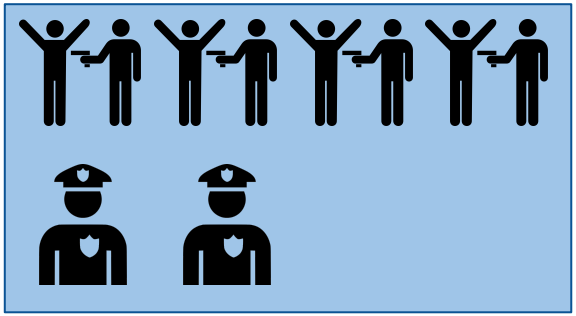
Forecasting Seattle Crime

Laura Borton

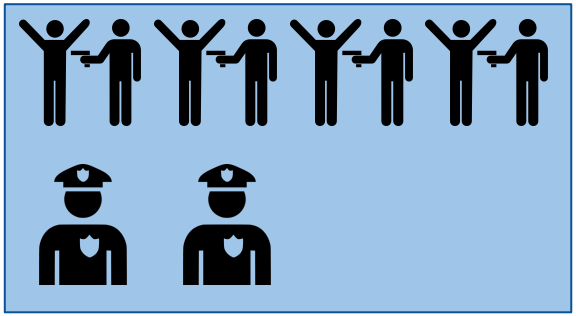
Help Law Enforcement Allocate Resources

Historical

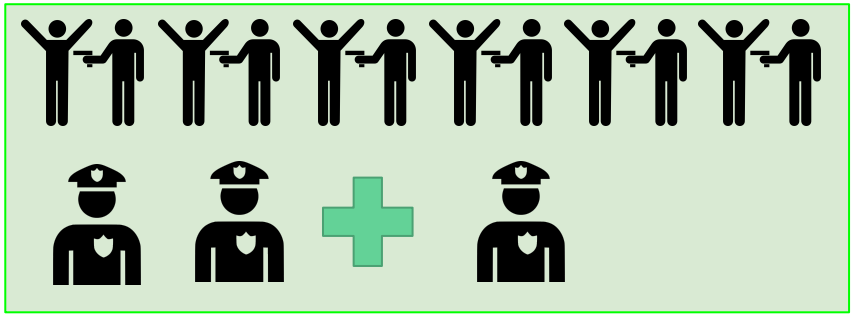


Help Law Enforcement Allocate Resources

Historical

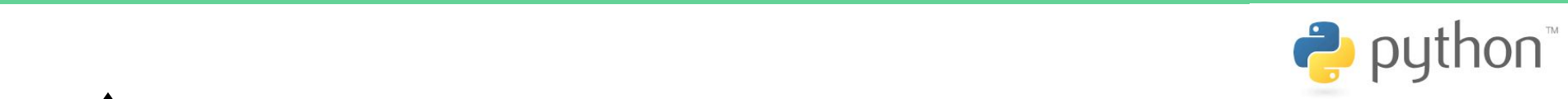


Next Week



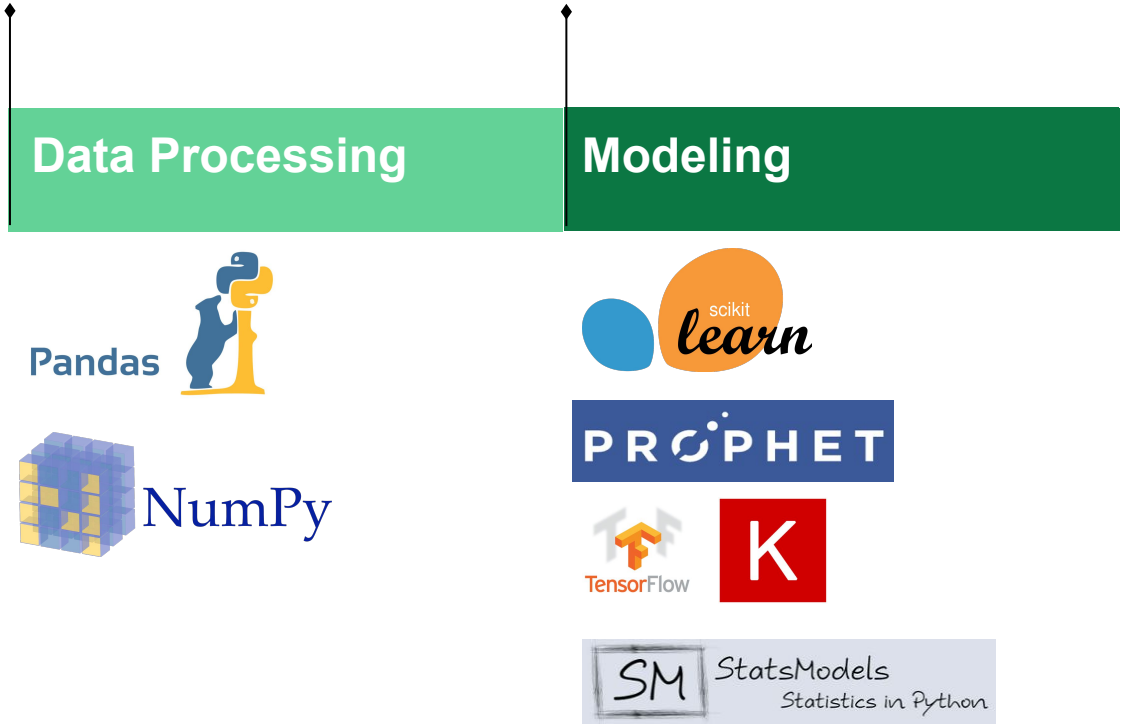
Workflow and Tools

Workflow and Tools

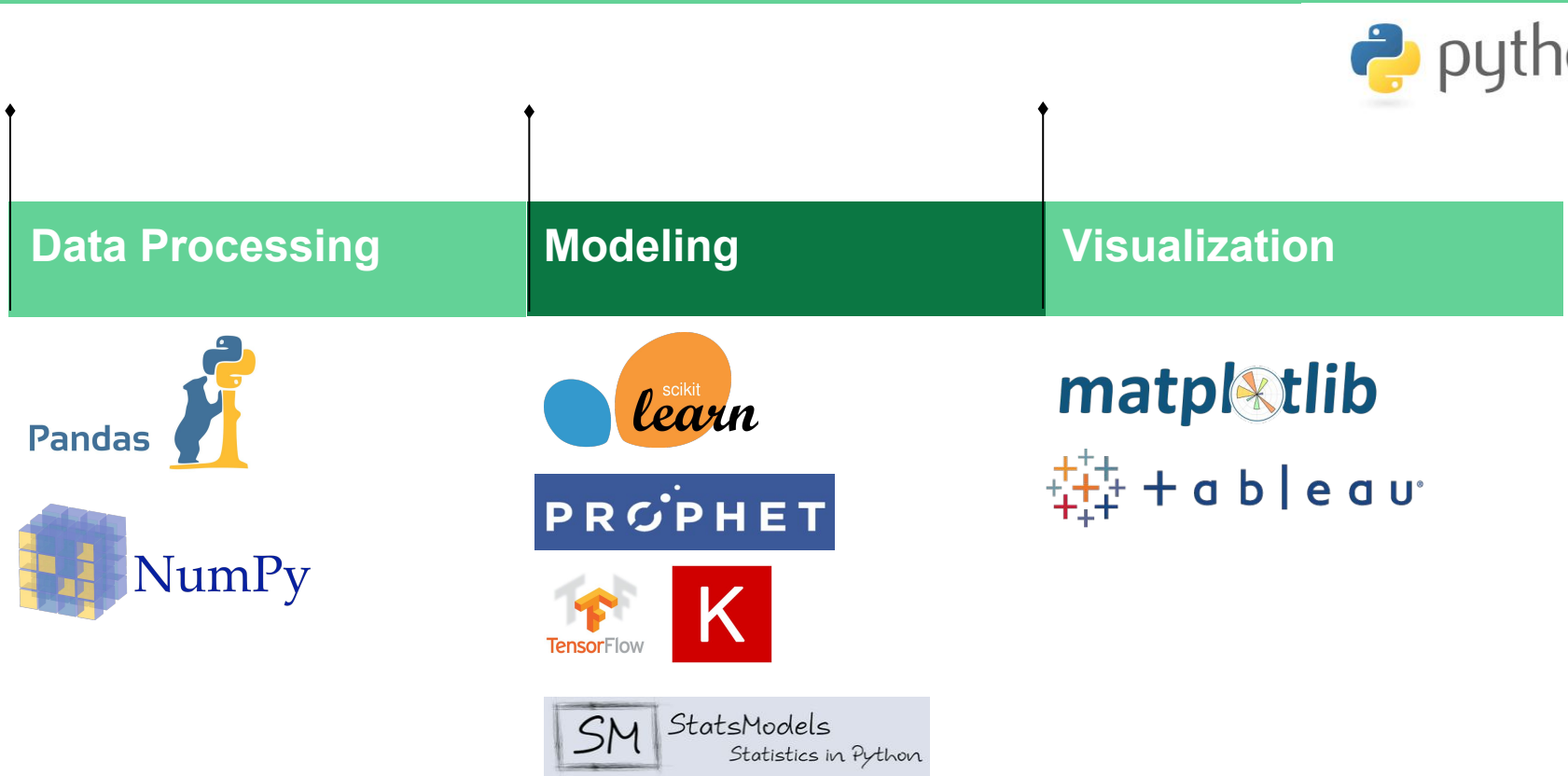


Data Processing

Workflow and Tools



Workflow and Tools



Data

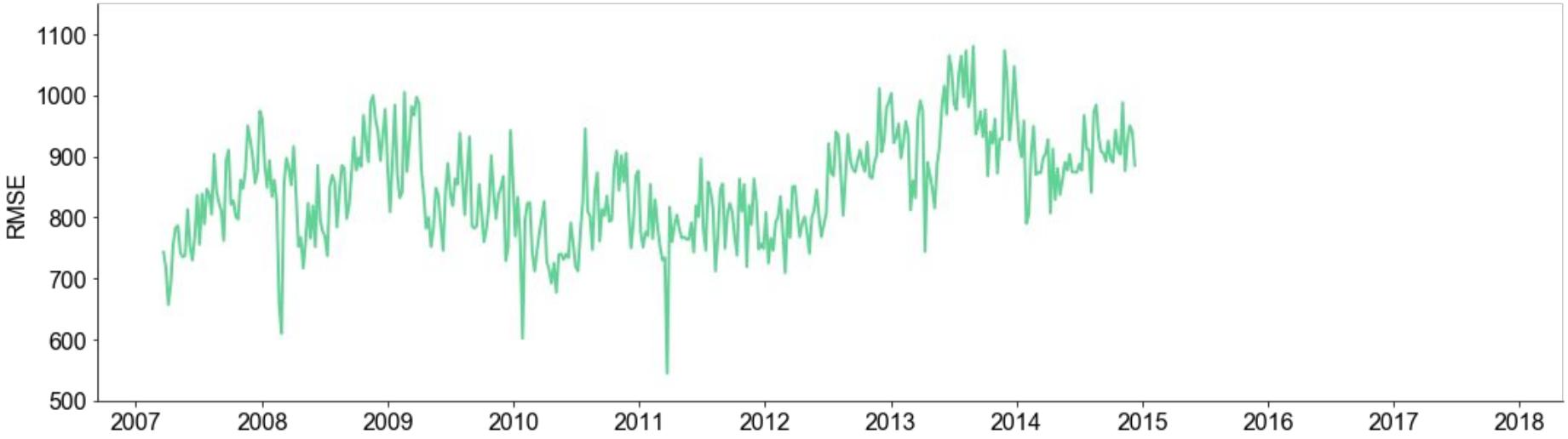
Date Range:
Jan 6, 2008 - August 26, 2018

