Predicting Crime in Chicago

Bliss Paik

Crime Statistics

- The city's overall violent crime rate is higher than U.S. average.
- Chicago was responsible for nearly half of 2016's increase in homicides in the U.S.

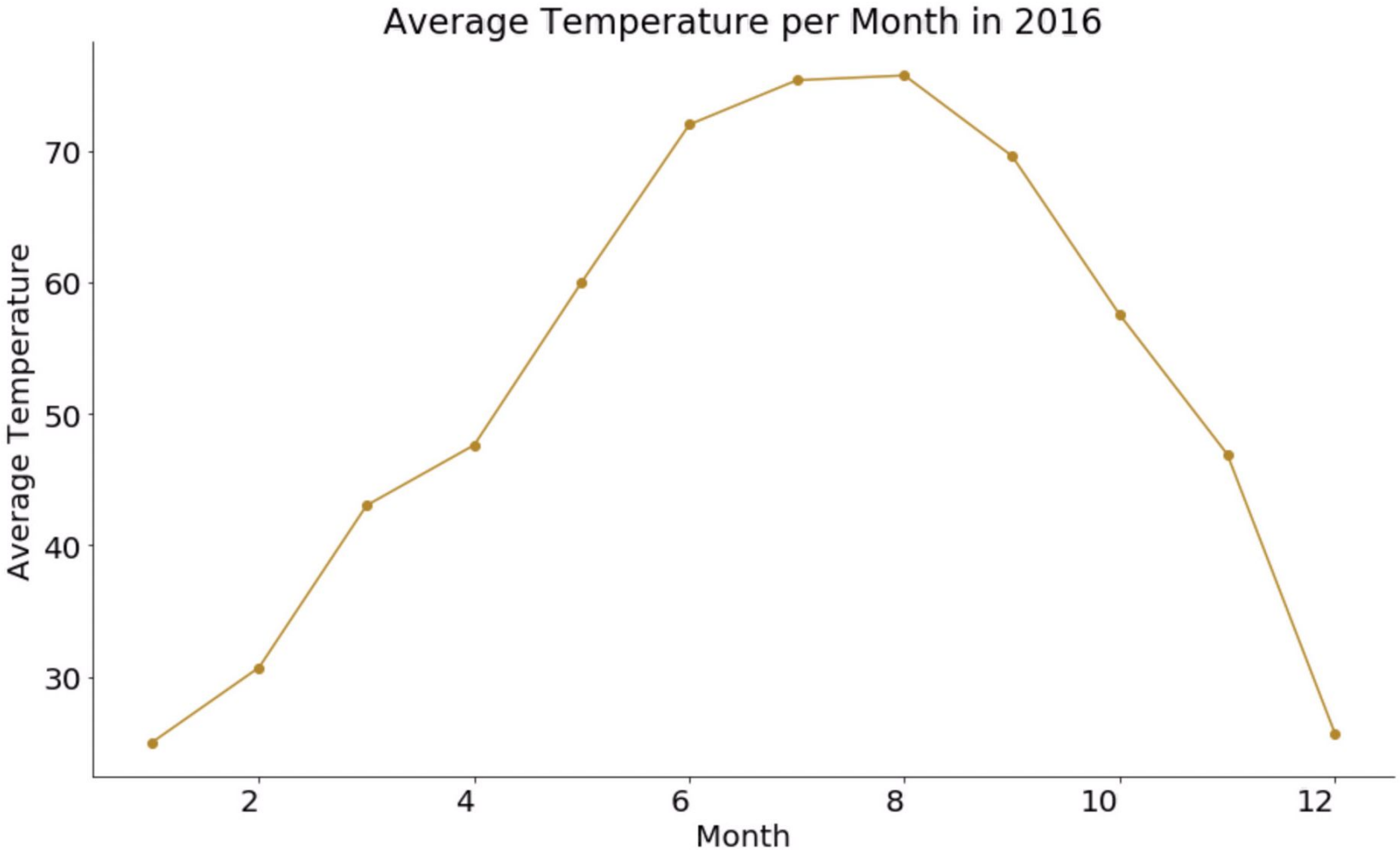
https://home.chicagopolice.org/wp-content/uploads/2018/10/1_PDFsam_CompStat-Public-2018-Week-41.pdf
https://en.wikipedia.org/wiki/Crime_in_Chicago#cite_note-11

As weather changes, what happens to the rate of violent crimes?

