



# Predicting Car Accidents in Chicago

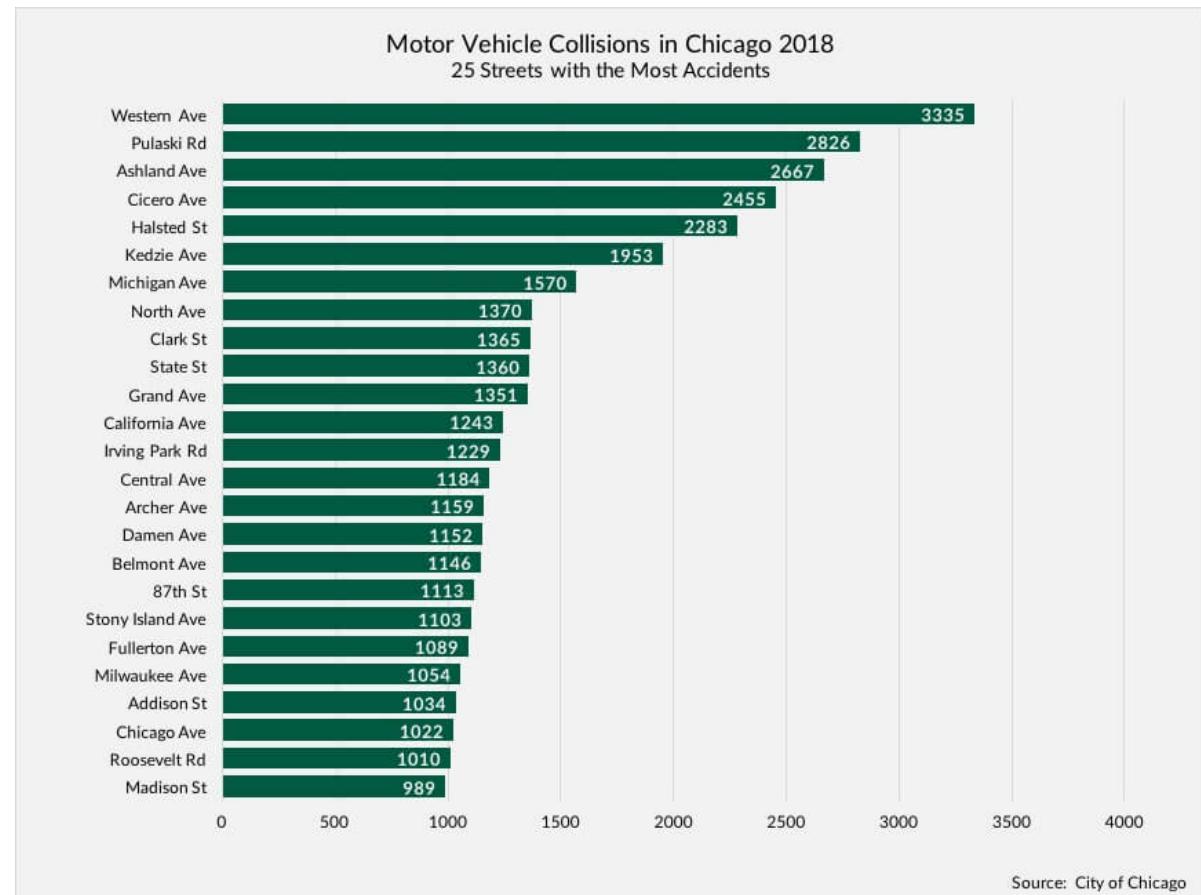
Duc Vu

Motivation

Methodology

Takeaways

- **281,788** crashes involving motor vehicles
- **772** car wrecks every day
- **2** people killed every day in vehicle accidents.
- **9** people injured every hour in a car wreck.



# 1. Motivation

- Can machine learning help us understand the causes and the factors that affect car crash?
- Can machine learning help save lives?

