Chicago Violent Crime

George Ferre

TABLE OF CONTENTS

BUSINESS PROBLEM

———

The city of Chicago wants to confront the epidemic of violence. One way to tackle this is to determine the amount of violent crime that is to be expected on a given day, and if possible try to figure out where the crime will occur.

Data and Methodology

DATA

DATA SET
20k victims of violent crime in Chicago from 2016 to 2021

VARIABLES
Date, Fatality, Area, Age, Sex, Race and Nearby Street Outreach

TRANSFORMATION
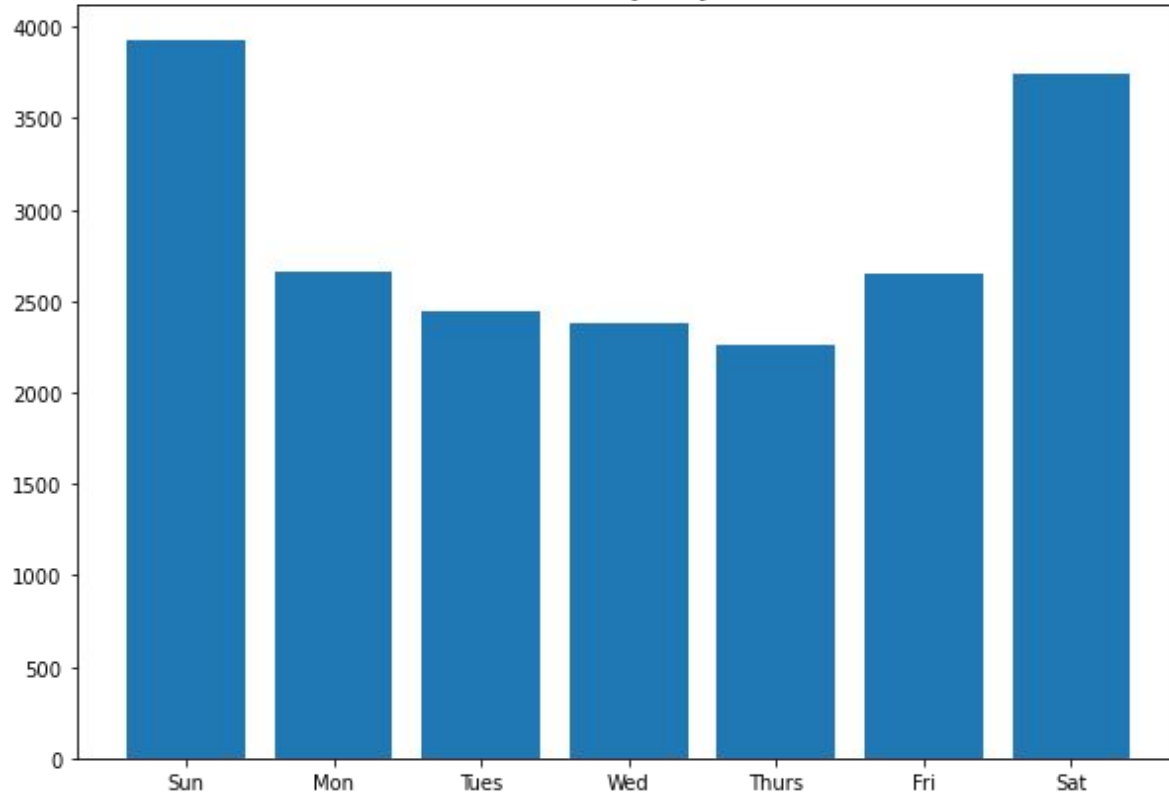Grouped by date to get sum of victims by date
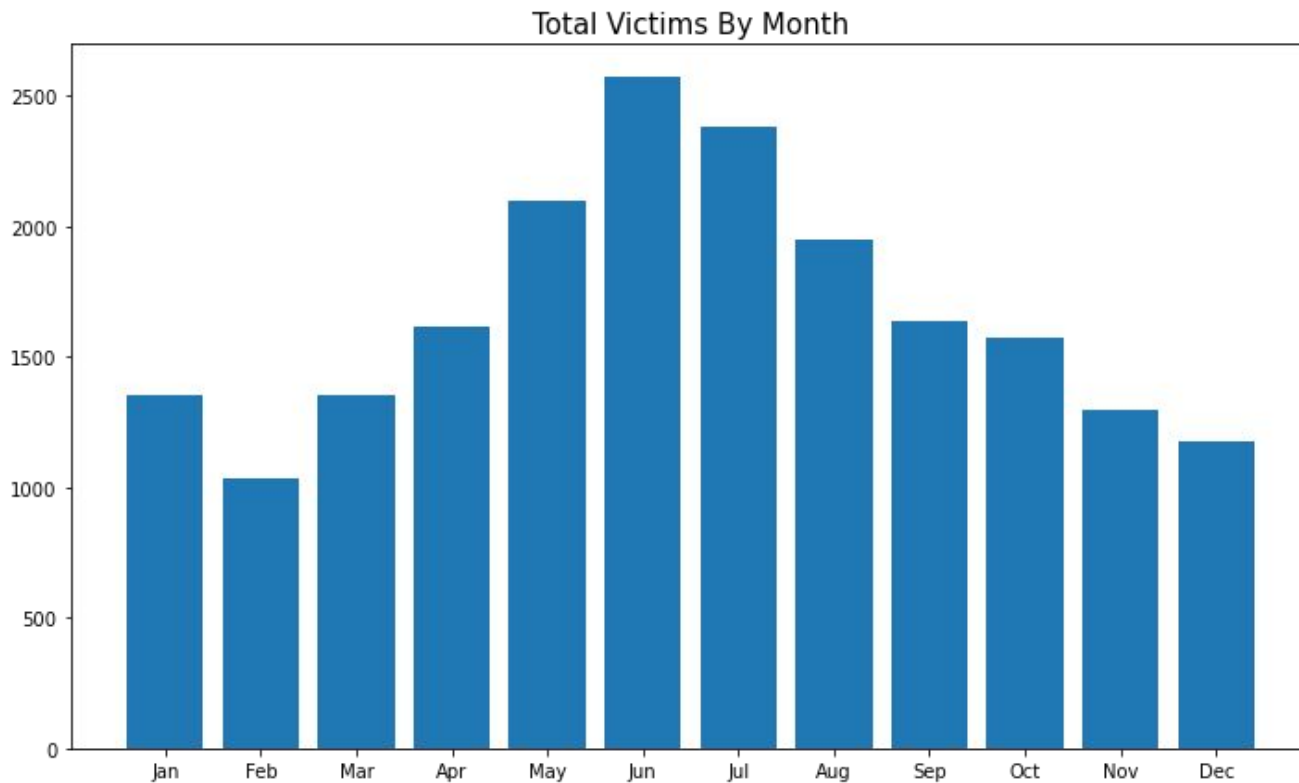
CLEANING
Grouping variables (ex grouping homicides together)
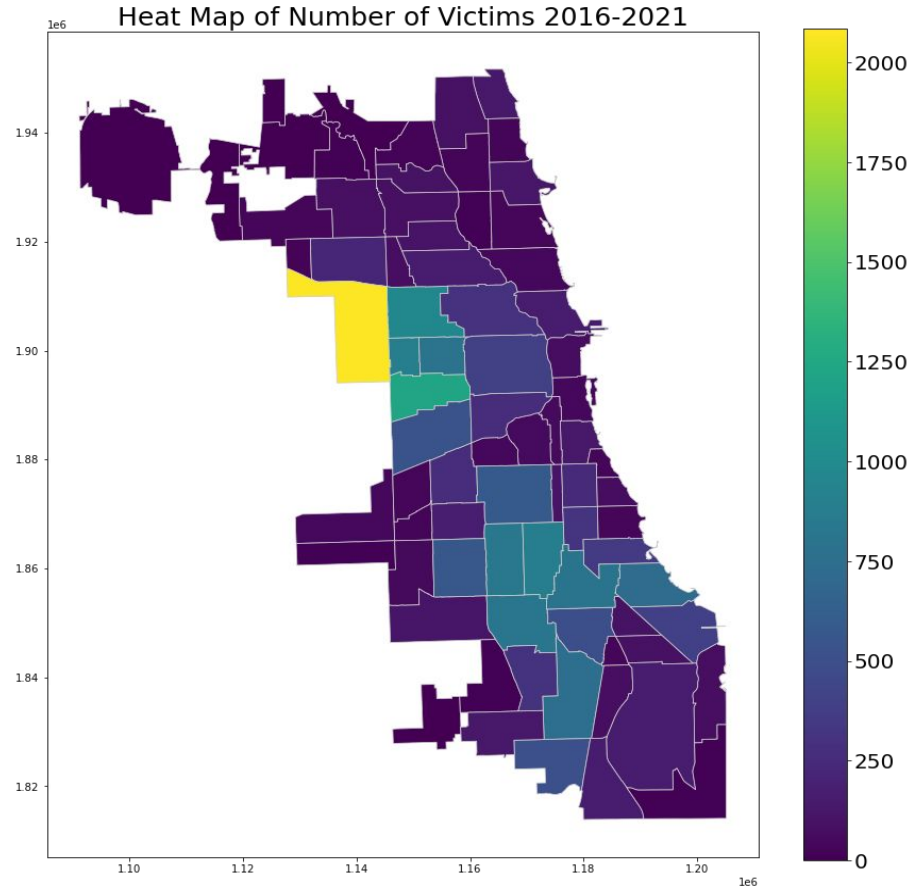
Victims By Day of Week

