NYPD Criminal Complaints

Time Series Analysis

Data Collection

NYC OpenData

NYPD Complaint Data 2006-2017, 6 million observations

Scope narrowed to 2014-2017 (after Stop-and-Frisk policy change)

1.9 million observations

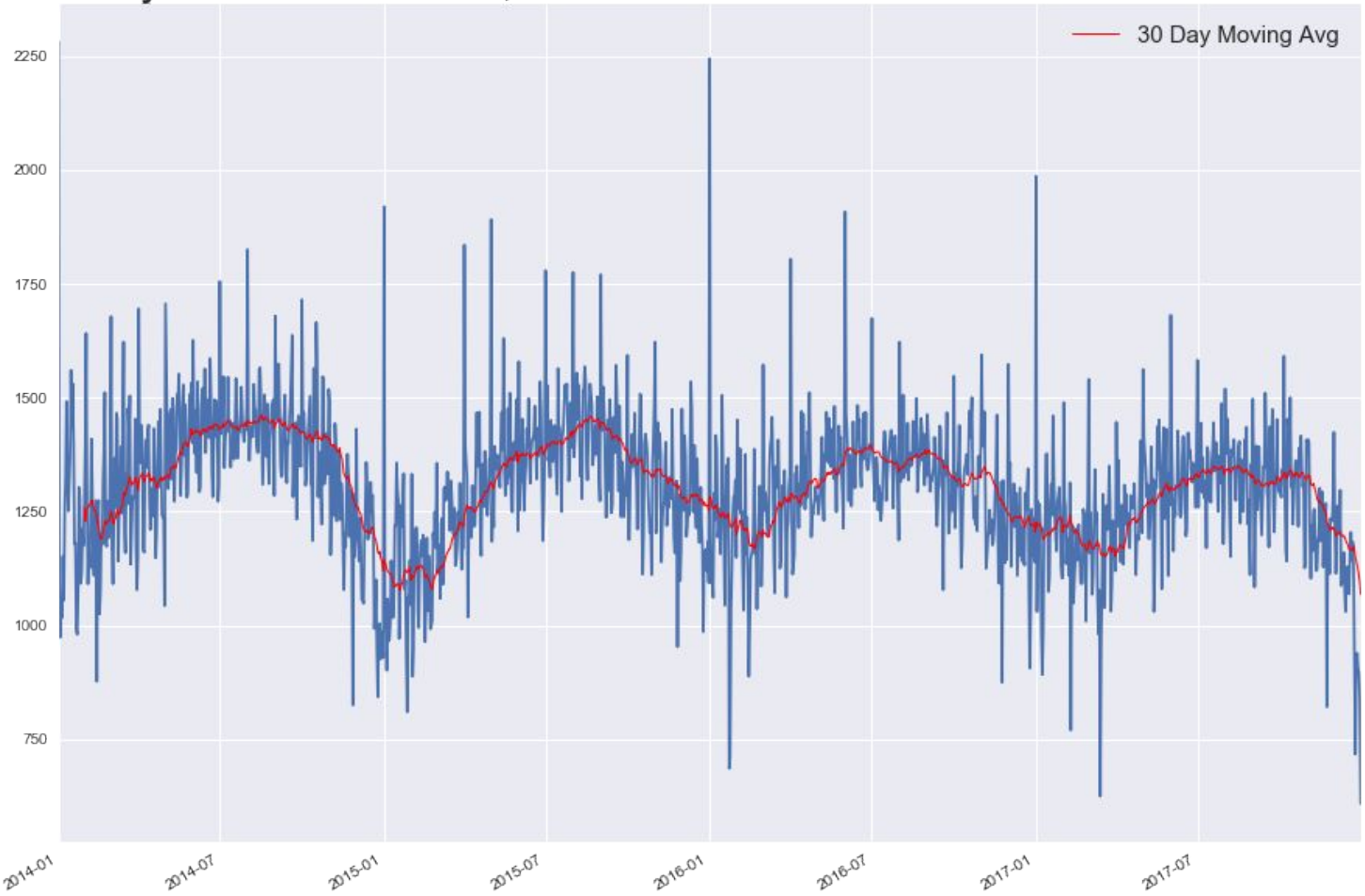
NOAA

Central Park Daily Weather Data

