Metro Traffic Volume in Minneapolis

Time Series Analysis Final Project

The Dataset

- UC Irvine Machine Learning Repository
- Documents hourly interstate Westbound traffic volume for station 301
- Monitors weather and holiday features to observe volume impact

Feature	Data type	Description
holiday	categorical	US National holidays plus regional holiday, Minnesota State Fair
temp	numeric	Average temp in kelvin
rain_1h	numeric	Amount in mm of rain that occurred in the hour
snow_1h	numeric	Amount in mm of snow that occurred in the hour
clouds_all	numeric	Percentage of cloud cover
weather_main	categorical	Short textual description of the current weather
weather_description	categorical	Longer textual description of the current weather
date_time	datetime	Hour of the data collected in local CST time
traffic_volume	numeric	Hourly I-94 ATR 301 reported westbound traffic volume

Volume estimates are a major indicator of traffic flow

1. Determination of necessary tax revenues depending on future need

2. Design of new development within the metro system and local roadway systems, ie location, number and type of lanes, pavement thickness, etc.

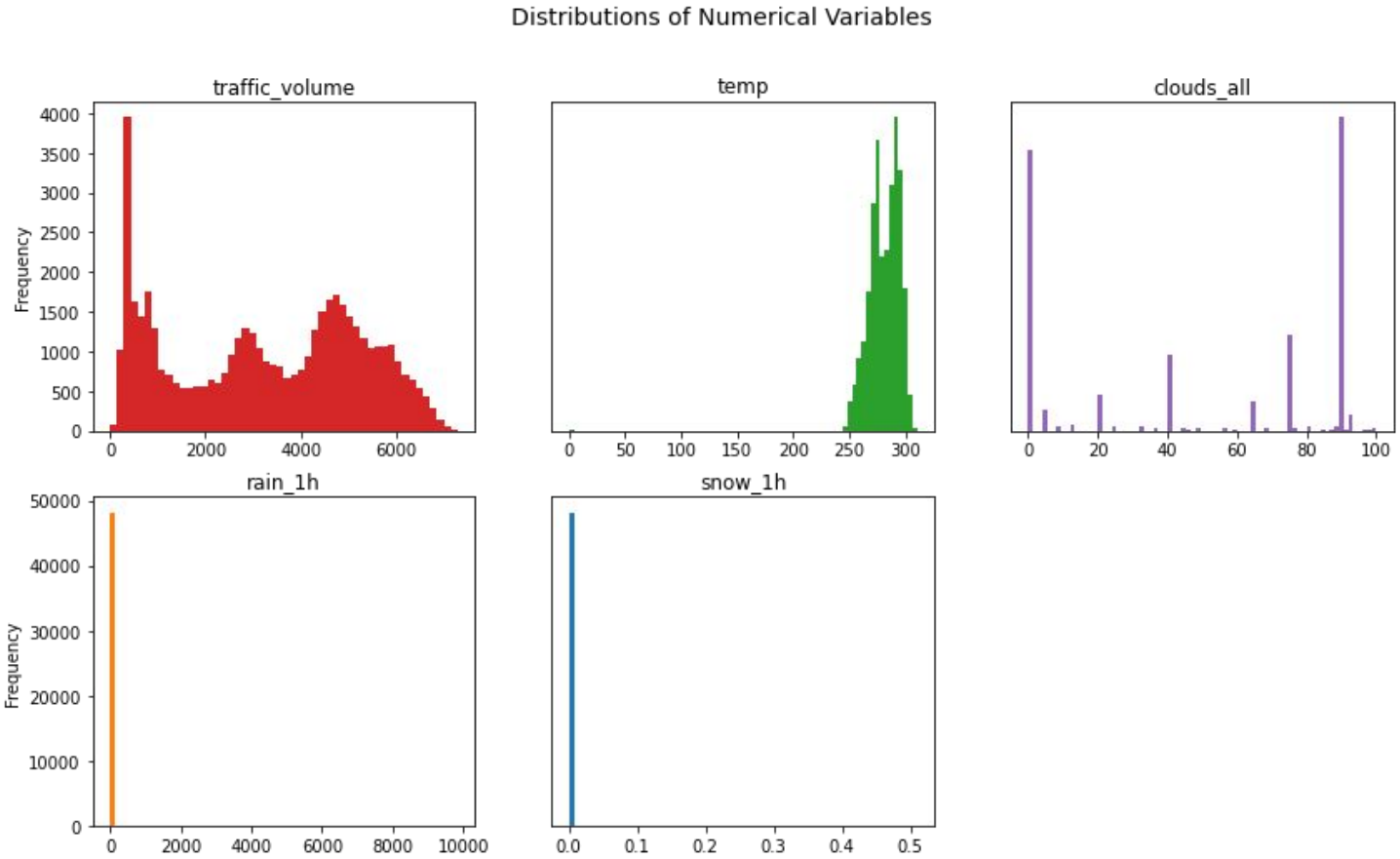
3. Serve as a key indicator for various impact assessments– air and noise pollution, energy conservation, economics, etc.

