Appliances Energy

Sarah Gates

DATS 6450 11 – Time Series Analysis & Modeling

Final Project Presentation

Agenda

- Introduction
- About the Data
- Time Series EDA
- Model Creation
- Final Model selection

Appliances Energy Prediction

Goal: Build a model using the algorithms learned this semester to predict energy usage in Kilowatt Hour (Wh) of appliances of a low energy house

- Metrics logged: temperature and humidity in various rooms of the house and outdoor conditions
- This house is located in Belgium. Outdoor conditions were recorded at the Chèvres Airport weather station near by
- The metrics were collected every 10 minutes (19,735 data points)
  - Refactored the data to 30-minute intervals (6579 data points)
    - Took the sum of Appliance usage
    - Took the average of metrics logged

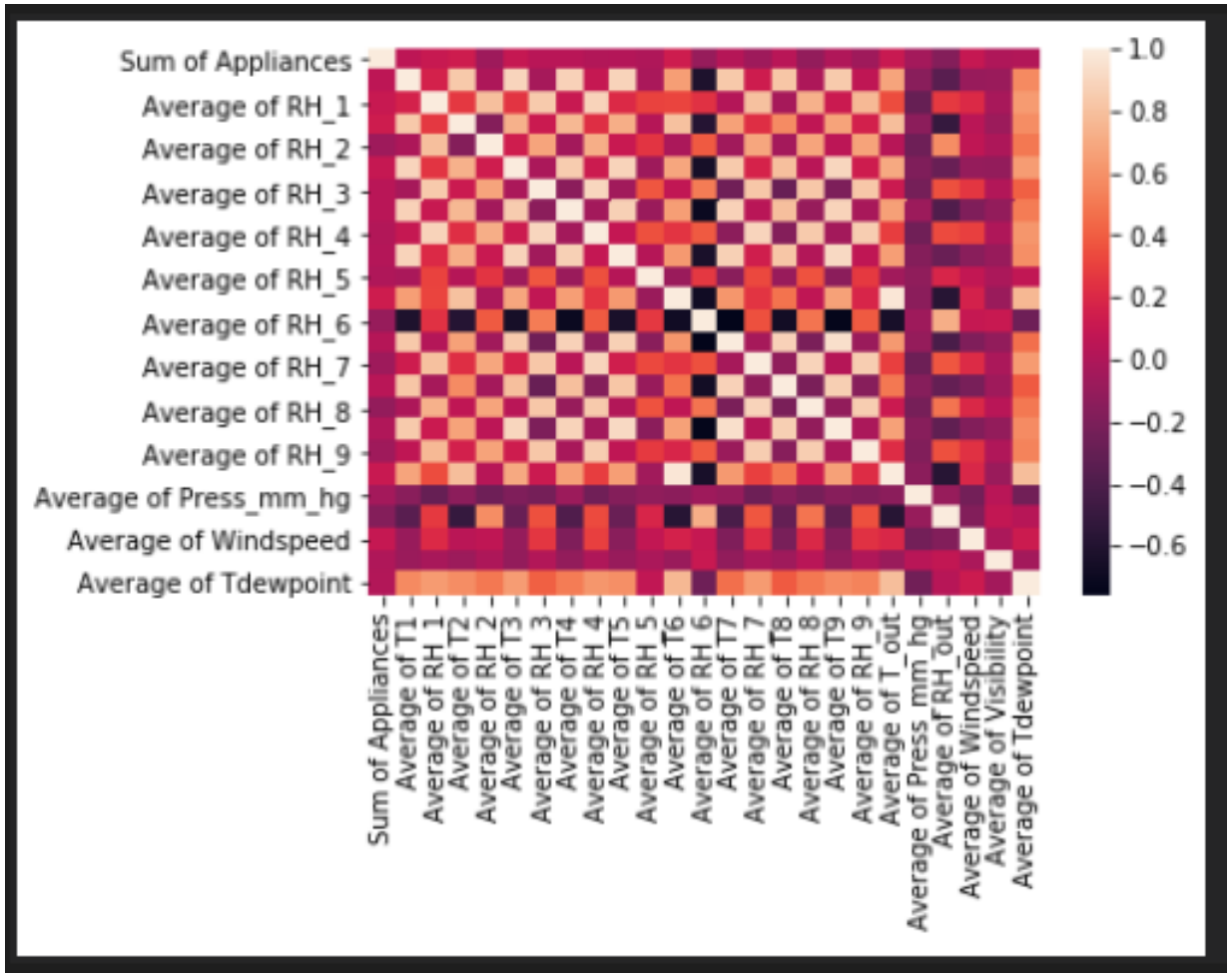
About the Data

Variable	Description
date	date of recording
half_hour	half hour interval of recording
Sum of Appliances	energy use in Wh
Sum of lights	energy use of light fixtures in the house in Wh
Average of T1	Temperature in kitchen area, in Celsius
Average of RH_1	Humidity in kitchen area, in %
Average of T2	emperature in living room area, in Celsius
Average of RH_2	Humidity in living room area, in %
Average of T3	Temperature in laundry room area
Average of RH_3	Humidity in laundry room area, in %
Average of T4	Temperature in office room, in Celsius
Average of RH_4	Humidity in office room, in %
Average of T5	Temperature in bathroom, in Celsius
Average of RH_5	Humidity in bathroom, in %
Average of T6	Temperature outside the building (north side), in Celsius
Average of RH_6	Humidity outside the building (north side), in %
Average of T7	Temperature in ironing room , in Celsius
Average of RH_7	Humidity in ironing room, in %
Average of T8	Temperature in teenager room 2, in Celsius
Average of RH_8	Humidity in teenager room 2, in %
Average of T9	Temperature in parents room, in Celsius
Average of RH_9	Humidity in parents room, in %
Average of T_out	Temperature outside (from Chievres weather station), in Celsius
Average of Press_mm_hg	Humidity outside (from Chievres weather station), in %
Average of RH_out	Humidity outside (from Chievres weather station), in %
Average of Windspeed	Wind speed (from Chievres weather station), in m/s
Average of Visibility	Visibility (from Chievres weather station), in km
Average of Tdewpoint	Tdewpoint (from Chievres weather station)