Crimes : 488,000+

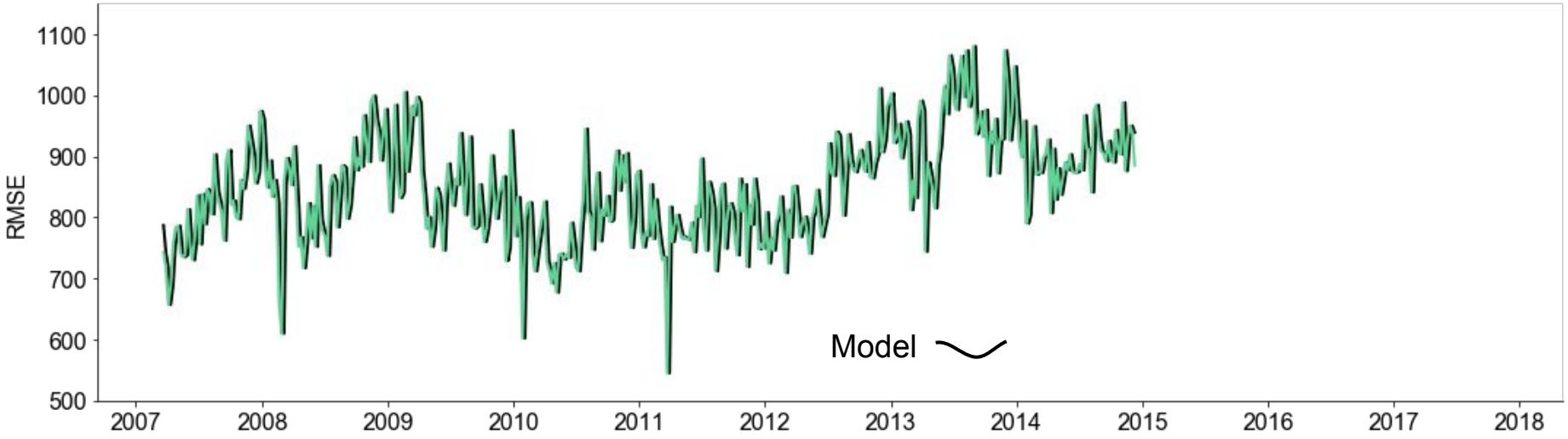
Seattle Open Data Program

	Crime Subcategory	date
2438	ROBBERY-COMMERCIAL	2008-01-13
2439	ROBBERY-STREET	2008-01-13
2440	AGGRAVATED ASSAULT	2008-01-13
2441	CAR PROWL	2008-01-13
2442	THEFT-BUILDING	2008-01-13