4. Continuous monitoring of traffic performance

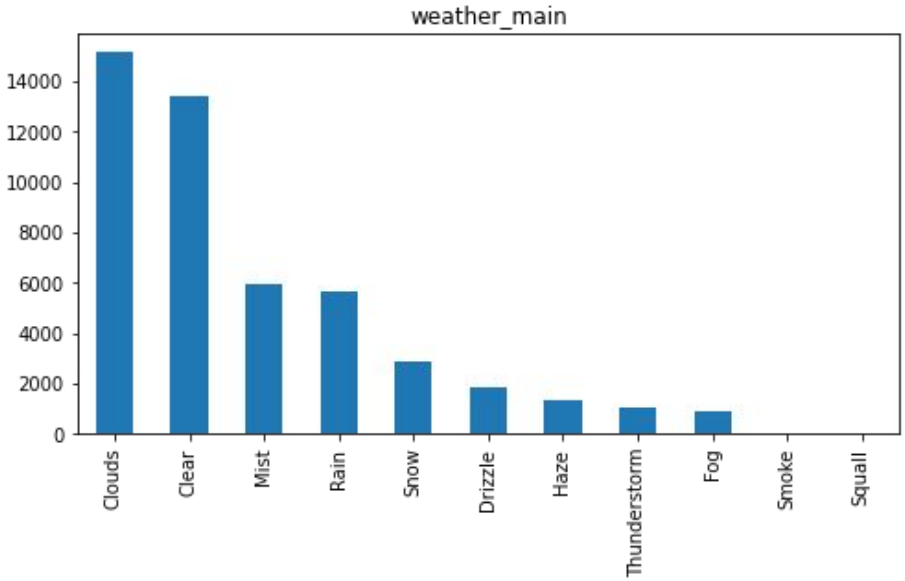
1

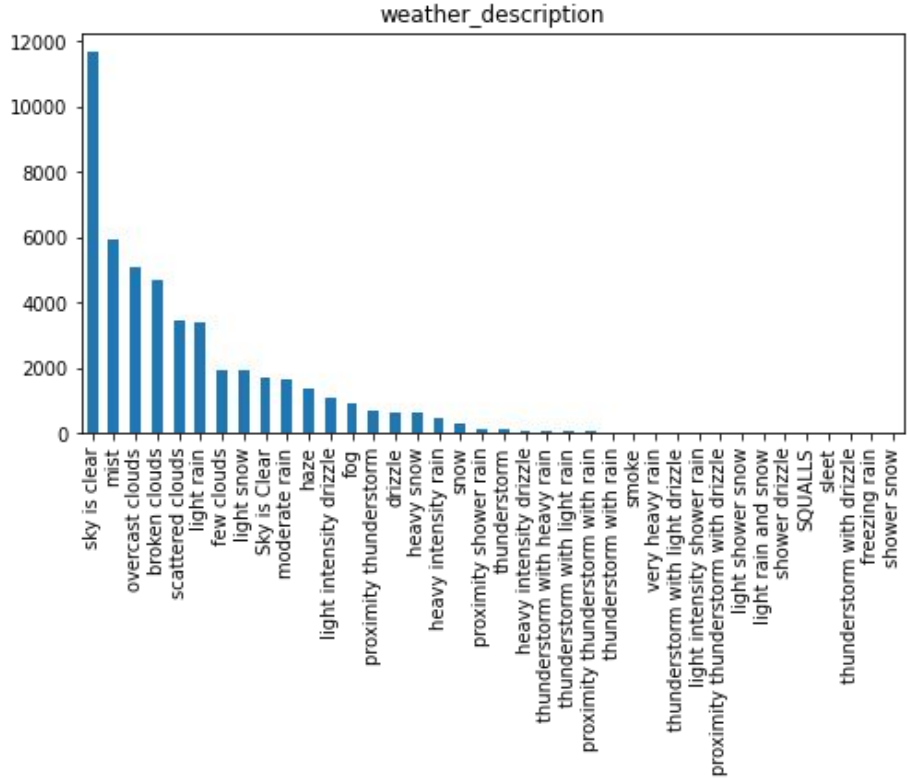
Exploratory Data Analysis

Numeric Features

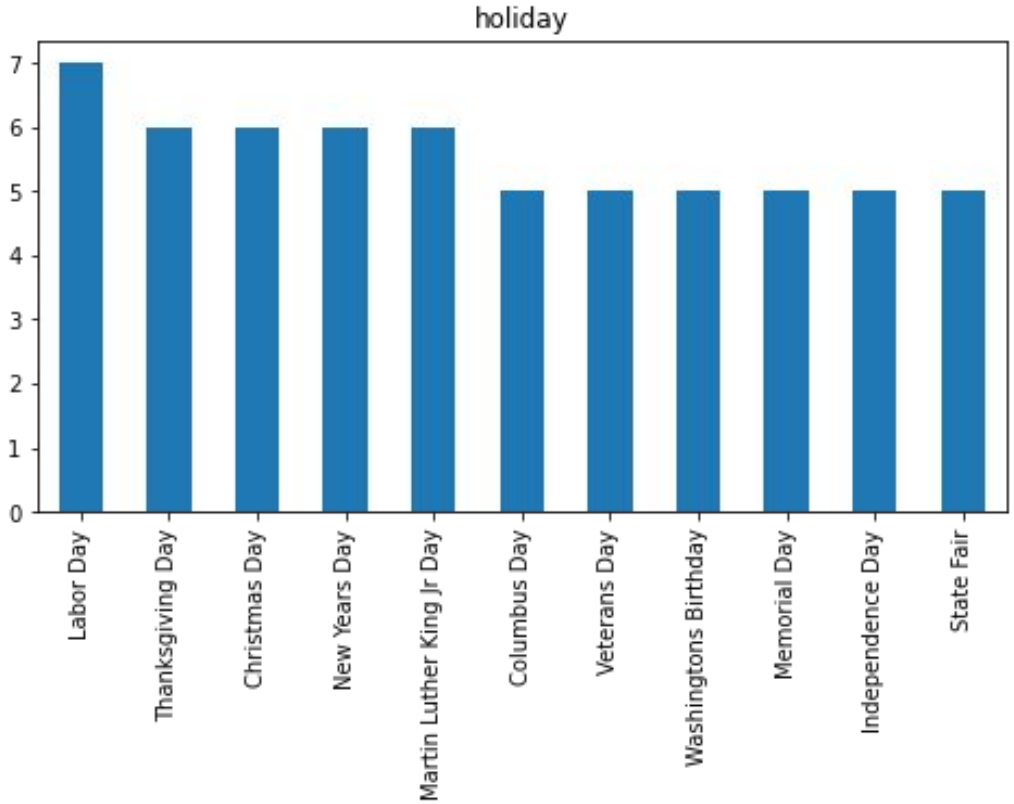


Categorical Features

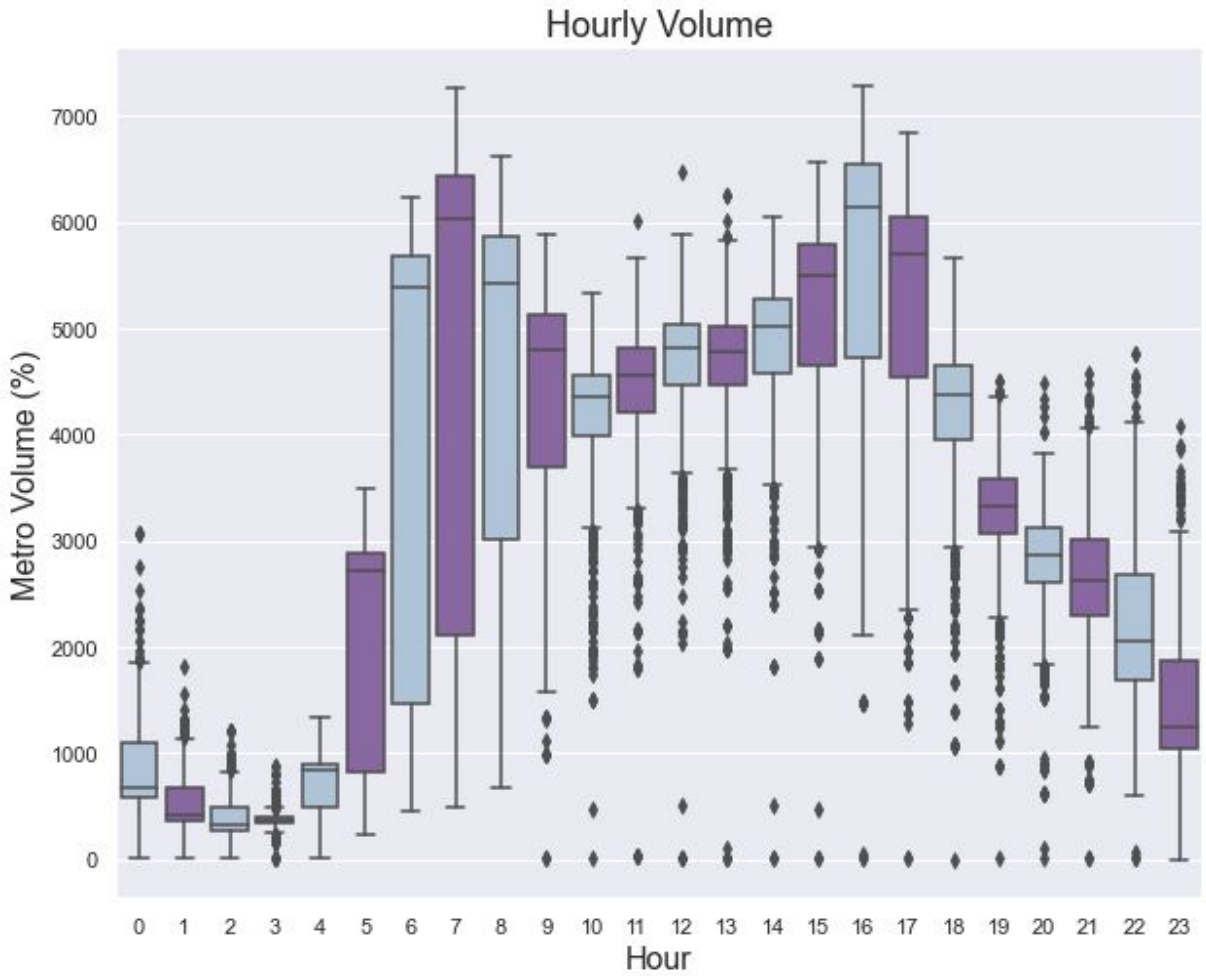