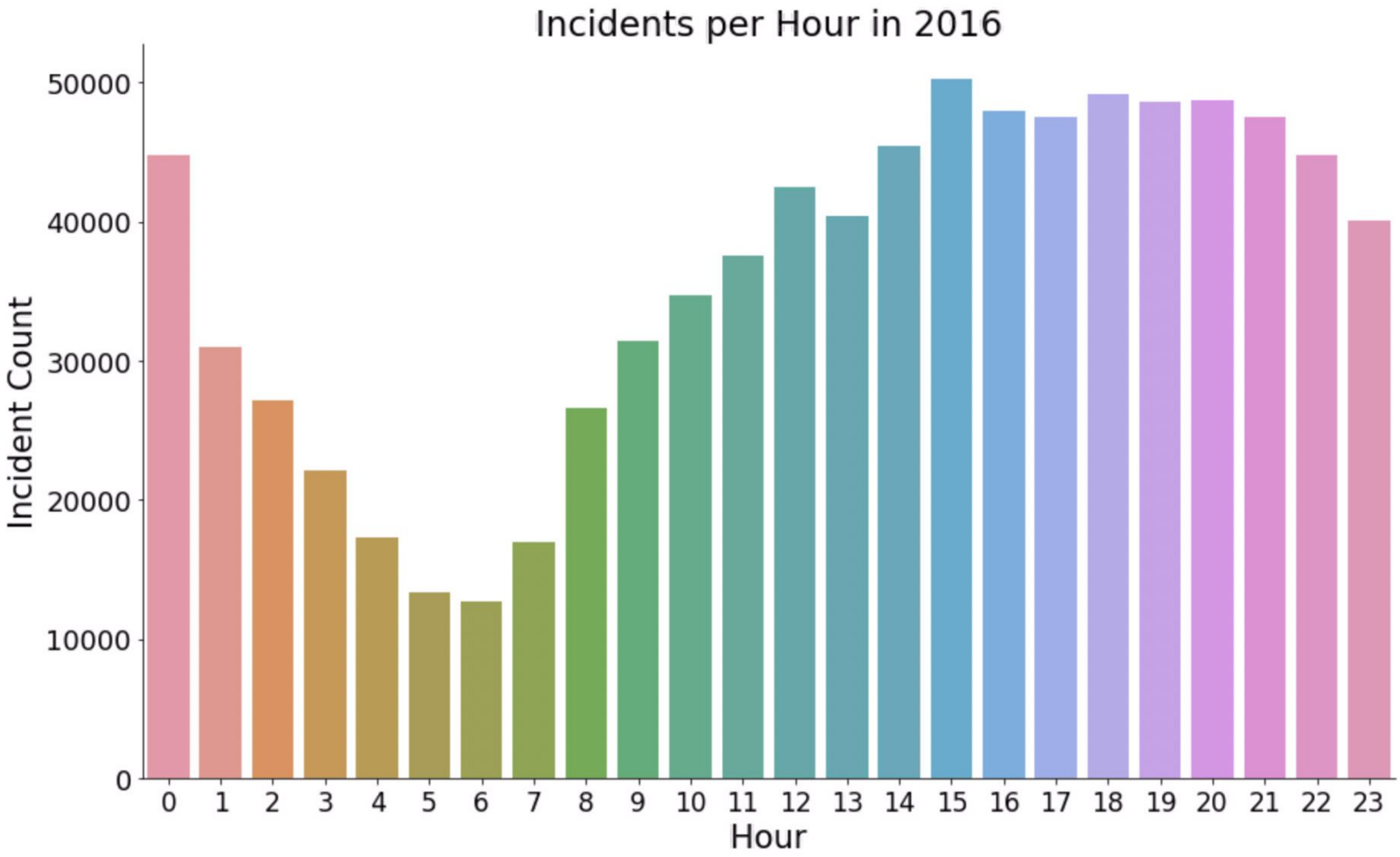
“Hot weather is known to be associated with increases in aggression and violence, as well as a lower general mood.”

https://www.livescience.com/21431-hot-temperatures-mood.html

Which weather features are most indicative for predicting crime?


Average Temperature per Month in 2016
Average Temperature
70
60
50
40
30
2
4
6
8
10
12
Month


Incidents per Month in 2016
Incident Count
0
10000
20000
30000
40000
50000
60000
70000
80000
1
2
3
4
5
6
7
8
9
10
11
12
Month


Incidents per Hour in 2016
Incident Count
50000
40000
30000
20000
10000
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
Hour