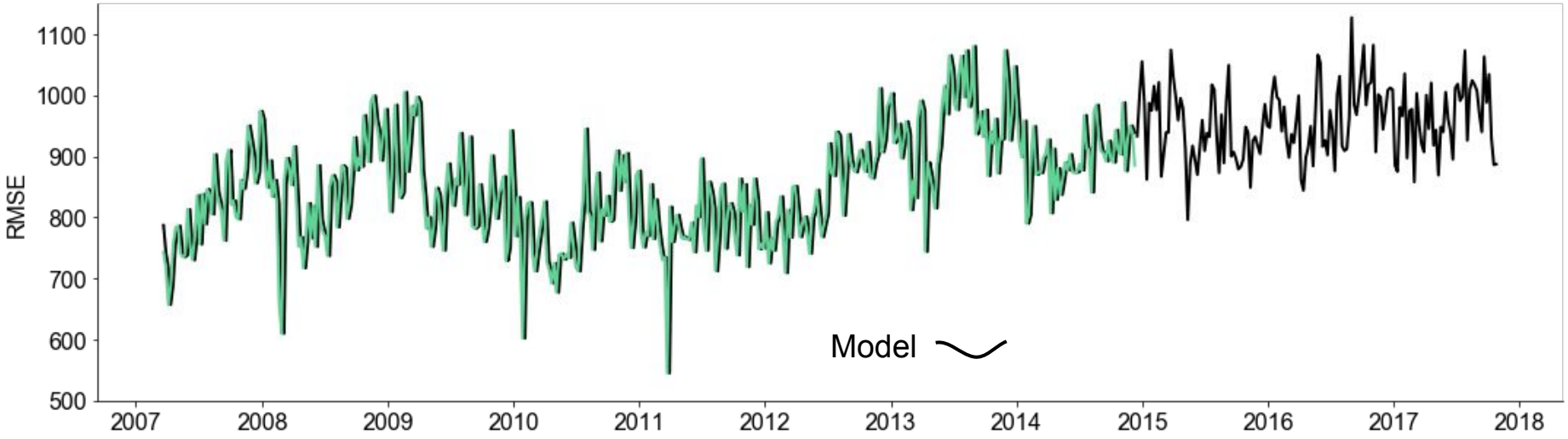
Time Series Methods: Naive



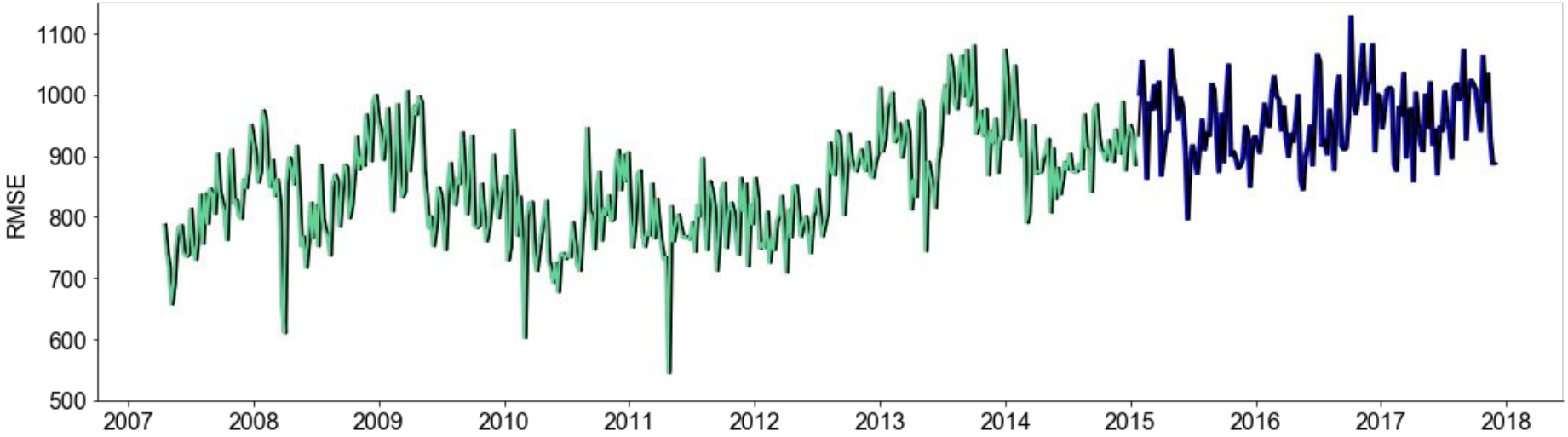
Time Series Methods: Naive



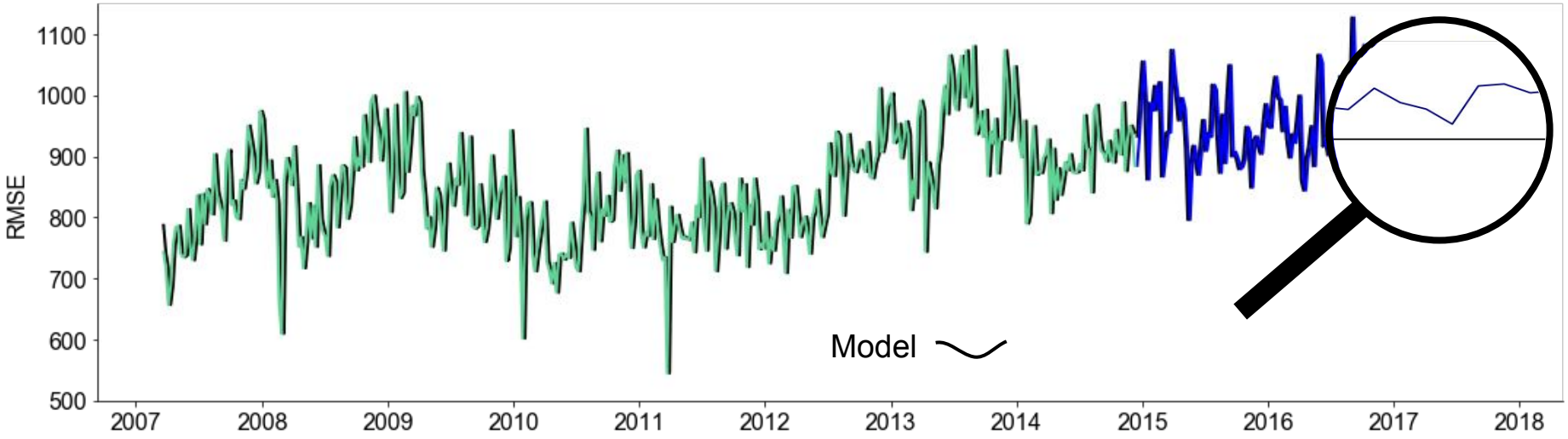
Time Series Methods: Naive



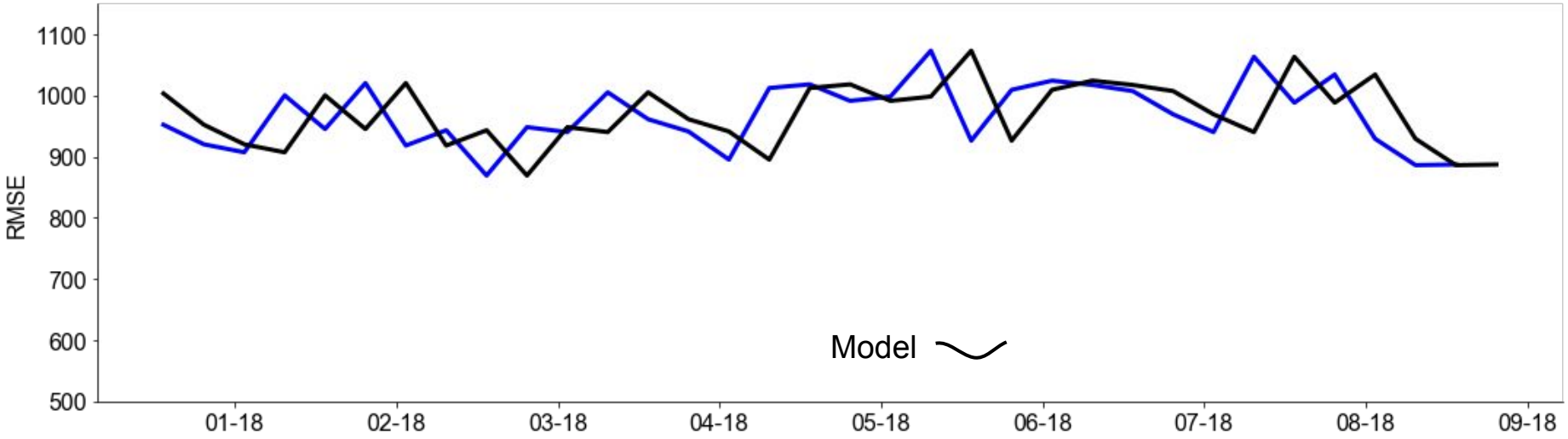
Time Series Methods: Naive

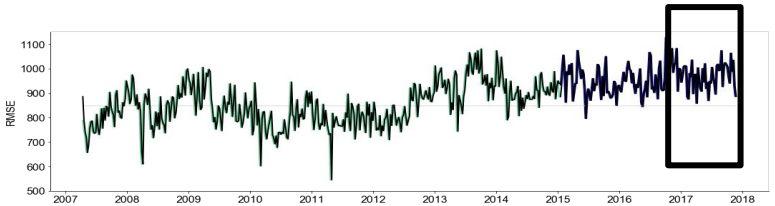


Time Series Methods: Naive



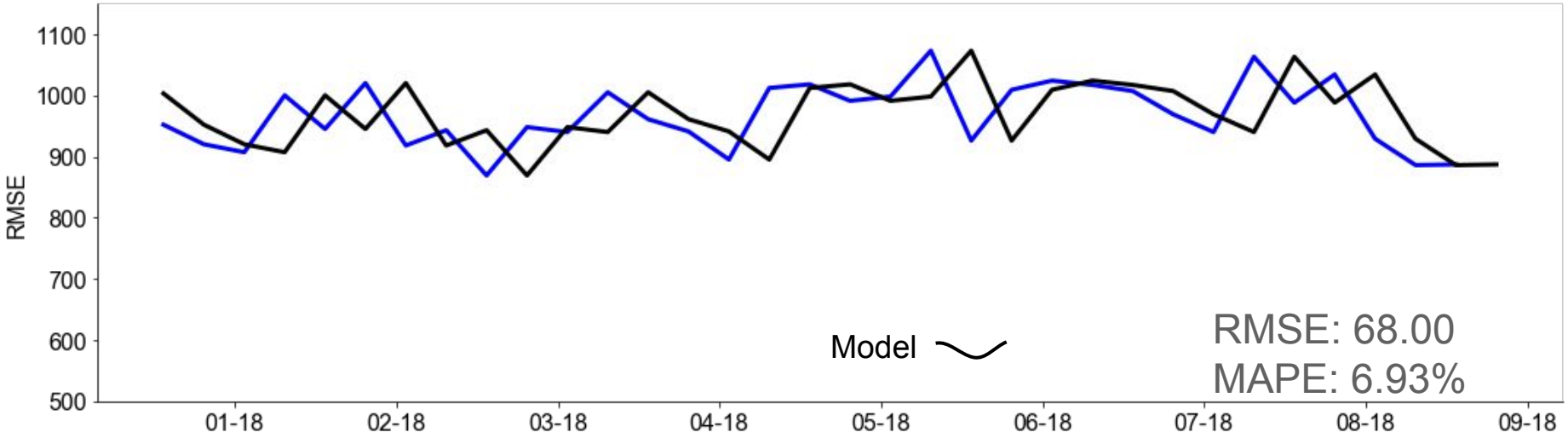
Time Series Methods: Naive

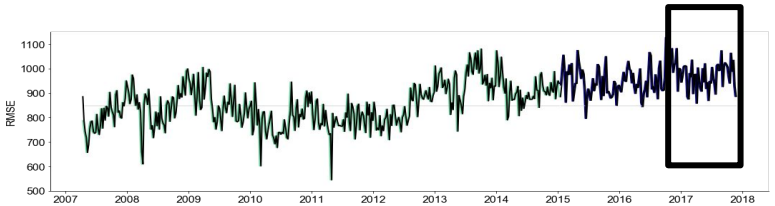




Data Processing | Modeling | Visualization

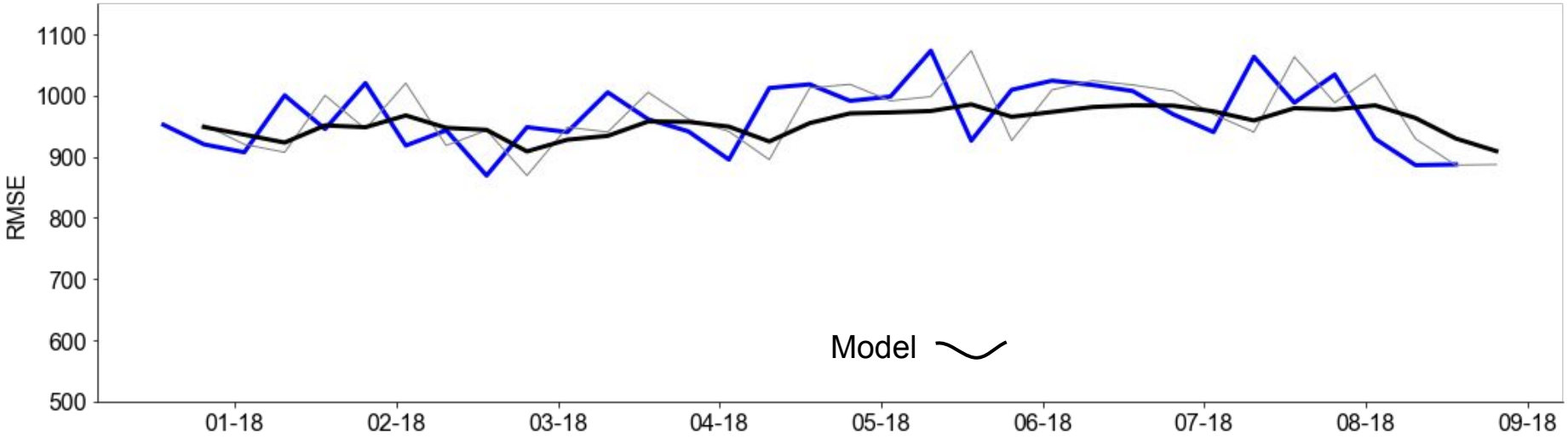
