Chicago Crime and Housing Values

JARED HAMMERNIK

Goals

- Broadly to understand the effects of crime and different crime types on housing values.
- To understand what crime types exert the strongest effect on housing values.
- To be able to predict future housing values based on past crime values and trends.

Crime Data

- Crime data from Chicago Police Department (https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2)
- 7 million crimes from 2001 to 2019
- Includes type of crime and location

Housing Value Data

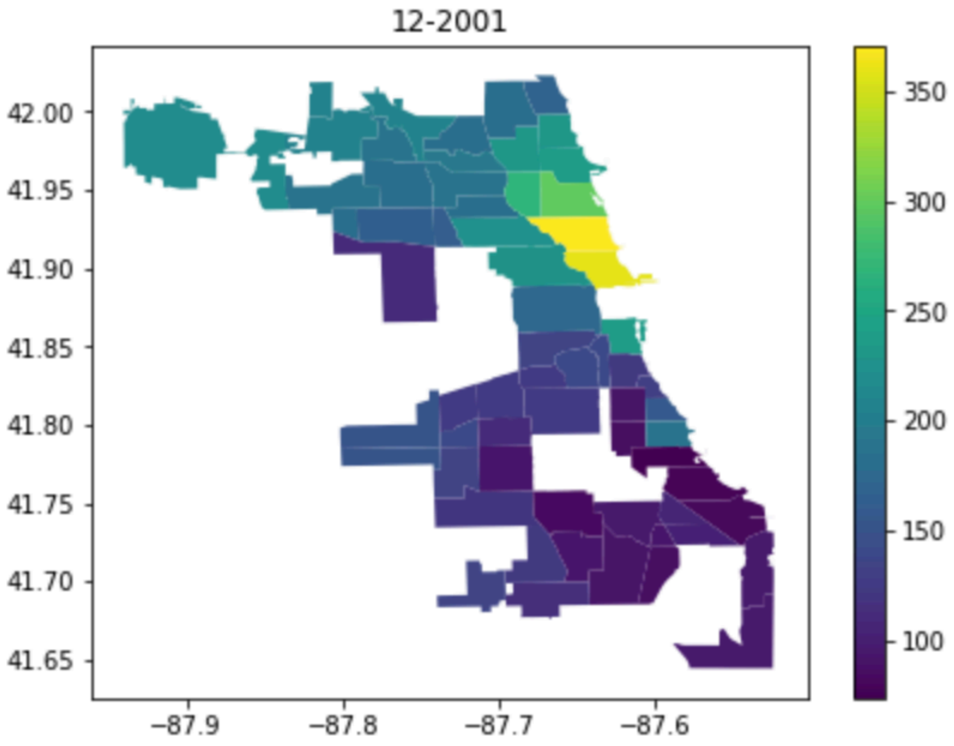
- Data from Zillow (https://www.zillow.com/research/data/)
- Median Housing Value per Square Foot
- Monthly values by Chicago neighborhood, using values from 2001-2019