Total Victims By Day of Week

Violent Crimes by Month

Total Victims By Month

Daily Victims by Area

Heat Map of Number of Victims 2016-2021

Rolling Data

Some features required a rolling count of victims over a period of time (ex total victims in last 3 days)

Single Category Data

Other features needed to be discrete category (ex: day of the week)

RESULTS

Final models ended up being 2 features:
- Total recent victims
- Day of week

Models

Linear Regression

CV Score:
.75

Support Vector

Test Score:
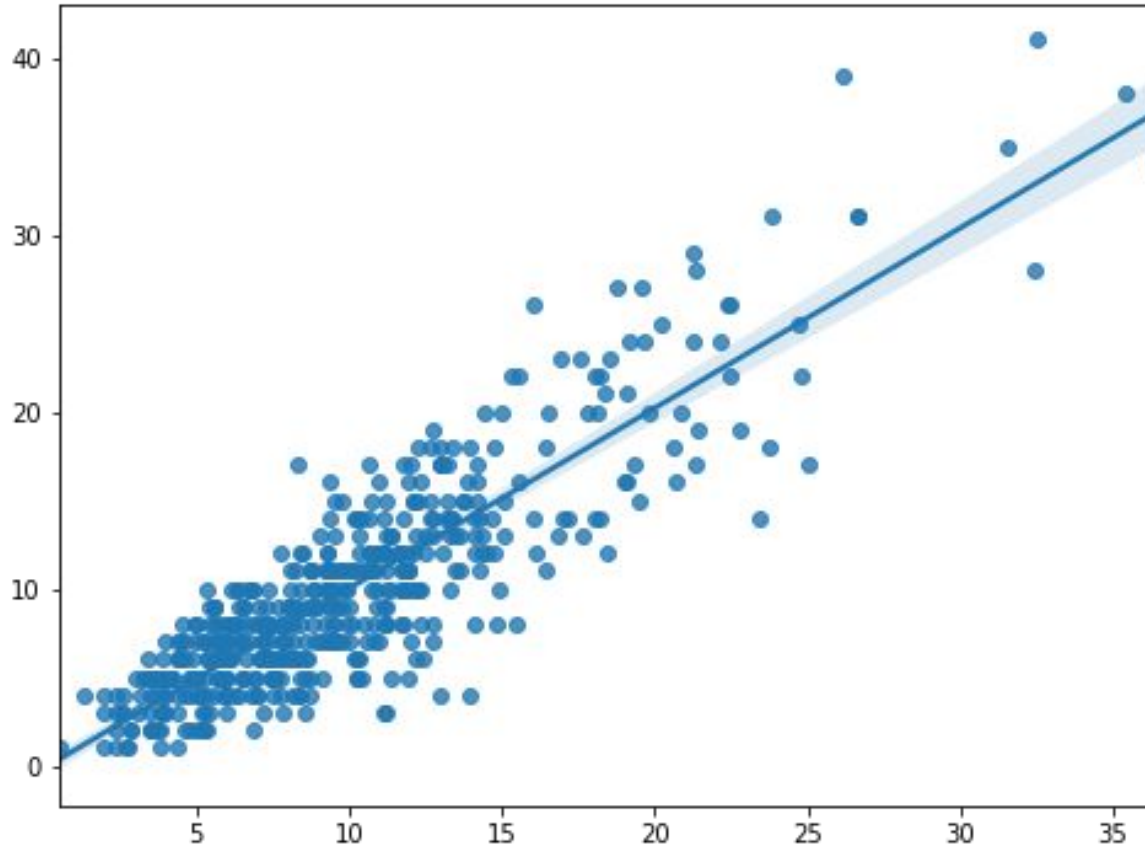.75

SGD Regression

Test Score:
.75

Conclusion

So far less has been more

- Started rolling sum at 7 days, ended at 2
- Sum of victims and day of week are most important by far
- Month does not seem to be a good predictor, perhaps temperature is better?

PREDICTIONS AND RESULTS

Predictions vs Test Results

Next Steps

Continue Optimization	Investigate Time Series	Build Dashboard
Look at Temperature data	Search for New Data	Include Hospital information

THANK YOU

Do you have any question?

georgeaferre@gmail.com
linkedin.com/georgeferre
github.com/georgeferre