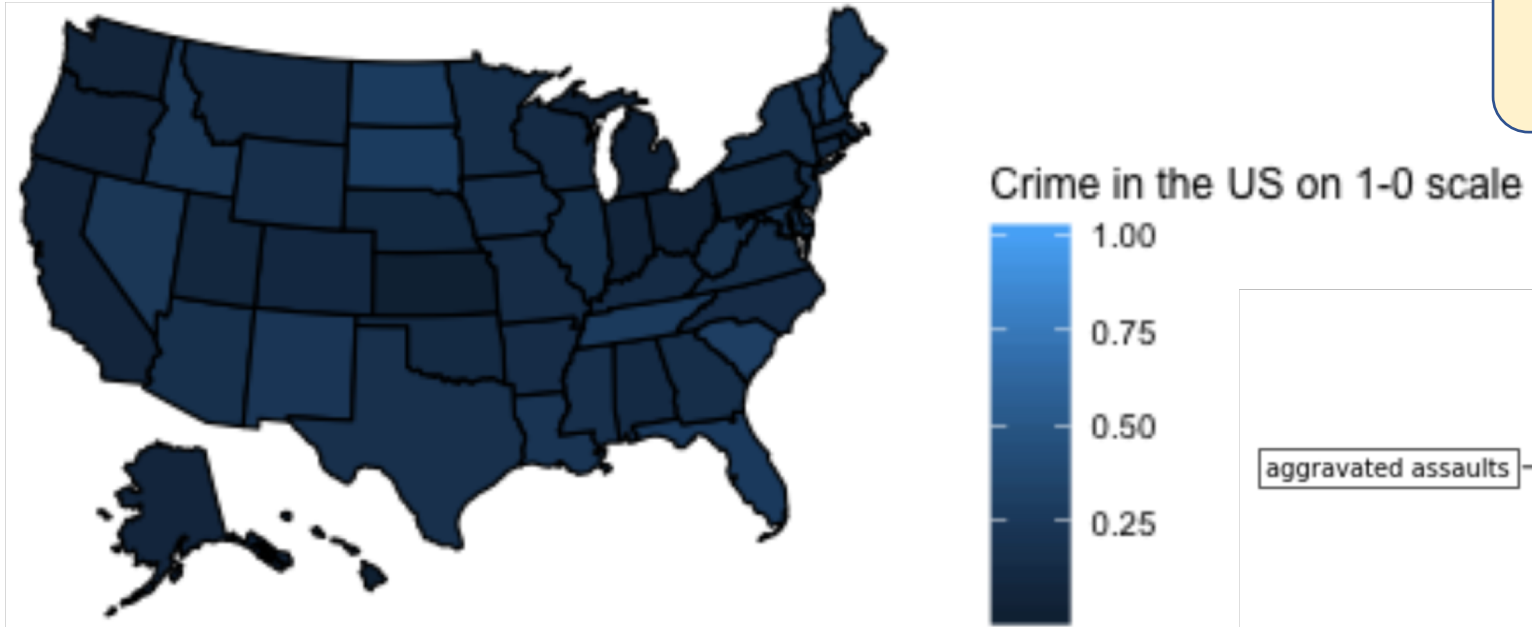


Predictive analysis of crime rate

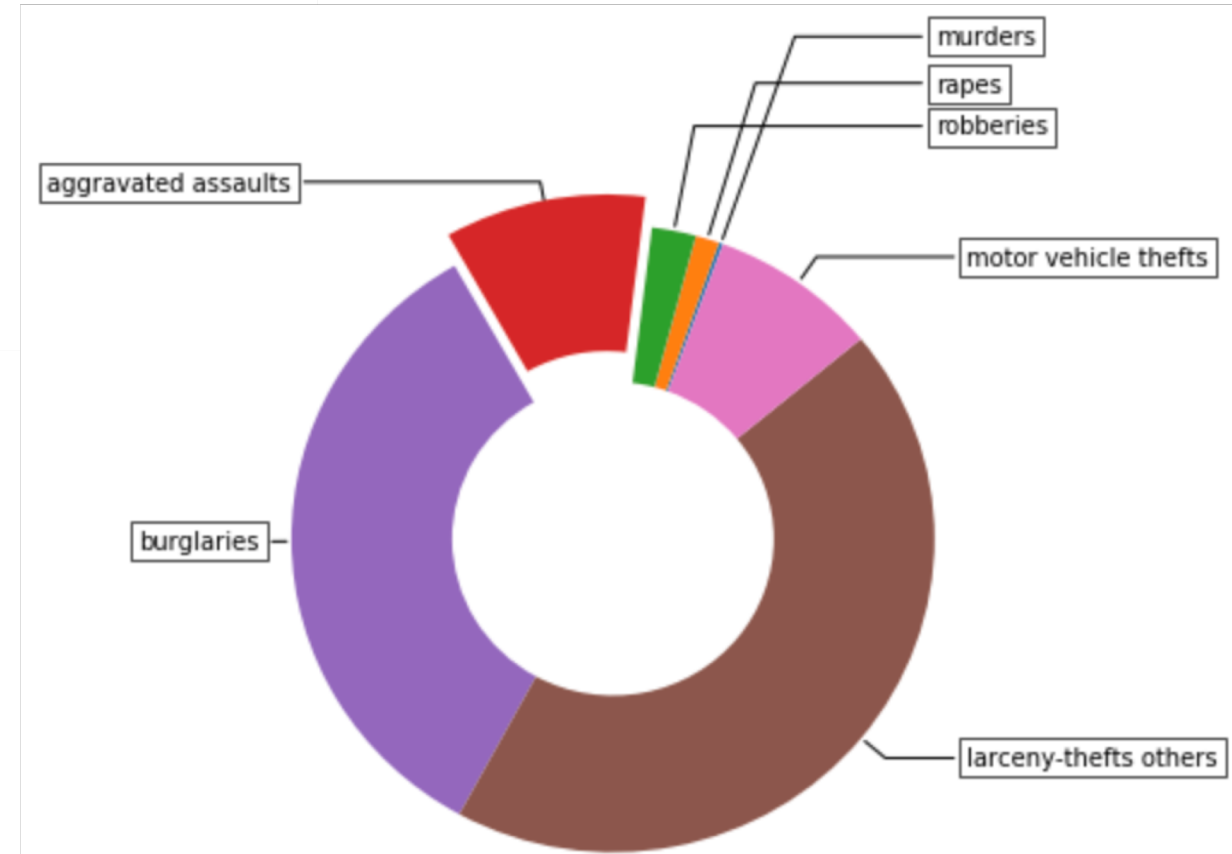
Yaqi Zhang

Predict **crime rate** from other **social factors** on **county level**



Dataset:

- USA Counties: 2011 -- U.S. Census Bureau
- 1.2 GB
- 51 categories. Crime is one of them.
- It contains >6000 feature columns
- Each row contains info of a county



Preprocess of the data

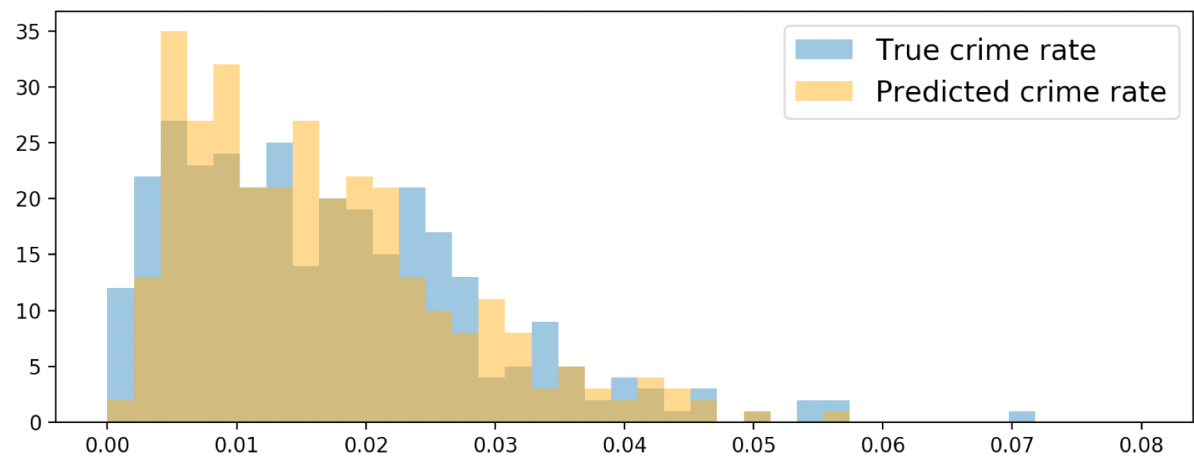
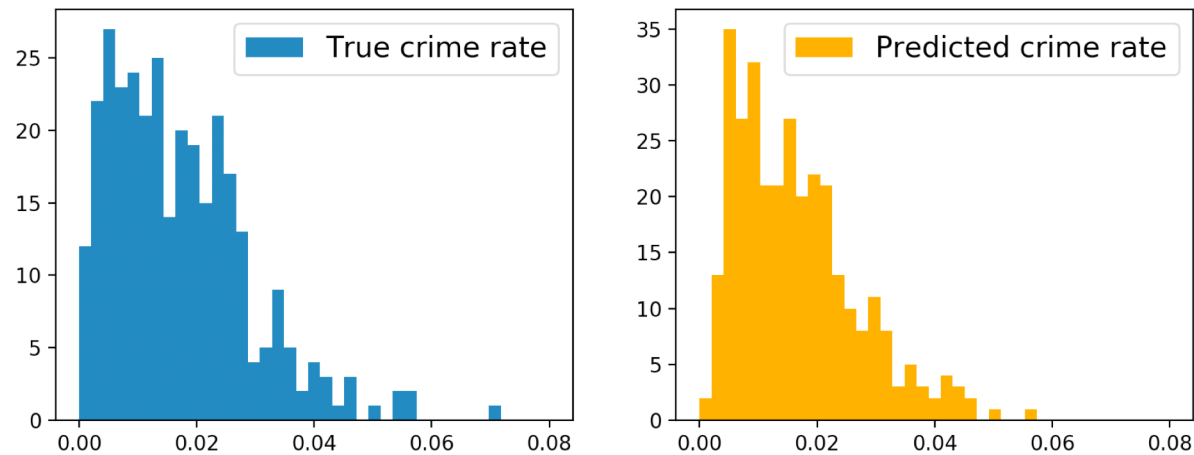
- Select meaningful columns and rows
- Replace the same info of different years with the mean value
- Normalize each column by total population if necessary

Crime data:

- Normalize by total population to get the **crime rate**

Result of the LightGBM model

LightGBM model, total crime rate



	R-square train	R-square test	avg train percent error	avg test percent error
Total crime	0.93	0.82	22%	19%
Assaults	0.89	0.69	33%	28%

Feature importance of LightGBM model

50 categories
1896 features

Total crime rate

Aggravated assault rate