Time Series Methods: Naive

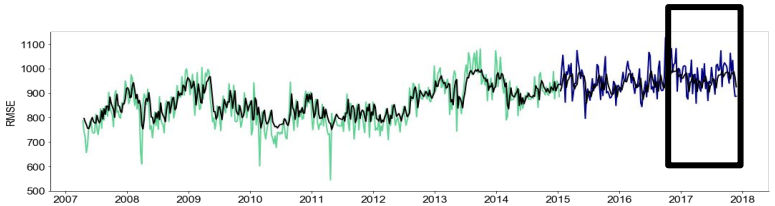




Data Processing | Modeling | Visualization

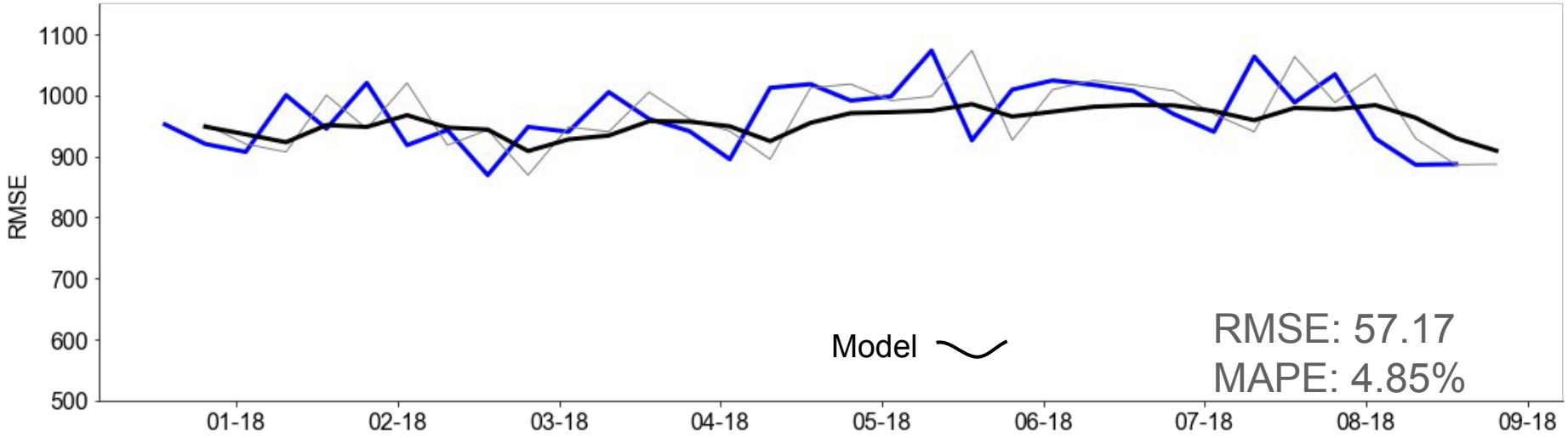
Time Series Methods: LSTM

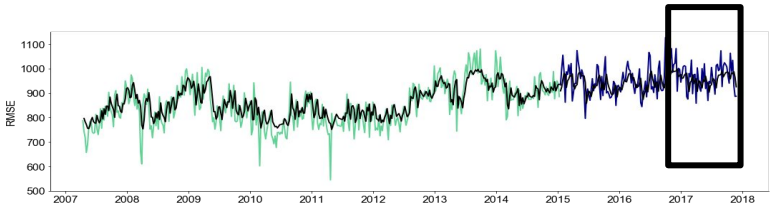




Data Processing | Modeling | Visualization

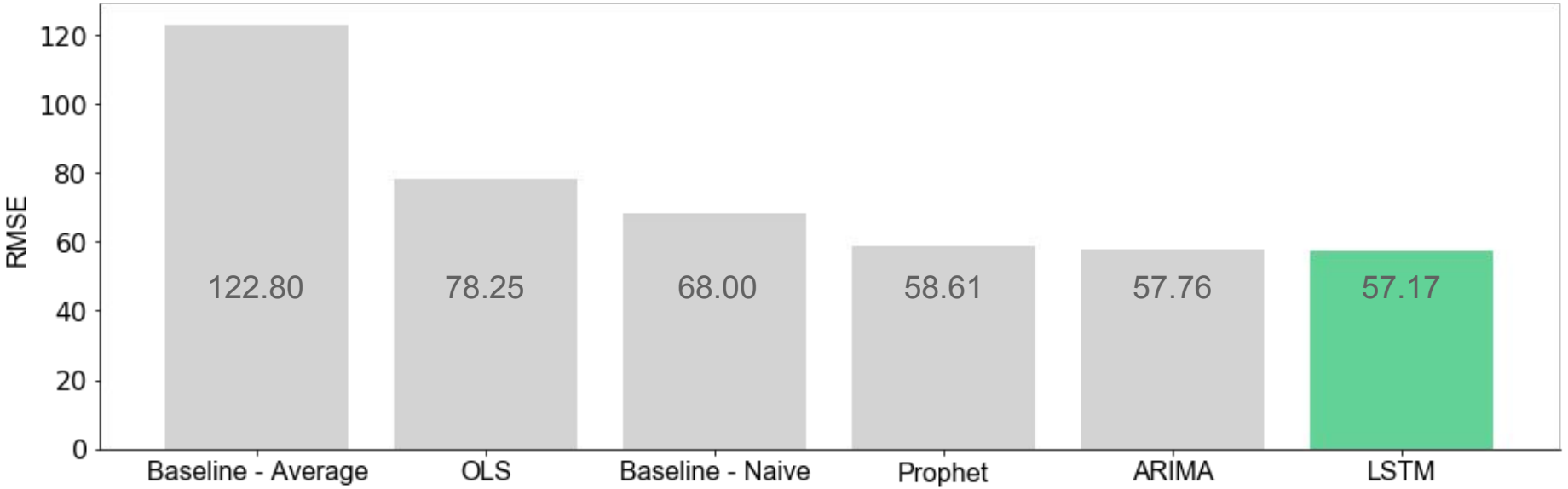
Time Series Methods: LSTM



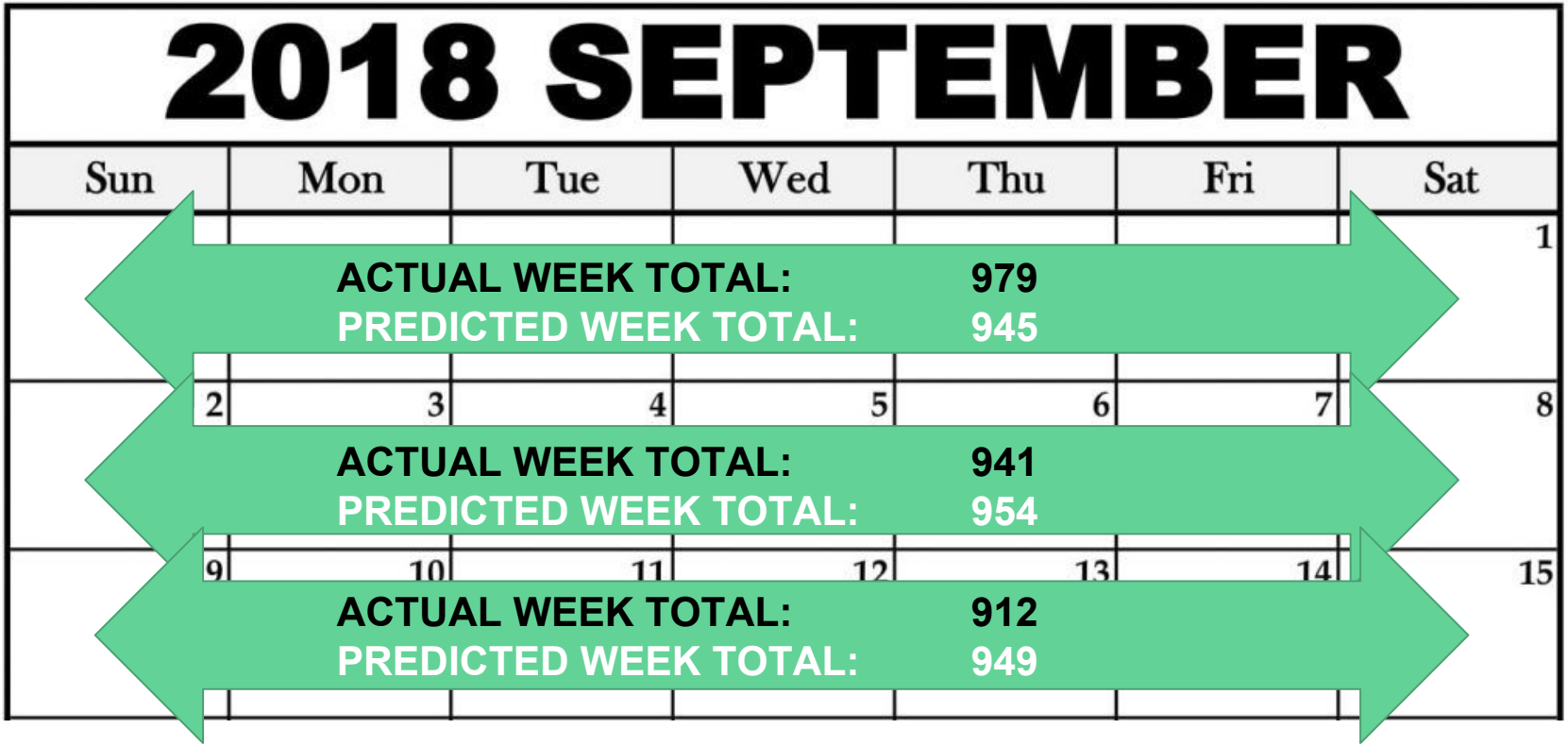


Data Processing | Modeling | Visualization

RMSE Histogram

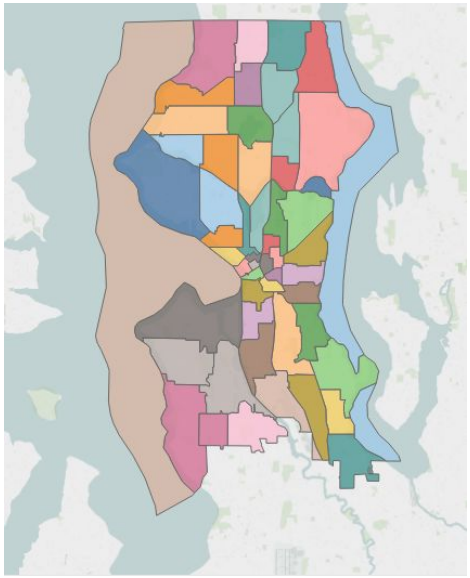


True Model Forecasts



Going forward...

