Groningen Municipality Crime Data 1996-2021

Evalution of crime data, corelation and fitting a LSTM model for predictions.

Alex de Vries, Feb 2023

Introduction

1) A presentation targeted at stakeholders
2) A presentation targeted at data science peers

a presentation targeted at stakeholders

The use case

Is the crime rate going down or up over time?

Crime rates in Groningen, Netherlands

Level of crime	18.60	Very Low
Crime increasing in the past 3 years	55.56	Moderate
Worries home broken and things stolen	22.60	Low
Worries being mugged or robbed	16.57	Very Low
Worries car stolen	9.94	Very Low
Worries things from car stolen	16.05	Very Low
Worries attacked	22.26	Low
Worries being insulted	17.63	Very Low
Worries being subject to a physical attack because of your skin color, ethnic origin, gender or religion	14.74	Very Low
Problem people using or dealing drugs	41.99	Moderate
Problem property crimes such as vandalism and theft	30.43	Low
Problem violent crimes such as assault and armed robbery	15.86	Very Low
Problem corruption and bribery	12.66	Very Low

Index

Crime Index: 21.84

Safety Index: 78.16

Safety in Groningen, Netherlands

Safety walking alone during daylight	87.35	Very High
Safety walking alone during night	73.50	High

https://www.numbeo.com/crime/in/Groningen

The data set

	Year	theft car	theft motorcycle/scooter	theft moped	theft bike	theft from car	theft from company	theft from home	shoplifting	pickpocketing	robbery (including bag theft)	other property crimes	abuse/violence	sex crime	theft/human trafficking	destruction property damage
0	1996	327	28	438	2842	2657	157	3162	1094	295	241	28	780	137	256	2014
1	1997	428	30	404	319	2712	1559	2908	982	193	217	3229	870	135	246	2183
2	1998	351	26	418	3058	2918	1761	2787	1188	387	227	2983	926	117	302	2428
3	1999	334	28	560	3322	3087	1481	281	1241	375	271	3206	894	93	305	2244
4	2000	262	19	301	3512	3058	1852	2279	1169	383	248	3037	975	140	386	2249

Crime data for the municipality of Groningen from 1996 - 2021.
It is broken down by crime type. The crime type and subtypes are available both in Dutch and in English. English translations done by Google translate.

https://www.kaggle.com/datasets/megan3/groningen-crime-data-19962021

The data set

	Year	theft car	theft motorcycle/scooter	theft moped	theft bike	theft from car	theft from company	theft from home	shoplifting	pickpocketing	robbery (including bag theft)	other property crimes	abuse/violence	sex crime	theft/human trafficking	destruction property damage
count	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.00000	26.000000	26.000000	26.000000	26.000000
mean	2008.500000	159.961538	30.384615	365.346154	2881.769231	1533.576923	740.923077	1263.807692	938.692308	398.730769	133.538462	2447.00000	887.307692	99.115385	456.961538	1716.615385
std	7.648529	124.587152	12.185489	86.701761	795.686738	1010.707680	593.331336	822.015816	228.652709	145.681998	76.237120	919.63558	128.993882	19.279682	99.096511	675.941008
min	1996.000000	17.000000	13.000000	188.000000	319.000000	103.000000	150.000000	137.000000	113.000000	84.000000	23.000000	28.00000	524.000000	68.000000	246.000000	124.000000
25%	2002.250000	67.750000	23.250000	312.500000	2799.250000	890.000000	346.500000	668.000000	862.000000	301.750000	60.000000	2087.75000	867.750000	88.250000	425.000000	1569.500000
50%	2008.500000	106.000000	28.000000	351.500000	3148.000000	1272.000000	414.500000	1187.500000	950.000000	410.500000	132.000000	2709.50000	928.000000	97.500000	469.000000	1906.500000
75%	2014.750000	283.000000	34.250000	414.500000	3332.500000	2595.500000	1241.750000	1601.000000	1073.250000	484.250000	198.000000	3023.50000	957.000000	107.250000	528.750000	2204.750000
max	2021.000000	428.000000	64.000000	560.000000	3829.000000	3251.000000	1852.000000	3162.000000	1241.000000	687.000000	271.000000	3544.00000	1103.000000	140.000000	612.000000	2442.000000

No missing values. 26 years of data for all features.