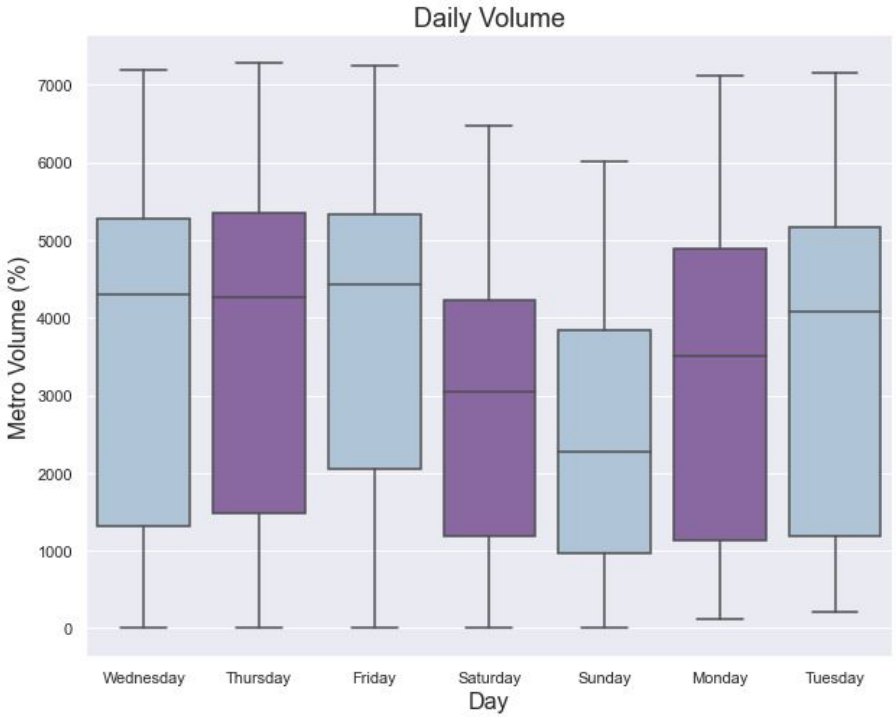
Categorical Features

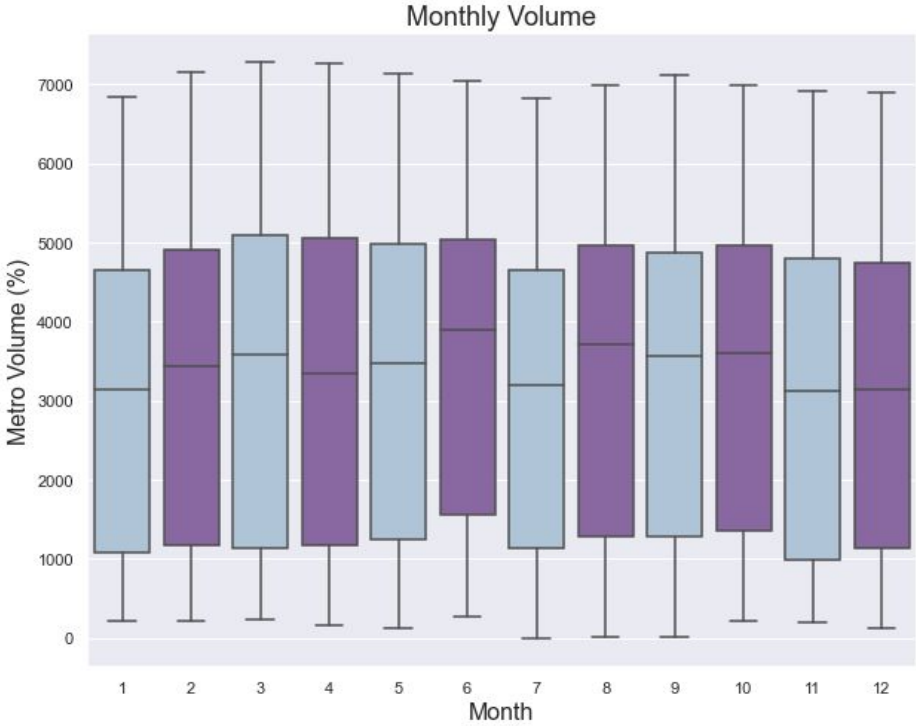


Seasonality

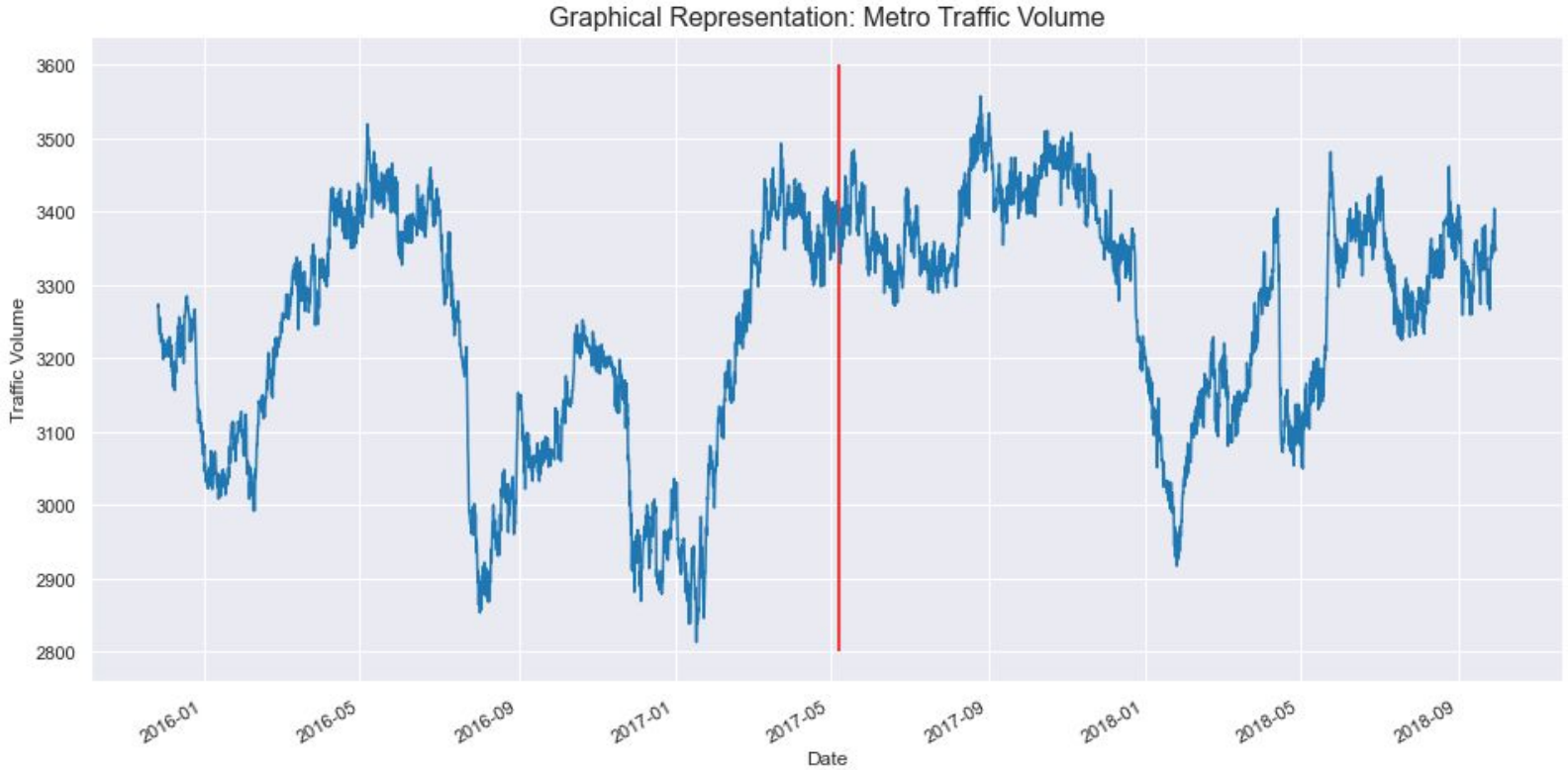


Seasonality



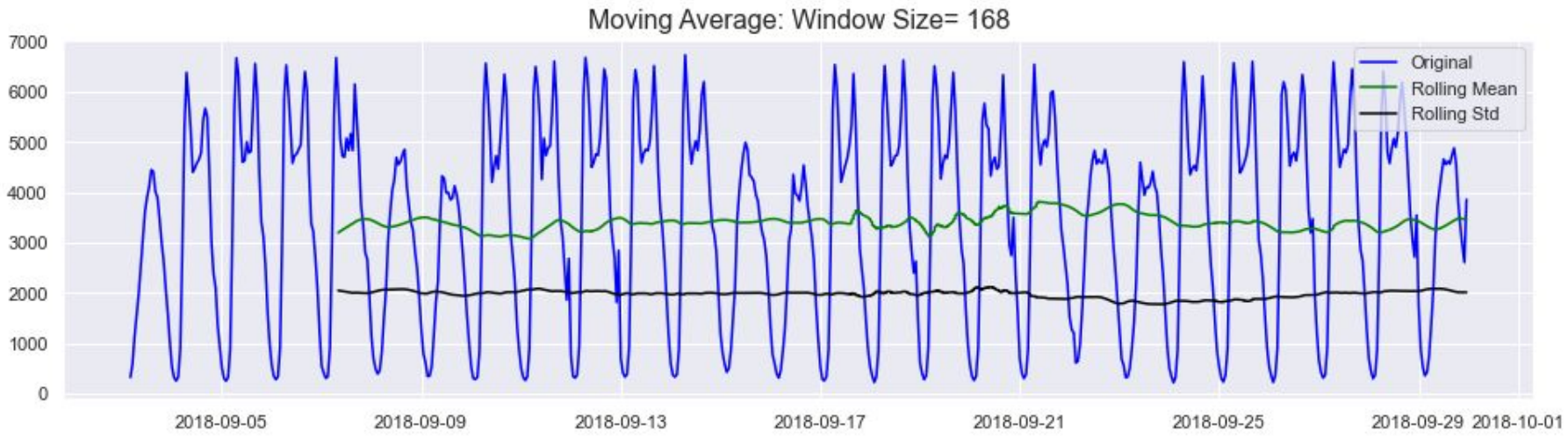


Graphical Representation



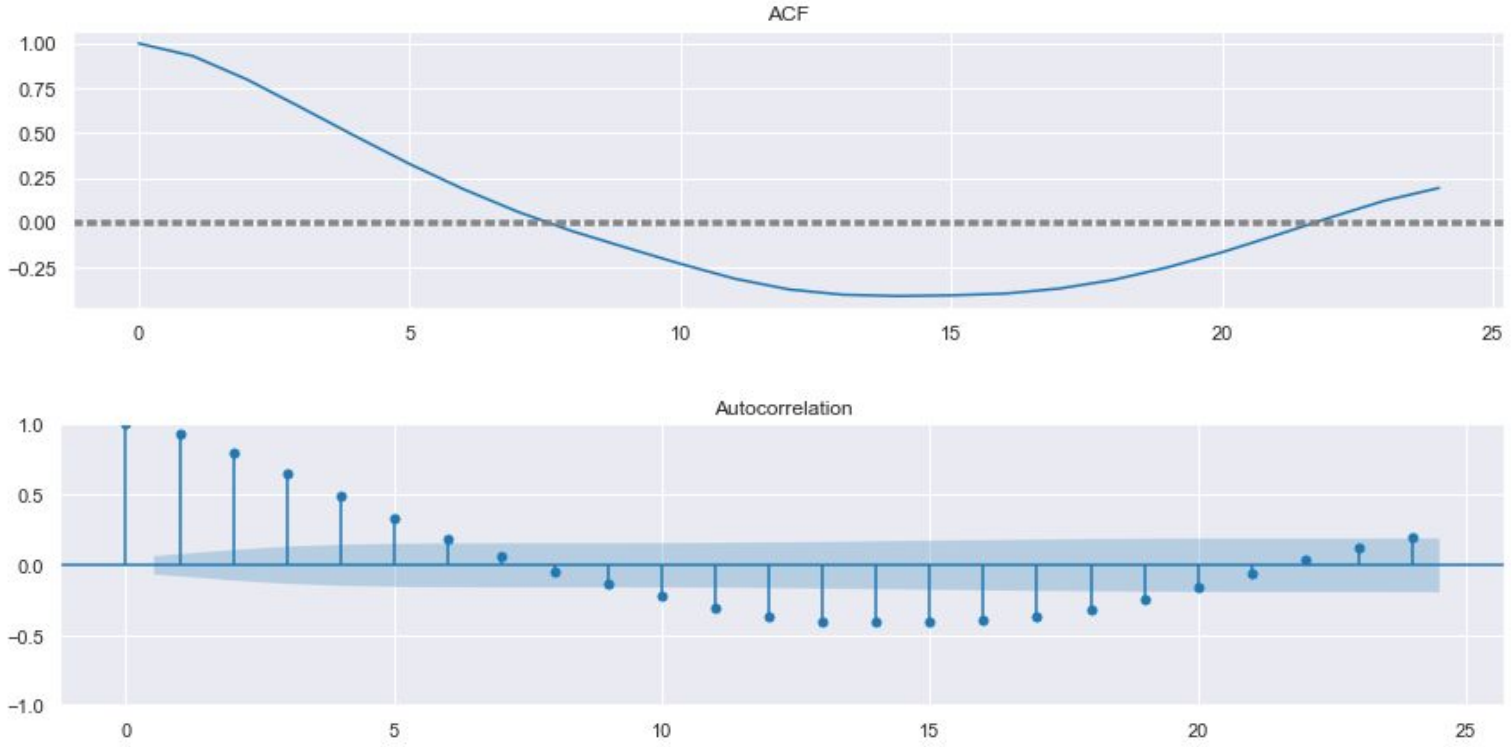
