THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Time Series Forecasting

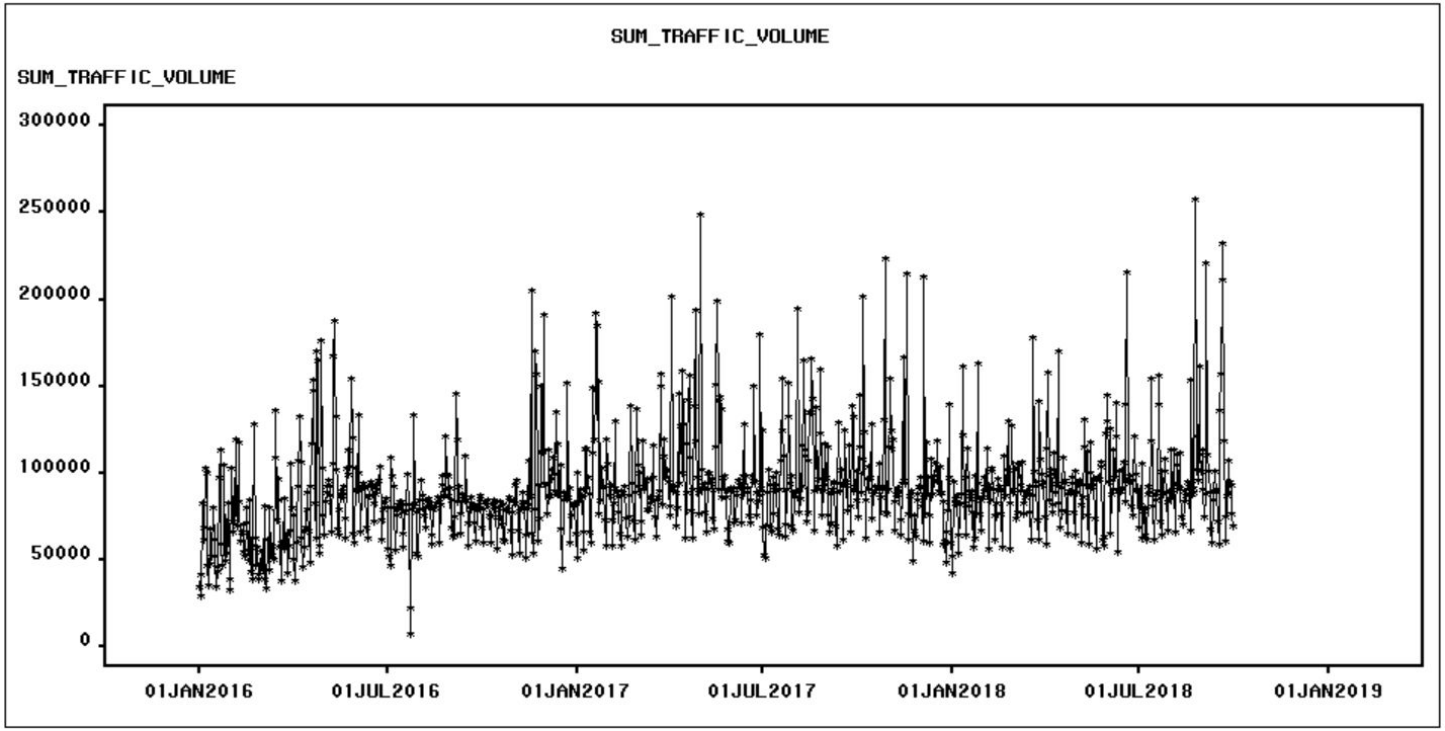
Modeling Daily Traffic Volume on I-94

Yuzhou Ding
Viktoriia Dunleavy
Guotian Kan
Jennifer Nguyen
Ci Song

Introduction to Dataset and Description of the Project

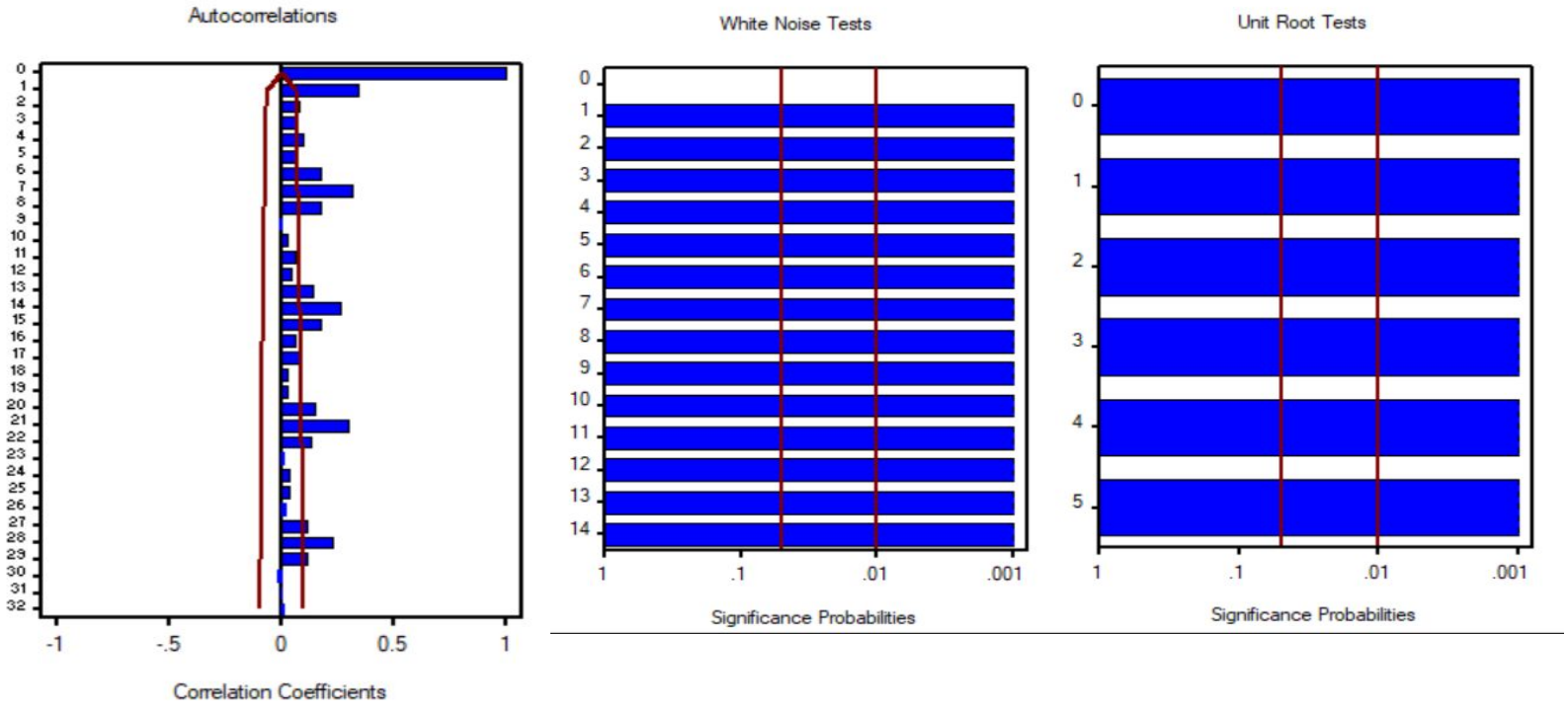
- Daily Interstate 94 Westbound traffic volume for MN DoT ATR station 301, roughly midway between Minneapolis and St Paul, MN.
- 1004 observations from 1/1/2016 to 9/30/2018.
- Main variable of interest: sum_traffic_volume (Numeric): Daily I-94 ATR 301 reported westbound traffic volume.
- The data set could be used to analyze the relationship between traffic volume and time, as well as to analyze the relationship with other variables.
- Hold out sample: 150