# Where to Collect Data?



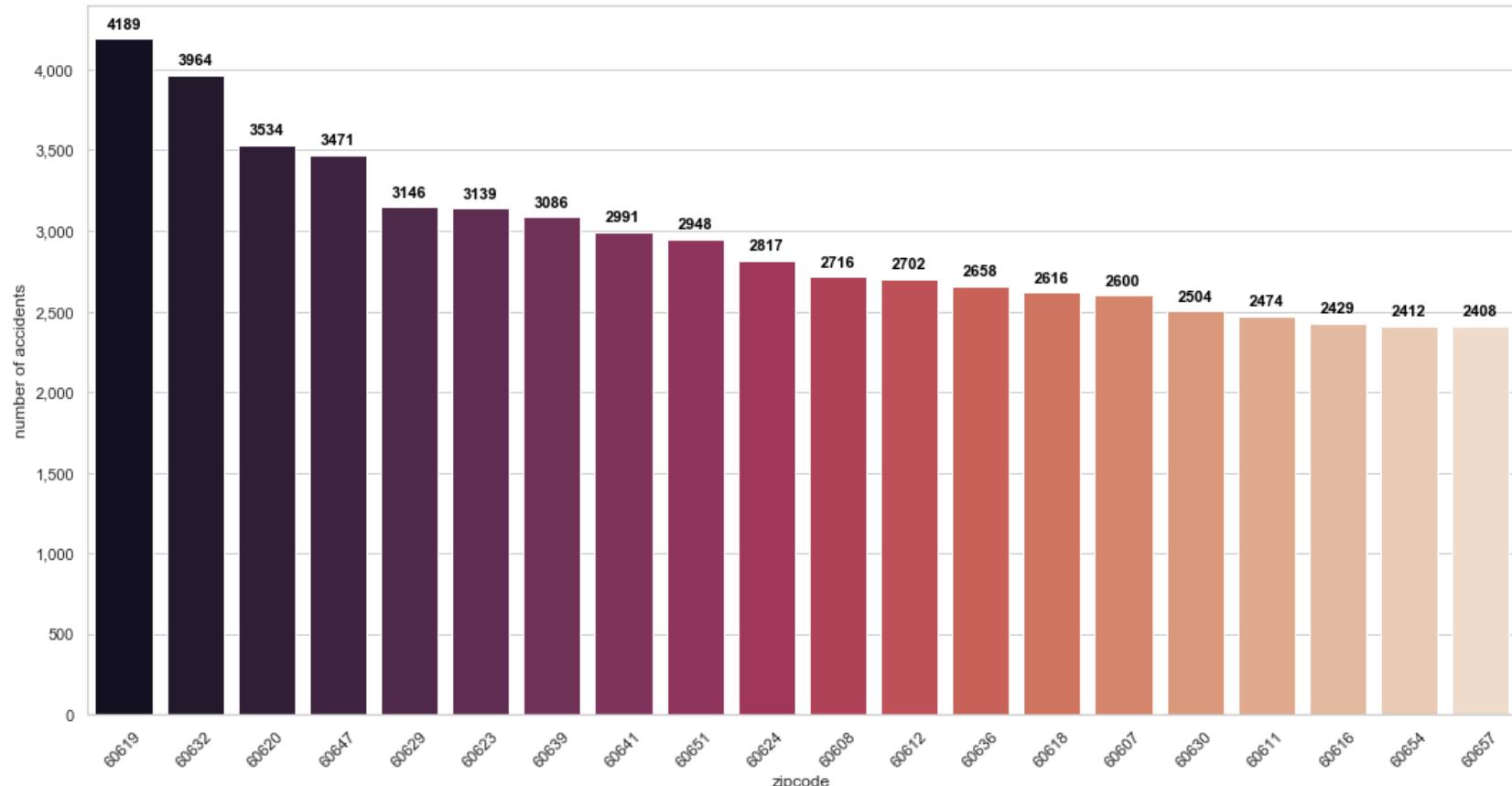
HARVARD  
Dataverse

<https://dataverse.harvard.edu>

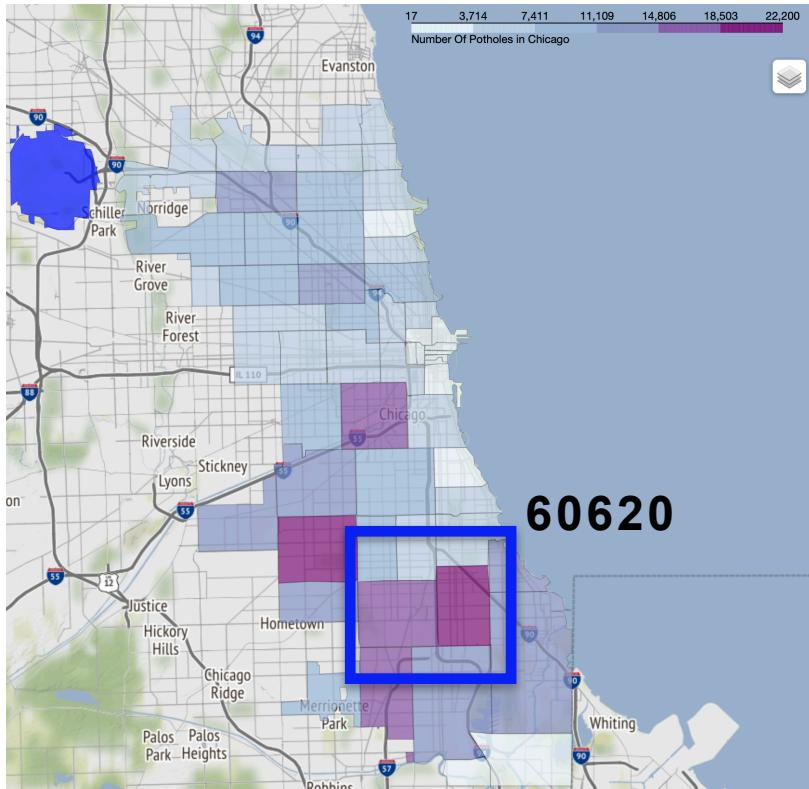