mean1= 3203.011119, mean2= 3310.810224 and var1= 3878882.025234, var2= 3910121.223393

Augmented Dickey Fuller Test

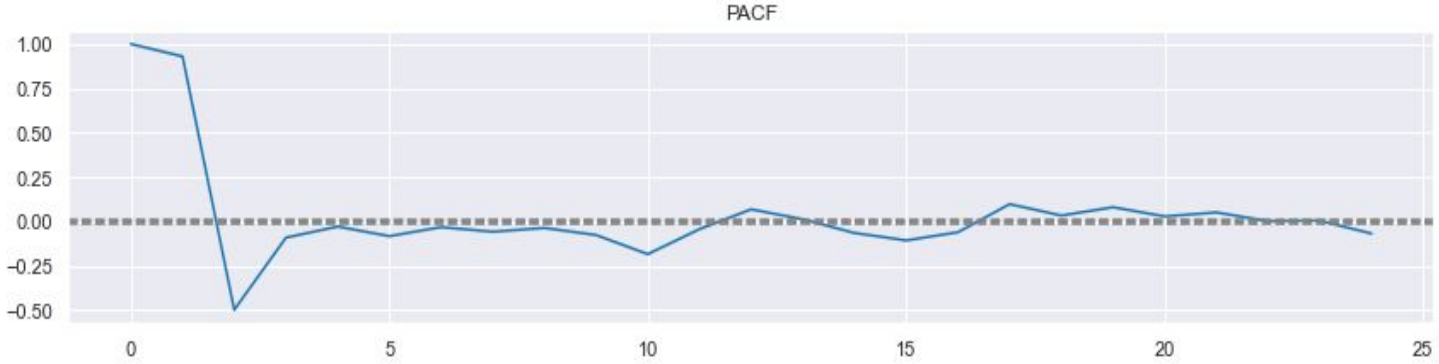


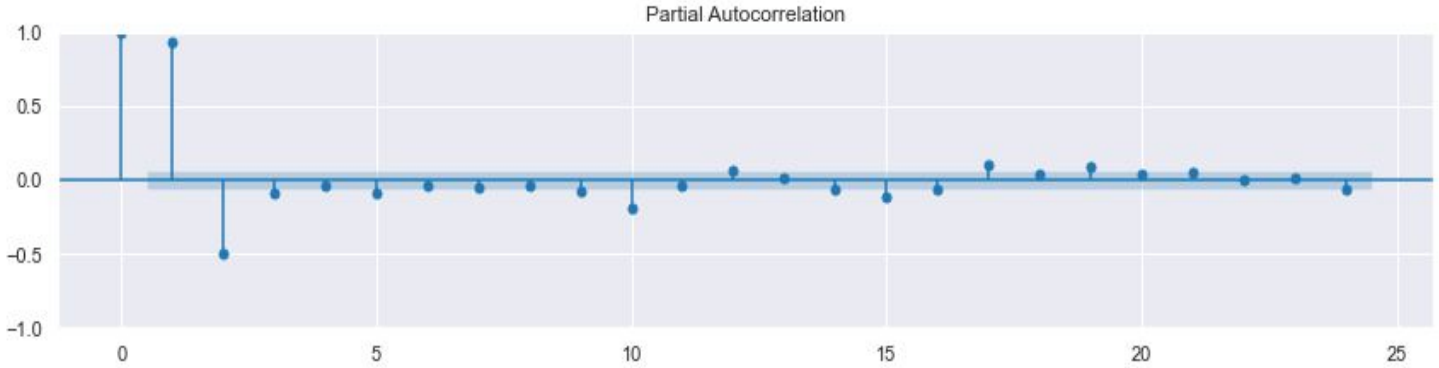
p-value = 5.3e-9 < 0.05, therefore ts is likely stationary