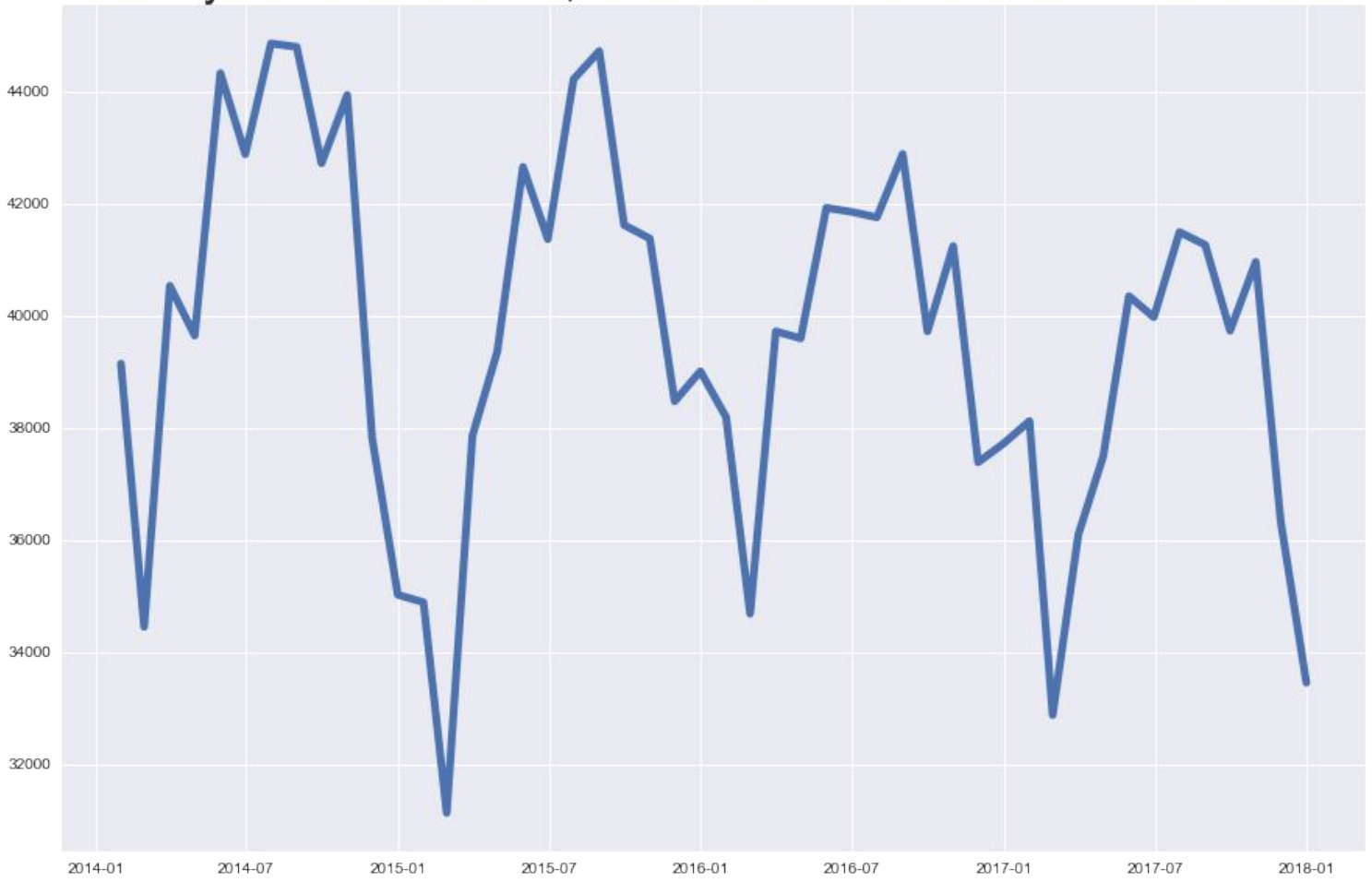
Maximum Temperature

Data Exploration

Daily Crime Incidents, 2014-2017



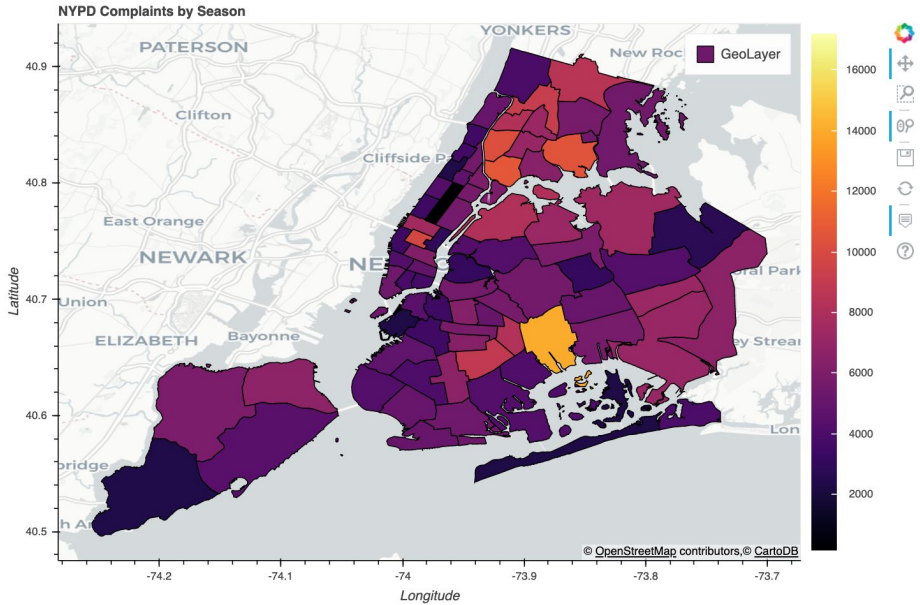
Monthly Crime Incidents, 2014-2017



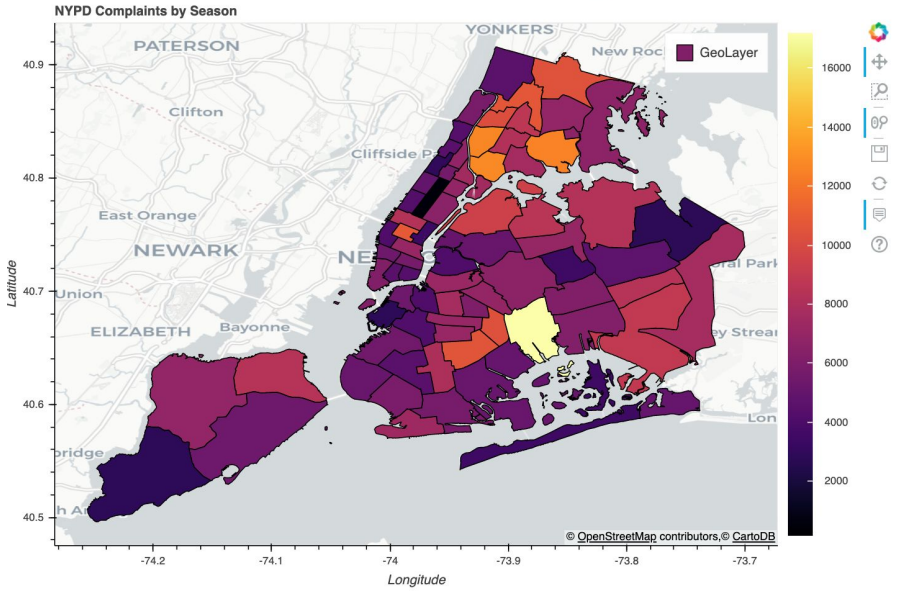
Top Precincts

Crime by Precinct, Seasonal Comparison

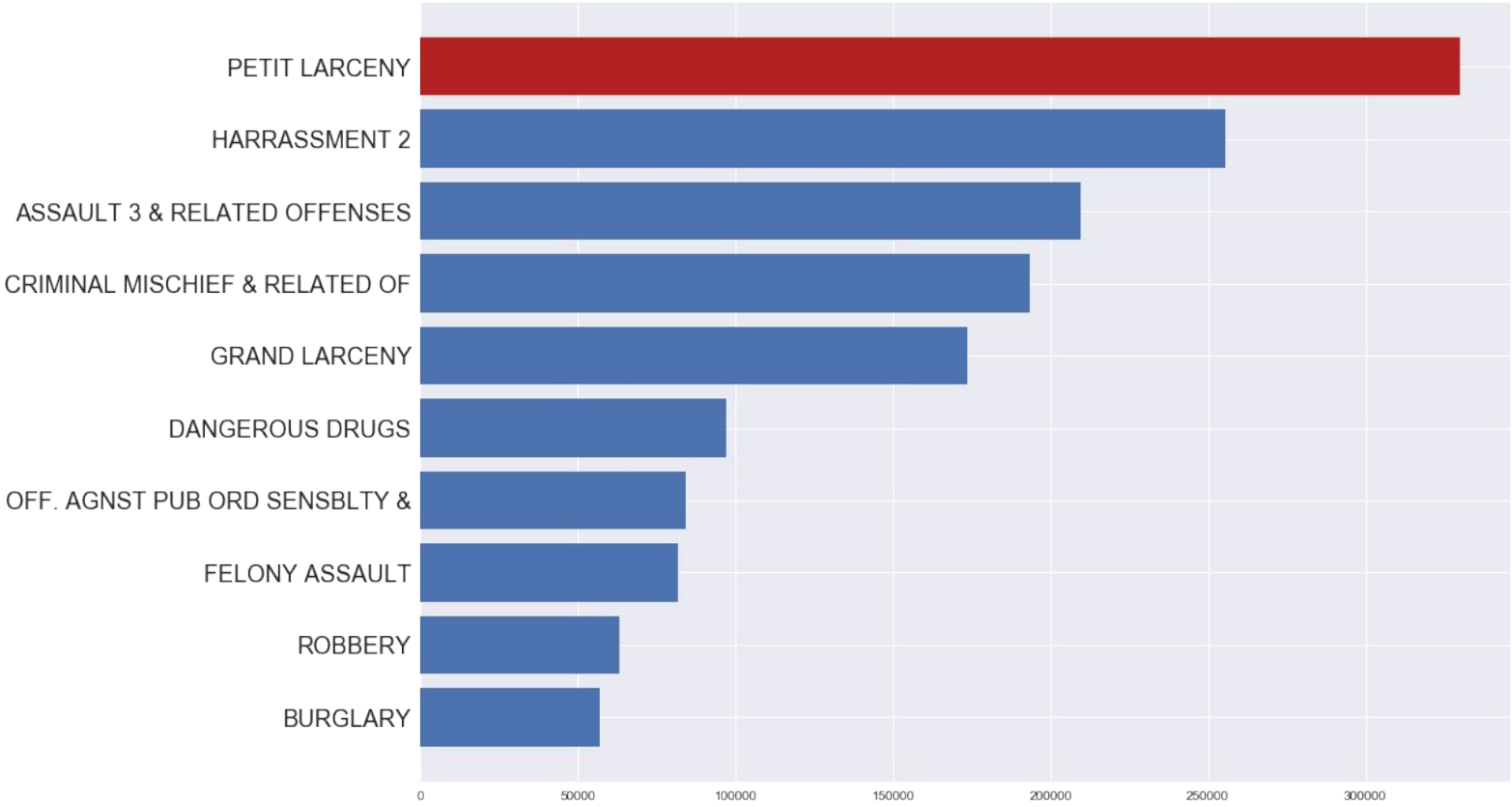
Winter



