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

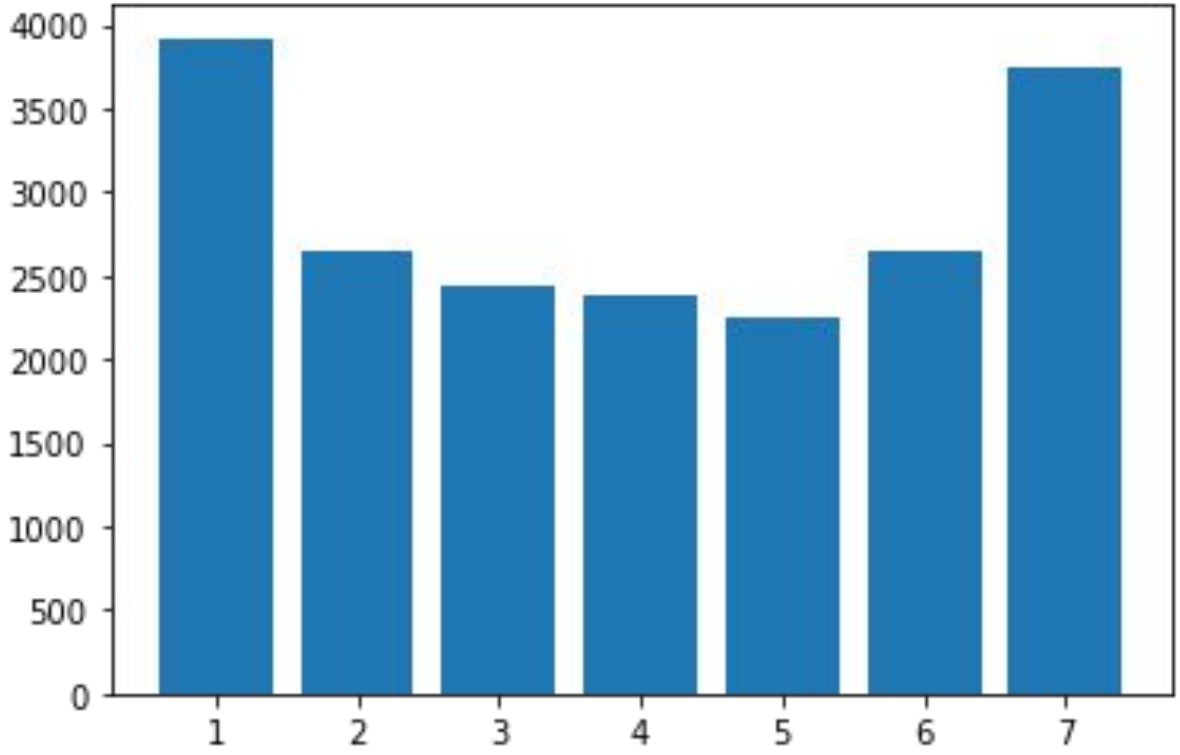
TRANSFORMATION

Grouped by date to get sum of victims by date

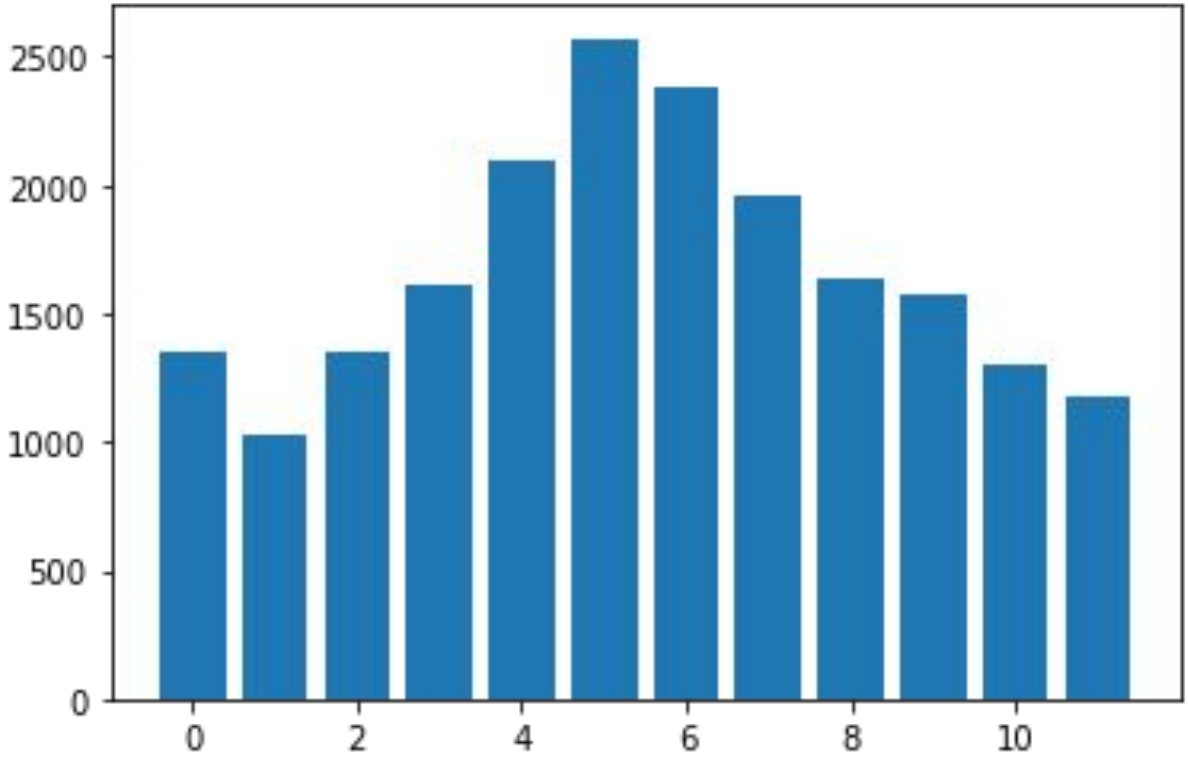
CLEANING

Removed data during 2020 onward due to COVID-19