Autocorrelation

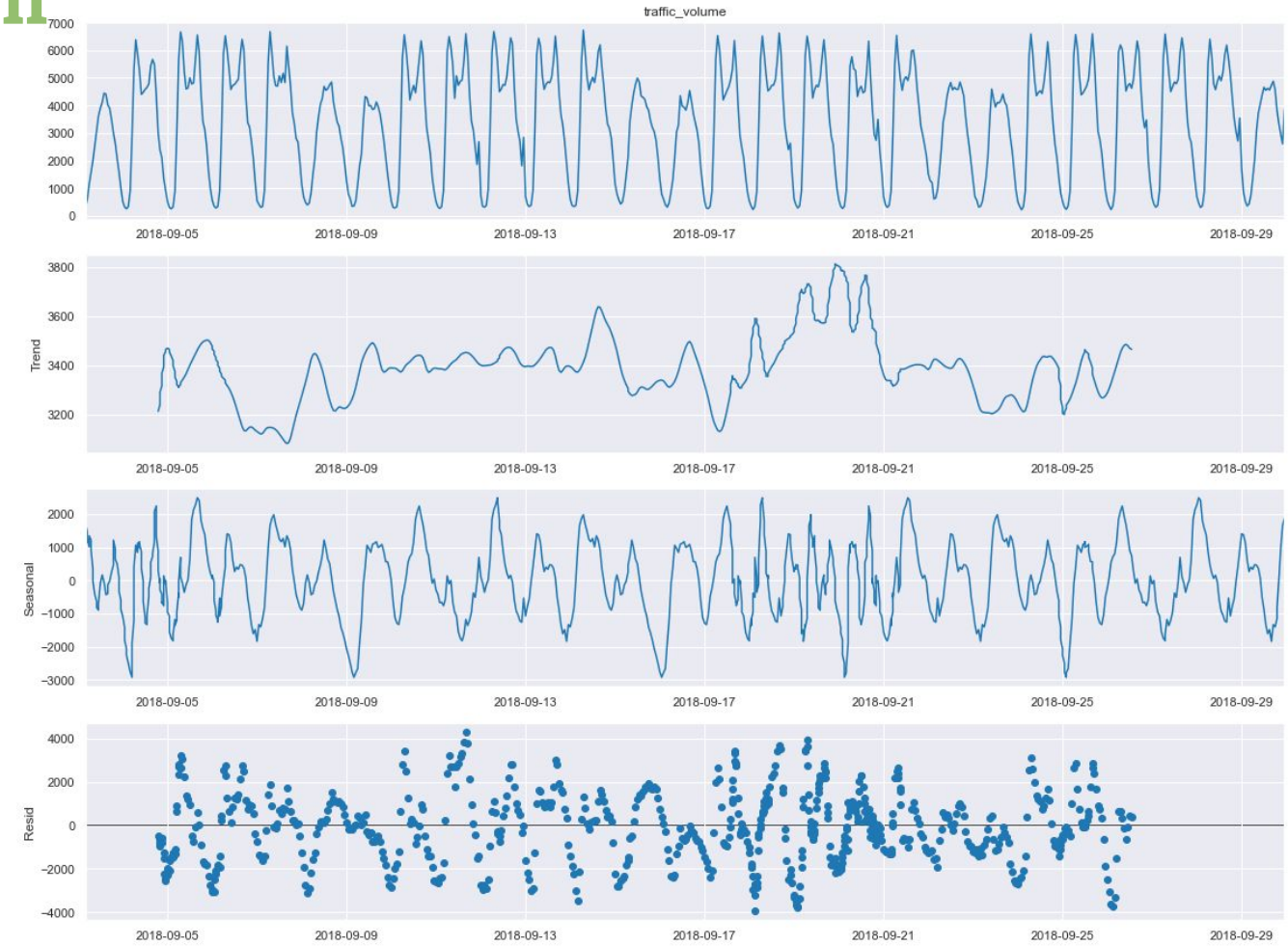


Partial Autocorrelation



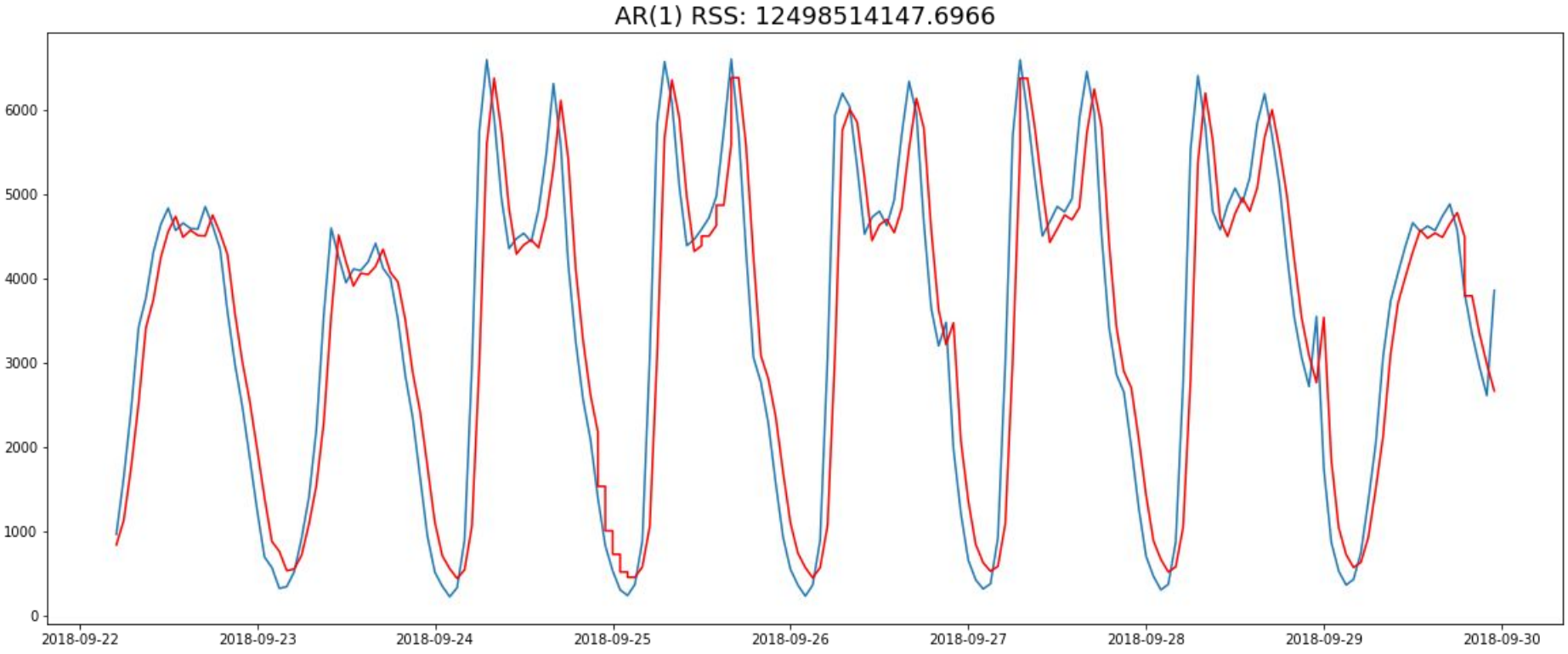


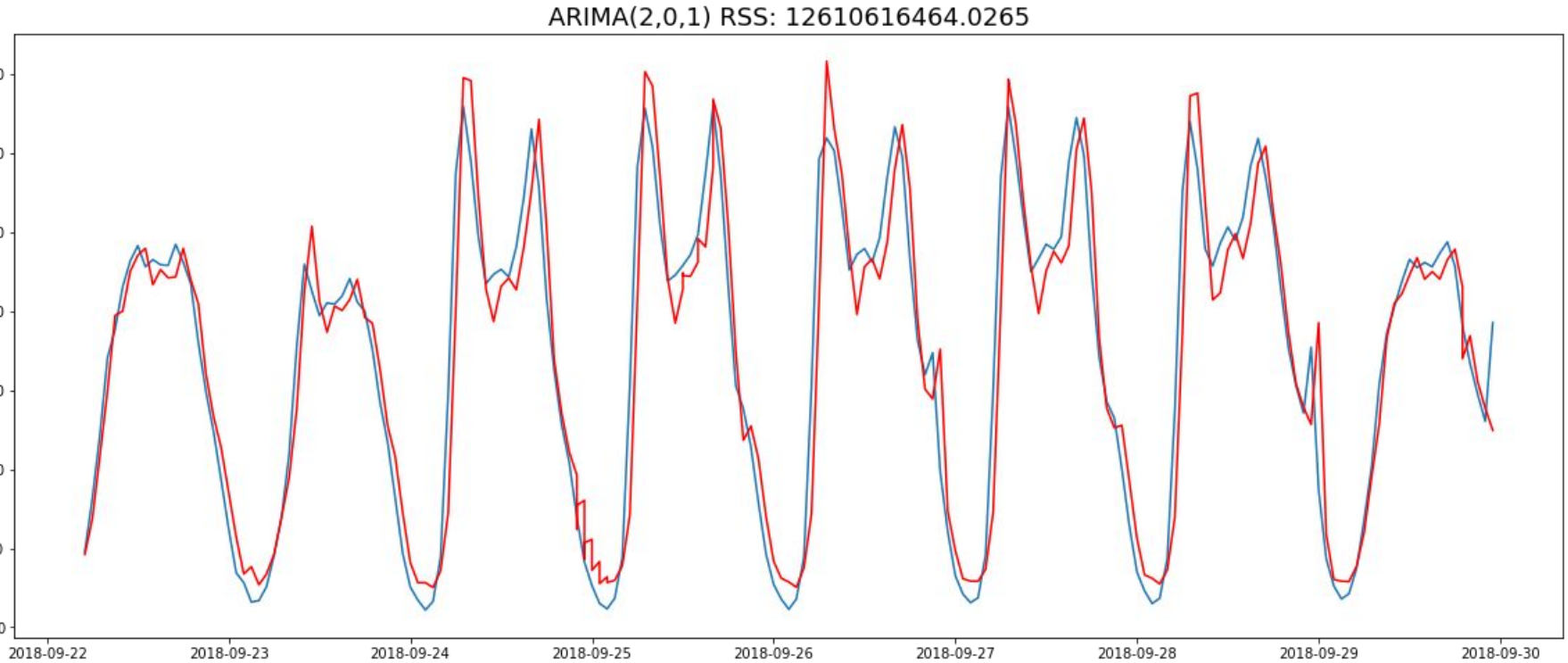
Decomposition



2

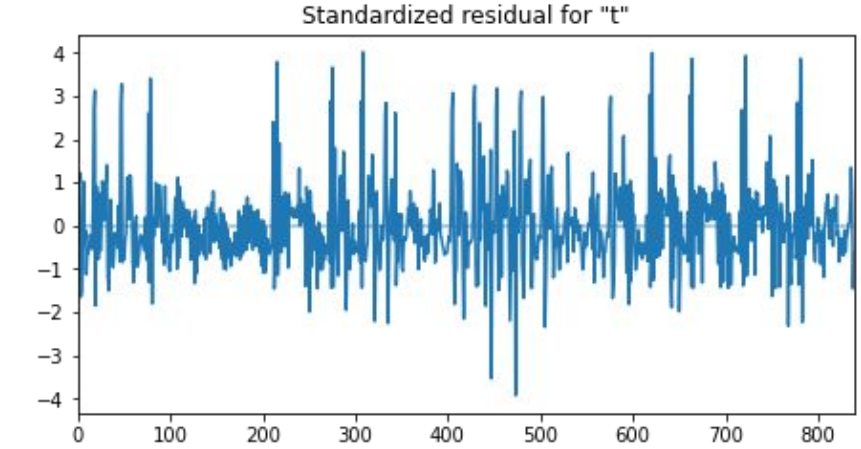
Prediction

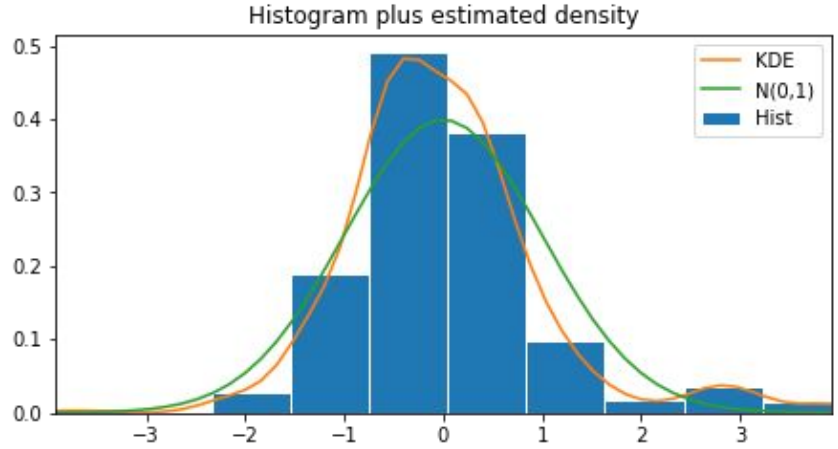

AR(1) RSS: 12498514147.6966
6000
5000
4000
3000
2000
1000
0
2018-09-22
2018-09-23
2018-09-24
2018-09-25
2018-09-26
2018-09-27
2018-09-28
2018-09-29
2018-09-30

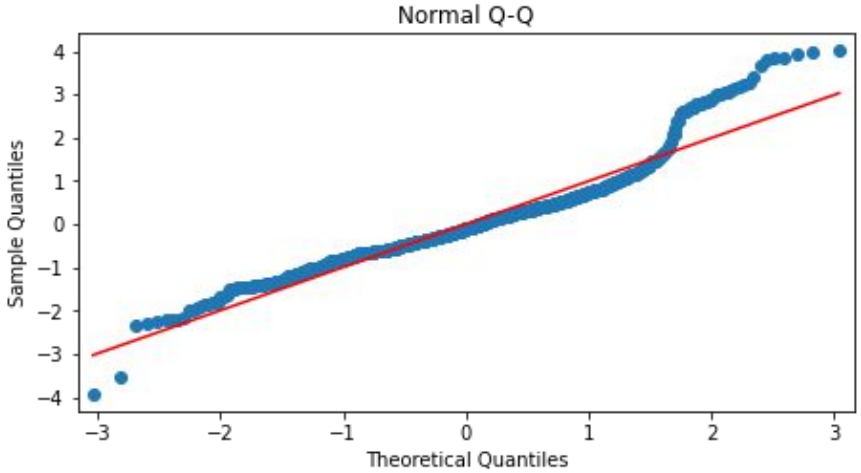

ARIMA(2,0,1) RSS: 12610616464.0265
7000
6000
5000
4000
3000
2000
1000
0
2018-09-22
2018-09-23
2018-09-24
2018-09-25
2018-09-26
2018-09-27
2018-09-28
2018-09-29
2018-09-30

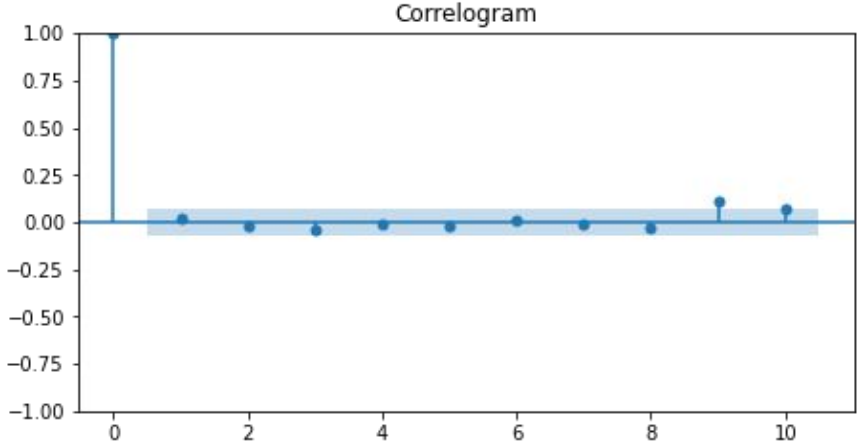
Model	*RSS*	*Log likelihood/ AIC/BIC/HQIC*	*Coef P-values*	*Ljung-Box*	*Jarque-Bera*
MA(3)	1.17e10	LL -6749.456 AIC 13508.912 BIC 13532.579 HQIC 13517.983	ma.L1 0.000 ma.L2 0.000 ma.L3 0.000 Sigma2 0.000	39.22 – cannot reject null of white noise	698.48 (Prob: 0) – not normally distributed
AR(1)	1.25e10	LL -6718.620 AIC 13443.239 BIC 13457.439 HQIC 13448.682	Const 0.000 ar.L1 0.000 Sigma2 0.000	178.31 – cannot reject null of white noise	698.43 (Prob: 0) – not normally distributed
AR(2)	1.26e10	LL -6599.958 AIC 13207.915 BIC 13226.849 HQIC 13215.172	Const 0.000 ar.L1 0.000 ar.L2 0.000 sigma2 0.000	1.73 – cannot reject null of white noise	491.26 (Prob: 0) – not normally distributed
ARIMA(2,0, 1)	1.26e10	LL -6599.958 **AIC 13201.893** **BIC 13225.560** HQIC 13210.964	ar.L1 0.000 ar.L2 0.000 ma.L1 0.001 Sigma2 0.000	**0.03 – likely white noise**	375.18 (Prob: 0) – not normally distributed
ARIMA(2,1,1)	1.32e10	LL -6631.082 AIC 13270.164 BIC 13289.093 HQIC 13277.420	ar.L1 0.552 ar.L2 0.087 ma.L1 0.131 Sigma2 0.000	0.24 – cannot reject null of white noise	508.81 (Prob: 0) – not normally distributed

ARIMA(2,0,1)