https://www.kaggle.com/datasets/megan3/groningen-crime-data-19962021

the solution to the use case

Step 1: normalize each feature

Step 2: store the combined mean of all features for each year in a new feature

Step 3: Use this new feature as the indication of overal crime rate

Step 4: Apply LSTM on this non-linear time data.

the solution to the use case

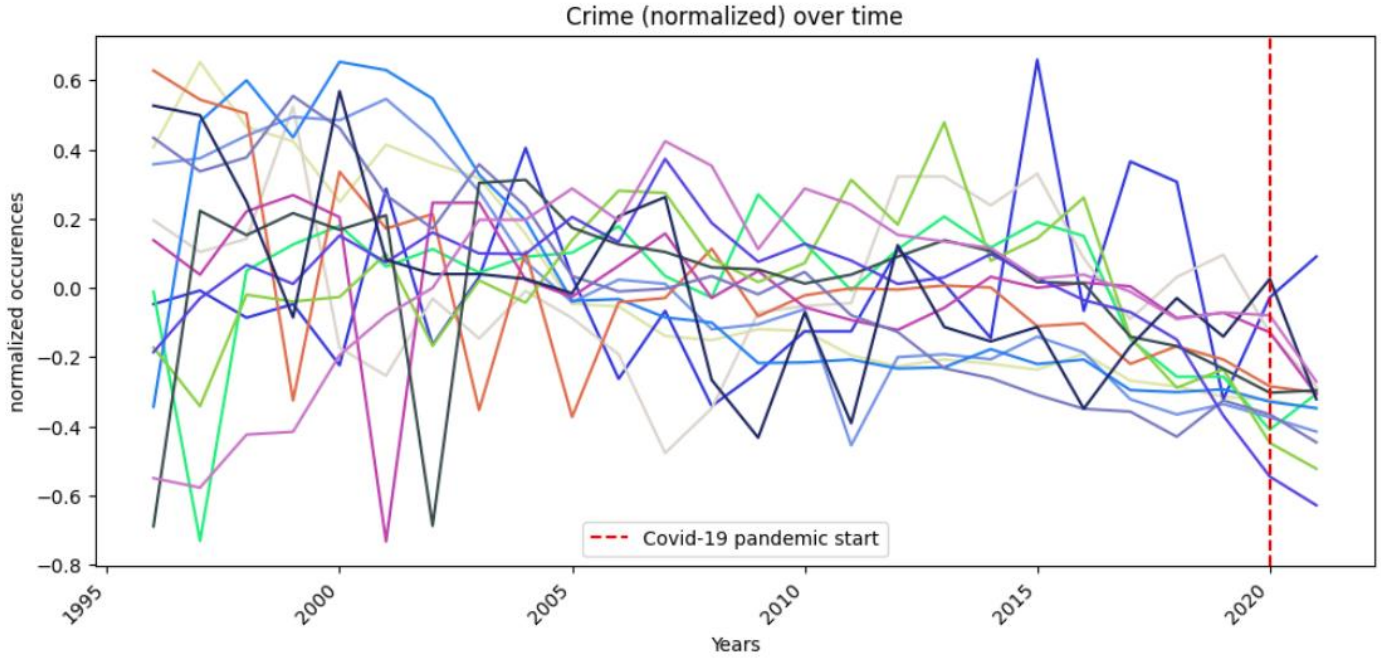


Figure 1: plot all crime features over time. Seems to go down over time, but unclear.