Introduction to the Response Variable



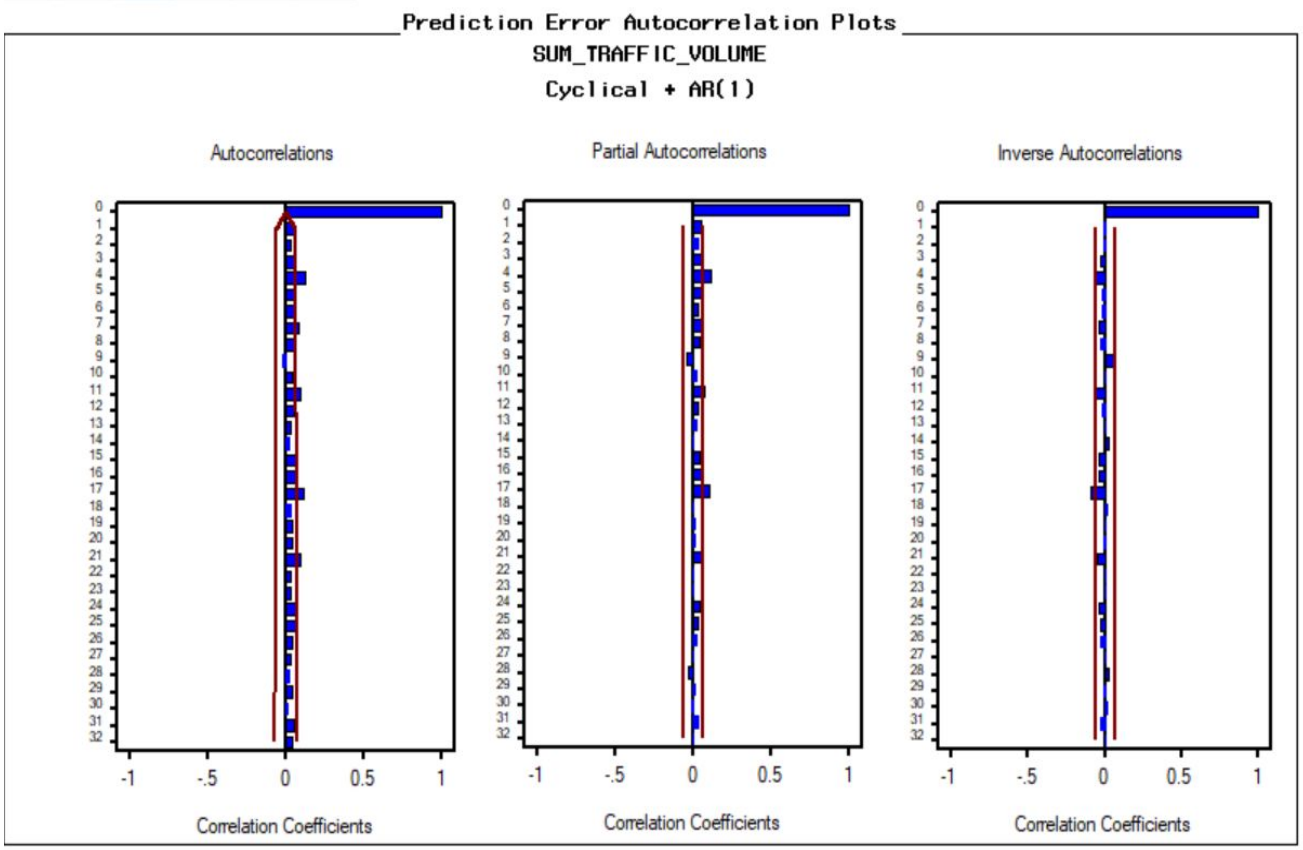
- Stationary with the probability of seasonality.
- No visual trend detected.

Introduction to the Response Variable



- Not WN but Non-seasonal stationary.
- Seasonality at lag 7, 14, 21, etc..

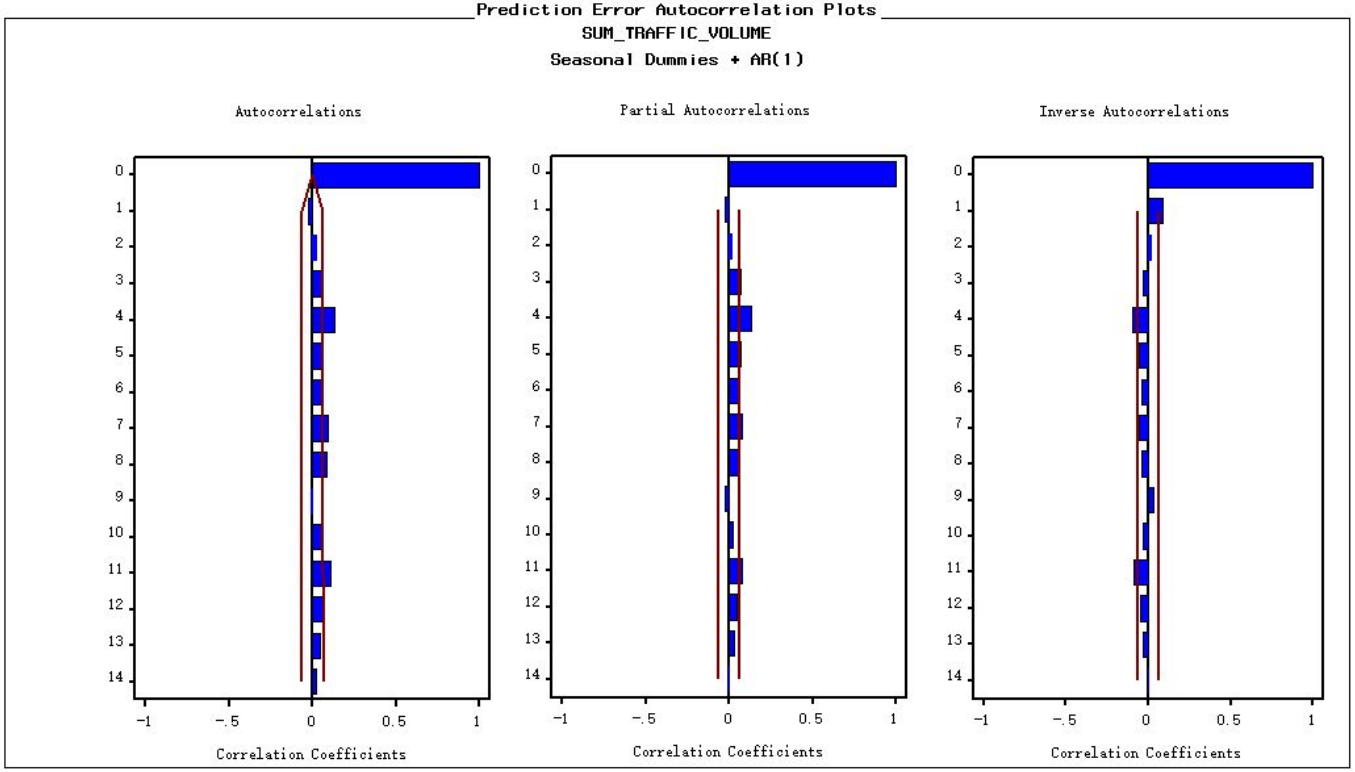
Model Building - Cyclical Trend Model



Cyclical + AR(1) Prediction Error Plot

- Normal Cyclical Trend Model ACF indicates an AR(1) noise model.
- However, after adding the noise model, it is still not WN.
- The test error is the worst among all considered deterministic models.
- Thus, we will not focus on the cyclical trend model.

Model Building - Seasonal Dummies + AR(1)

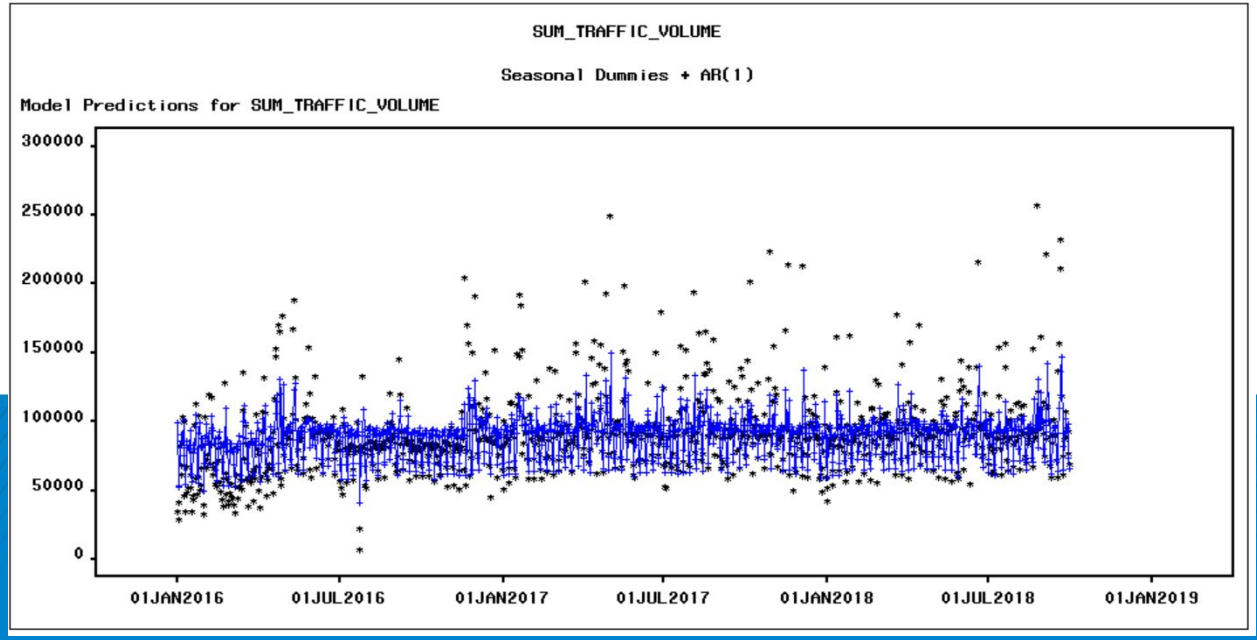


- Normal dummies model ACF and PACF suggests an AR(1) noise term.
- Though some out-of-bound ACFs, they are small enough to be considered as WN.