Forecast by:



Location

Day

Time

Thank You and Stay Safe!


Seattle's own crime fighter Phoenix Jones

bortonlauraj@gmail.com

ljborton

laura-borton-59615a16/

OUTLIERS

Seattle Times:

Memorable Seattle Storms

Jan 18, 2012

Nov. 22, 2010

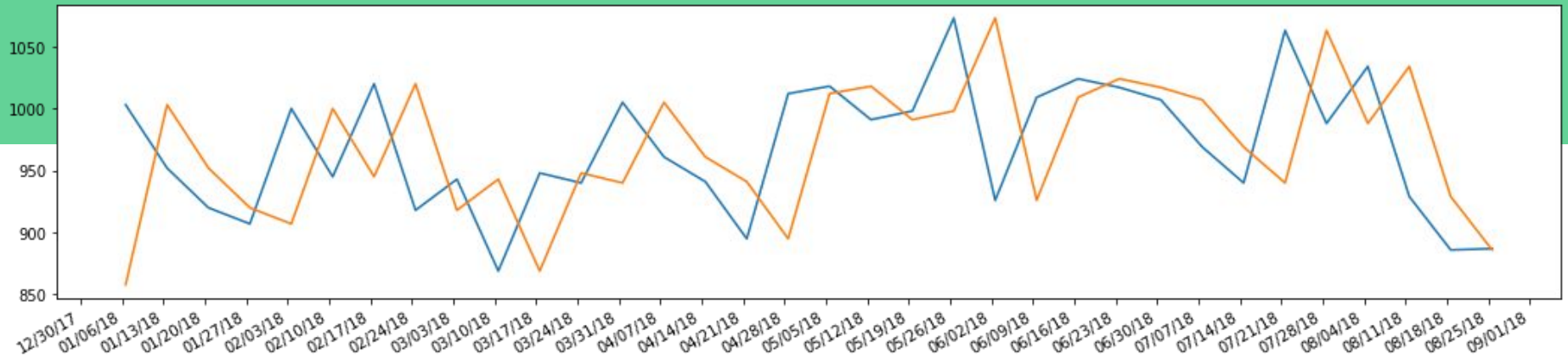
Dec. 19, 2008 (two weeks of bad weather)

Mean_absolute_percentage_error: Naive model = 6.92

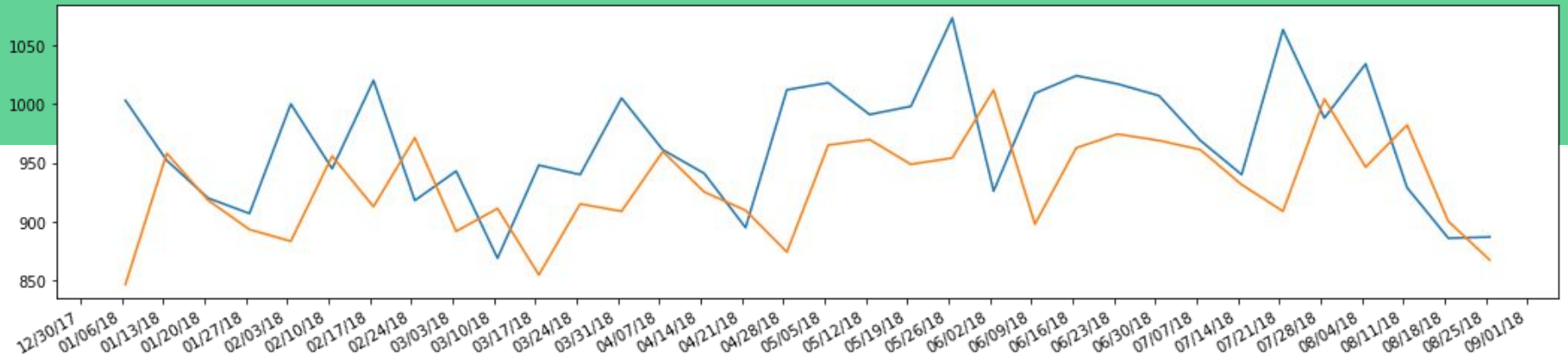
MODEL	RMSE
Baseline - Average	122.798
Baseline - Naive	68.003
Lag = 1 Regression	69.487
Lag = 3 Regression	63.483
Lag = 3 Time Step Regression (lag list)	65.191
Memory (stateful and no shuffle)	59.827
Stacked LSTM (sequence)	59.665
Stacked LSTM (sequence) - RMSProp	60.739
Stacked LSTM (sequence) - 1 lag	57.960
Stacked LSTM (sequence) - 1 lag, 150 epochs	**57.172**
Memory (stateful and no shuffle) - 1 lag	57.279
Memory (stateful and no shuffle) - 1 lag, 150 epochs	57.272
Multivariate	58.661

MODEL	RMSE
Baseline - Average	122.798
Baseline - Naive	68.003
Lag = 1 Regression Difference	64.501
Lag = 3 Regression Difference	59.859
Lag = 3 Time Step Regression (lag list) Difference	60.382
Memory (stateful and no shuffle) - 3lag Difference	**59.832**
Memory (stateful and no shuffle) - 1 lag Difference	63.796
Stacked LSTM (sequence) - Difference	61.887
Stacked LSTM (sequence) - 1 lag, Difference	66.091
Multivariate	**56.830**

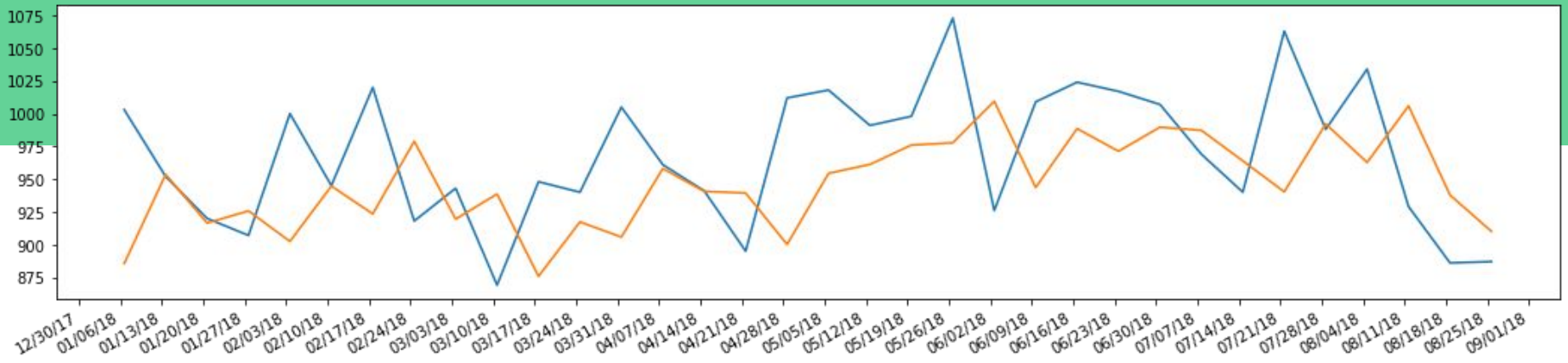
MODEL	RMSE
Baseline - Average	122.798
Baseline - Naive	68.003
Prophet	**58.611**
Prophet with temp	61.199
OLS with variables	**70?**
OLS - no variables	**65.998**
ARIMA	**57.755**


1050
1000
950
900
850
12/30/17
01/06/18
01/13/18
01/20/18
01/27/18
02/03/18
02/10/18
02/17/18
02/24/18
03/03/18
03/10/18
03/17/18
03/24/18
03/31/18
04/07/18
04/14/18
04/21/18
04/28/18
05/05/18
05/12/18
05/19/18
05/26/18
06/02/18
06/09/18
06/16/18
06/23/18
06/30/18
07/07/18
07/14/18
07/21/18
07/28/18
08/04/18
08/11/18
08/18/18
08/25/18
09/01/18

Naive

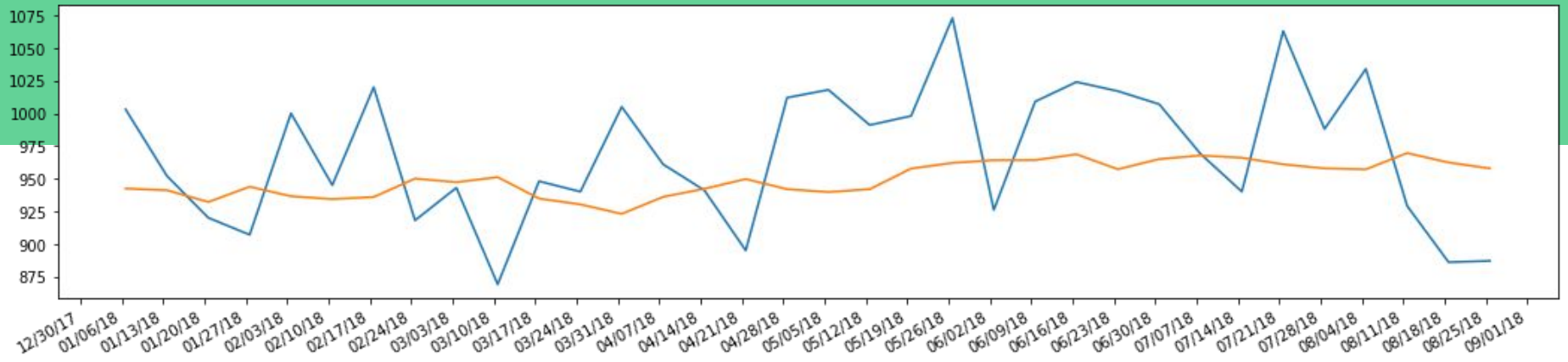

1050
1000
950
900
850
12/30/17
01/06/18
01/13/18
01/20/18
01/27/18
02/03/18
02/10/18
02/17/18
02/24/18
03/03/18
03/10/18
03/17/18
03/24/18
03/31/18
04/07/18
04/14/18
04/21/18
04/28/18
05/05/18
05/12/18
05/19/18
05/26/18
06/02/18
06/09/18
06/16/18
06/23/18
06/30/18
07/07/18
07/14/18
07/21/18
07/28/18
08/04/18
08/11/18
08/18/18
08/25/18
09/01/18

Lag = 1


1075
1050
1025
1000
975
950
925
900
875
12/30/17
01/06/18
01/13/18
01/20/18
01/27/18
02/03/18
02/10/18
02/17/18
02/24/18
03/03/18
03/10/18
03/17/18
03/24/18
03/31/18
04/07/18
04/14/18
04/21/18
04/28/18
05/05/18
05/12/18
05/19/18
05/26/18
06/02/18
06/09/18
06/16/18
06/23/18
06/30/18
07/07/18
07/14/18
07/21/18
07/28/18
08/04/18
08/11/18
08/18/18
08/25/18
09/01/18