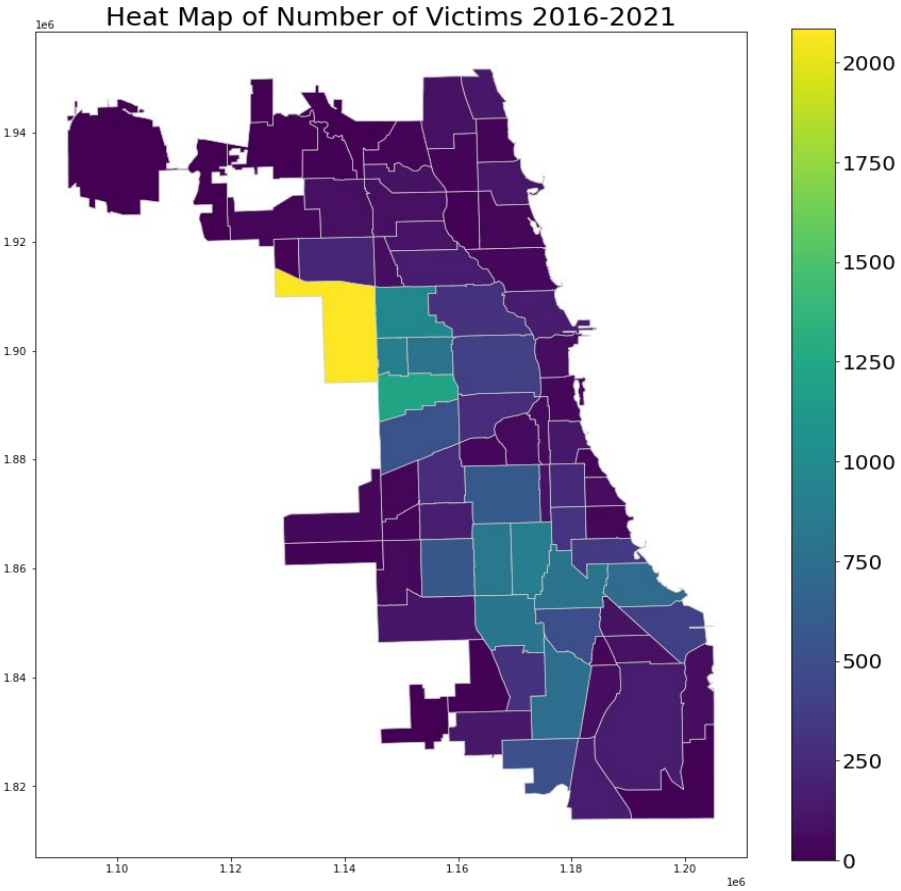
Victims By Day of Week



Violent Crimes by Month



Daily Victims by Area



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

Regression models tuned to best predict number of victims

Models

Linear Regression

Score: 69%