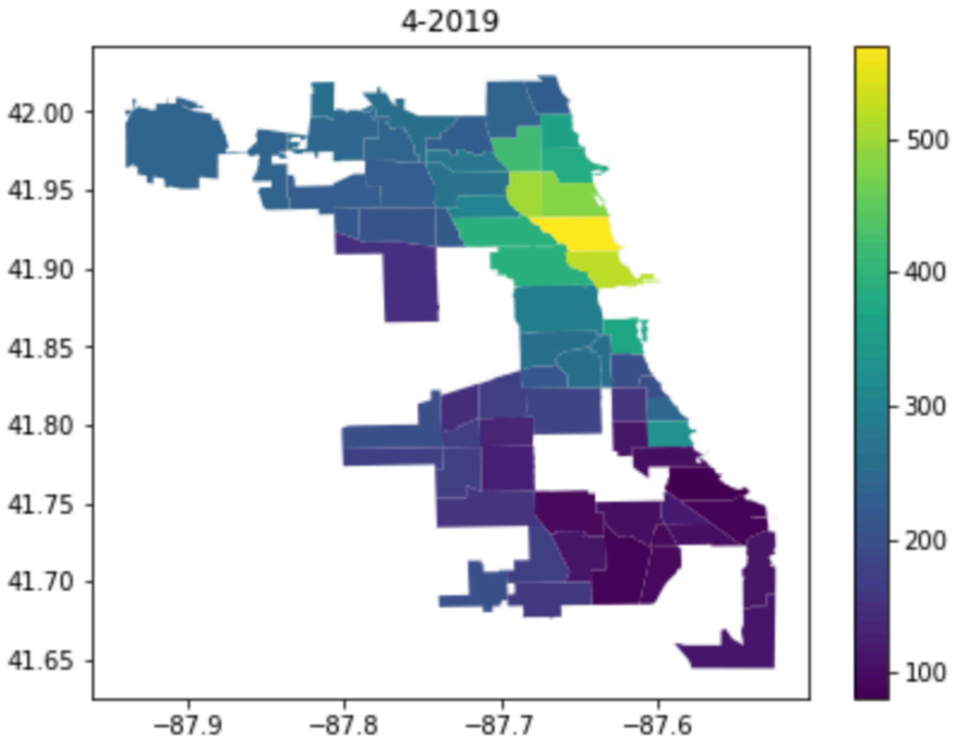
Data Wrangling

- Converted data into Dataframe with Chicago community-area-months as rows (~12700) and count of crime type in that area as columns (~35)
- Divided crime counts by population values to get crime rates per 100,000 people
- To adjust for seasonality, made each crime value a rolling 12-month average
- Added Zillow housing values as target column
- Also added a column with the next months housing value to predict on, as well as an "adjusted" housing value column that is the difference from the average housing value of the city (to adjust for effects of housing crisis and inflation)

EDA – Housing Values Maps

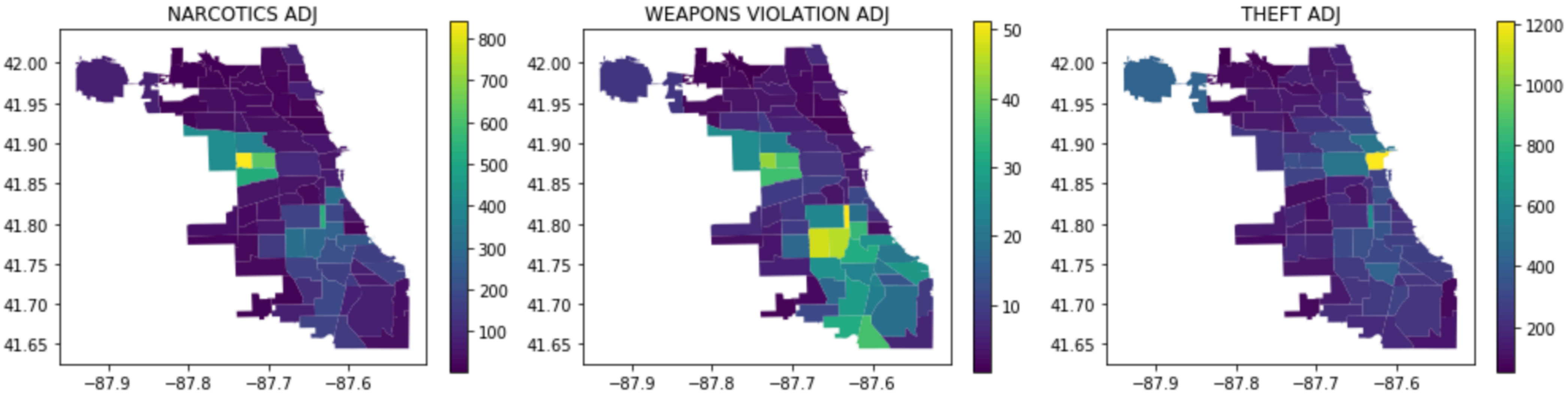
*some community areas had no housing value data