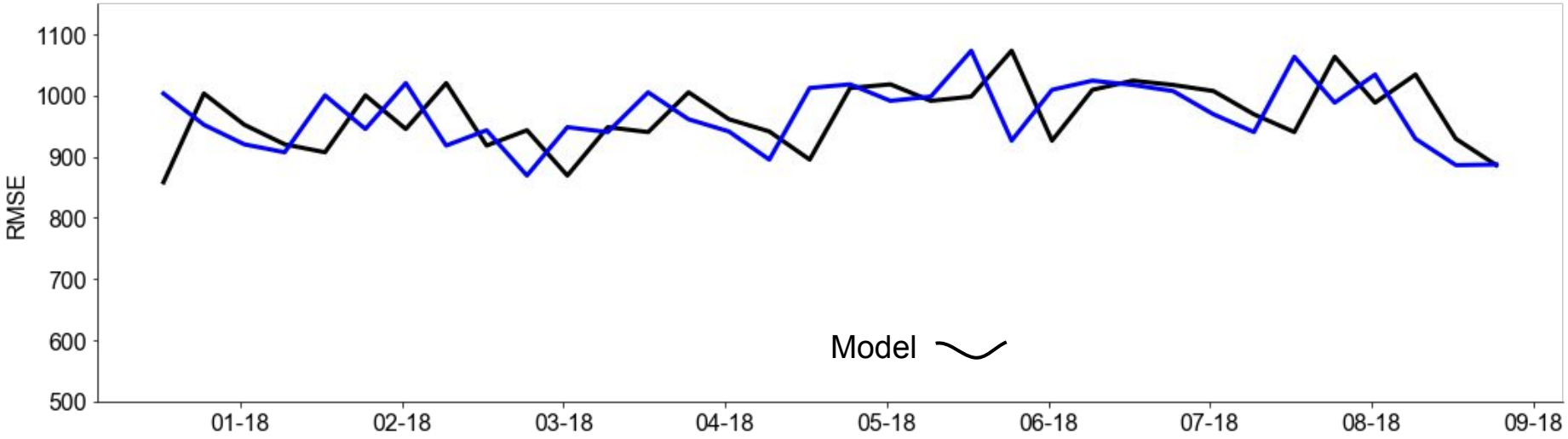
Lag = 3

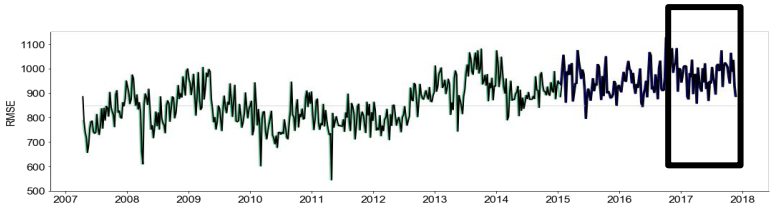
Lag = 3 timestep



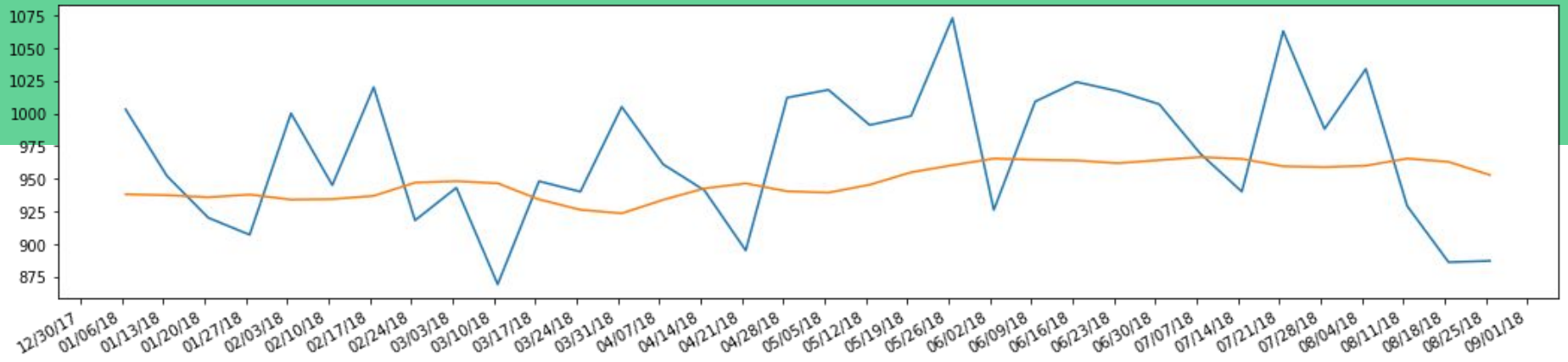
Lag = 3 timestep with memory

Time Series Methods: Naive

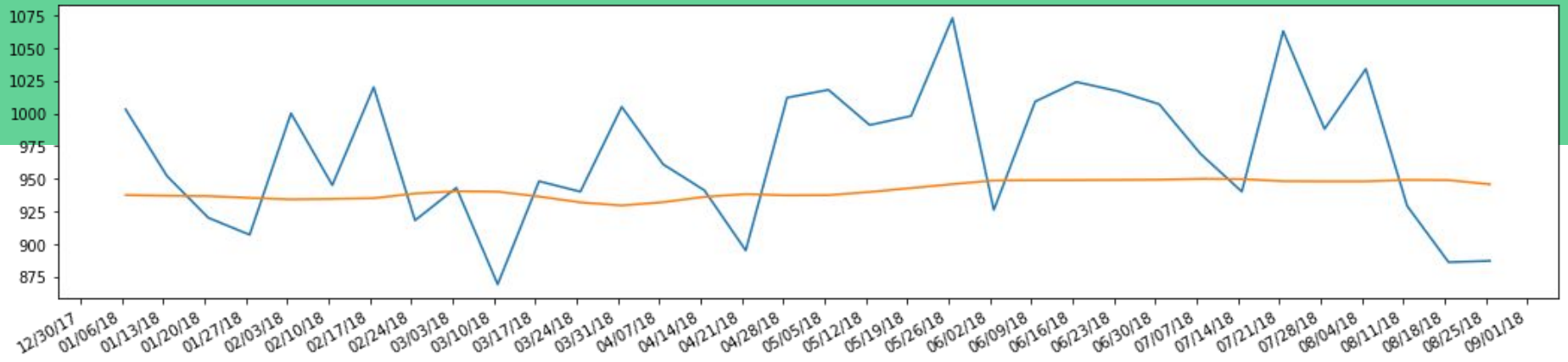




Data Processing | Modeling | Visualization

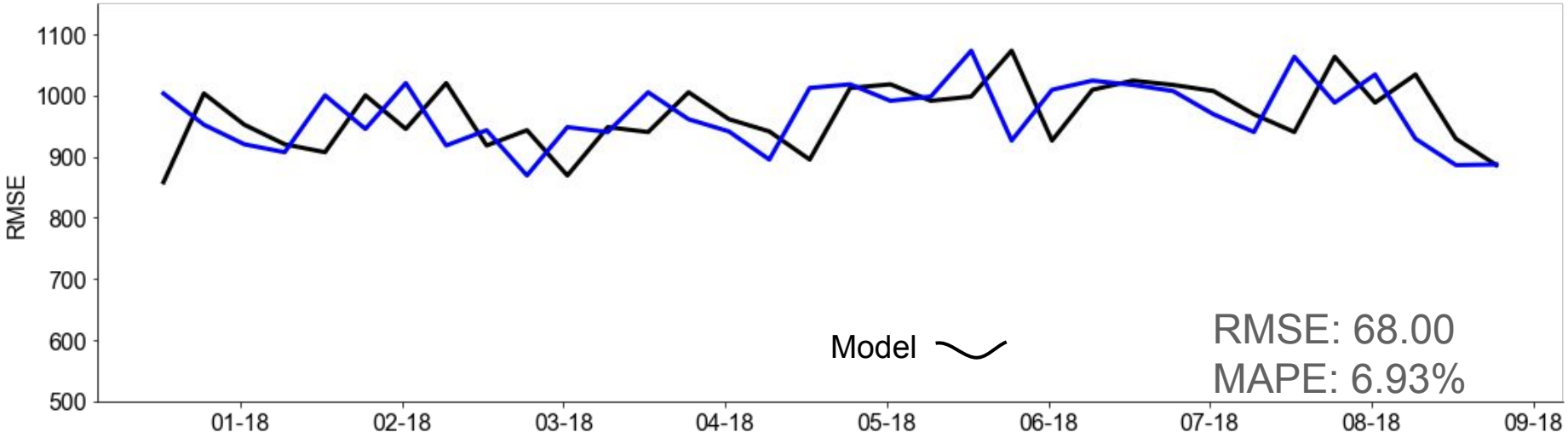


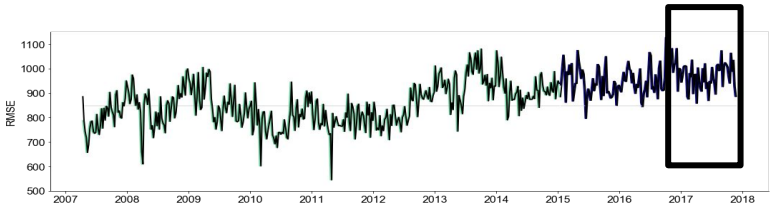
Lag = 3 timestep with stacked memory



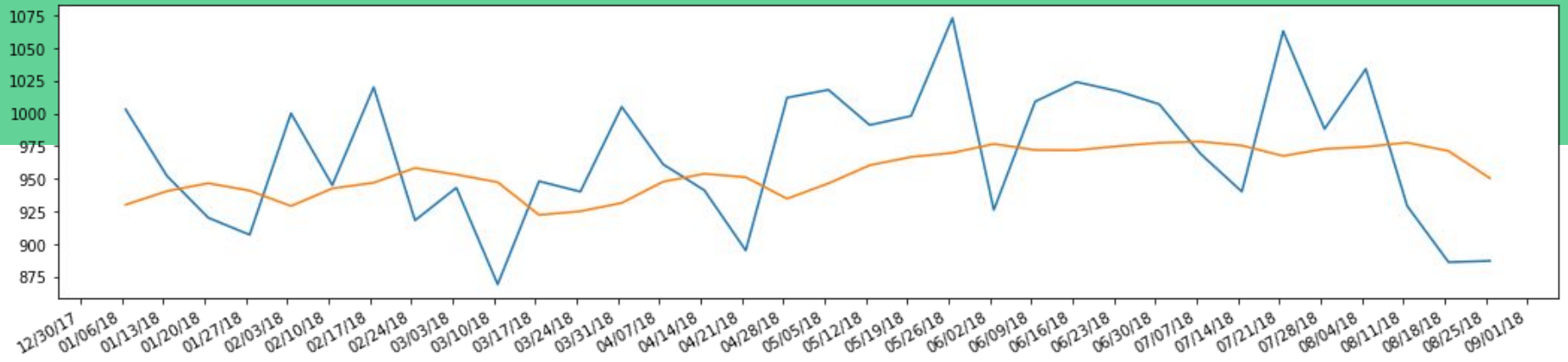
Lag = 3 timestep with stacked memory, RMSProp

Time Series Methods: Naive

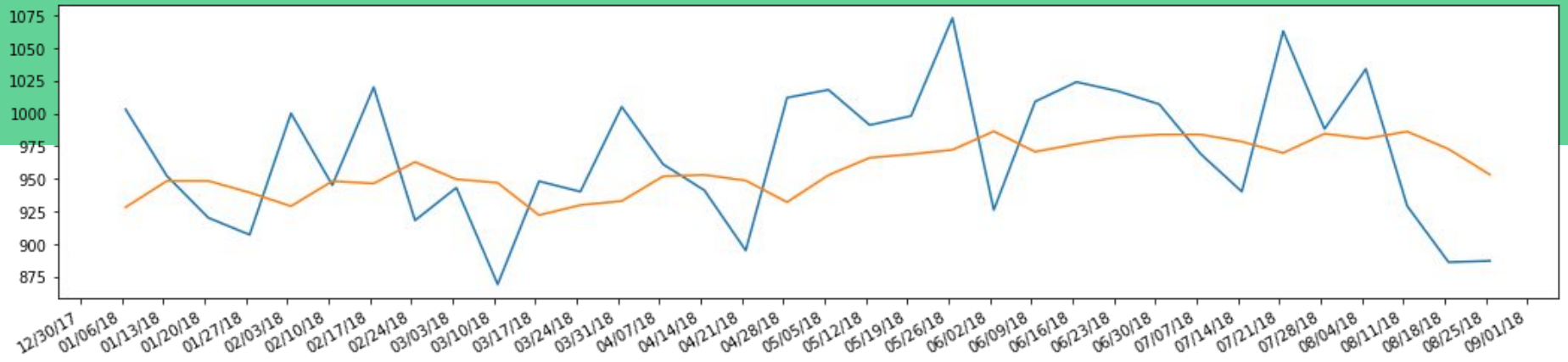




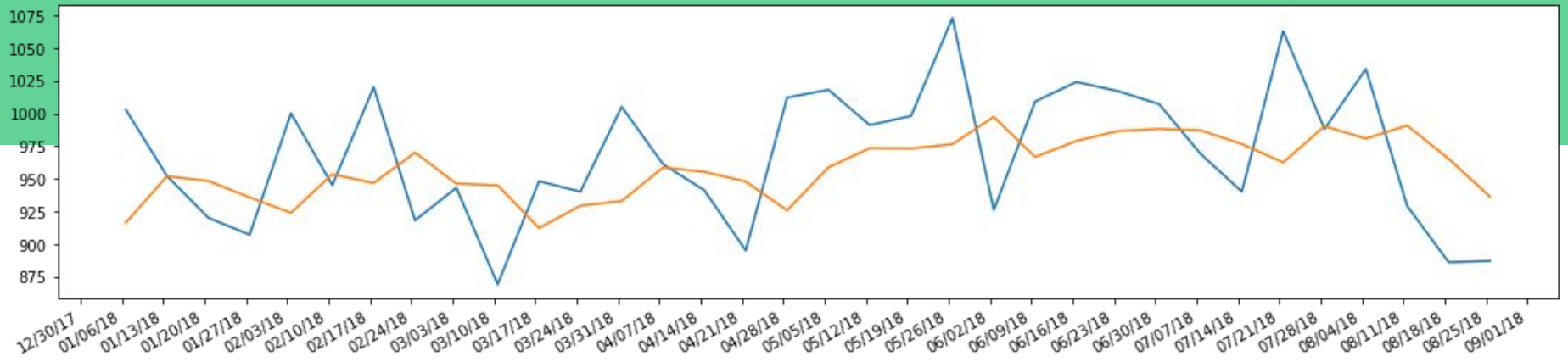
Data Processing | Modeling | Visualization