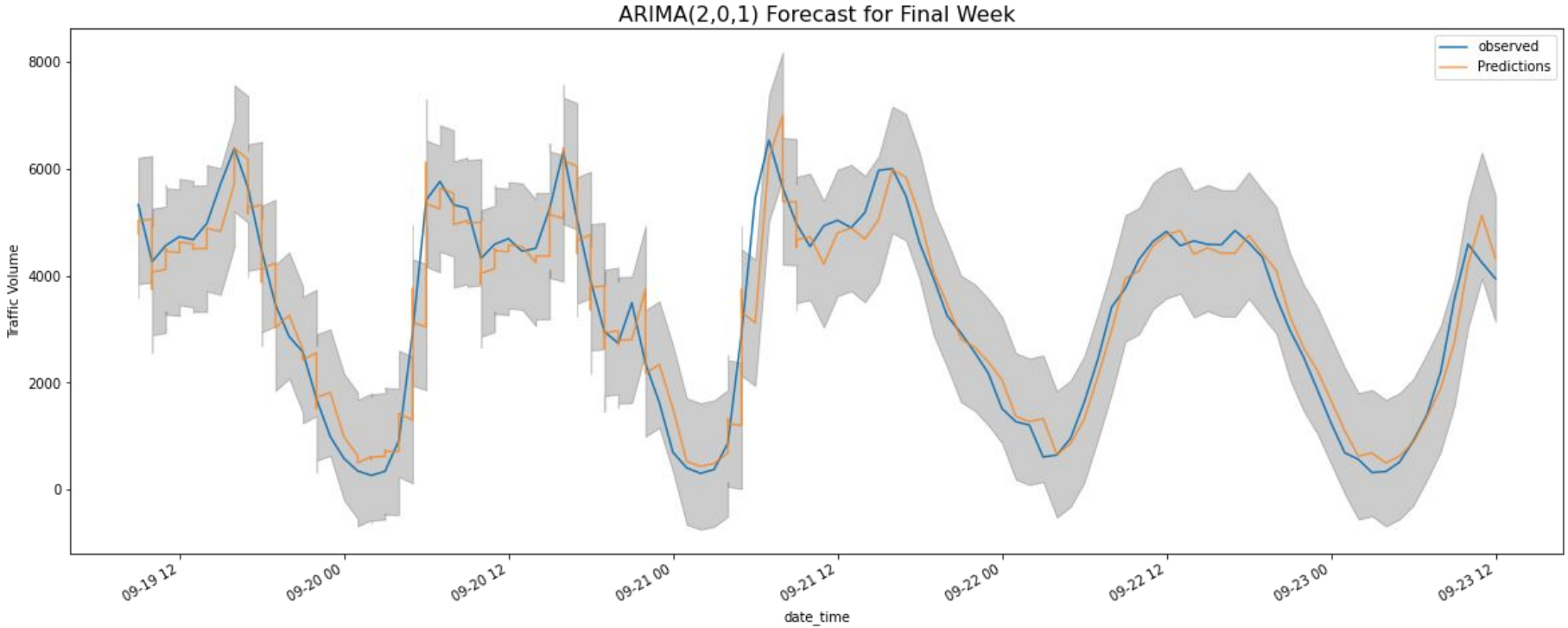

Standardized residual for "t"


Histogram plus estimated density
KDE
N(0,1)
Hist


Normal Q-Q
Sample Quantiles
Theoretical Quantiles

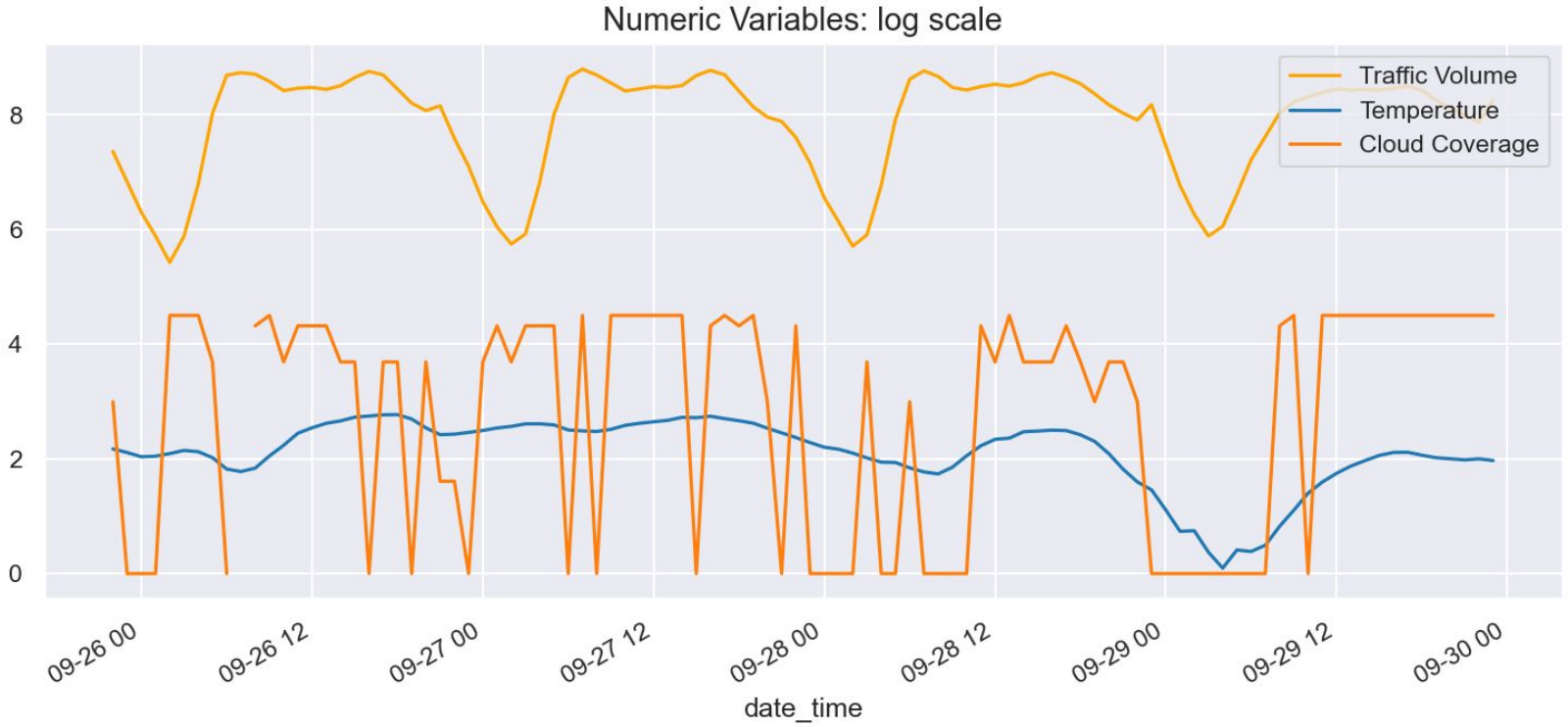

Correlogram

Validation

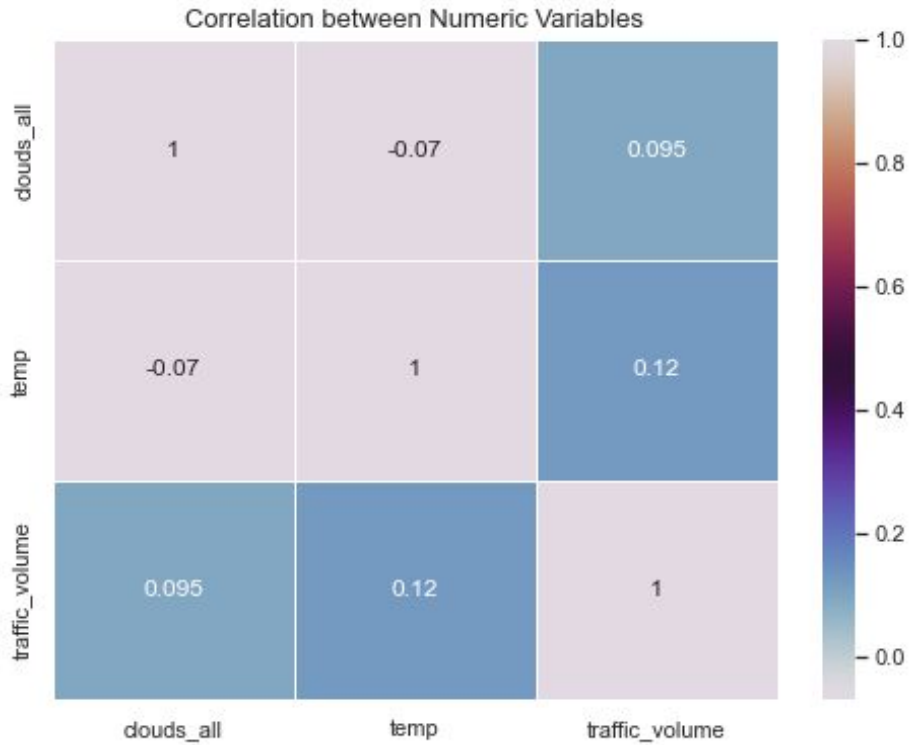

ARIMA(2,0,1) Forecast for Final Week
observed
Predictions
Traffic Volume
8000
6000
4000
2000
0
09-19 12
09-20 00
09-20 12
09-21 00
09-21 12
09-22 00
09-22 12
09-23 00
09-23 12
date_time

4

Multivariate Analysis


Numeric Variables: log scale
Traffic Volume
Temperature
Cloud Coverage
8
6
4
2
0
09-26 00
09-26 12
09-27 00
09-27 12
09-28 00
09-28 12
09-29 00
09-29 12
09-30 00
date_time

Correlation between Numeric Variables



Granger Causality Test

- Tests the null hypothesis that the coefficients of past values in the regression equation is zero
- If a given p-value < 0.05 then the corresponding column X causes row Y

	traffic_volume_x	temp_x	clouds_all_x
traffic_volume_y	1.000	0.0000	0.0118
temp_y	0.000	1.0000	0.0002
clouds_all_y	0.001	0.0016	1.0000