Model Building - Seasonal Dummies

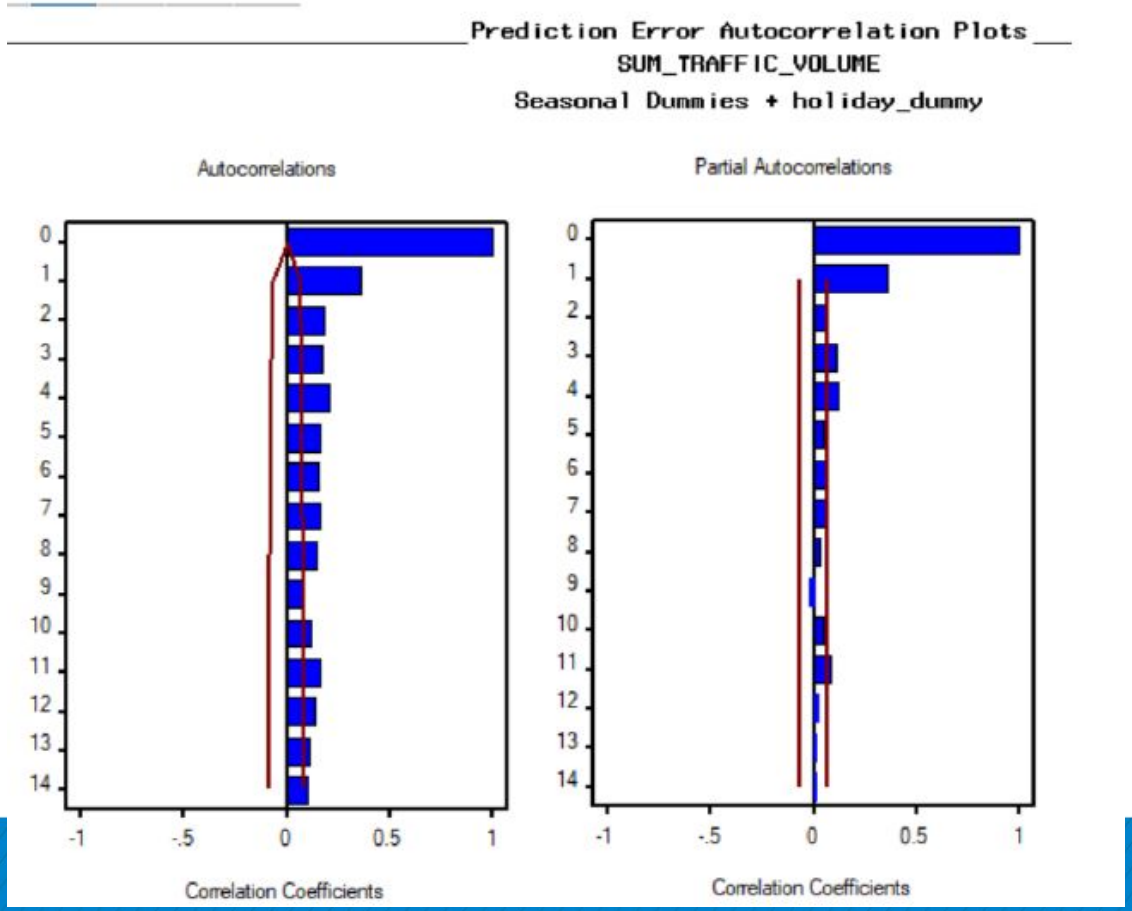
Parameter Estimates
SUM_TRAFFIC_VOLUME
Seasonal Dummies + AR(1)

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	75457	2401	31.4258	<.0001
Autoregressive, Lag 1	0.35242	0.0322	10.9558	<.0001
Seasonal Dummy 1	-10320	2729	-3.7813	0.0002
Seasonal Dummy 2	18204	3168	5.7455	<.0001
Seasonal Dummy 3	19833	3294	6.0209	<.0001
Seasonal Dummy 4	22573	3294	6.8519	<.0001
Seasonal Dummy 5	21269	3169	6.7111	<.0001
Seasonal Dummy 6	23786	2729	8.7166	<.0001
Model Variance (sigma squared)	615367462	.	.	.



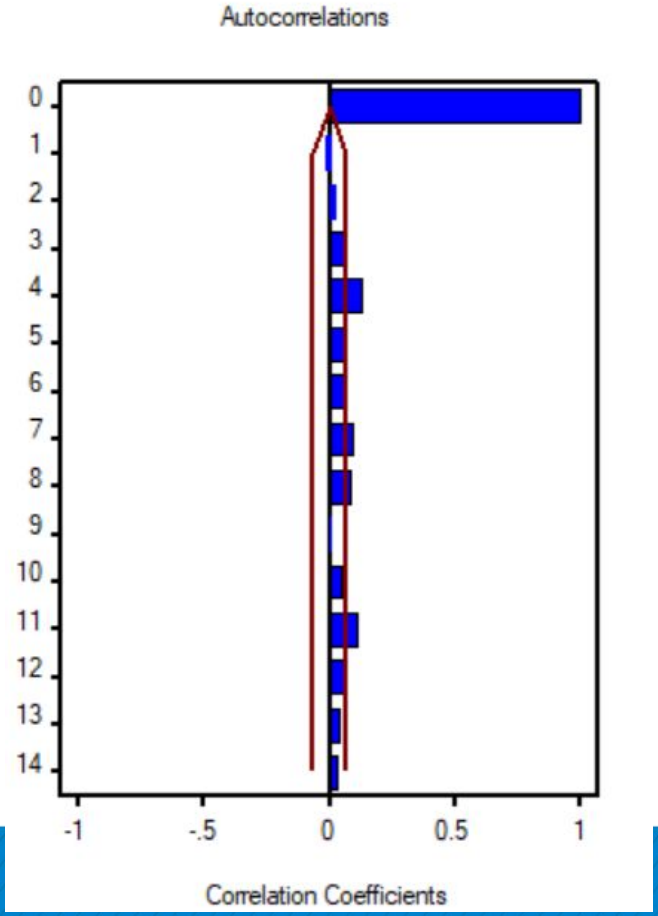
- All predictors are significant.
- Reference date: Sat;
- Dummy 1: Sun; etc...
- The model captured seasonal trend but did not capture observations that are far away from the general trend.

Adding Holiday Dummy - 1



- ACF decays and PACF is chopped off after lag 1.
- We should use AR(1) error model.

Adding Holiday Dummy - 2



- The ACF of the residuals suggests that the series are WN. There are a few lags that are out of 2 se bounds but are considered insignificant.

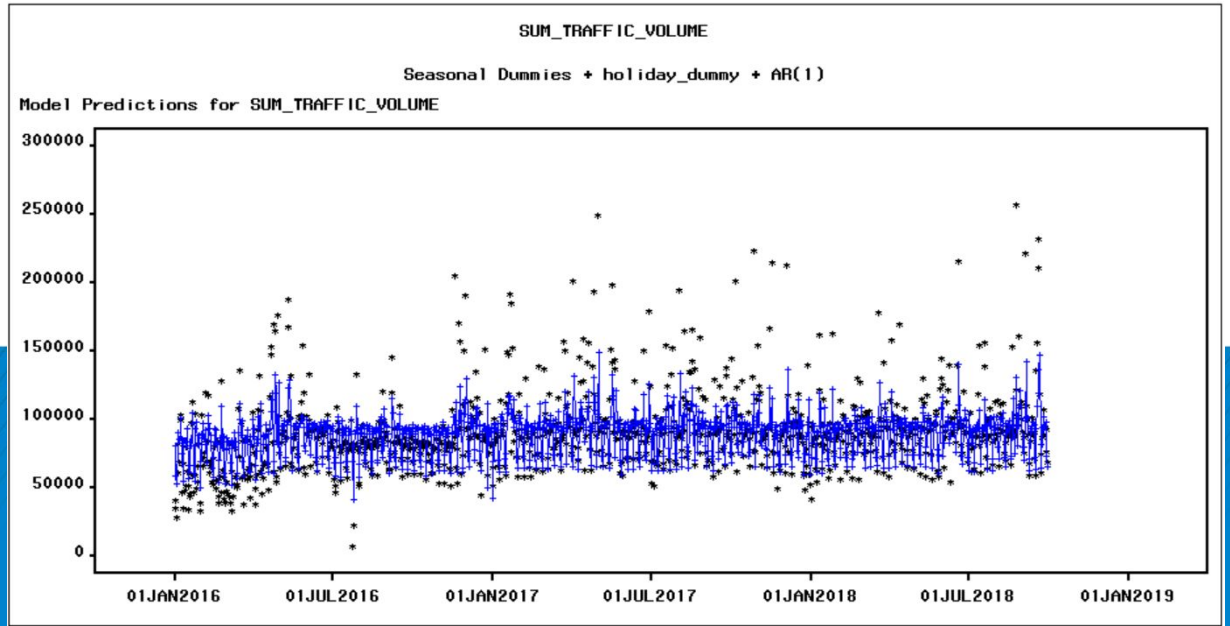
Adding Holiday Dummy - 3

Parameter Estimates

SUM_TRAFFIC_VOLUME

Seasonal Dummies + holiday_dummy + AR(1)

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	75457	2380	31.7087	<.0001
Autoregressive, Lag 1	0.35079	0.0322	10.8898	<.0001
Seasonal Dummy 1	-10011	2709	-3.6946	0.0003
Seasonal Dummy 2	20520	3198	6.4162	<.0001
Seasonal Dummy 3	19994	3266	6.1212	<.0001
Seasonal Dummy 4	22592	3266	6.9164	<.0001
Seasonal Dummy 5	21940	3148	6.9698	<.0001
Seasonal Dummy 6	24247	2711	8.9456	<.0001
holiday_dummy	-18817	4833	-3.8934	0.0002
Model Variance (sigma squared)	605238917	.	.	.