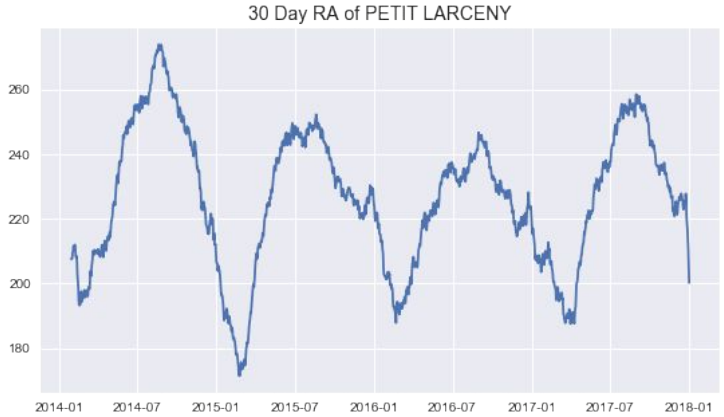
Summer



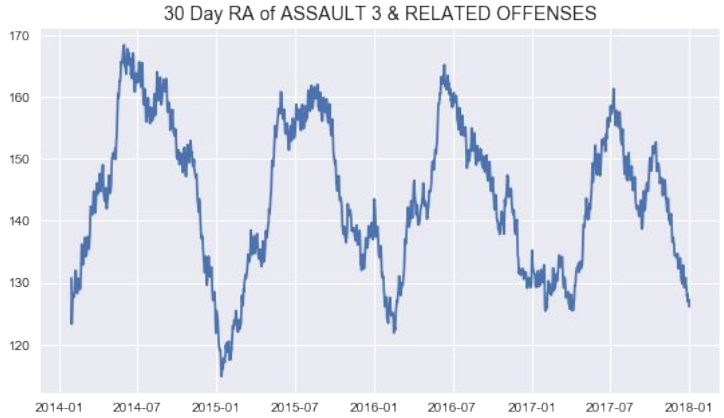
Top Crimes


PETIT LARCENY
HARRASSMENT 2
ASSAULT 3 & RELATED OFFENSES
CRIMINAL MISCHIEF & RELATED OF
GRAND LARCENY
DANGEROUS DRUGS
OFF. AGNST PUB ORD SENSBLTY &
FELONY ASSAULT
ROBBERY
BURGLARY
0
50000
100000
150000
200000
250000
300000