the solution to the use case

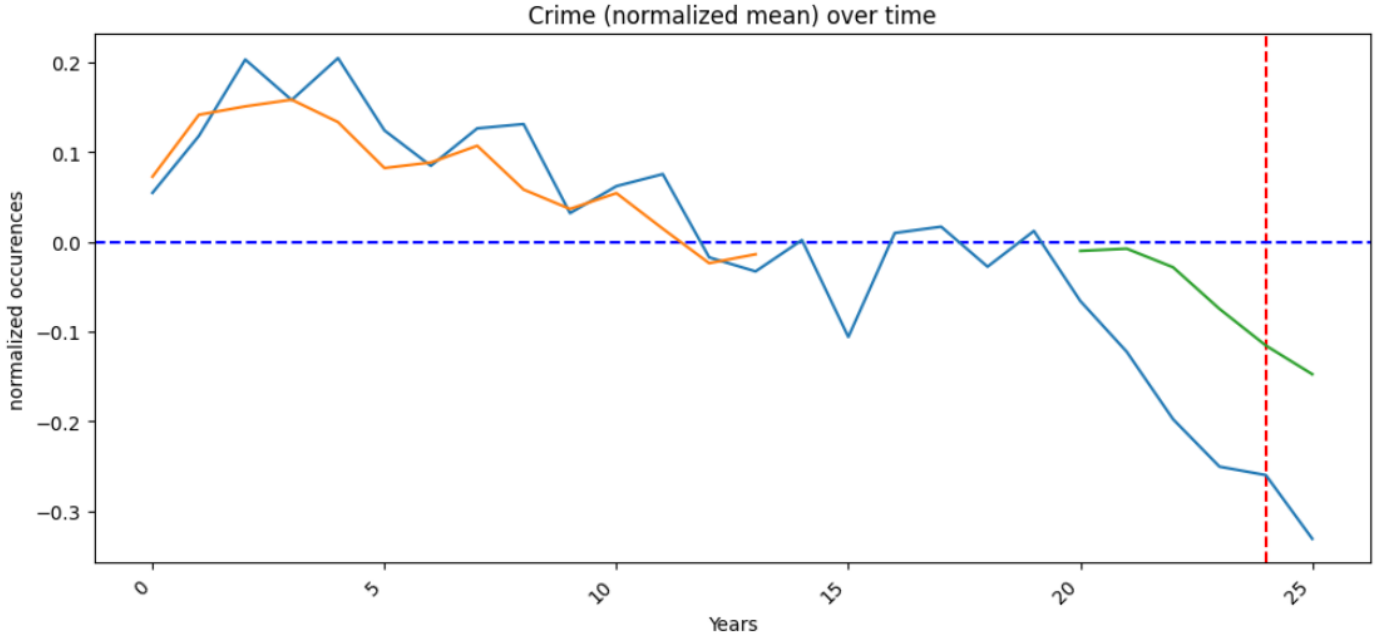


Figure 2: Blue line is the normalized mean crime rate. Below 0 indicates a reduction. Orange line is the training data set, Green the test dataset. Clearly crime rate goes Down over time.

summary

- Keeping it simple for stakeholders
- Crime goes down over time.
- The decrease is accelarating.
- The LSTM is predicting well, but can be improved with more data.

A presentation targeted at data science peers

architectural choices

- Why LSTM modelling
  - LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series.

```
# create and fit the LSTM network
model = Sequential()
model.add(LSTM(4, input_shape=(1, look_back)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
history = model.fit(trainX, trainY, epochs=100, batch_size=1, verbose=2,)

# Training Phase
model.summary()
```

data quality assessment

- data pre-processing & feature engineering
  - Check for null values -> none
  - Transpose data
  - Create year column
  - Normalize each feature
  - New feature -> mean of sum of all normalized features. This will be used as Crime rate indication.