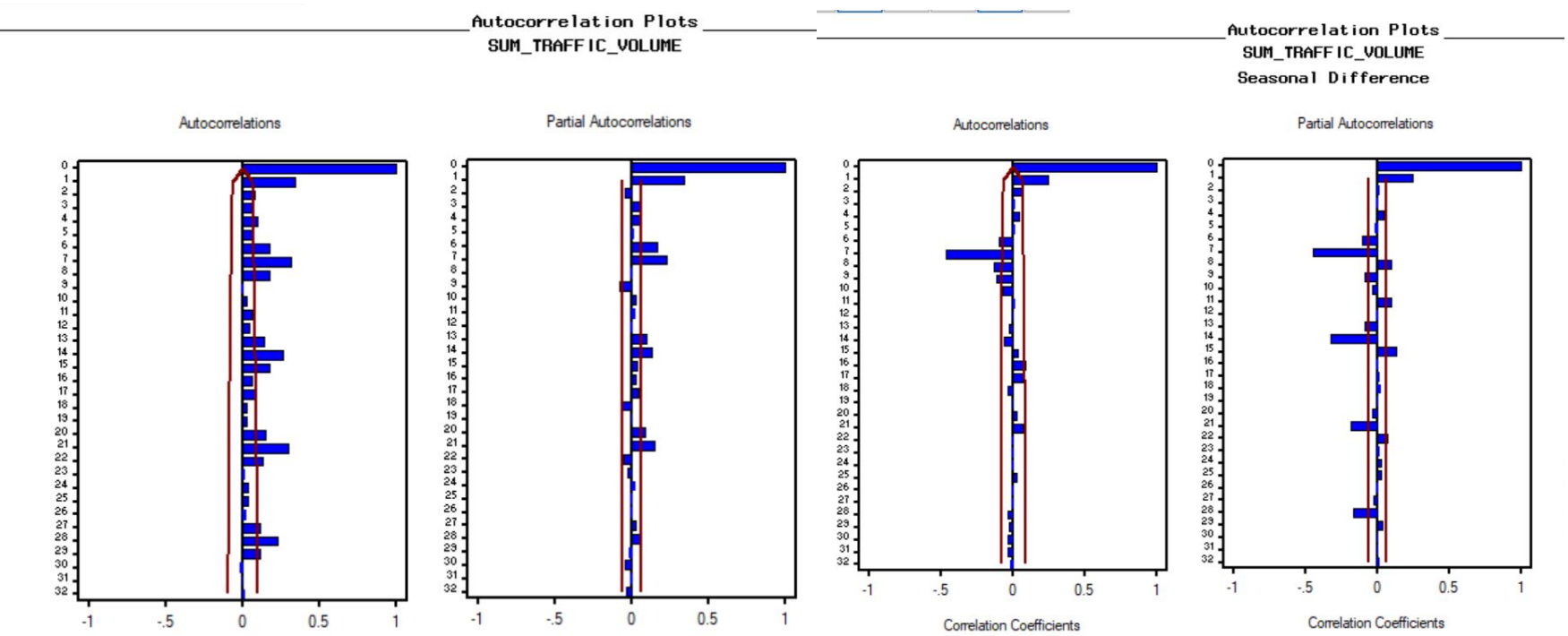


Deterministic Model Selection

		Test Error	
	Model Fit (Square Root of Model Variance)	Validation RMSE	Validation MAPE
Seasonal Dummies	26482.6	30550.0	14.035
Seasonal Dummies + AR(1)	24806.6	28615.1	14.690
Seasonal Dummies + Holiday Dummy + AR(1)	24601.6	27928.4	14.462
Cyclical trend	24701.6	34651.9	18.092
Cyclical trend + AR(1)	23933.2	31291.8	15.404

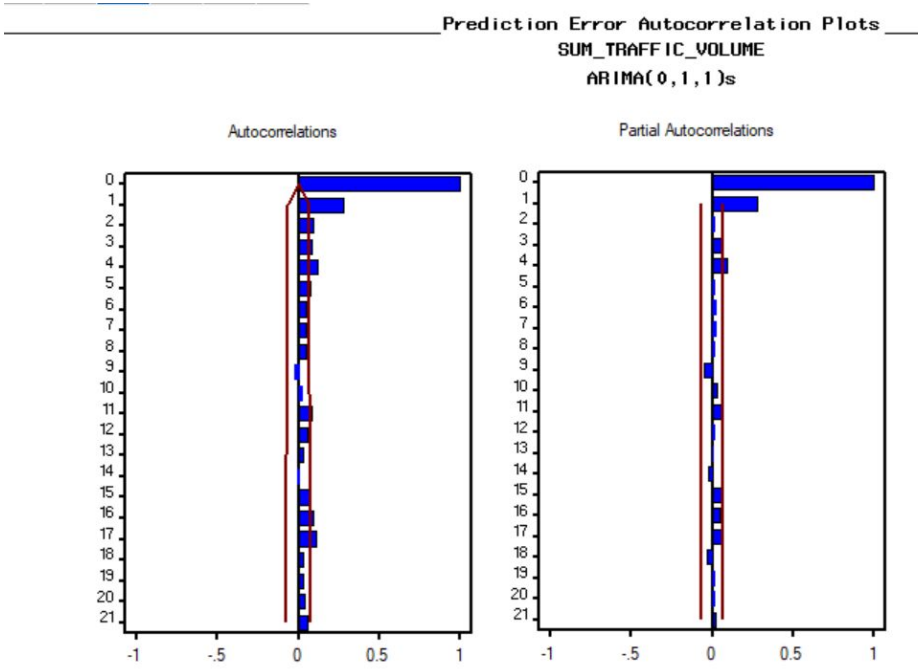
Seasonal Dummies + Holiday Dummy + AR(1) model has the lowest validation error thus, it is selected to be compared against the ARIMA model.

Model Building - ARIMA



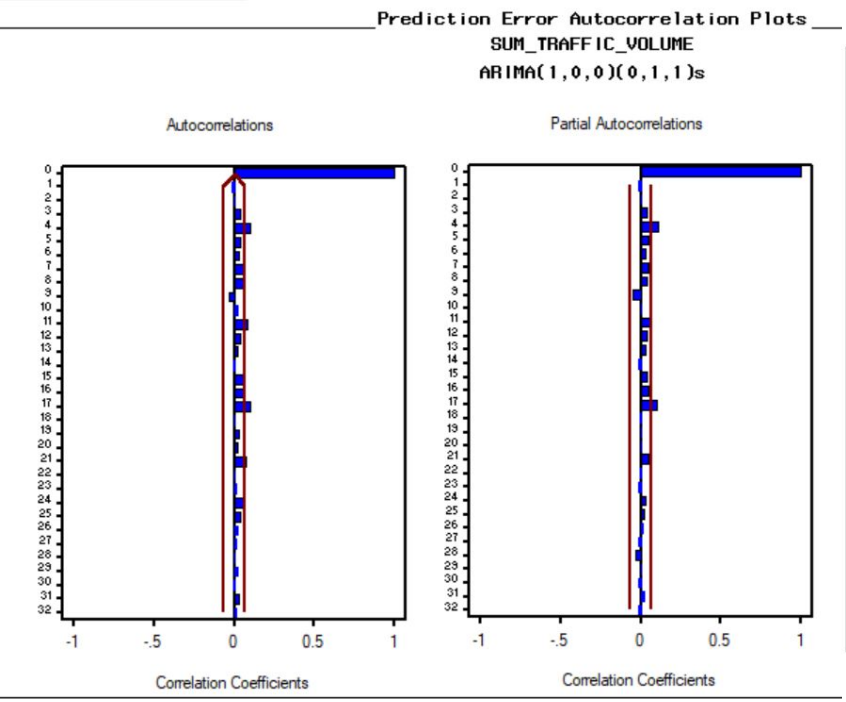
P = 0, D = 1, Q = 1
ARIMA(0,0,0)(0,1,1)

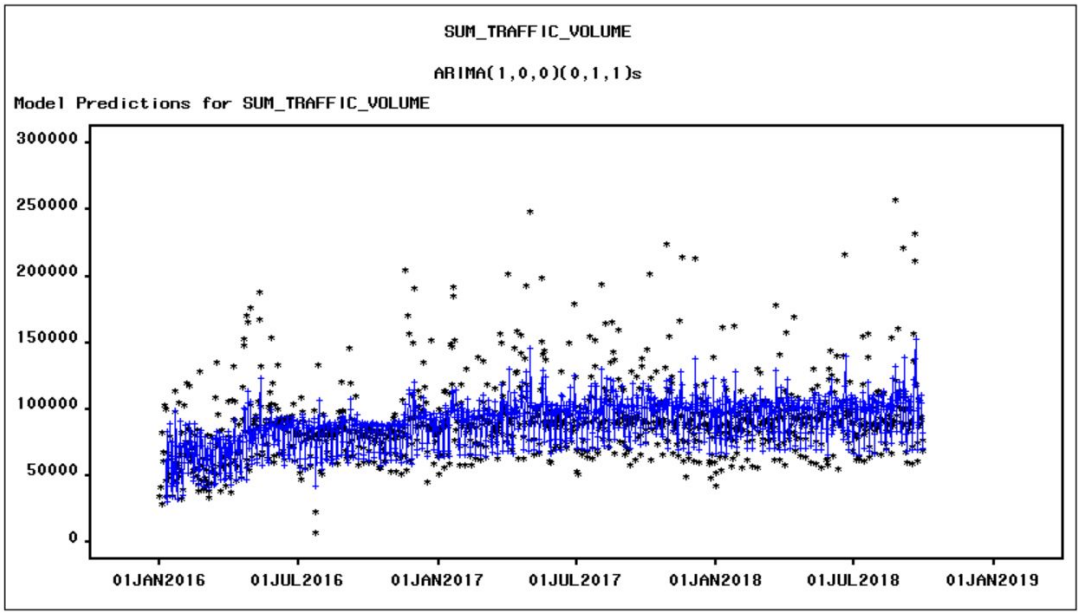
Model Building - ARIMA



- ACF decays over lags
- PACF chopped off after lag 1
- An AR(1) is selected for the non-seasonal part.
- SARIMA(1,0,0)(0,1,1)

Model Building - ARIMA





WN Residuals. All coefficients are significant. -- Reasonable SARIMA model.