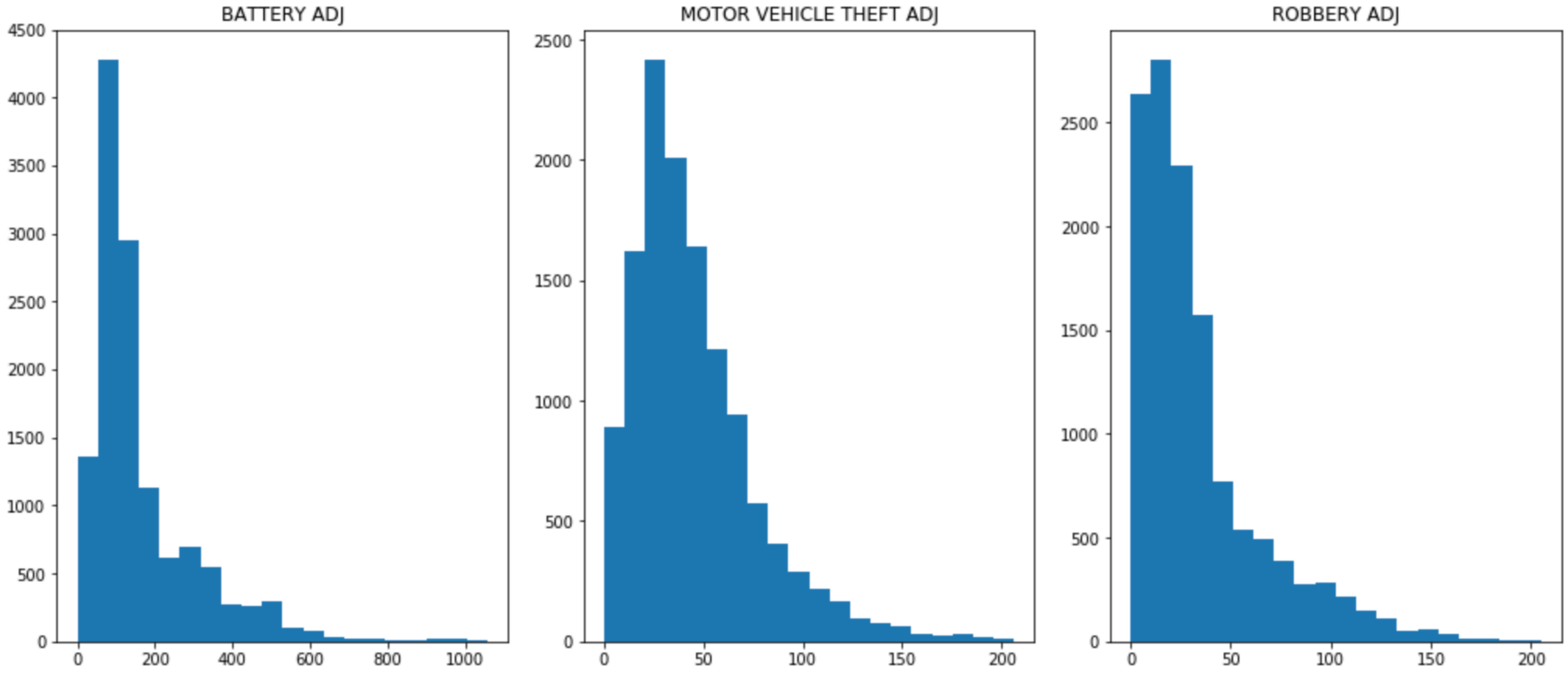
Here we can see from 2001 to 2019 there has not been much difference in the *relative* housing values, although housing values in general have certainly increased.