Data set overview

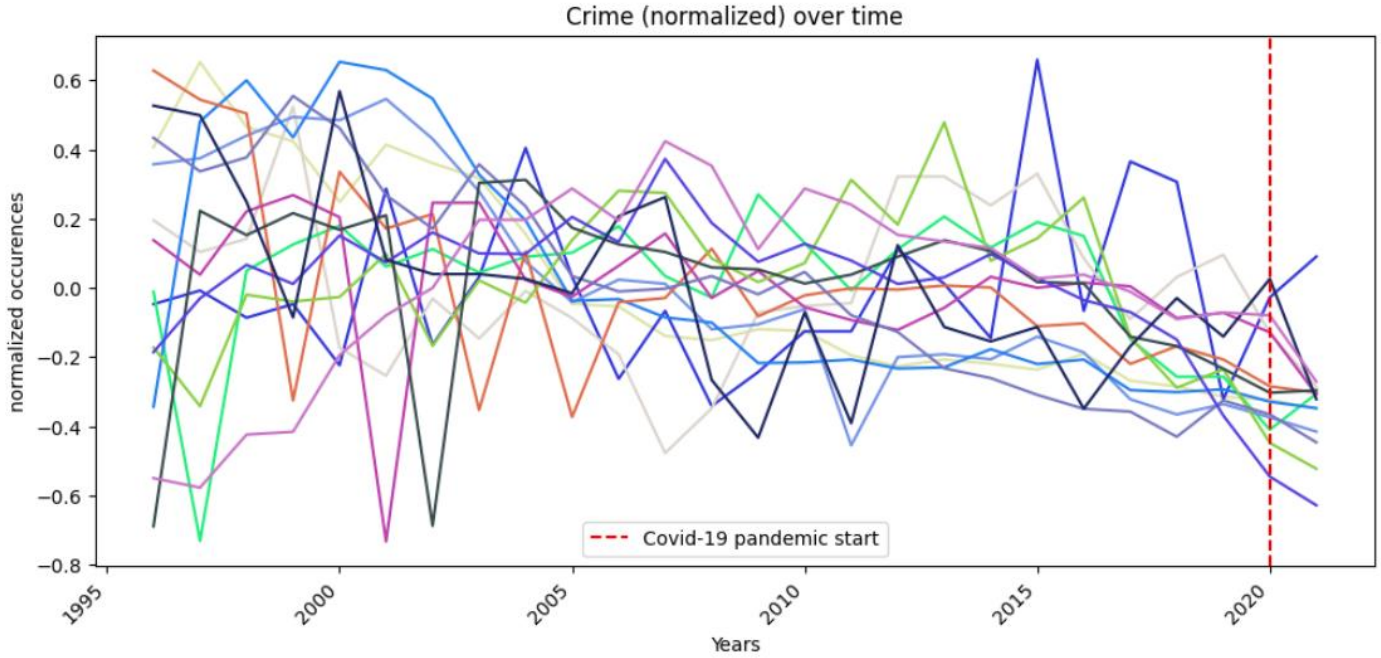