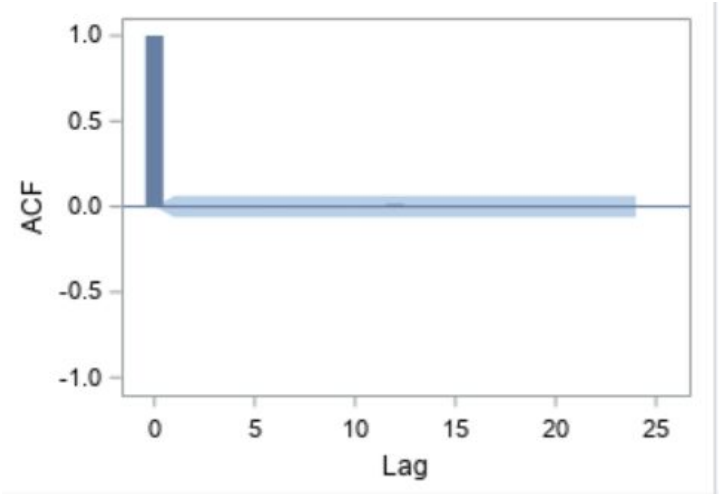
Model Building - Univariate Models Comparison

		Test Error	
	Model Fit (Square Root of Model Variance)	Validation RMSE	Validation MAPE
Seasonal Dummies + Holiday Dummy + AR(1)	24601.6	27928.4	14.462
ARIMA(1,0,0)(0,1,1)s	24602.3	27903.0	17.811

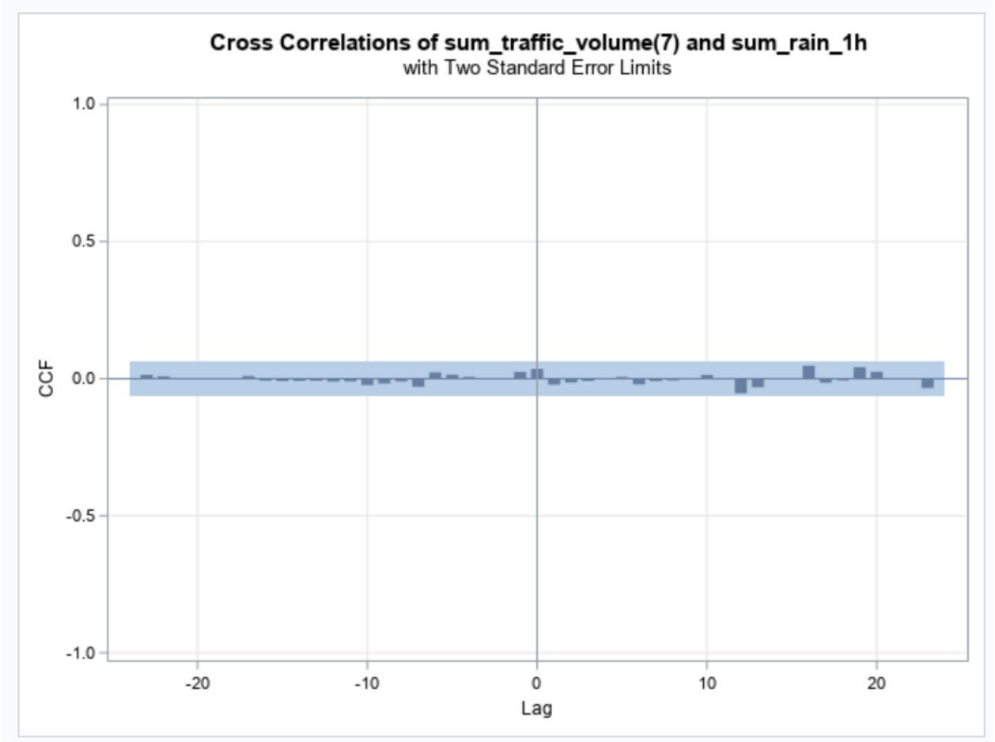
The **Seasonal Dummies + Holiday Dummy + AR(1)** is the better model based on MAPE, as RMSE and Model fit performance of the two models are quite similar.

Model Building - TF Model Variables

Sum_rain:

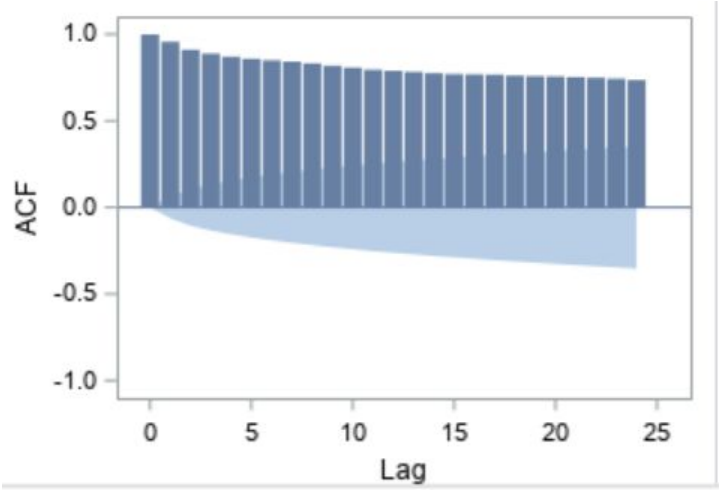


Unrelated to the response variable -- not included in the TF model.



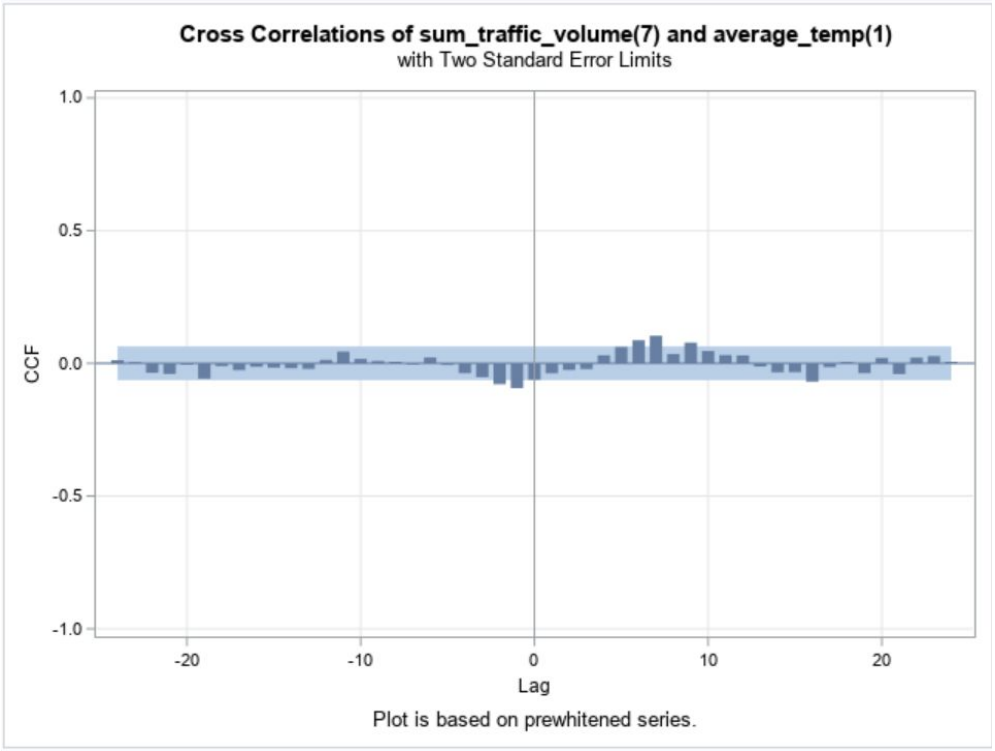
Model Building - TF Model Variables

Average-Temp: RW series



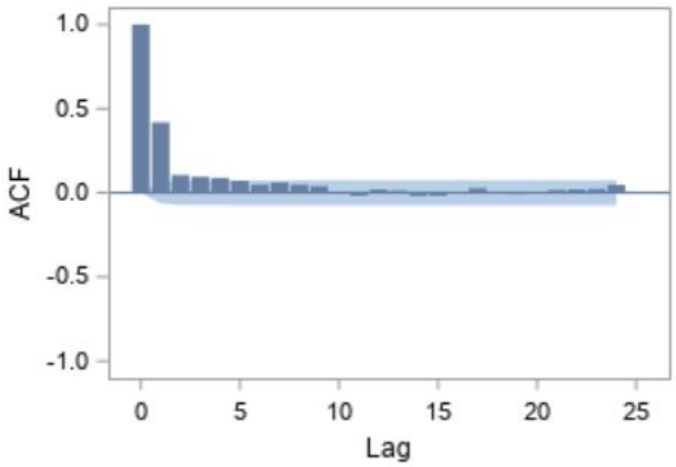
Prewhitening: First difference +ARIMA(1,1,2)