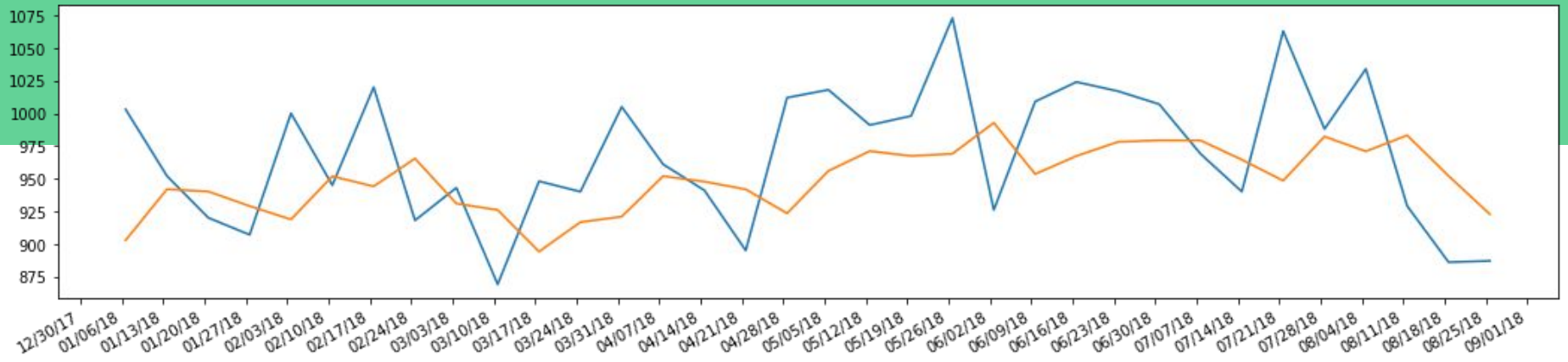
Lag = 1 timestep with stacked memory



Lag = 1 timestep with memory

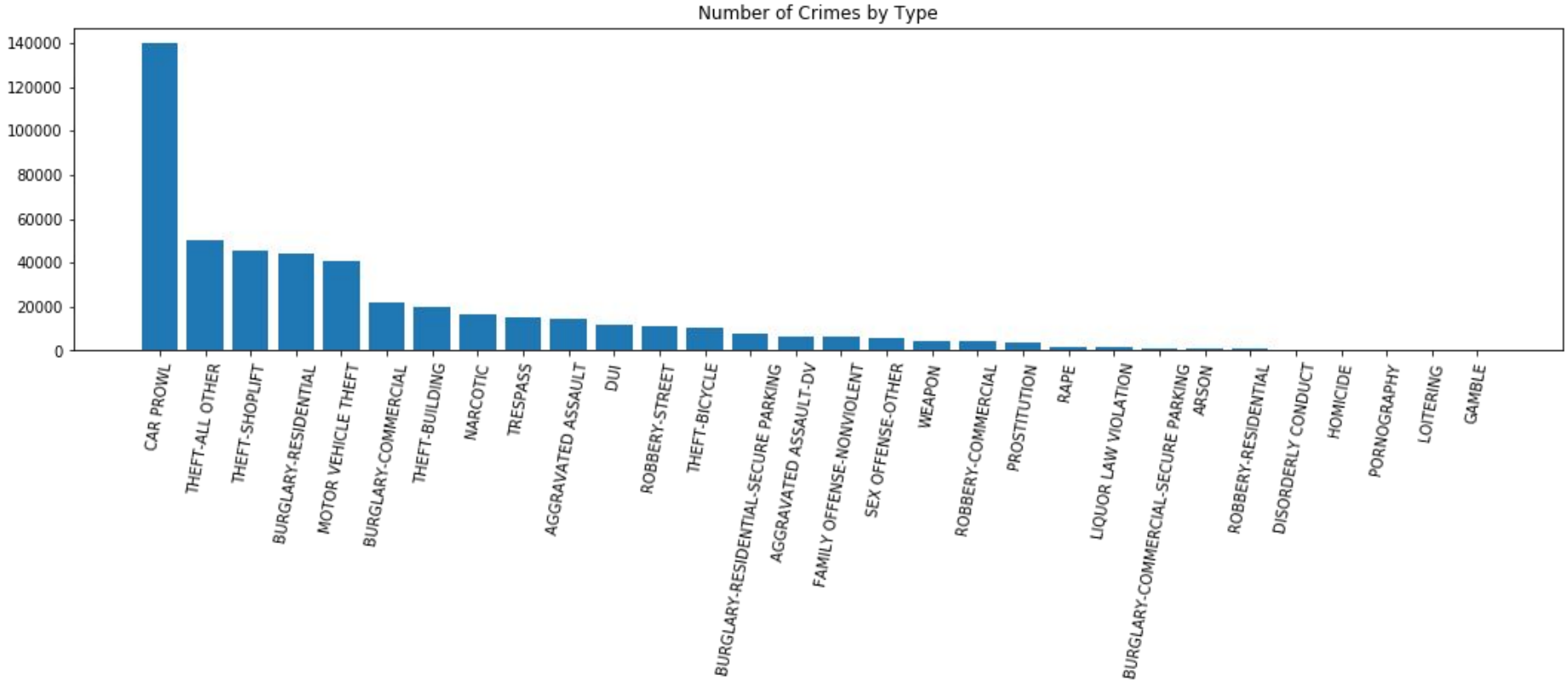


Lag = 1 timestep with memory, 150 epochs

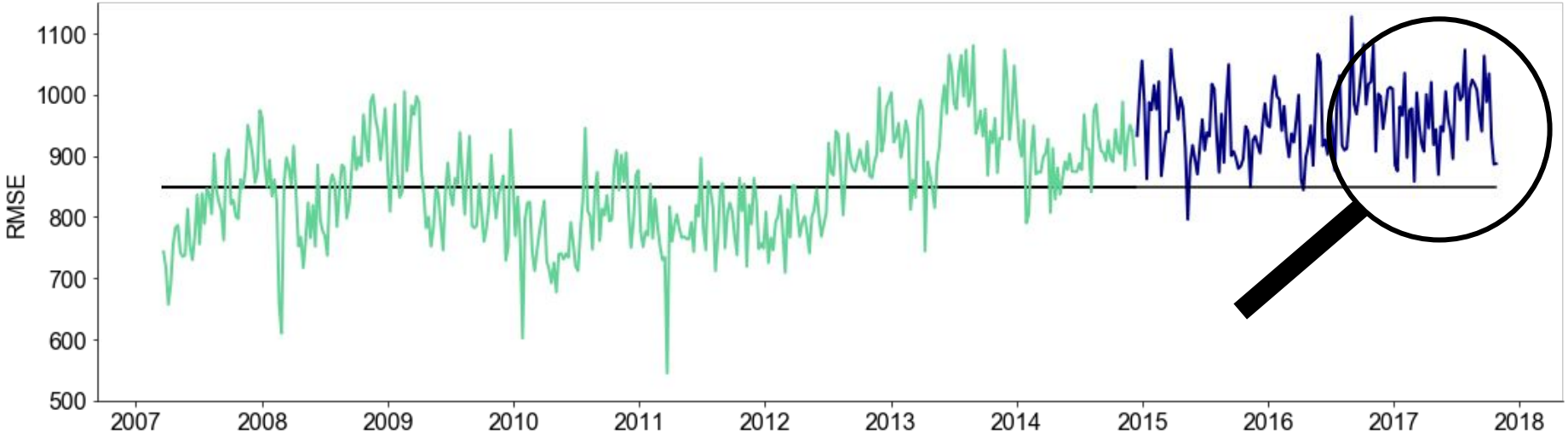


Lag = 1 timestep with memory, Multivariate (precip, temp, time)

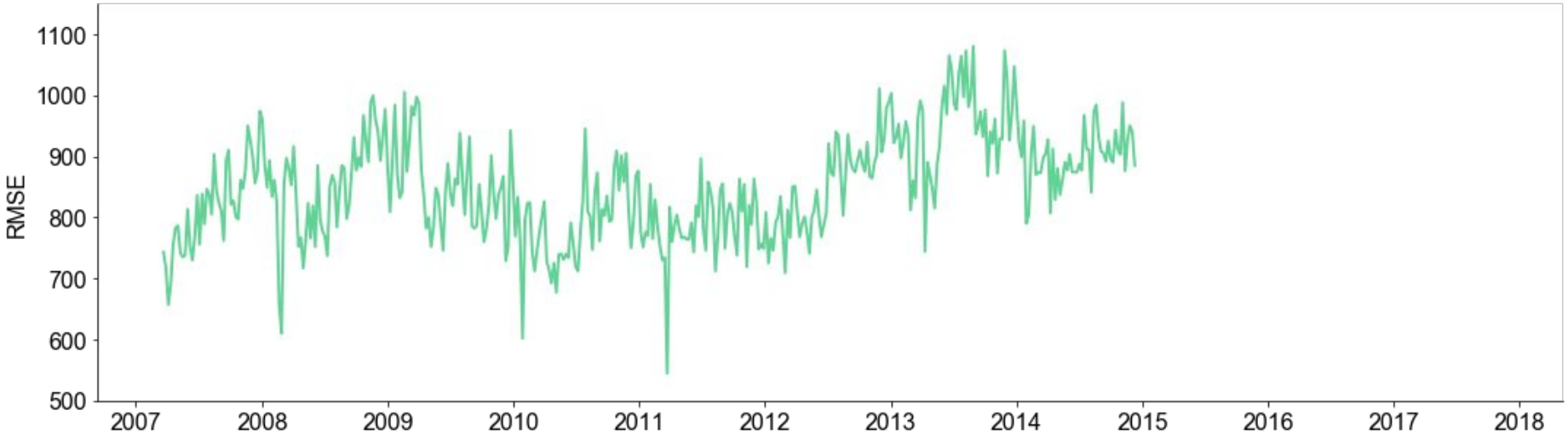
Total Number of Crimes: Jan 2008 - August 2018



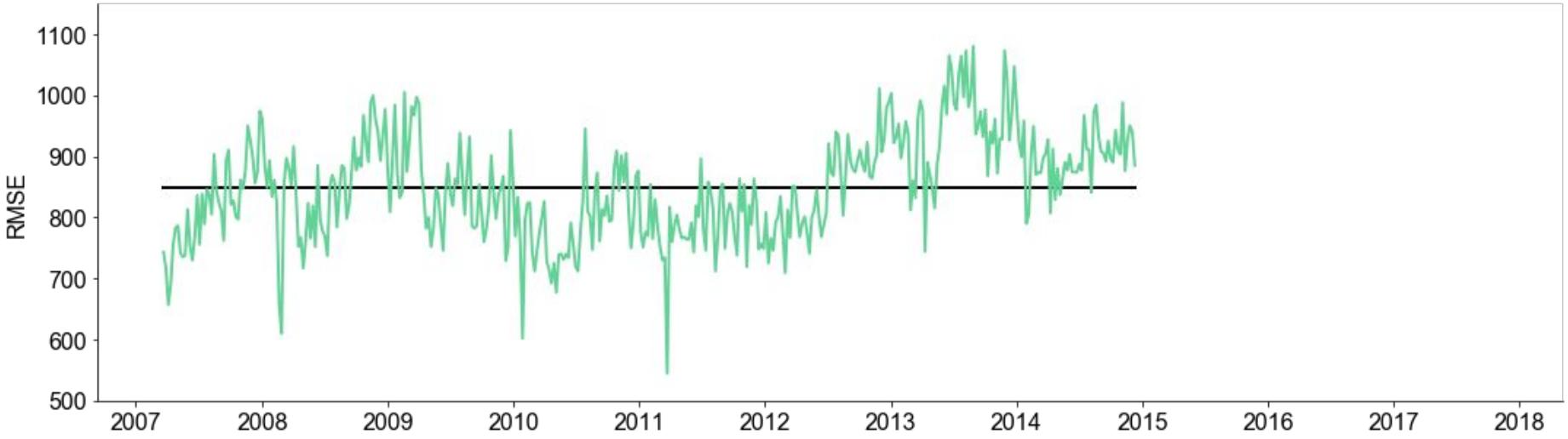
Time Series Methods:



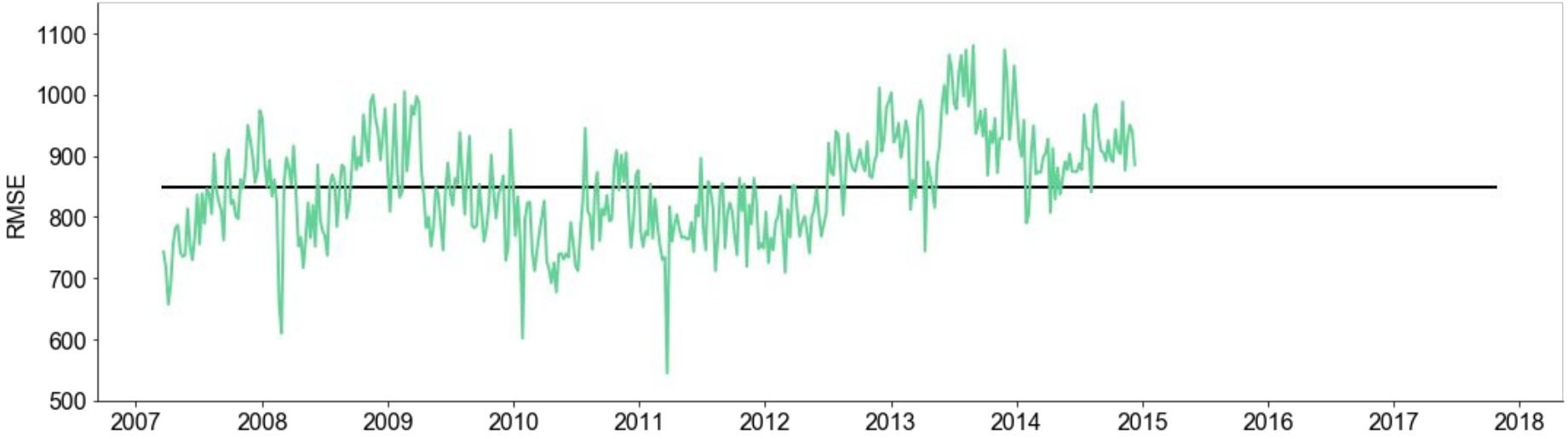
Time Series Methods: Average



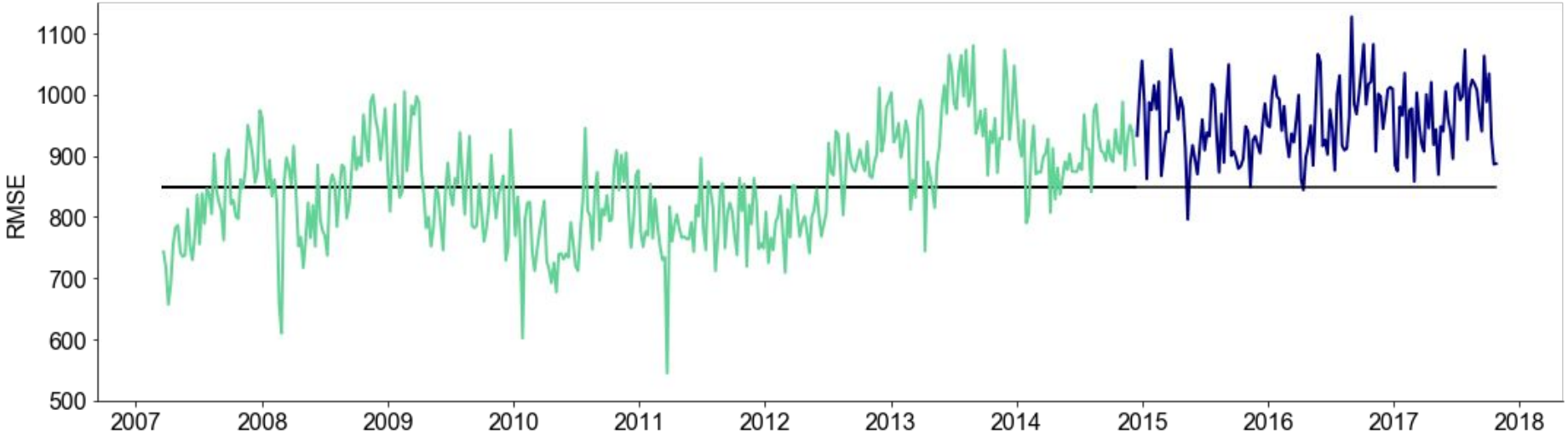
Time Series Methods: Average



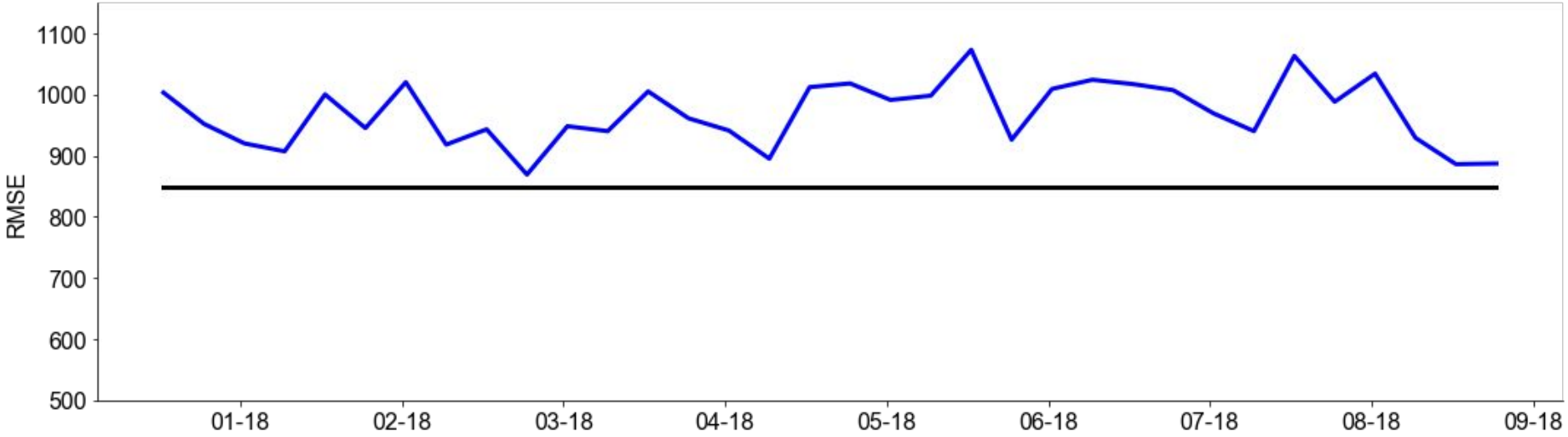
Time Series Methods: Average

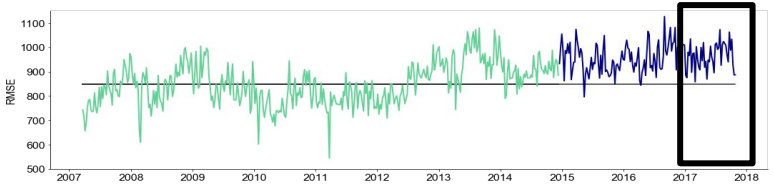


Time Series Methods: Average

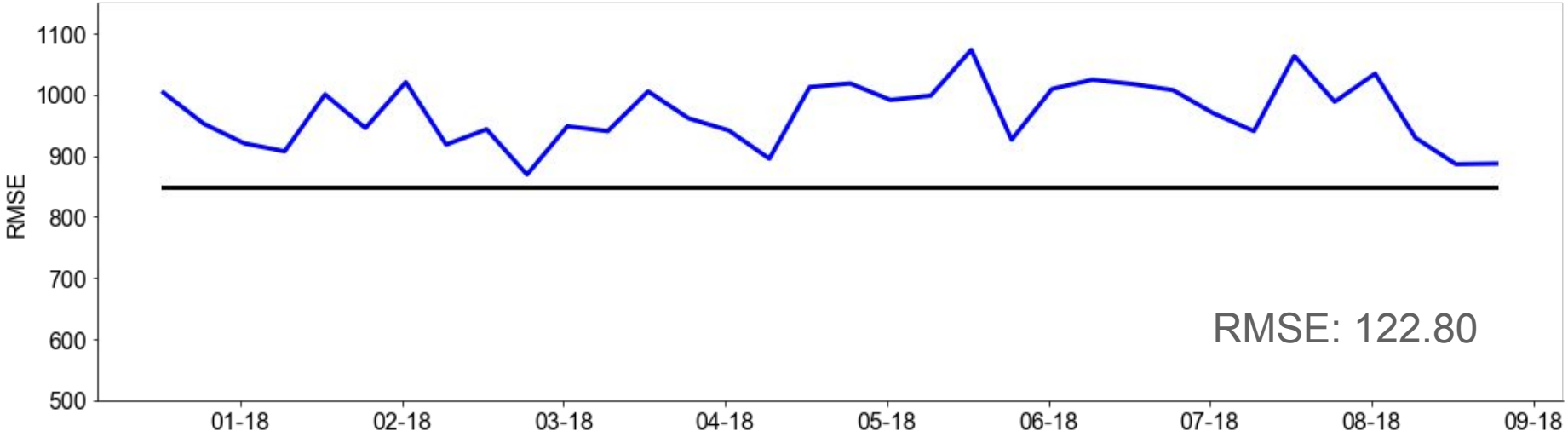


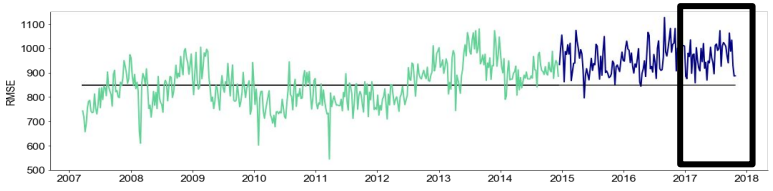
Time Series Methods: Average



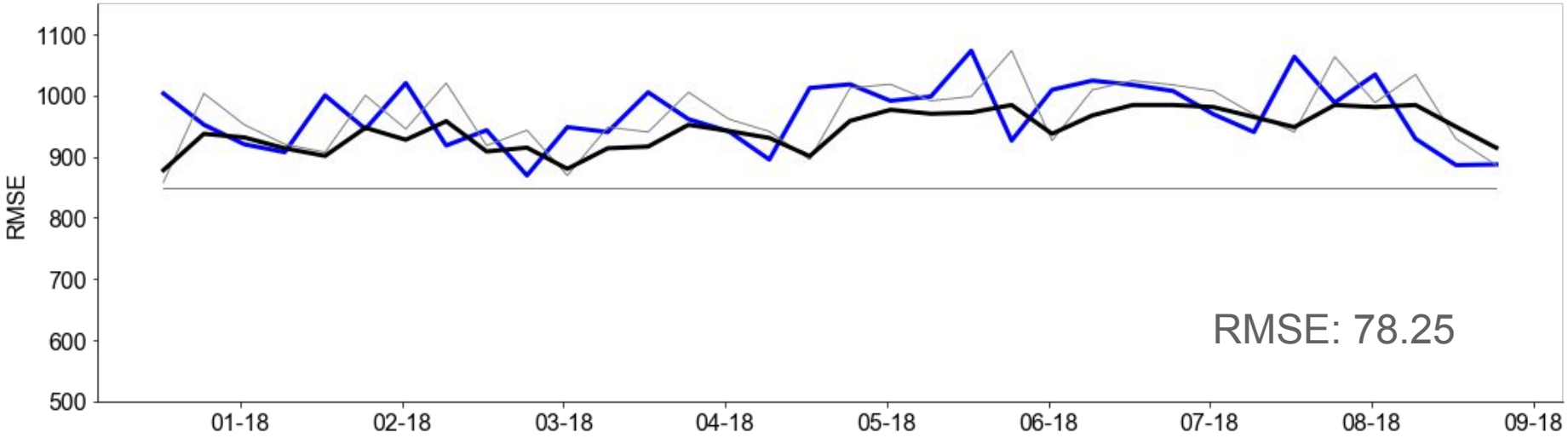


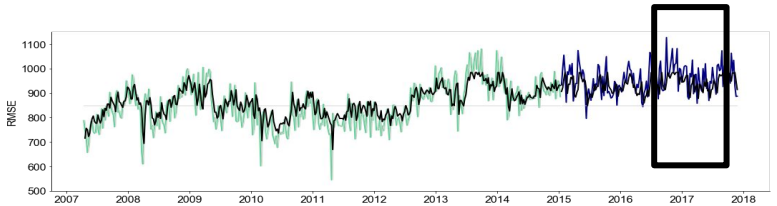
Time Series Methods: Average



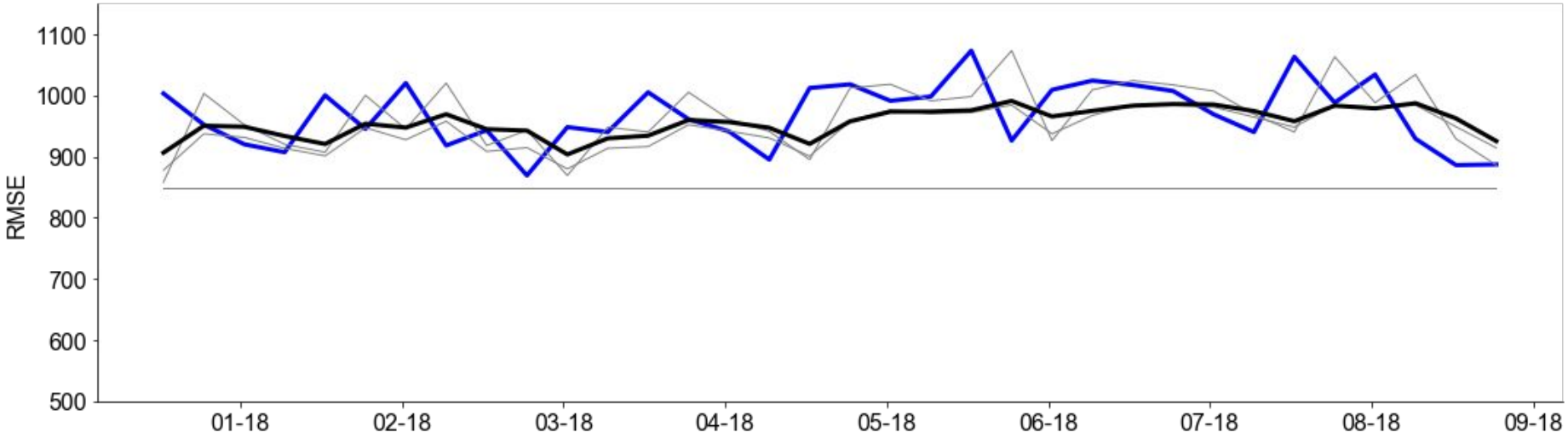


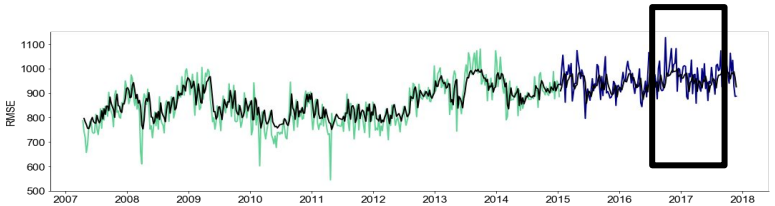
Time Series Methods: OLS



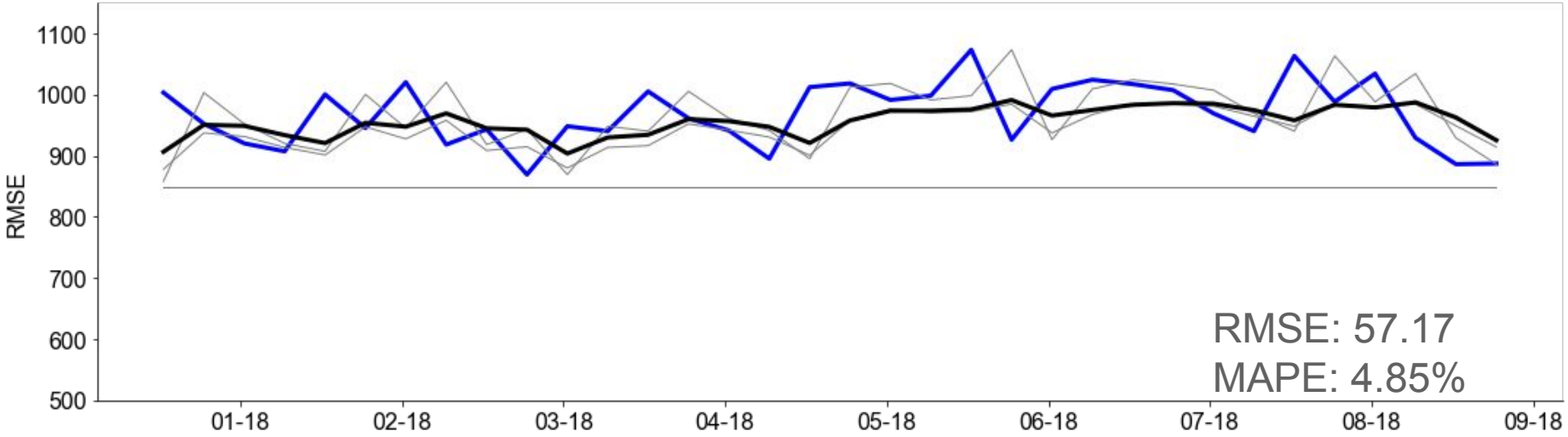


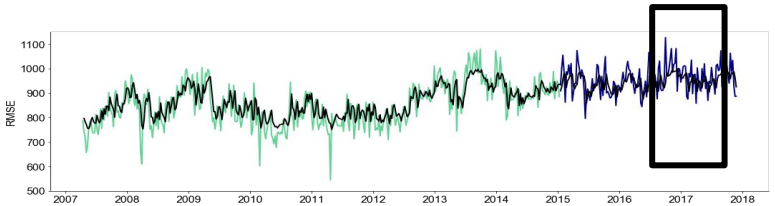
Time Series Methods: LSTM





Time Series Methods: LSTM

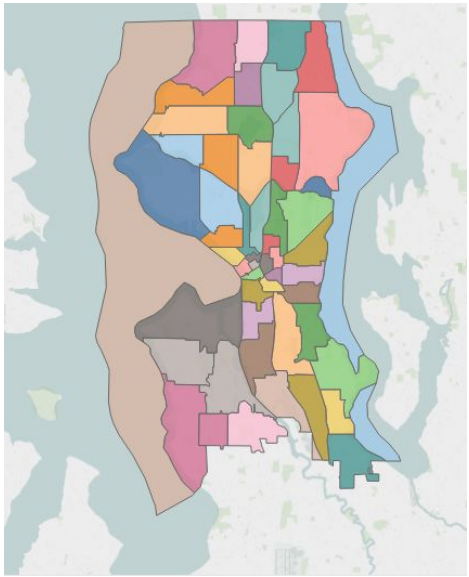




Data Processing | Modeling | Visualization

Going forward...

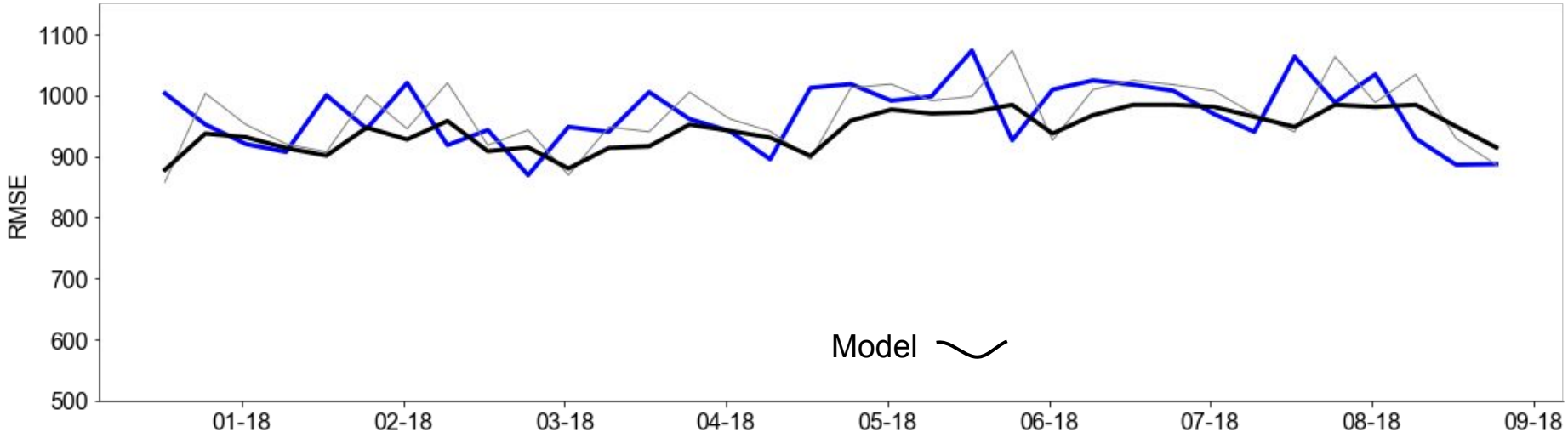
Forecast by:

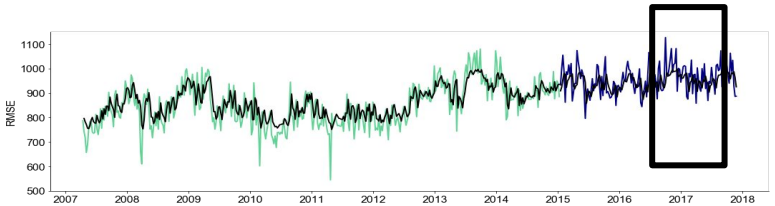


Location

Time

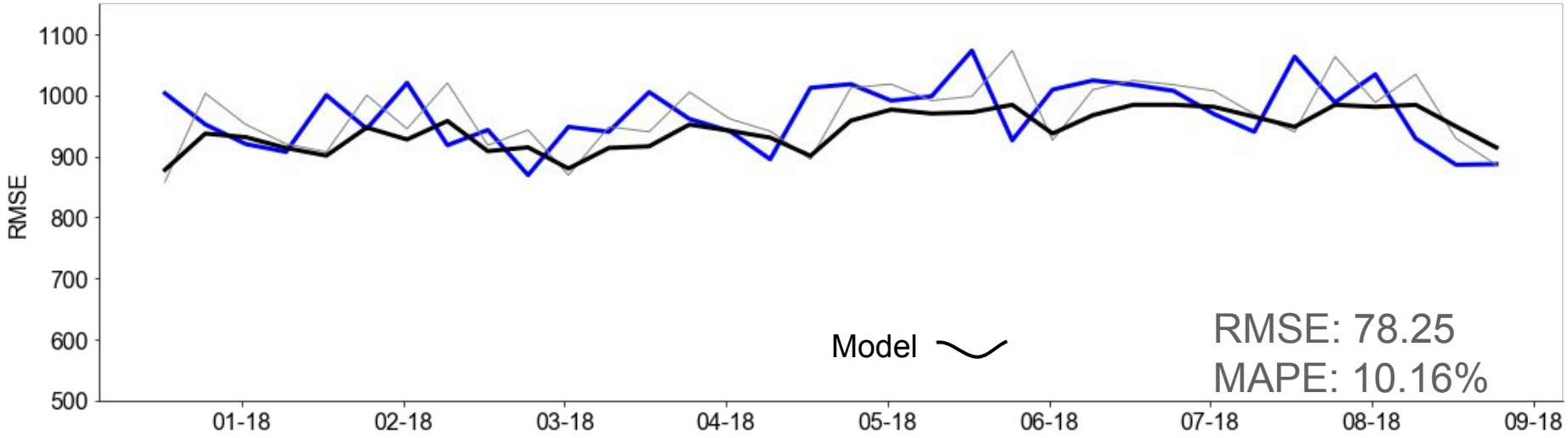
Time Series Methods: OLS

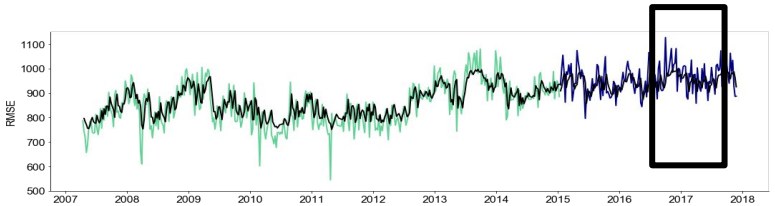




Data Processing | Modeling | Visualization

Time Series Methods: OLS





Data Processing | Modeling | Visualization