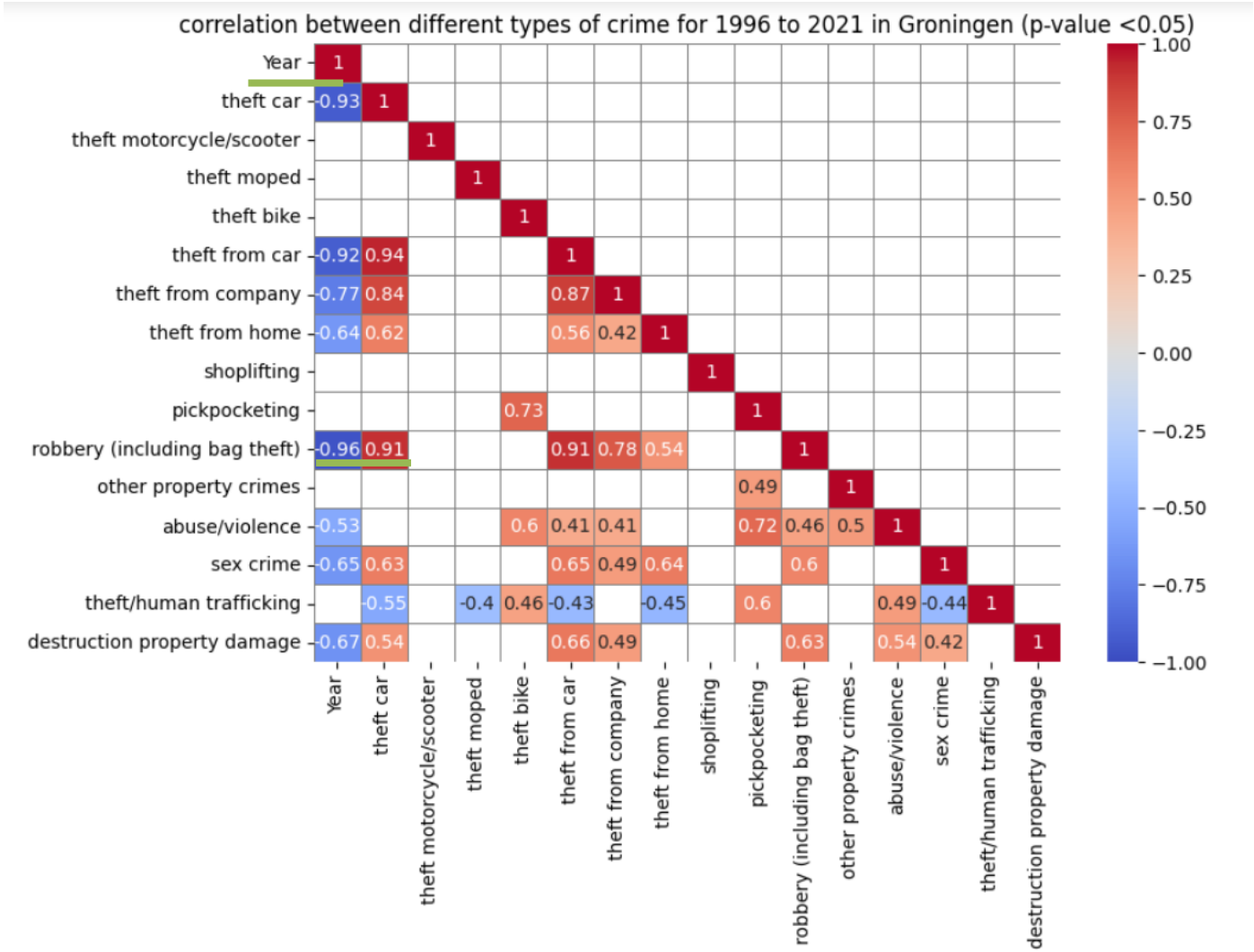
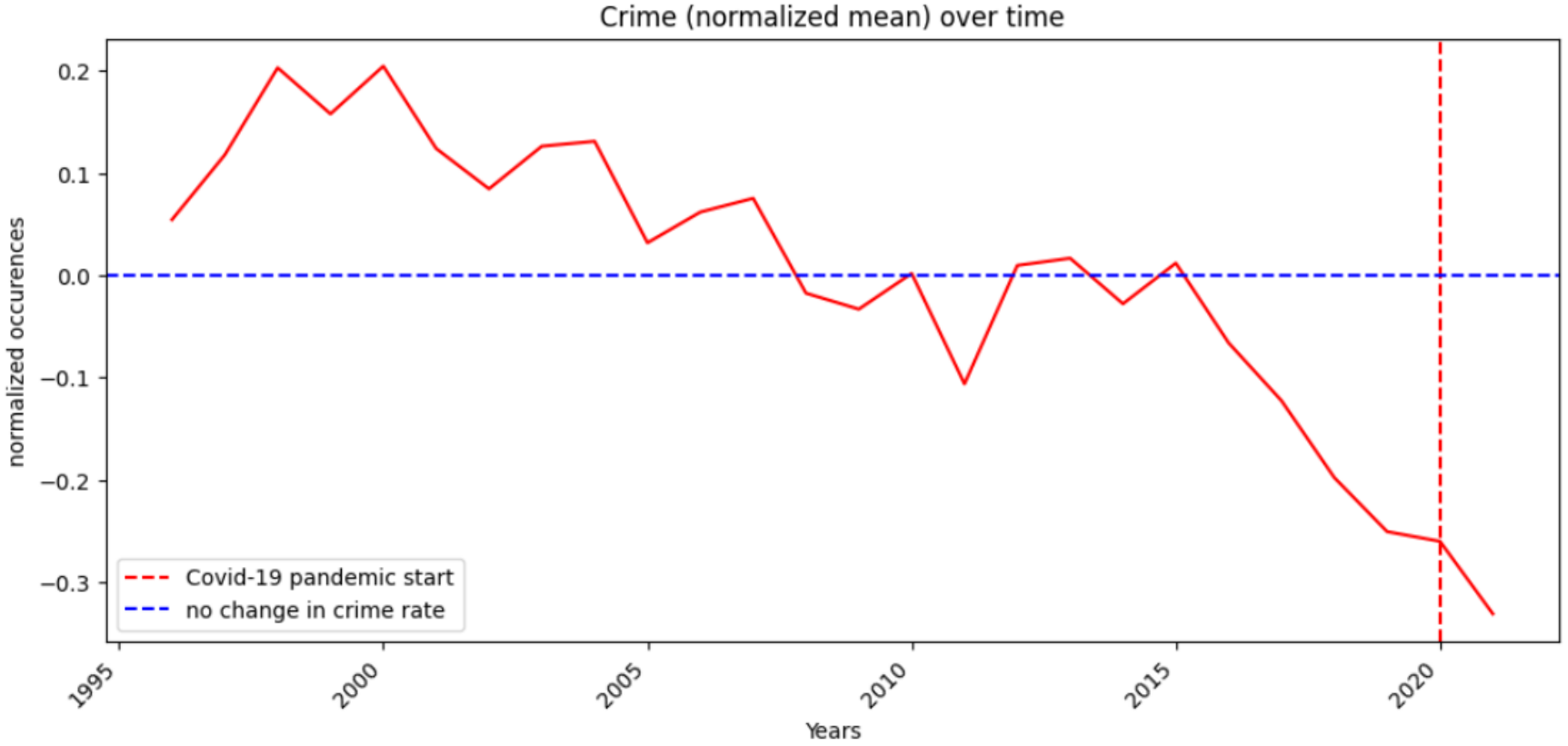