<https://data.cityofchicago.org>

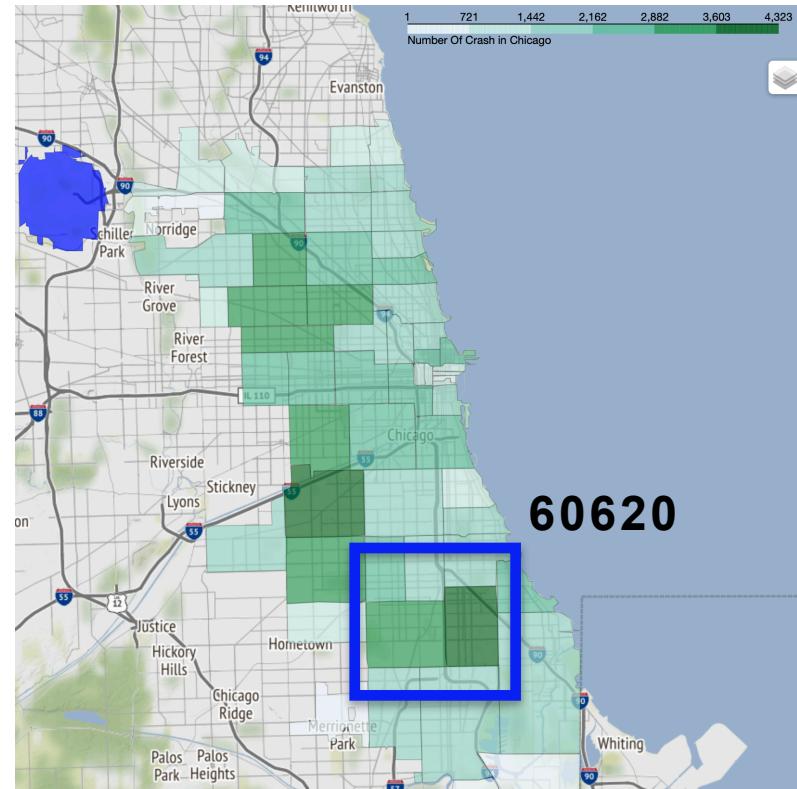
# The Top 20 Zip Codes With the Highest Rate of Car Accidents