Finished Tuning

Random Forest Regression

Score: 55%
Starting Tuning

SGD Regression

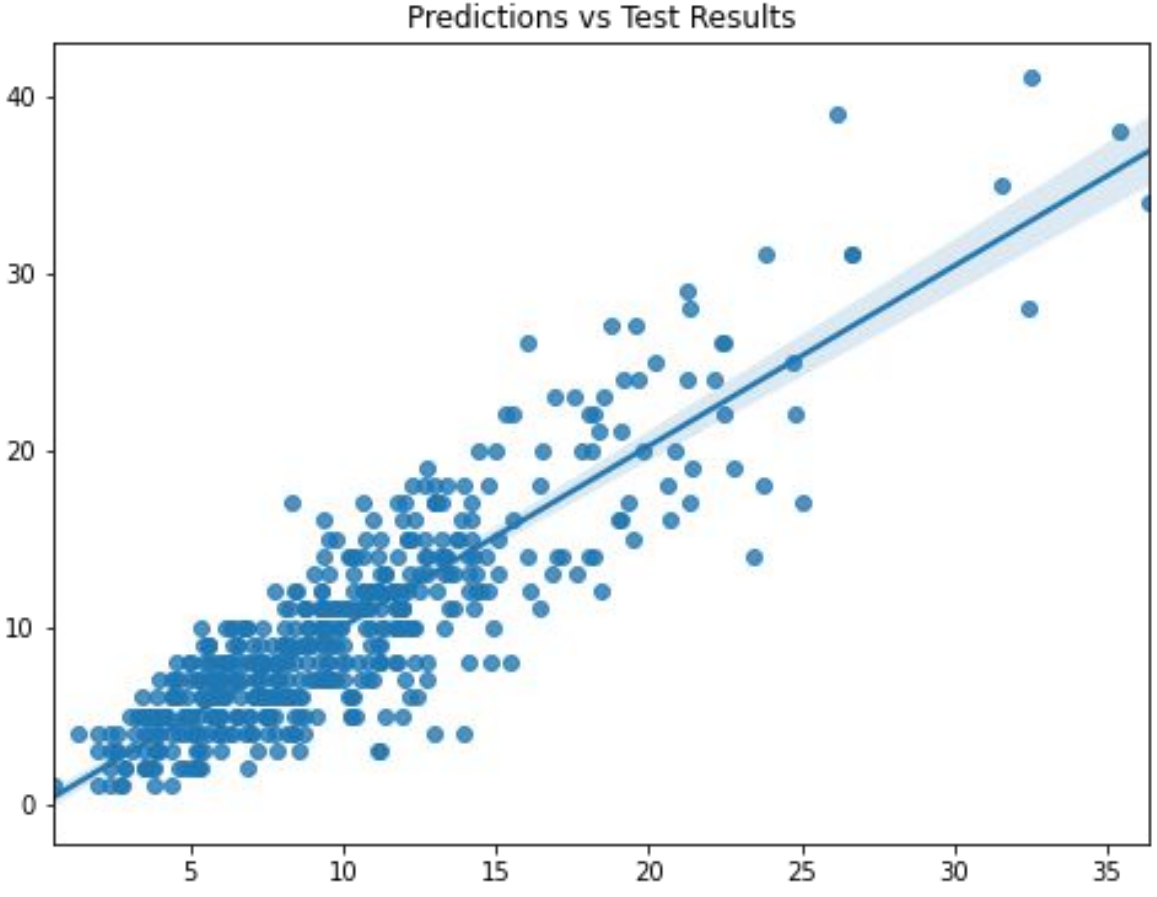
Score: TBH
Coming up

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

Predictions and 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