Top Crimes

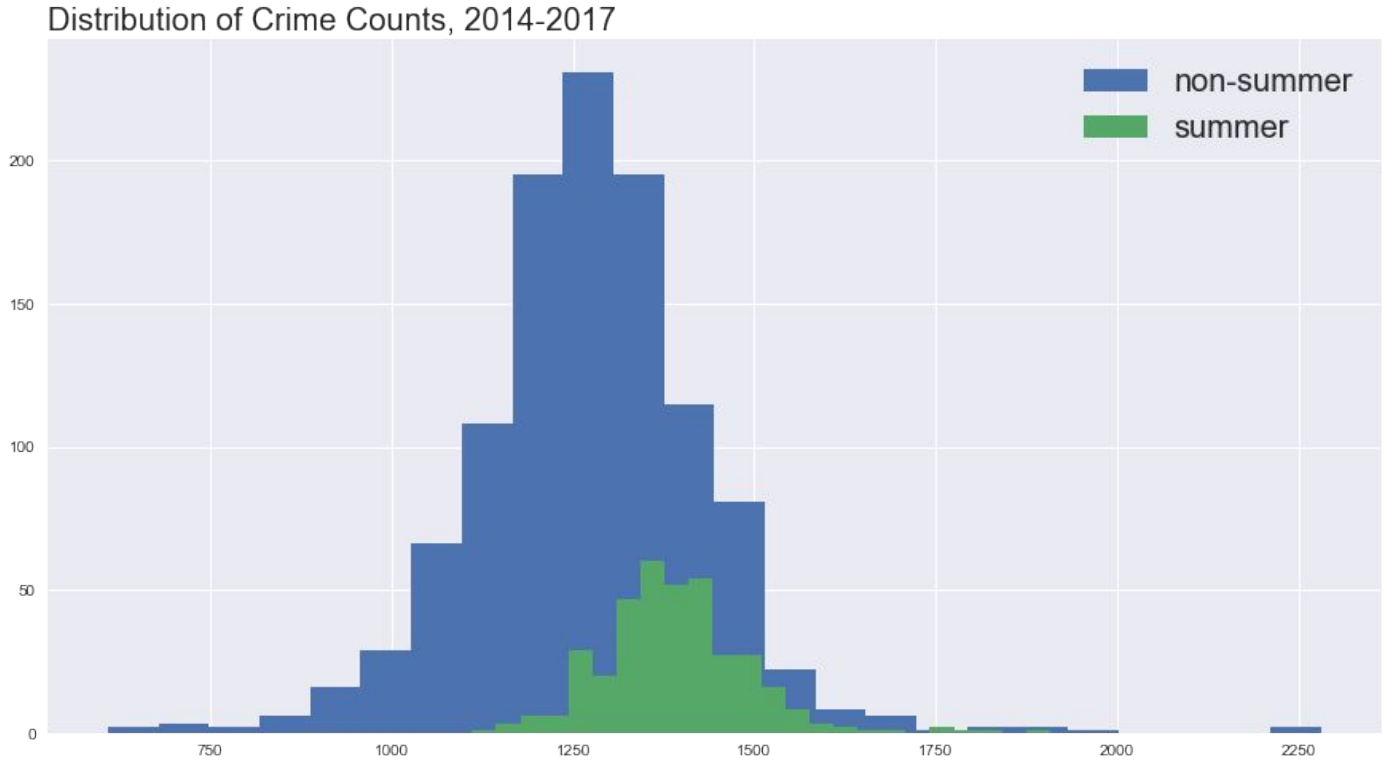

30 Day RA of PETIT LARCENY
260
240
220
200
180
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01


30 Day RA of HARRASSMENT 2
210
200
190
180
170
160
150
140
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01


30 Day RA of ASSAULT 3 & RELATED OFFENSES
170
160
150
140
130
120
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01


30 Day RA of CRIMINAL MISCHIEF & RELATED OF
150
140
130
120
110
100
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01

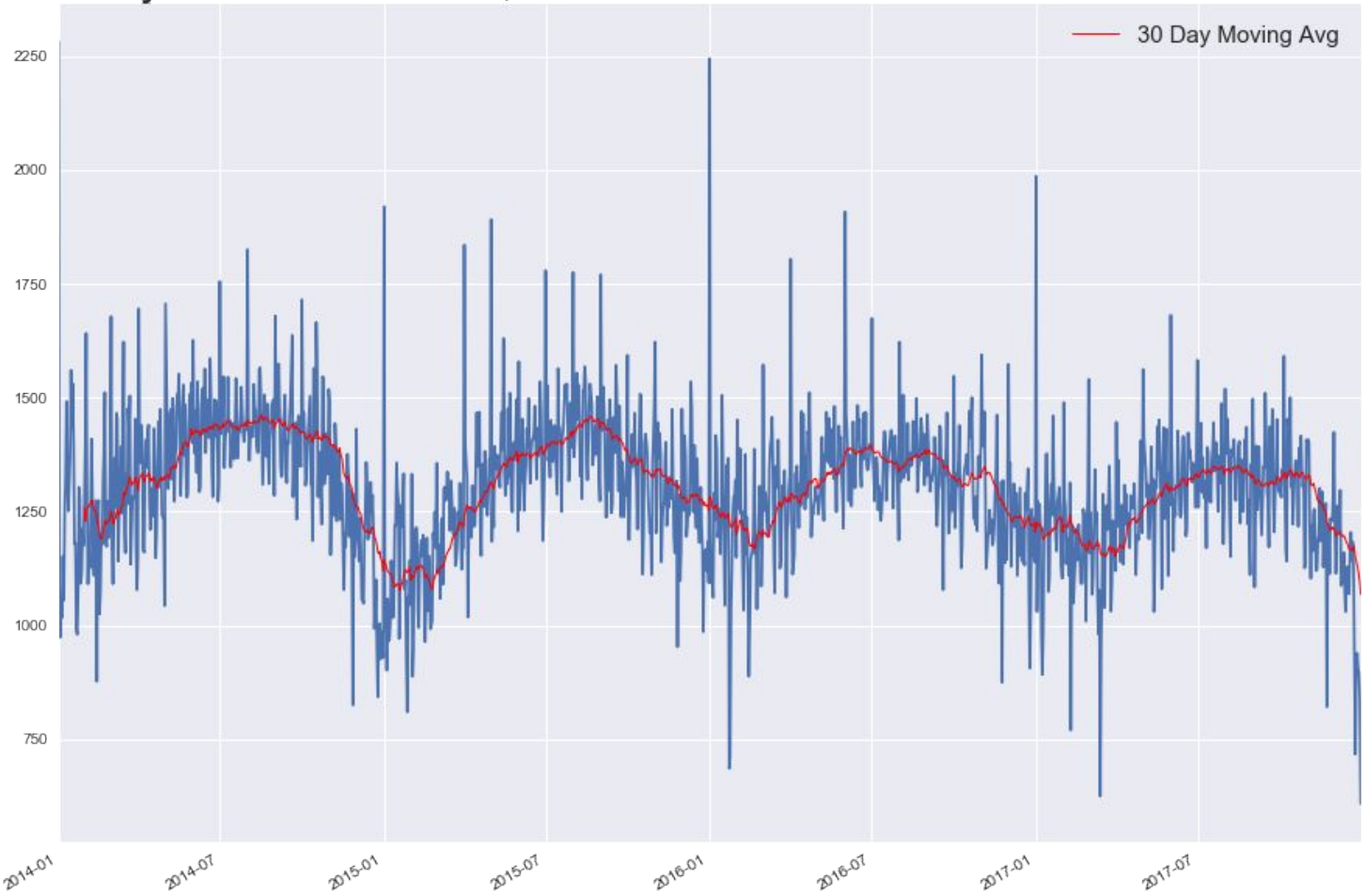
T-Test: Difference of Summer/Non-Summer Crime