# The Zip Code 60620 Has One of the Highest Accident Densities

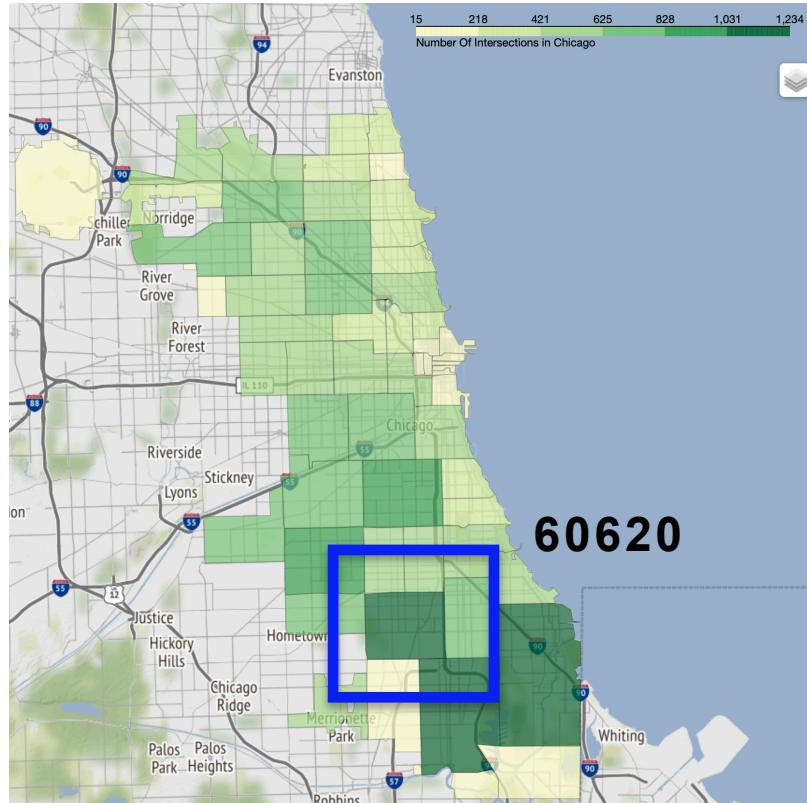


Intersection Density

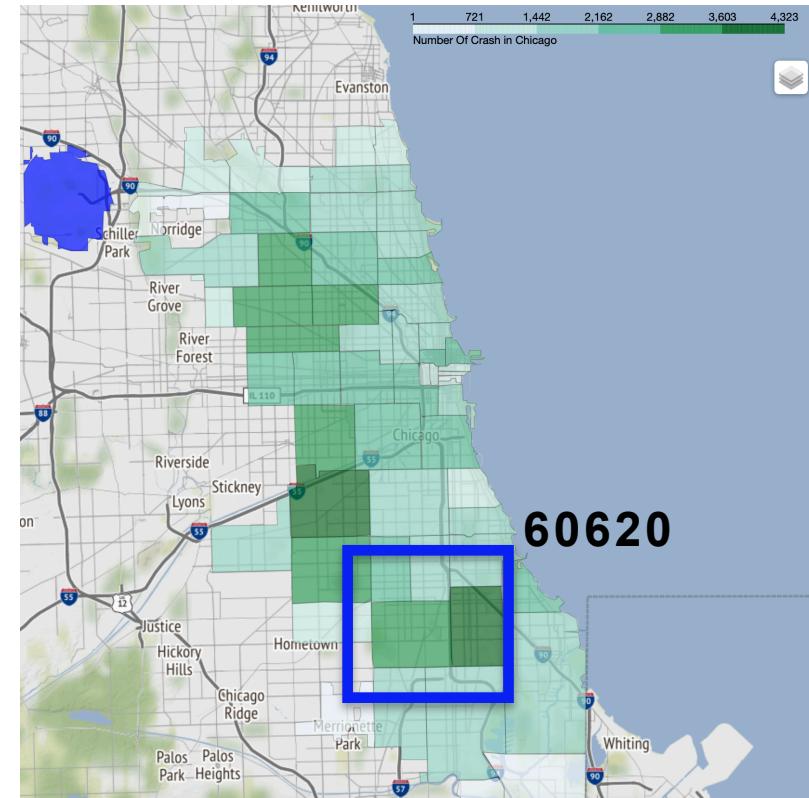


Accident Density

# Possibility Pothole Hell in 60620

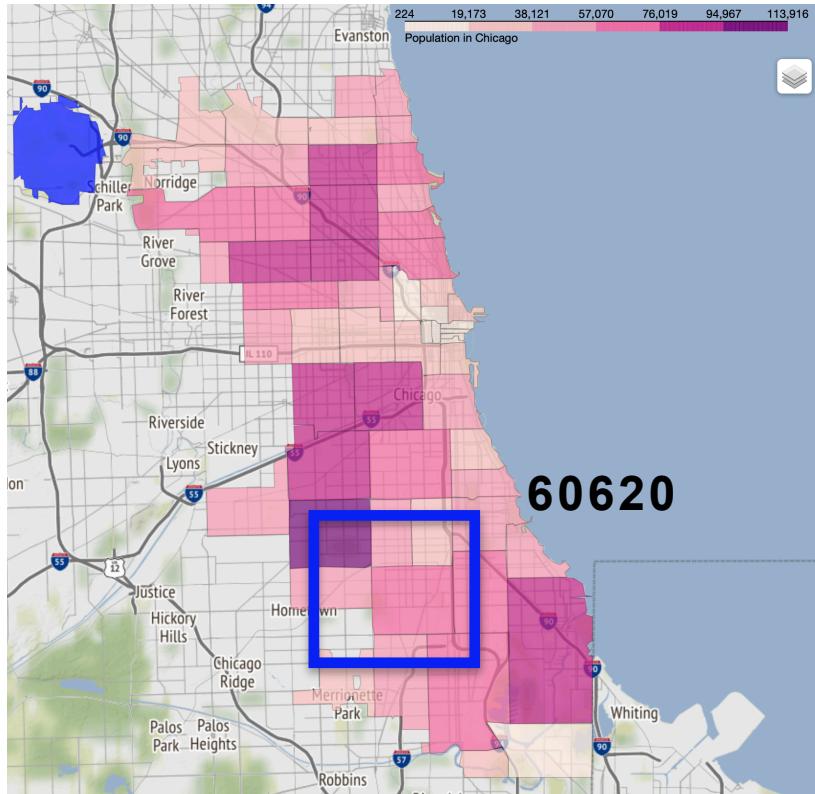


**Pothole Density**

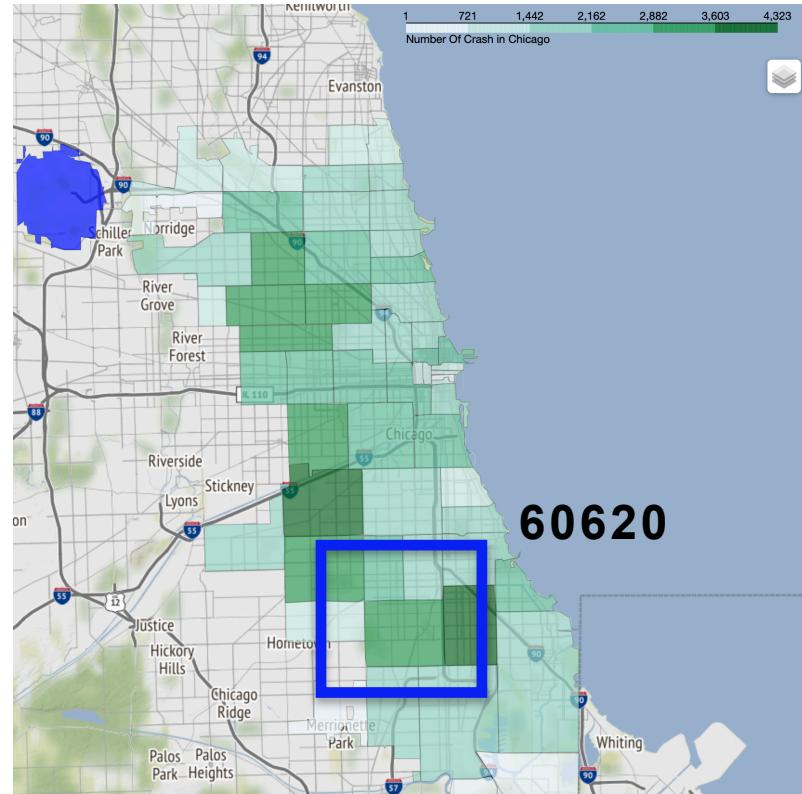


