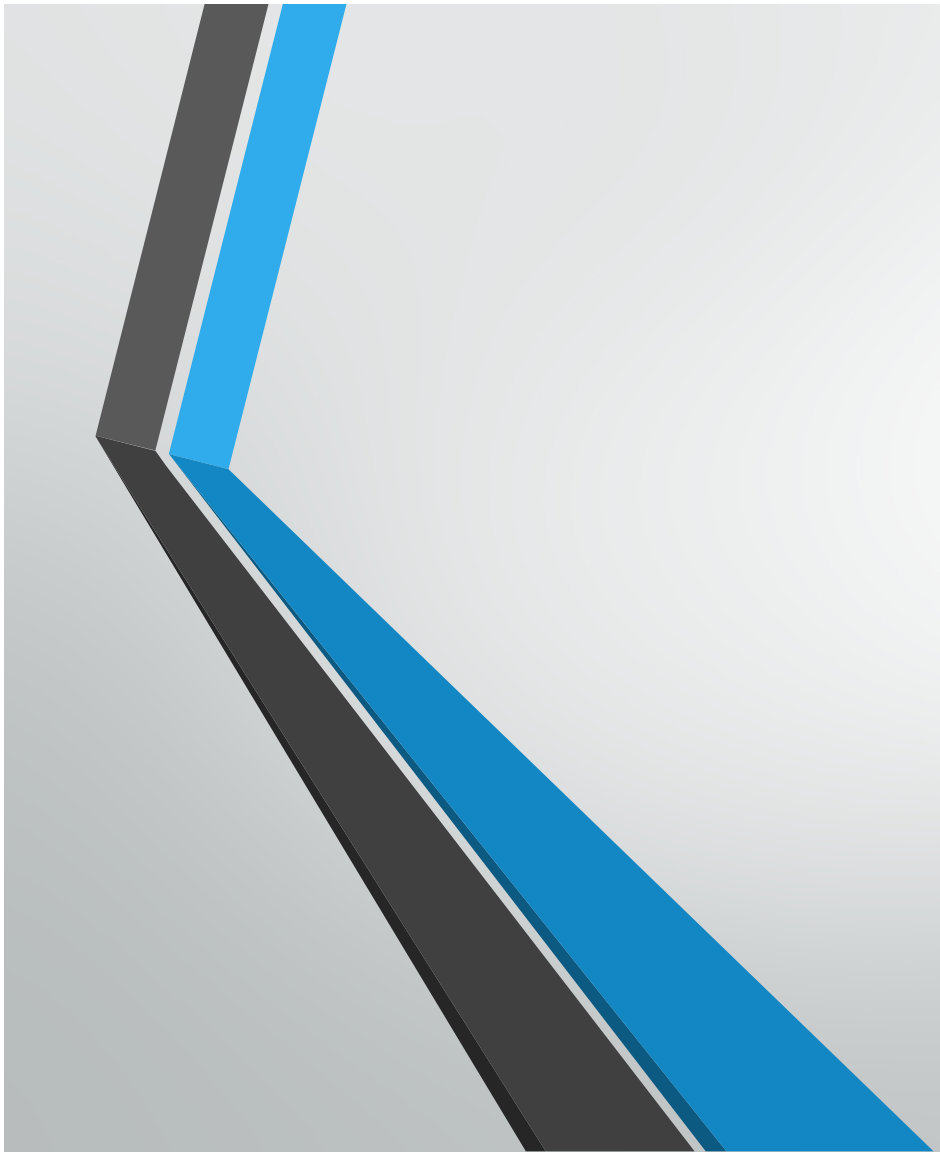


Caveats

- Small sample size (n=27)
- Quirks in Google data

Project Roadmap

- Gather more data!!
- Deploy as a web app
  - Traffic predictions
  - Dashboard of upcoming events
- Re-train and re-tune models
- Try regression with more data



Thank you!

michael.j.sanders@gmail.com

michaeljsanders.com