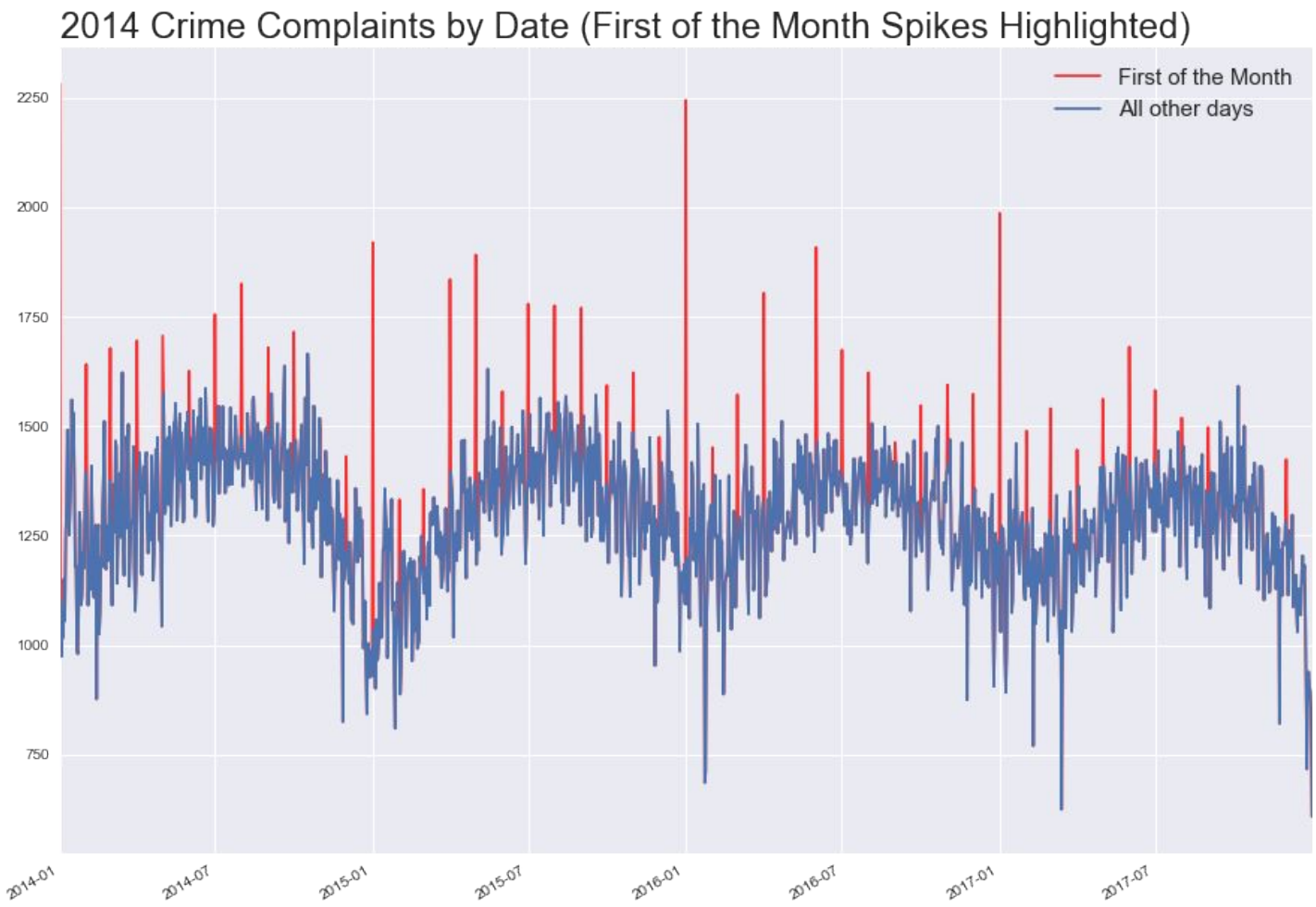


p value:

4.184e-39

Daily Crime Incidents, 2014-2017

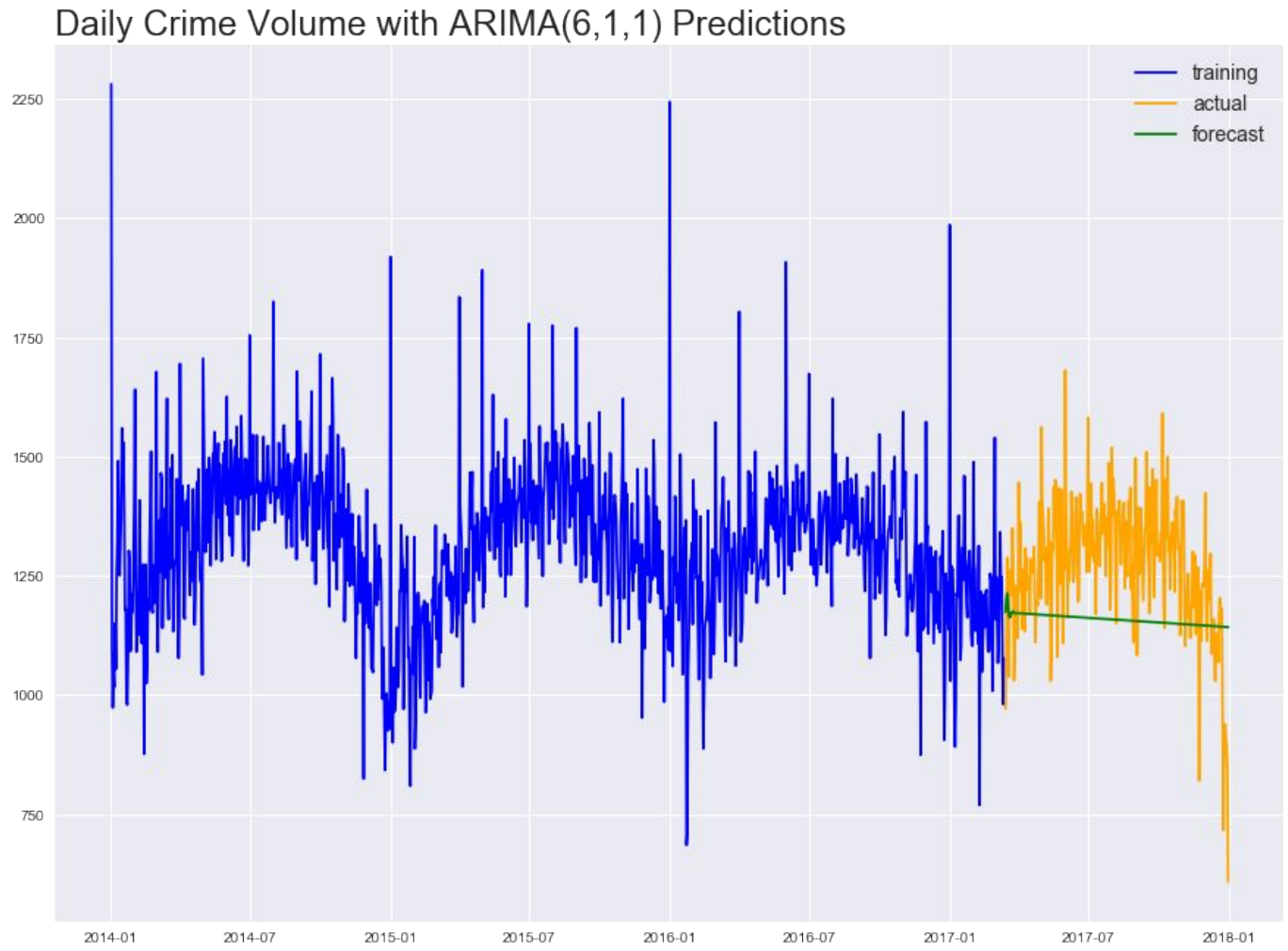



2014 Crime Complaints by Date (First of the Month Spikes Highlighted)
First of the Month
All other days
2250
2000
1750
1500
1250
1000
750
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07

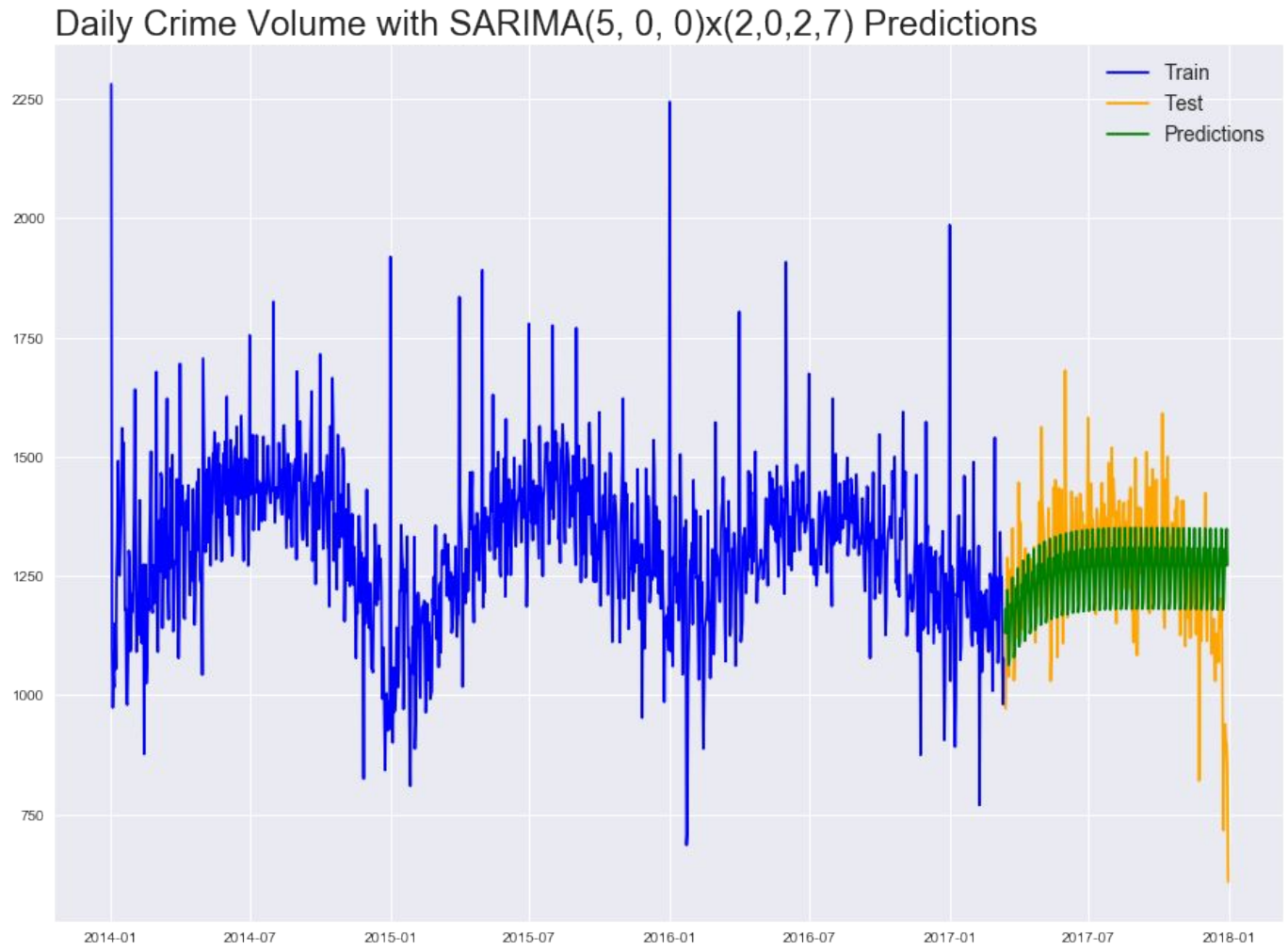

Average Crime Complaints Per Day, 2014-2017
1600
1400
1200
1000
800
600
400
200
0
0
5
10
15
20
25
30