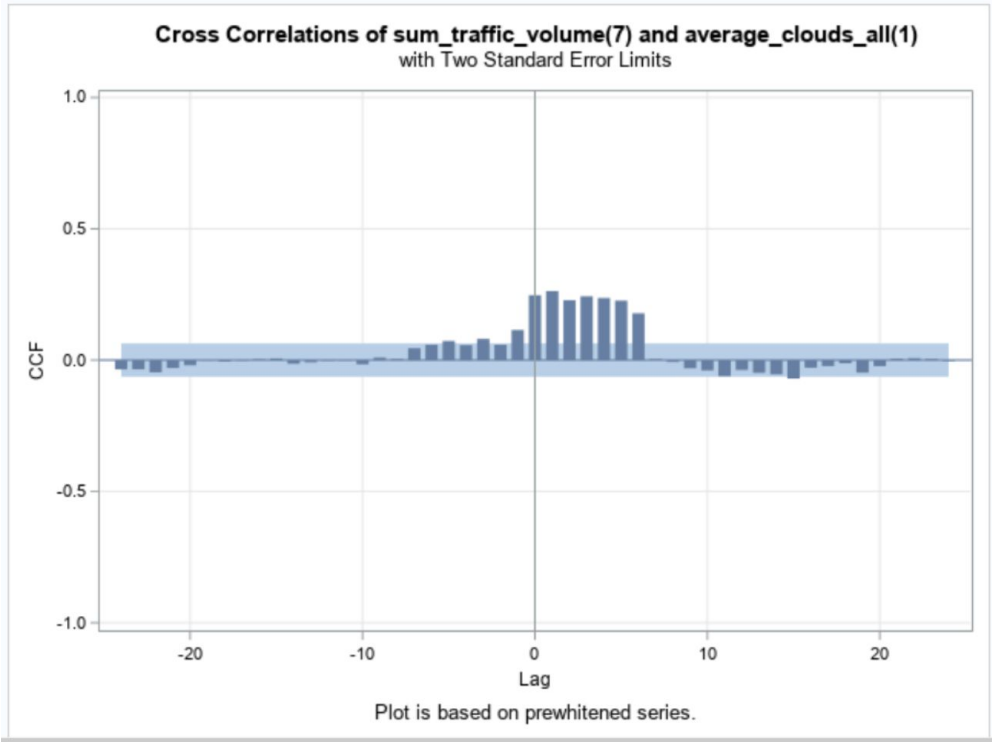
CCF shows very weak response, thus it is not included in the final model.



Model Building - TF Model Variables

Average_clouds_all:

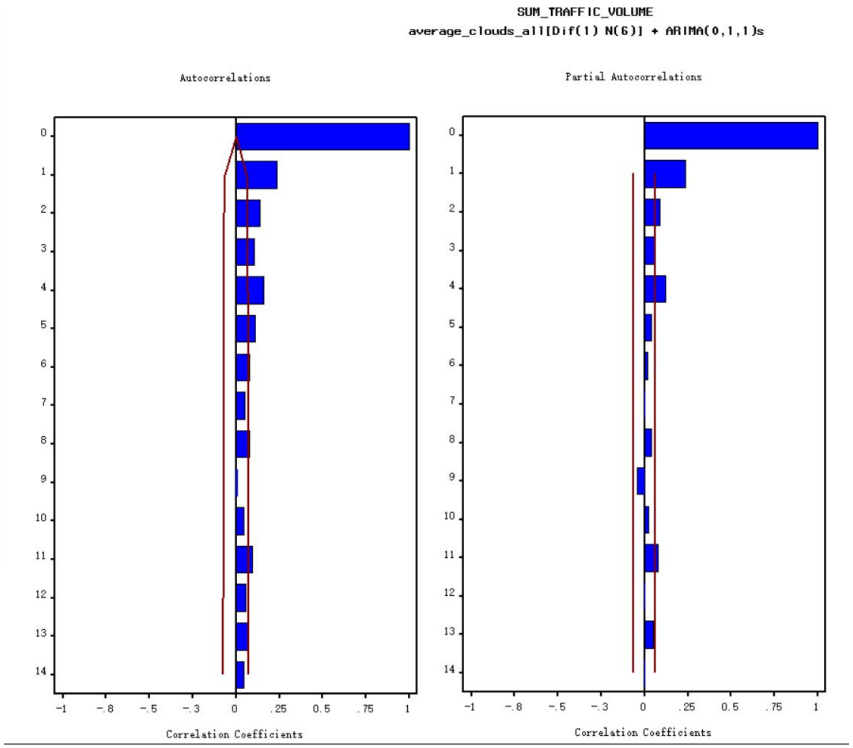




Prewhitening: First difference+MA(2)
TF: b = 0, s = 6, r = 0

Model Building - TF Model

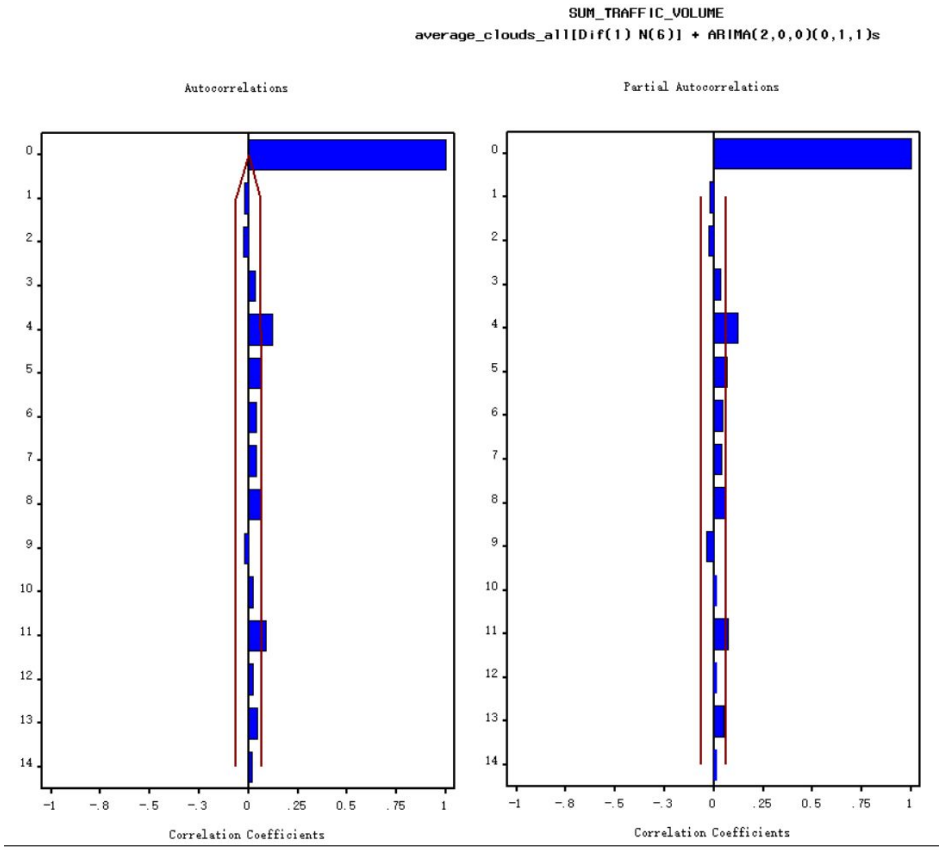
Crosscorrelation Check of Residuals with Input average_clouds_all									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	.	0	.	-0.001	0.009	-0.048	-0.017	0.030	-0.045
11	17.31	5	0.0039	-0.024	-0.040	-0.003	-0.093	-0.030	0.001
17	23.72	11	0.0140	0.024	-0.041	-0.033	-0.025	0.048	0.011
23	32.09	17	0.0147	0.009	-0.058	0.018	0.059	-0.026	0.022
29	33.75	23	0.0688	-0.027	-0.017	-0.004	-0.021	0.012	-0.007
35	43.61	29	0.0399	-0.024	-0.039	-0.038	0.034	-0.021	-0.069
41	49.36	35	0.0545	0.025	0.003	-0.014	-0.034	-0.031	-0.053
47	51.67	41	0.1228	0.022	-0.035	-0.022	0.010	0.002	-0.005



The CCF of the residuals with input indicates WN: adequate TF model.

The ACF of the prediction residual suggests an AR(2) noise model.