EDA – Crime Maps

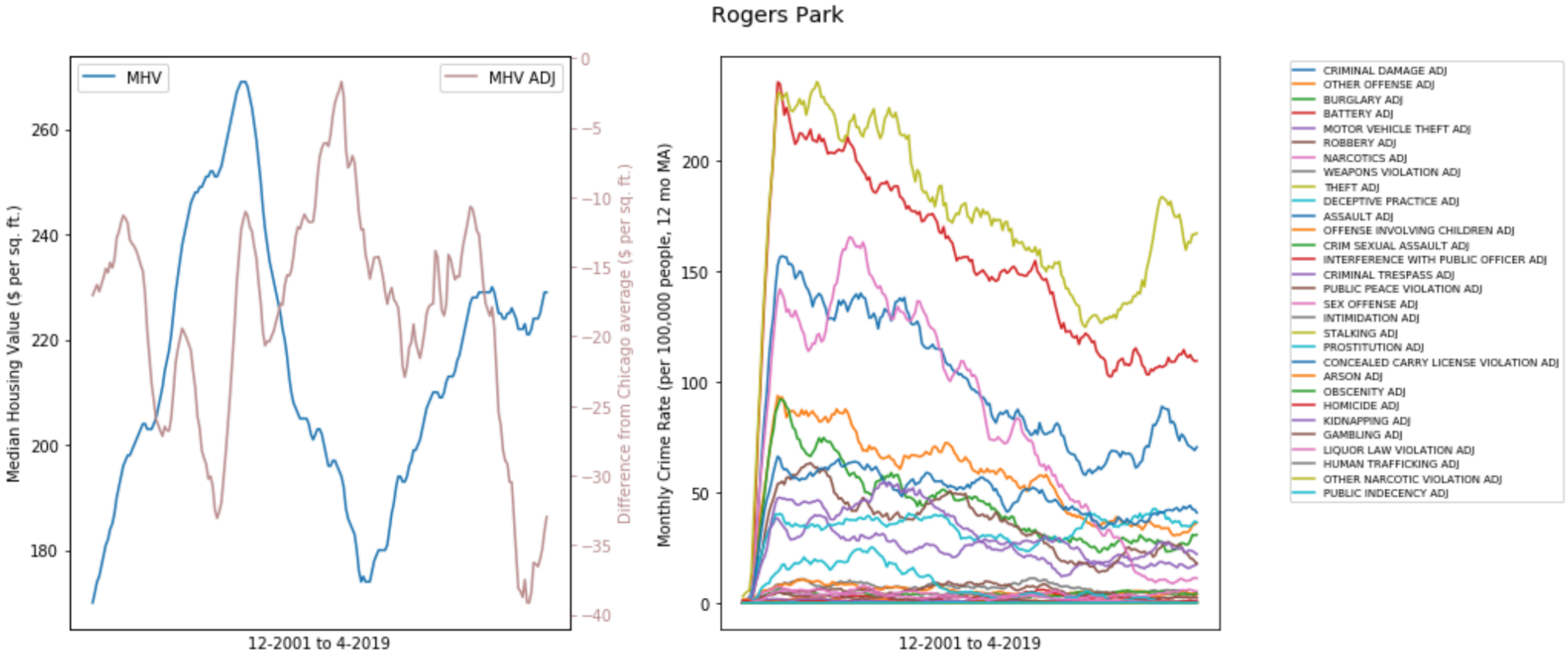


Average crime rates for each crime type from 2001 – 2019. One can see from these samples that while most crime types have been concentrated in the south and west sides, there are some types of crime that are more concentrated in the loop and other areas.

EDA – Distributions



EDA – Time Series



Example of one community area. One can see all crime tends to go down over time, and first couple months produce strange values that we may want to remove. Also, all community areas exhibit same pattern moving upward, then dipping, then recovering (due to 2008 housing crisis).

Predicting Current Month

Used current months crime values to predict housing values

- Random Forest - R^2: 0.971, RMSE: 15.019
- Gradient Boosting - R^2: 0.975, RMSE: 14.084
- KNN - R^2: 0.991, RMSE: 8.011
- Linear - R^2: 0.590, RMSE: 56.530
- Voting Regressor - R^2: 0.988, RMSE: 9.489

Predicting Current Month - Adjusted

Used current months crime values to predict housing values on the adjusted value. One can see the models perform slightly better.

- Random Forest - R^2: 0.974, RMSE: 13.821
- Gradient Boosting - R^2: 0.979, RMSE: 12.373
- KNN - R^2: 0.992, RMSE: 7.537
- Linear - R^2: 0.675, RMSE: 48.879
- Voting Regressor - R^2: 0.989, RMSE: 8.627

Feature Importances

Random Forest

feature	importance
OFFENSE INVOLVING CHILDREN ADJ	0.254353
DECEPTIVE PRACTICE ADJ	0.115295
THEFT ADJ	0.091560
WEAPONS VIOLATION ADJ	0.089229
HOMICIDE ADJ	0.059157
ASSAULT ADJ	0.037735
OTHER OFFENSE ADJ	0.037482
CRIMINAL TRESPASS ADJ	0.030624
MOTOR VEHICLE THEFT ADJ	0.029278
BATTERY ADJ	0.026109
PROSTITUTION ADJ	0.026072
NARCOTICS ADJ	0.025793
LIQUOR LAW VIOLATION ADJ	0.023061
CRIMINAL DAMAGE ADJ	0.017871
BURGLARY ADJ	0.017142
ROBBERY ADJ	0.016430
INTERFERENCE WITH PUBLIC OFFICER ADJ	0.014889
CRIM SEXUAL ASSAULT ADJ	0.012904
PUBLIC PEACE VIOLATION ADJ	0.011463
SEX OFFENSE ADJ	0.010267

Gradient Boosting

feature	importance
OFFENSE INVOLVING CHILDREN ADJ	0.309454
DECEPTIVE PRACTICE ADJ	0.128464
THEFT ADJ	0.096319
WEAPONS VIOLATION ADJ	0.078450
HOMICIDE ADJ	0.048303
OTHER OFFENSE ADJ	0.040451
MOTOR VEHICLE THEFT ADJ	0.031124
LIQUOR LAW VIOLATION ADJ	0.027115
NARCOTICS ADJ	0.024668
PROSTITUTION ADJ	0.023001
CRIMINAL TRESPASS ADJ	0.021965
KIDNAPPING ADJ	0.019566
ROBBERY ADJ	0.016544
BATTERY ADJ	0.016277
ASSAULT ADJ	0.013501
INTIMIDATION ADJ	0.013062
CRIMINAL DAMAGE ADJ	0.012359
BURGLARY ADJ	0.010873
GAMBLING ADJ	0.010572
PUBLIC PEACE VIOLATION ADJ	0.010276

Both random forest and gradient boosting models showed similar feature importances. Interestingly, crimes against children had the strongest effect by a large degree.