Most predictive features

- Hour
- Temperature
- Daylight
- Beat label

Recommendations

- Specific recommendations per beat for each hour
- Budget reallocation

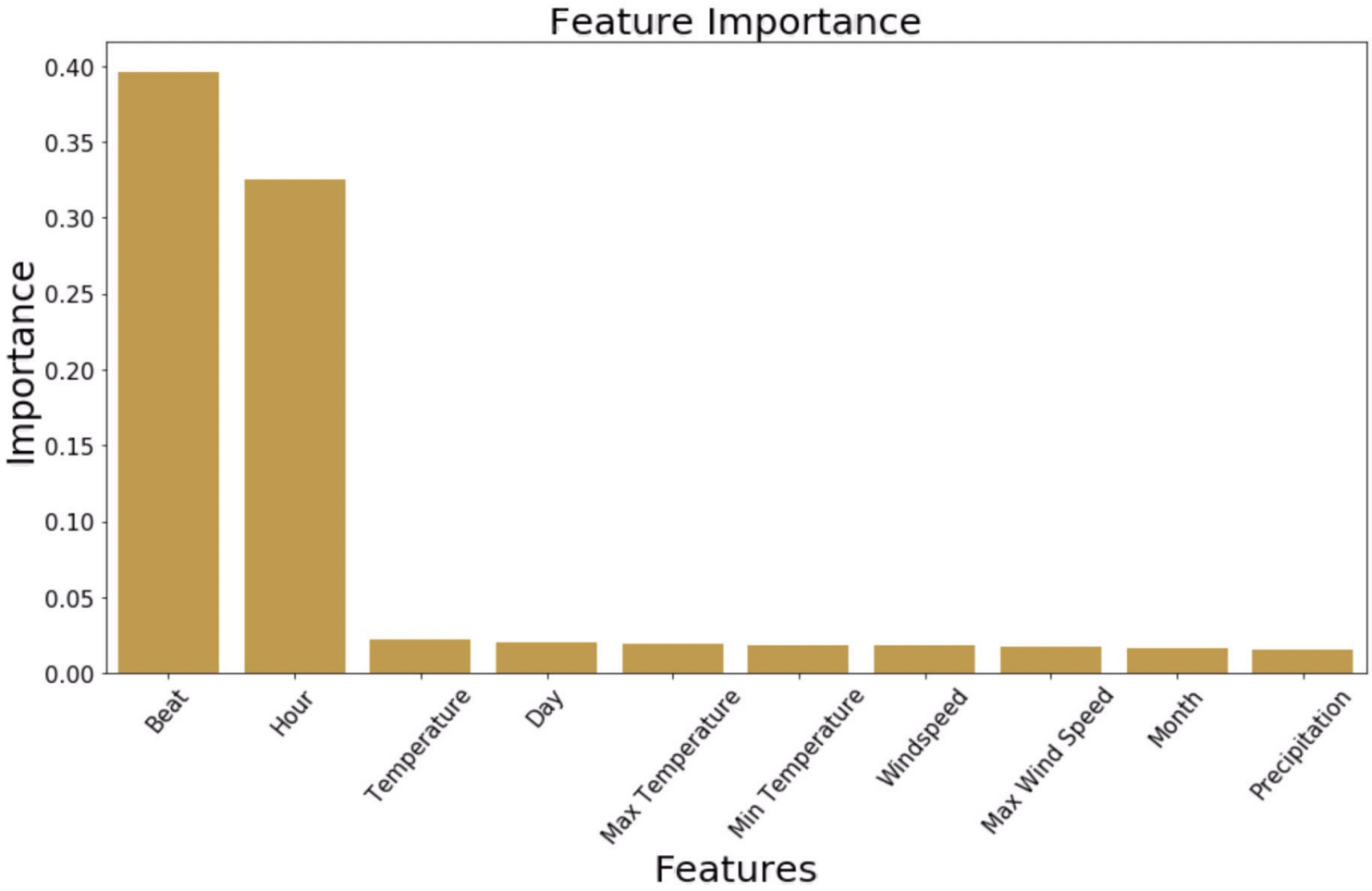
Questions?

Creating Negative Class

- Create hours dataframe for weather data so that it could be lined up with crime data
- Add beat label column to weather data
- Merge weather and crime dataframes
- Add target column for whether a crime occurred or not at that time
- Fill NaN's with 0 to create negative class

Feature Engineering

- Add features daylight and night time from scraped data
- Calculate rolling averages and sums for daylight, precip, temperature, and snow
- Bootstrap using SMOTEEN


Feature Importance
Importance
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Beat
Hour
Temperature
Day
Max Temperature
Min Temperature
Windspeed
Max Wind Speed
Month
Precipitation
Features

	coefficients
hr	0.295964
a_temp	0.096370
prev_7_day_avg_Temp	0.058088
prev_7_day_avg_Daylight	0.037458
sunrise	0.033768
a_wdsp	0.007811
prev_7_day_avg_Precip	0.002329
a_prcp	0.001929
prev_7_day_Rain_drizzle	0.000000
daylight	0.000000

	coefficients
a_max	0.000000
a_year	0.000000
nighttime	-0.000322
da	-0.000621
a_thunder	-0.002332
prev_7_day_Snow	-0.004757
mo	-0.013063
a_rain_drizzle	-0.027936
a_mxpsd	-0.028687
beat_label	-0.228698

Accuracy Scores

- Logistic Regression
  - **0.5866, 0.5449**
- Random Forest
  - **0.9884, 0.4318**
- Neural Network
  - **0.6437, 0.4638**

Classification Report

- Logistic Regression
  - Recall: **54%**
  - Specificity: **58%**
- Random Forest
  - Recall: **64%**
  - Specificity: **42%**
- Neural Network
  - Recall: **64%**
  - Specificity: **46%**

Next Steps

- Make model that can provide likelihood that a violent crime will occur looking into the future.
- Model would perform better by using additional years to build out broader analyses

Ethics

- Police recommendations for geospatial location can come with police bias and racial profiling
- Need to be careful in how recommendations are phrased and implemented to not have any harmful biases