Model Building - TF Model

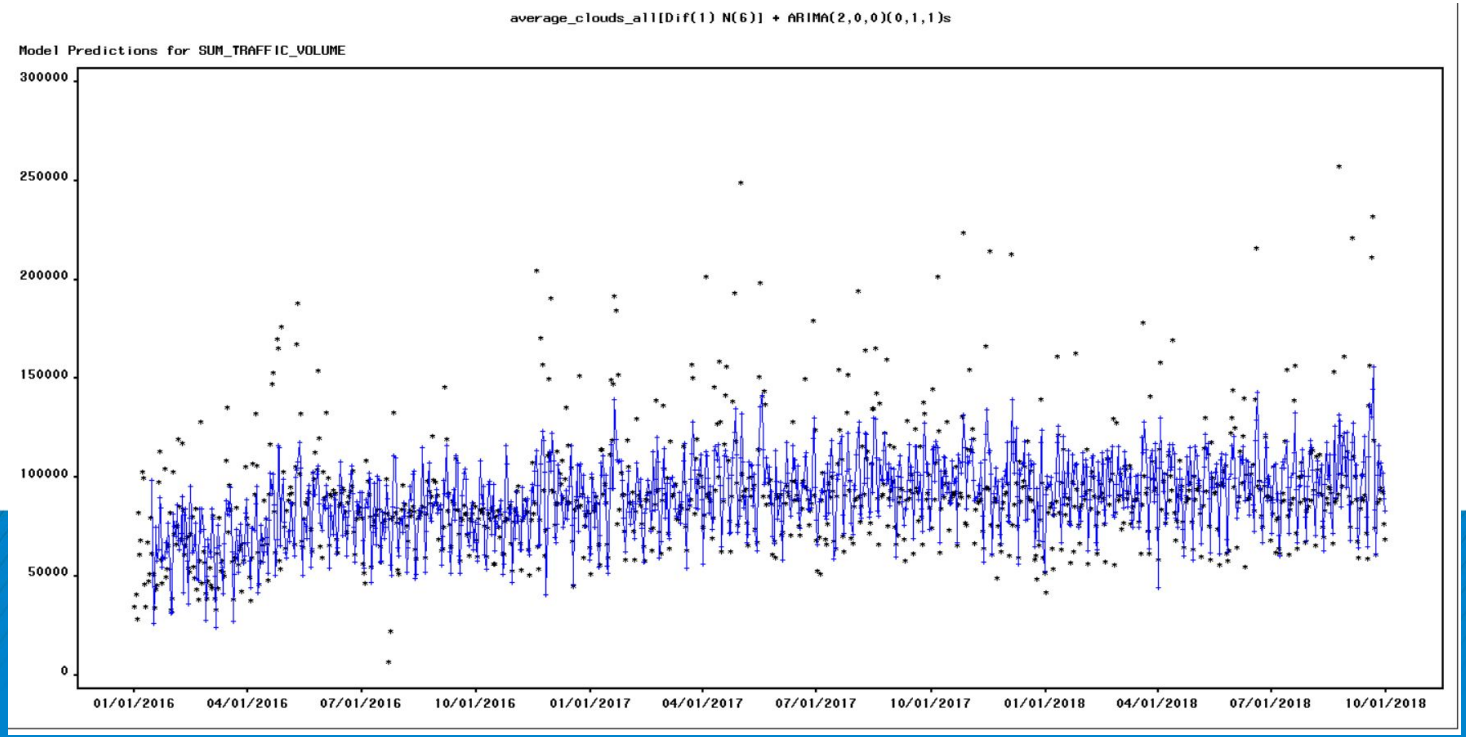


- Included Variable: Average_clouds_all
- b = 0, s = 6, r = 0;
- TF Noise: AR(2);
- Response Variable: Seasonal Difference + seasonal MA(1)
- WN Residual.

Summary

	Model Fit (Square Root of Model Variance)	Test Error	
		Validation RMSE	Validation MAPE
Seasonal Dummies + Holiday Dummy + AR(1)	24601.6	27928.4	14.462
ARIMA(1,0,0)(0,1,1)s	24602.3	27903.0	17.811
TF Average_clouds_all + ARIMA(2,0,0)(0,1,1)	22206.6	24061.9	14.737

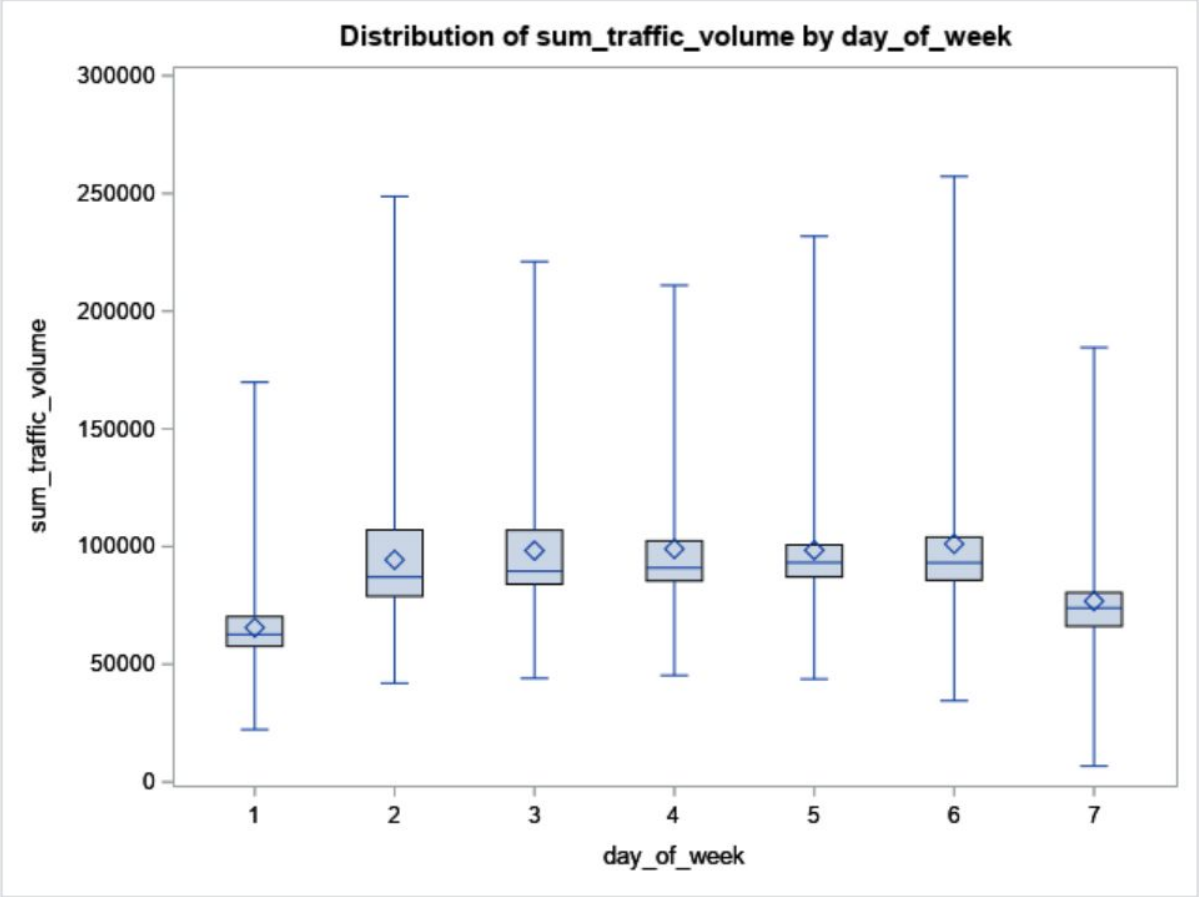
TF Model Plot:



Thank you for listening!

Any questions?

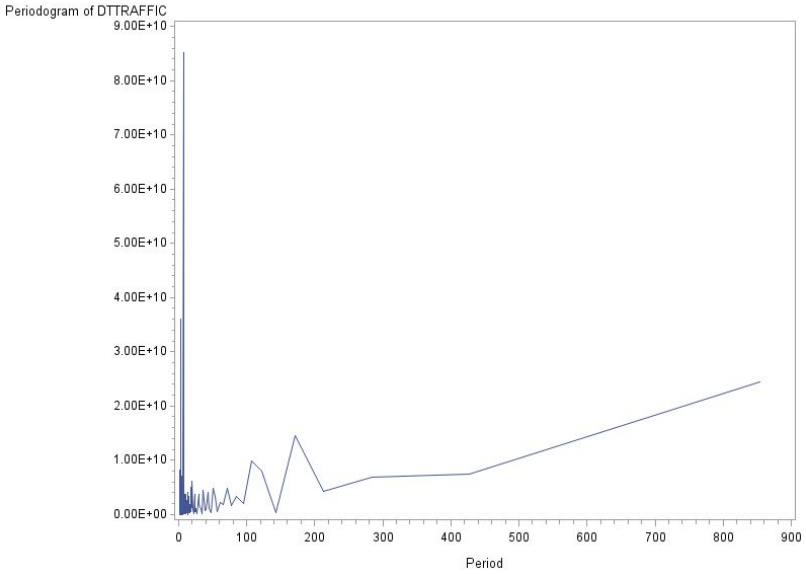
Appendix



1 = Sun, 2 = Mon, 3 = Tue, 4 = Wed, 5 = Thu, 6 = Fri, 7 = Sat

Appendix

Cyclical Trend Model parameters:



Obs	FREQ	PERIOD	P_01
123	0.8976	7	85108978893
245	1.7952	3.5	36098375968
2	0.00736	854	24422551155
6	0.03679	170.8	14586431983
9	0.05886	106.75	9843827894
367	2.69279	2.333	8209374529

	Training RMSE	Validation RMSE
Seasonal Dummies	26482.6	30550.0
Seasonal Dummies + AR(1)	24806.6	28615.1
Cyclical trend	24701.6	34651.9
Cyclical trend + AR(1)	23933.2	31291.8

Parameter Estimates

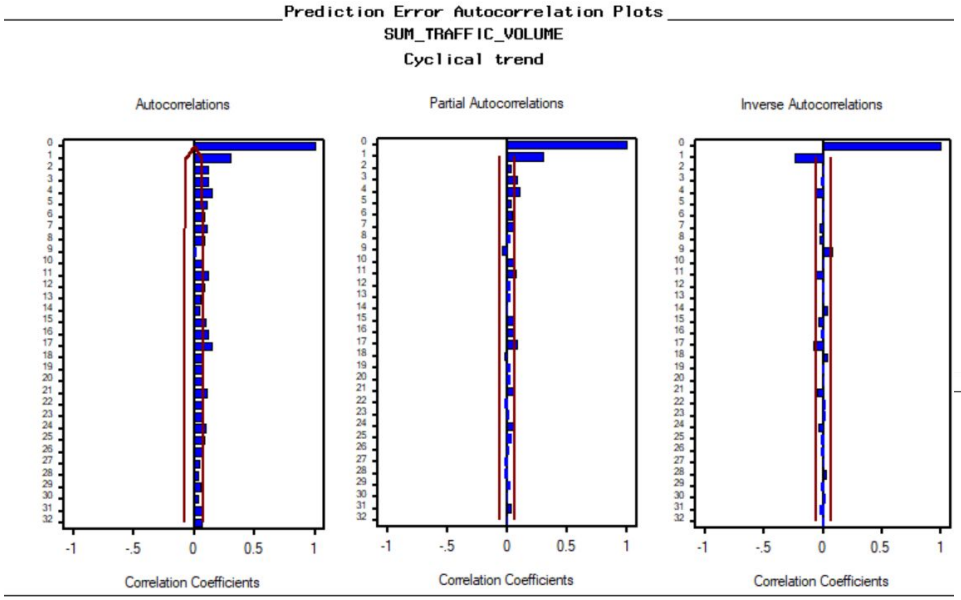
SUM_TRAFFIC_VOLUME

Cyclical + AR(1)