Predicting Future Month

Used current months crime values to predict future month housing values

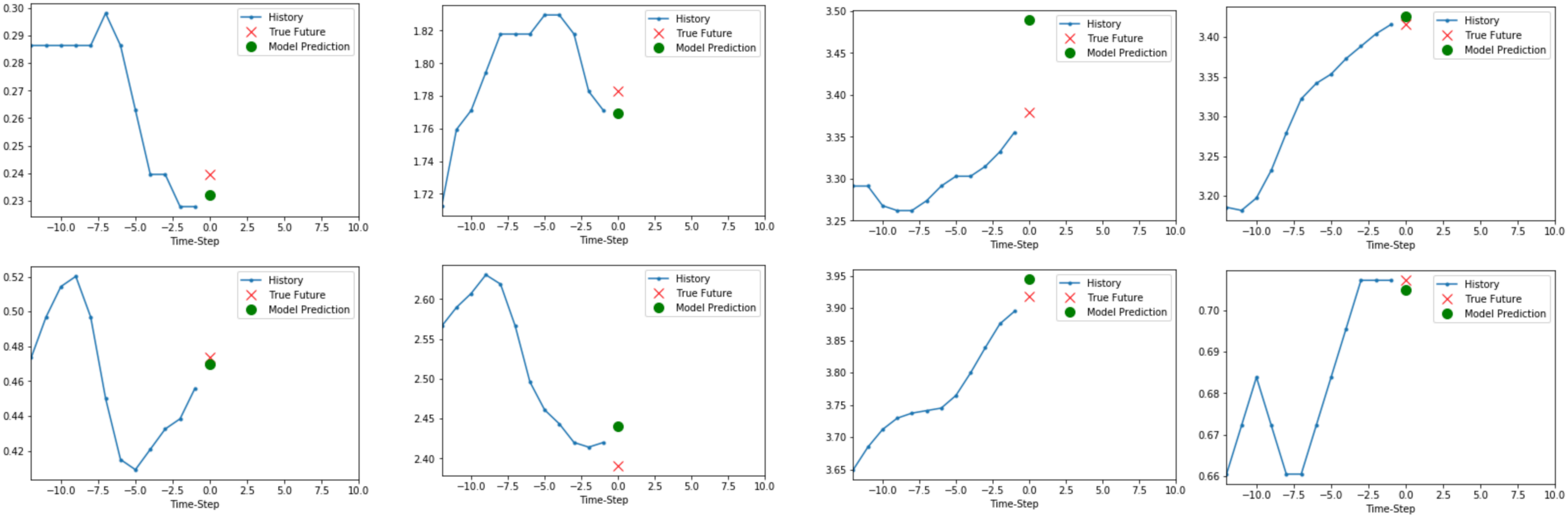
	R-squared	RMSE	%Directionality
Baseline	0.9995	1.947	54.5%*
Random Forest	0.9996	1.689	69.85
Gradient Boosting	0.9997	1.434	75.97
KNN	0.9979	4.011	66.71
Linear	0.9996	1.769	61.76
Voting	0.9997	1.517	78.96

*Baseline for directionality determined to be the balance of the data, i.e. the percent of the test points where the housing value increased.

LSTM

- Created a simple LSTM neural network to predict the future months housing value
- Used 12 month history, 4 neurons, ~100 epochs, 32 batch size
- Best RMSE was 6.837
- Best %Directionality was only 65% compared to 57.8% baseline
- Did not perform as well as other models

LSTM Examples



Conclusions

- Crimes against children exhibit strongest effect on housing values
- Direction of future housing value may be predicted decently well in some models (especially with Gradient Boosting and ensemble voting models), however further study is needed.

Next Steps

- Further tune LSTM model (different architectures, optimizers, etc.)
- Go in reverse, see if housing prices predict crime
- Try with rental prices
- Make interactive visualization (Bokeh)