**Accident Density**

# Populated area

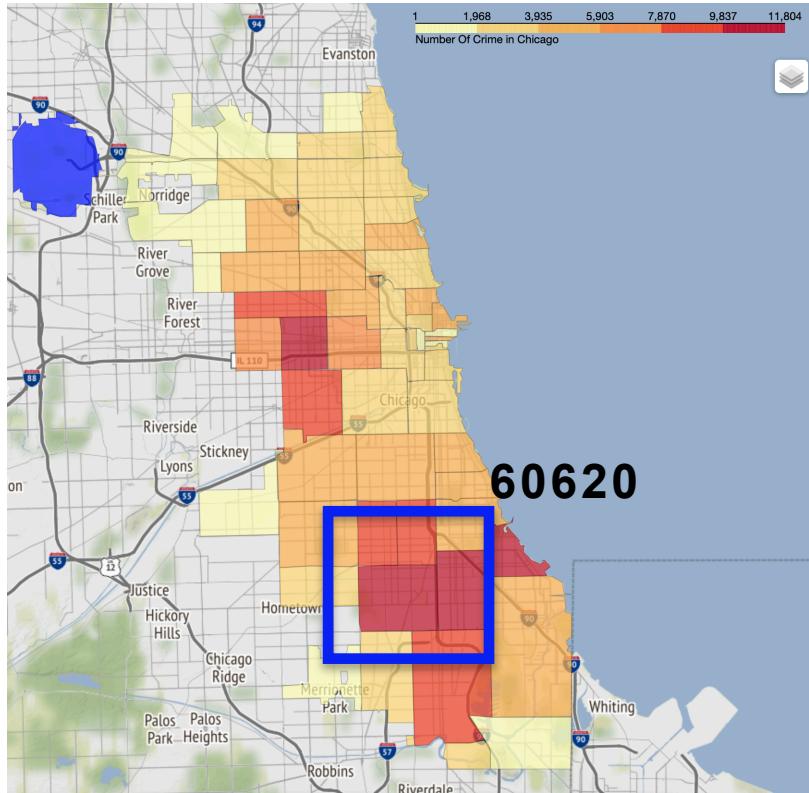


Population Density

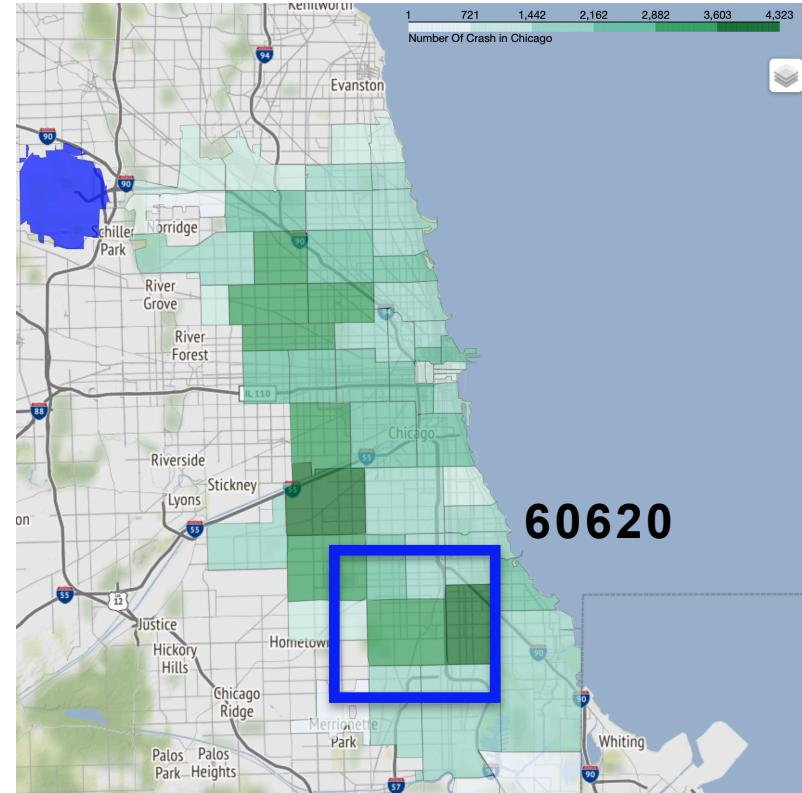


Accident Density

# “Chiraq” in 60620



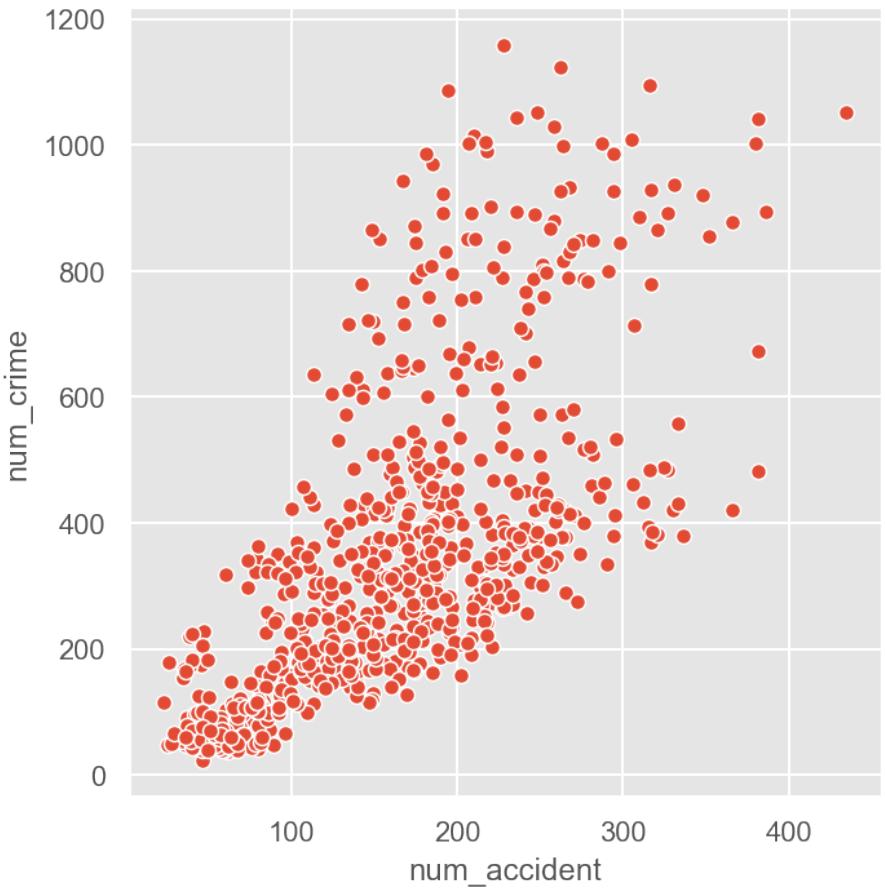
Crime Density



Accident Density