Johanson's Cointegration Test

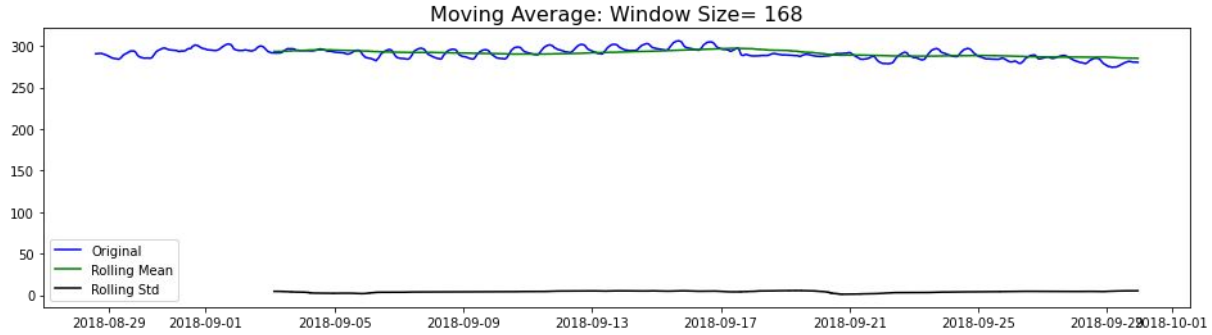
- When two or more TSs are cointegrated, they have a statistically significant relationship
- Null hypothesis: time series are not cointegrated

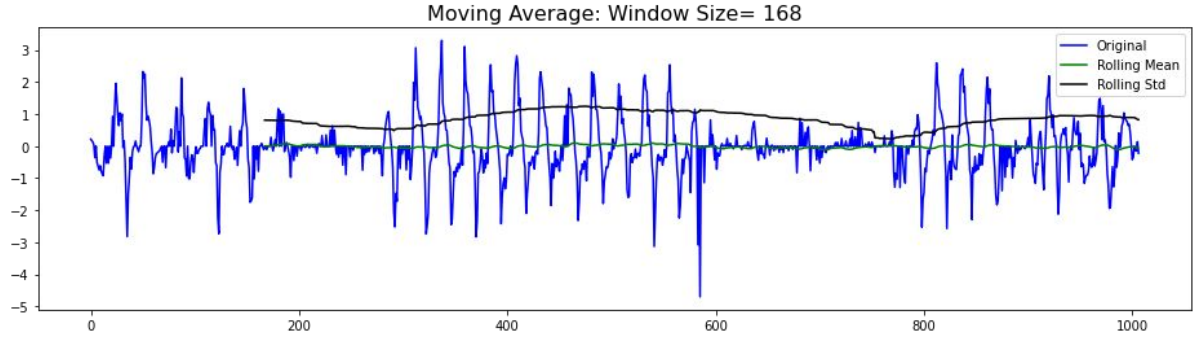
Johanson's Cointegration Test

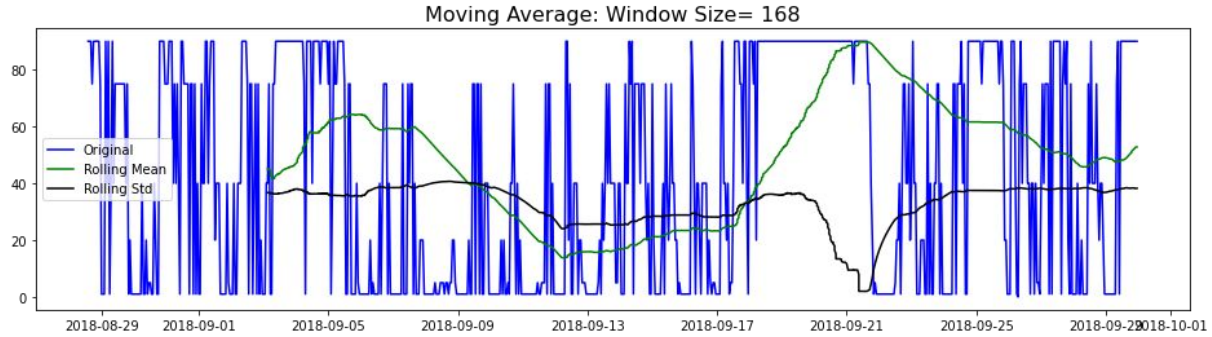
Name	::	Test Stat	>	C(95%)	=>	Signif
traffic_volume	::	154.35	>	24.2761	=>	True
temp	::	28.37	>	12.3212	=>	True
Clouds_all	::	0.07	>	4.1296	=>	False

ADF Tests

- Temp ADF
  - P-val: 0.183
  - Not-Stationary
- 1st diff Temp ADF
  - P-val: 0.000
  - Stationary
- Clouds ADF
  - P-val: 0.000
  - Stationary





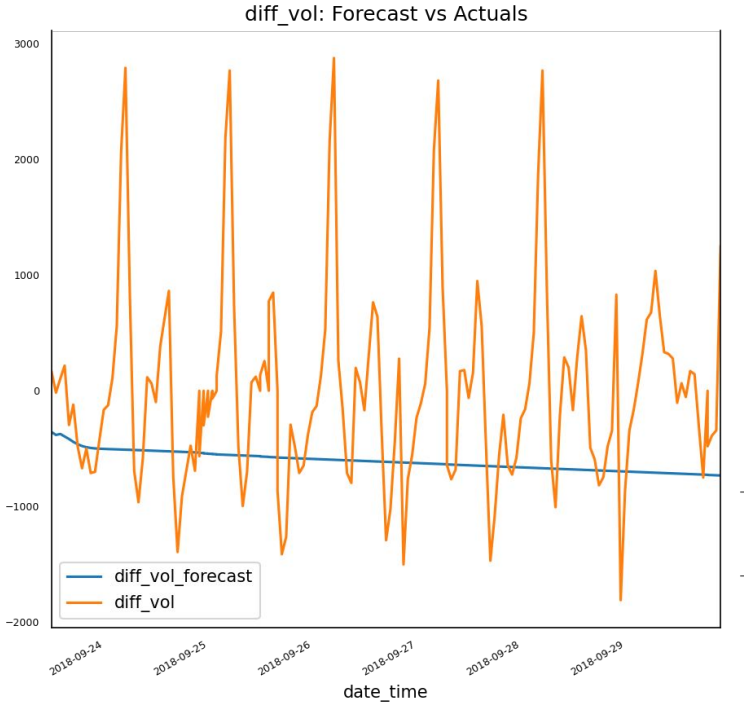


Selecting the VAR(p) model

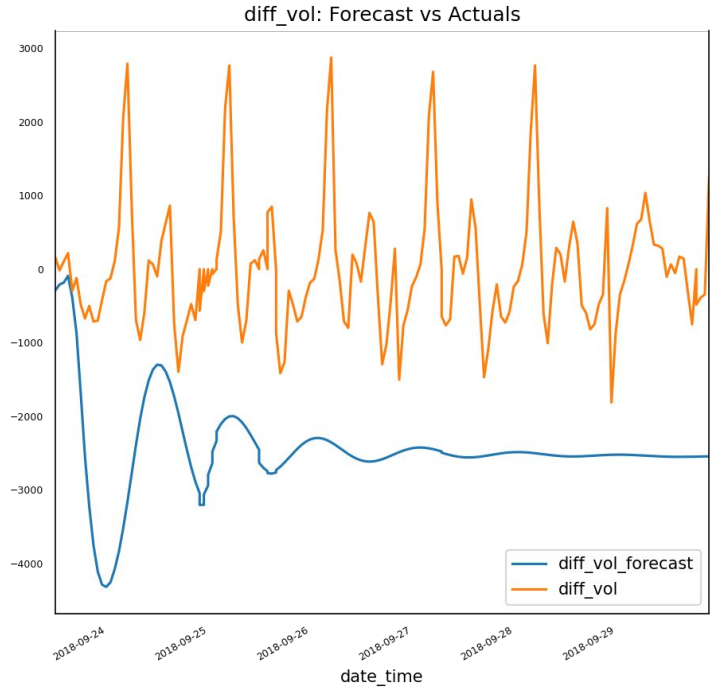
- Iteratively fit increasing order of VAR model (orders 1-10)
- VAR(10) performs best on AIC, FPE, and HQIC
- VAR(3) performs best on BIC

	AIC	BIC	FPE	HQIC
0	19.44	19.45	2.764e+08	19.44
1	18.30	18.36	8.899e+07	18.33
2	18.22	18.33	8.212e+07	18.26
3	18.14	18.28*	7.520e+07	18.19
4	18.10	18.29	7.264e+07	18.17
5	18.08	18.32	7.119e+07	18.17
6	18.05	18.33	6.916e+07	18.16
7	18.02	18.34	6.689e+07	18.14
8	17.98	18.35	6.421e+07	18.12
9	17.95	18.37	6.261e+07	18.11
10	17.93*	18.39	6.137e+07*	18.11*

VAR(p) forecast is still a bad fit…



VAR(3)



VAR(10)