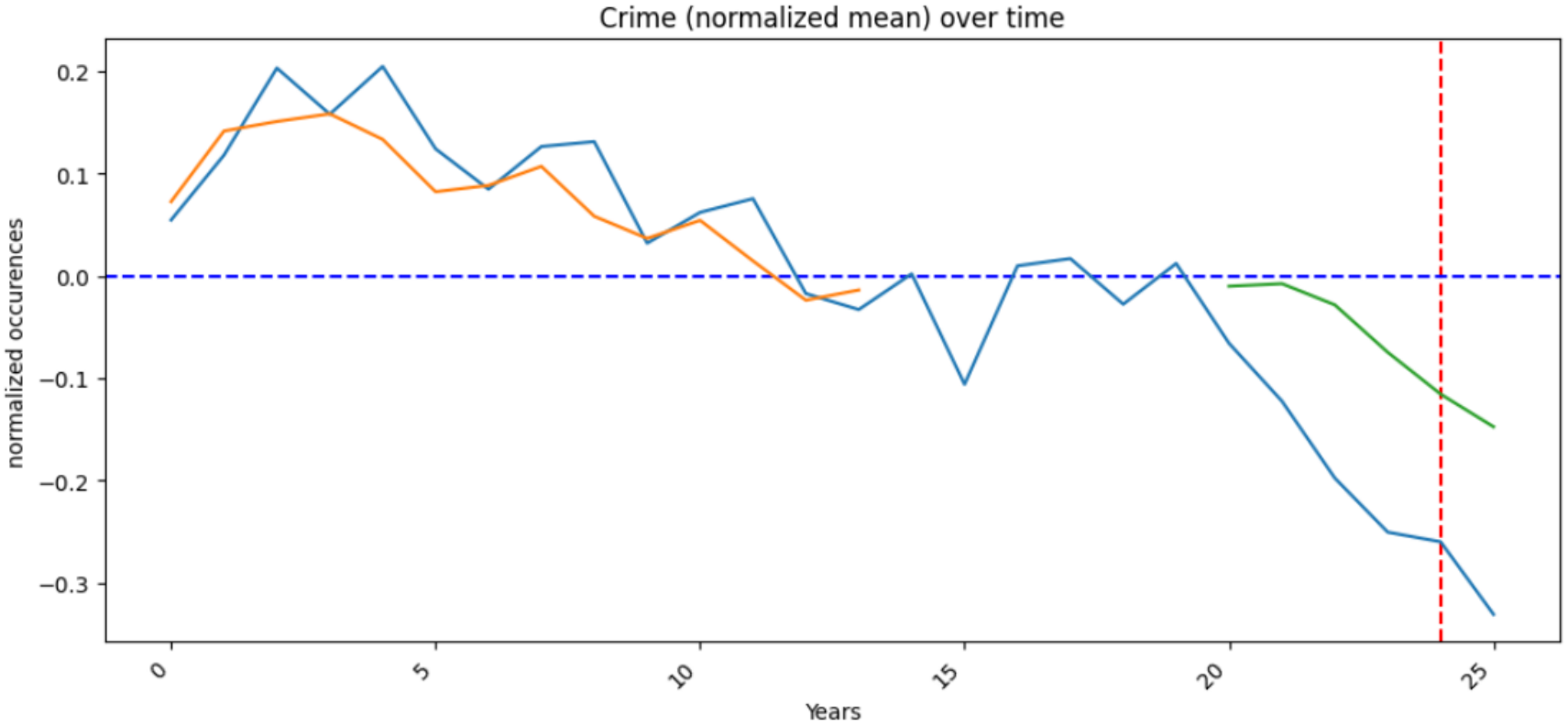
Figure 3: plot all crime features over time. Seems to go down over time, but unclear.