30 Day Moving Average Anomalies

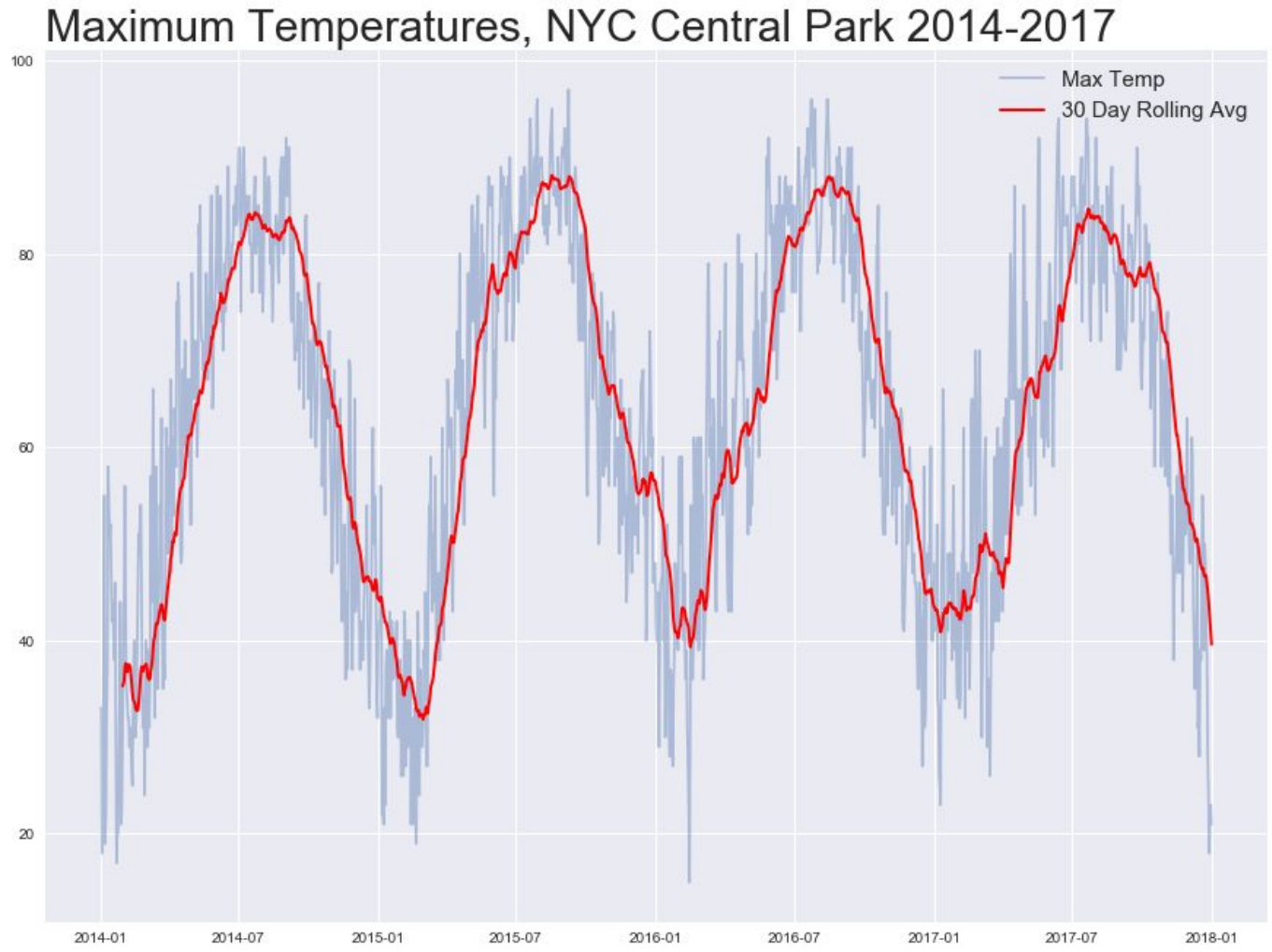
Modeling

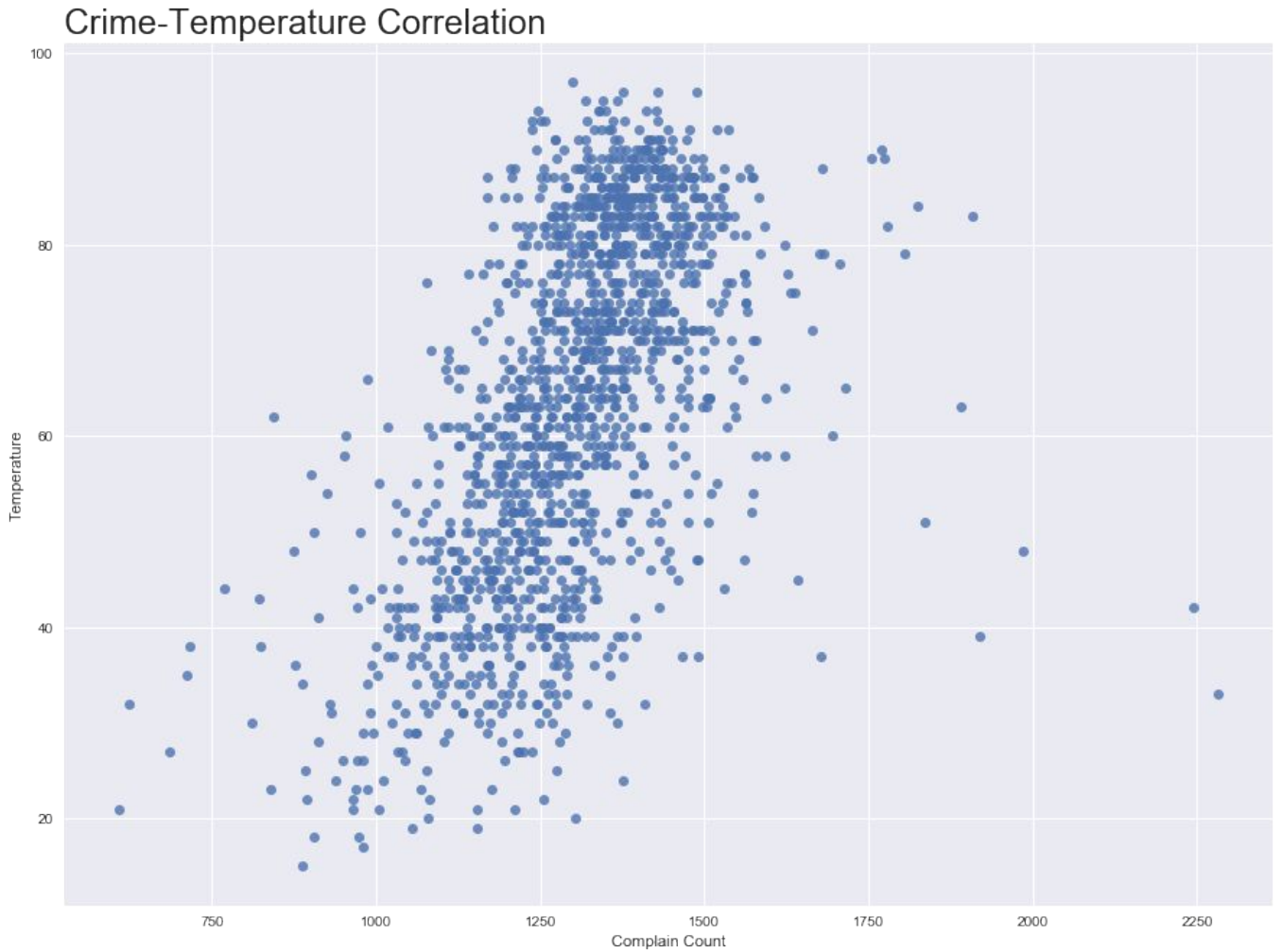

Daily Crime Volume with ARIMA(6,1,1) Predictions
training
actual
forecast
2250
2000
1750
1500
1250
1000
750
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01

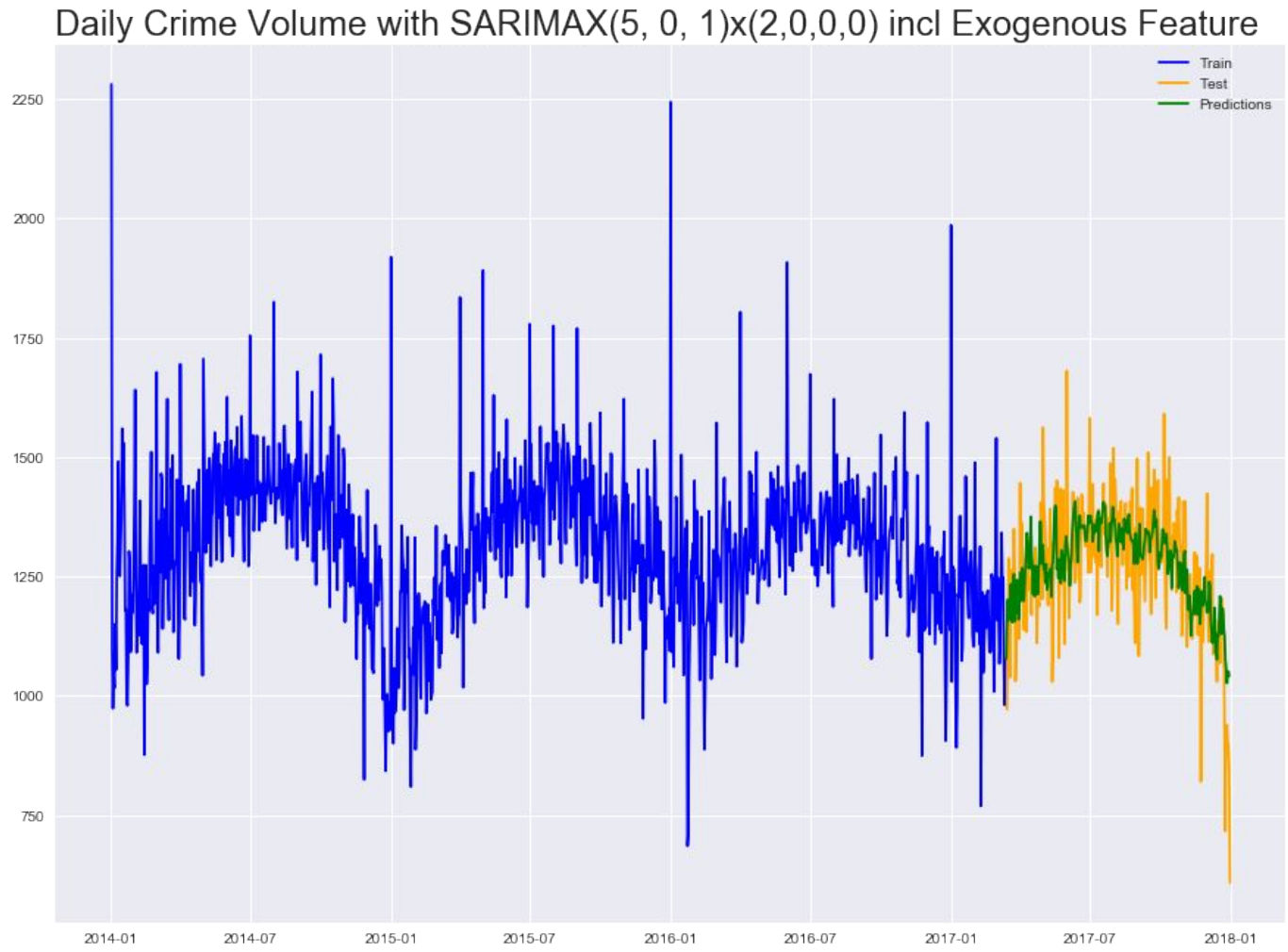
RMSE: 173.98


Daily Crime Volume with SARIMA(5, 0, 0)x(2,0,2,7) Predictions
Train
Test
Predictions
2250
2000
1750
1500
1250
1000
750
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01

RMSE: 142.43


Maximum Temperatures, NYC Central Park 2014-2017
Max Temp
30 Day Rolling Avg
100
80
60
40
20
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01


Crime-Temperature Correlation
Temperature
100
80
60
40
20
750
1000
1250
1500
1750
2000
2250
Complain Count


Daily Crime Volume with SARIMAX(5, 0, 1)x(2,0,0,0) incl Exogenous Feature
Train
Test
Predictions
2250
2000
1750
1500
1250
1000
750
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01

RMSE: 99.252

Conclusion

Exogenous features can greatly improve the performance of time series predictions.

By incorporating temperature into the model, RMSE for crime predictions was reduced by 30%.