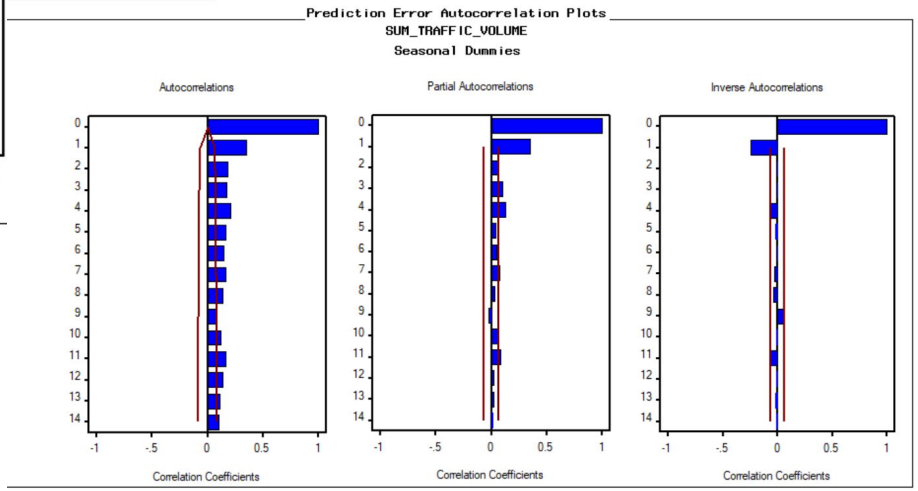
Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	89101	1092	81.6051	<.0001
Autoregressive, Lag 1	0.25024	0.0334	7.4915	<.0001
COS122	11068	1336	8.2824	<.0001
SIN122	-8800	1337	-6.5824	<.0001
COS244	-552.43426	1069	-0.5167	0.6062
SIN244	9166	1069	8.5745	<.0001
COS1	-7562	1543	-4.8994	<.0001
SIN1	-7968	1545	-5.1580	<.0001
COS5	1774	1543	1.1496	0.2523
SIN5	-7070	1544	-4.5783	<.0001
COS8	-2459	1542	-1.5941	0.1132
SIN8	3192	1544	2.0678	0.0406
COS366	-2819	941.8096	-2.9937	0.0033
SIN366	-3362	941.4113	-3.5718	0.0005
Model Variance (sigma squared)	572796829	.	.	.

Appendix

Cyclical Trend Model Error Plot:



Seasonal Dummies Model Error Plot:



Appendix

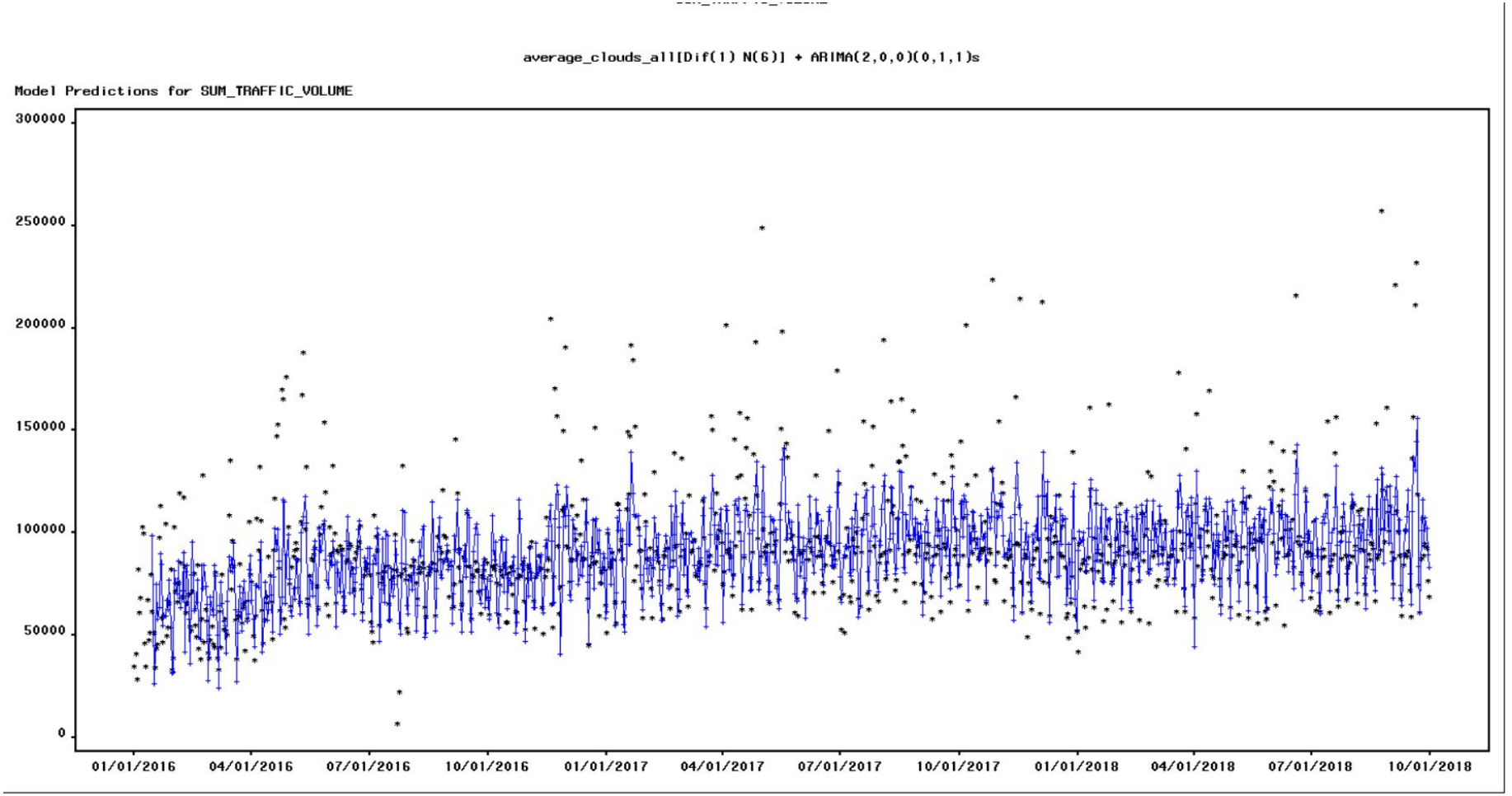
ARIMA Model Parameters::

Parameter Estimates

SUM_TRAFFIC_VOLUME

ARIMA(1,0,0)(0,1,1)s

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	194.43804	54.2452	3.5844	0.0005
Seasonal Moving Average, Lag 7	0.96974	0.0130	74.3782	<.0001
Autoregressive, Lag 1	0.29895	0.0326	9.1603	<.0001
Model Variance (sigma squared)	605275484	.	.	.


average_clouds_all[Dif(1) N(6)] + ARIMA(2,0,0)(0,1,1)s
Model Predictions for SUM_TRAFFIC_VOLUME
300000
250000
200000
150000
100000
50000
0
01/01/2016
04/01/2016
07/01/2016
10/01/2016
01/01/2017
04/01/2017
07/01/2017
10/01/2017
01/01/2018
04/01/2018
07/01/2018
10/01/2018

Appendix

TF Model Parameters:

SUM_TRAFFIC_VOLUME
average_clouds_all[Dif(1) N(6)] + ARIMA(2,0,0)(0,1,1)s

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	185.09308	54.7228	3.3824	0.0009
Seasonal Moving Average, Lag 7	0.96718	0.0130	74.2983	<.0001
Autoregressive, Lag 1	0.24423	0.0343	7.1302	<.0001
Autoregressive, Lag 2	0.09931	0.0344	2.8866	0.0045
AVERAGE_CLOUDS_ALL[Dif(1) N(6)]	381.27275	28.3474	13.4500	<.0001
AVERAGE_CLOUDS_ALL[Dif(1) N(6)] Num1	-345.70706	33.6672	-10.2684	<.0001
AVERAGE_CLOUDS_ALL[Dif(1) N(6)] Num2	-239.06781	39.1503	-6.1064	<.0001
AVERAGE_CLOUDS_ALL[Dif(1) N(6)] Num3	-203.36687	40.4027	-5.0335	<.0001
AVERAGE_CLOUDS_ALL[Dif(1) N(6)] Num4	-182.99626	39.1345	-4.6761	<.0001
AVERAGE_CLOUDS_ALL[Dif(1) N(6)] Num5	-106.92739	33.6646	-3.1763	0.0018
AVERAGE_CLOUDS_ALL[Dif(1) N(6)] Num6	-55.80491	28.1802	-1.9803	0.0496
Model Variance (sigma squared)	493131144	.	.	.