correlation between different types of crime for 1996 to 2021 in Groningen (p-value <0.05)
Year
theft car
theft motorcycle/scooter
theft moped
theft bike
theft from car
theft from company
theft from home
shoplifting
pickpocketing
robbery (including bag theft)
other property crimes
abuse/violence
sex crime
theft/human trafficking
destruction property damage
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

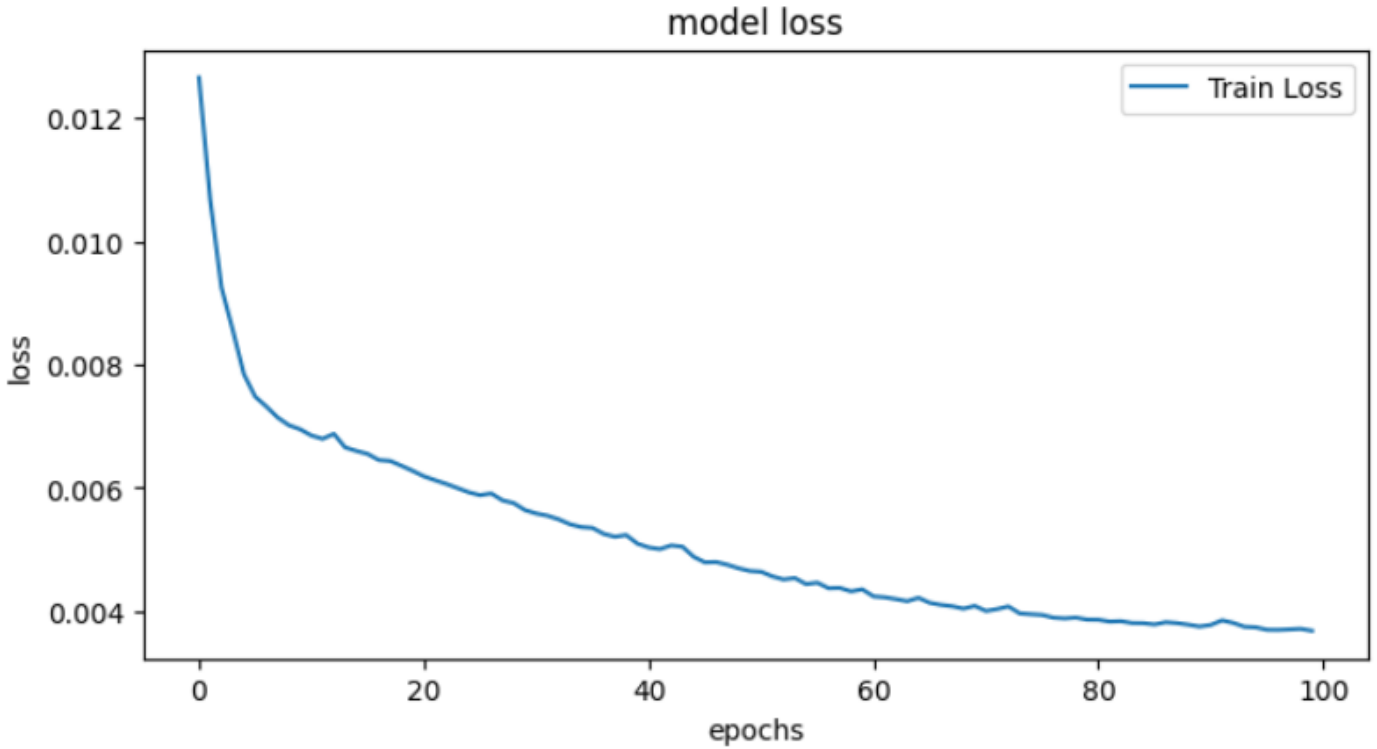
Data set overview



Data set overview



model performance indicators



summary

- Crime goes down over time.
- The decrease is accelarating.
- The LSTM is predicting well, but can be improved with more data.
- Multiple feature are correlated.