## 2. EDA

**Crime rate is highly correlated  
to car accident!!!**



# 3. Regression

- Removed collinear features and trained linear regression model
- No overfitting!!!
- Engineered features to improve R<sup>2</sup> score

## 4. Scoring



**R<sup>2</sup> score : 0.735**

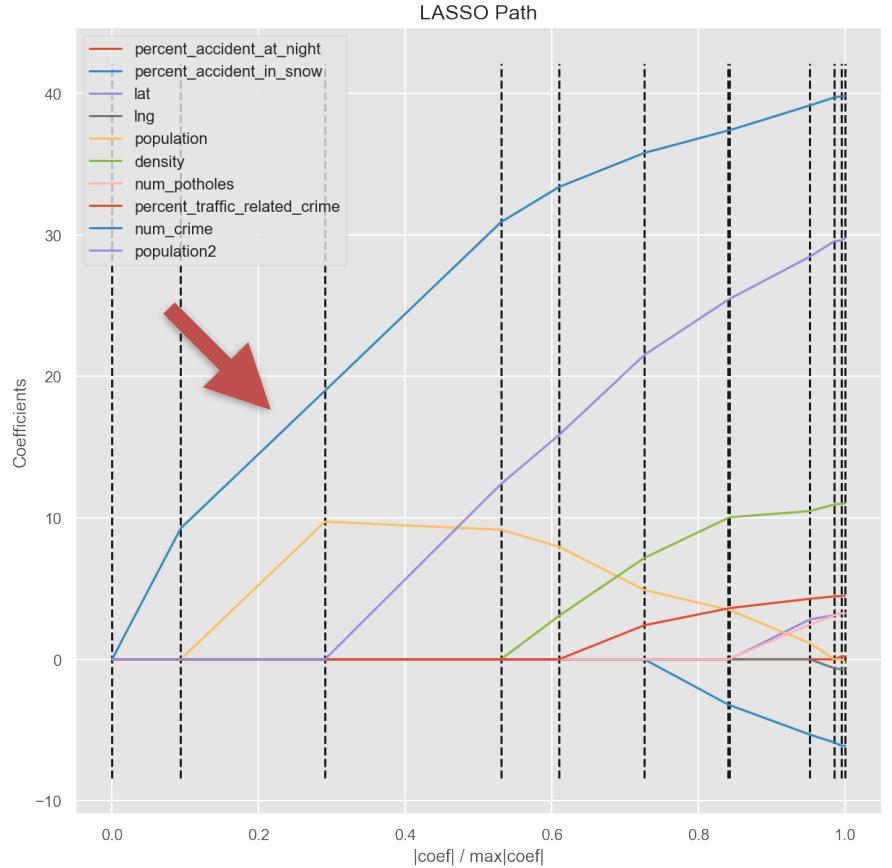
73.5% all observance data points around its mean, R-squared values in conjunction with residual plots



**RMSE : 34.389 crimes**

The observed number of accidents are 34 difference from my prediction model

# Significant Feature



- **Coefficient of num\_crime**  
0.1352
- **Every 8 crimes will contribute to  
1 crash accident in each zip code**

## 5. Further Research & Takeaways

- Adding ride-sharing, median income features to uncover the crash pattern in metropolitan area
- Reduce risk of crashes through enforcement.
- Where to place additional signage (e.g. warn for curves, speed limit)

# Questions