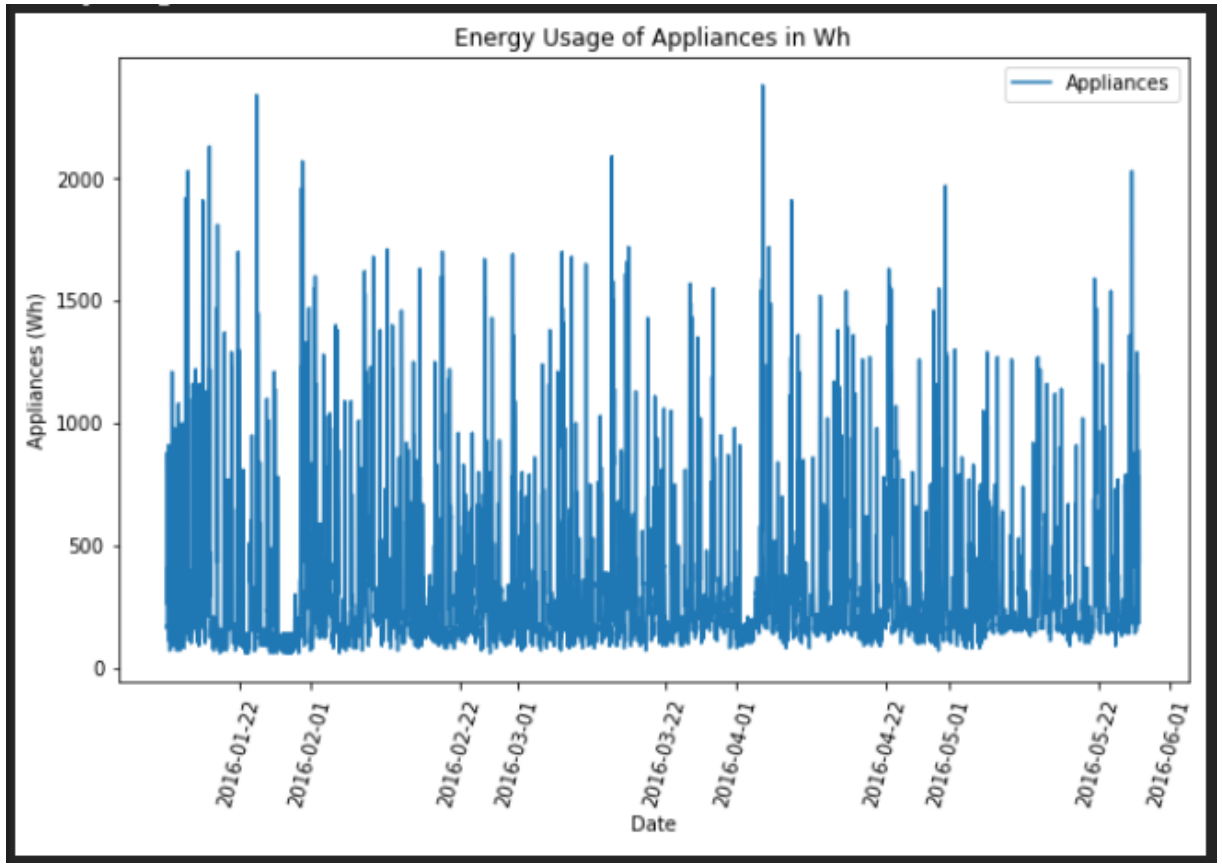
About the Data

	date	half_hour	Sum of Appliances	Average of T1	Average of RH_1	Average of T2	Average of RH_2	Average of T3	Average of RH_3	Average of T4	...	Average of RH_8	Average of T9	Average of RH_9	Average of T_out	Average of Press_mm_hg	Average of RH_out	Average of Windspeed	Average of Visibility	Average of Tdewpoint	datetime
0	1/11/2016	5:00:00 PM	170	19.890000	46.863333	19.200000	44.713056	19.790000	44.817778	18.975556	...	48.831111	17.033333	45.530000	6.483333	733.600000	92.000000	6.666667	59.166667	5.200000	2016-01-11 17:00:00
1	1/11/2016	5:30:00 PM	160	19.890000	46.142222	19.200000	44.540000	19.790000	44.977778	18.890000	...	48.590000	17.000000	45.363333	6.133333	733.900000	92.000000	5.666667	47.666667	4.900000	2016-01-11 17:30:00
2	1/11/2016	6:00:00 PM	180	19.845556	45.641389	19.200000	44.477778	19.750000	44.863333	18.890000	...	48.590000	17.000000	45.290000	5.916667	734.166667	91.833333	5.166667	40.000000	4.683333	2016-01-11 18:00:00
3	1/11/2016	6:30:00 PM	880	19.950000	46.116667	19.337778	44.400000	19.790000	44.863333	18.926667	...	48.604444	16.963333	45.290000	5.966667	734.366667	91.333333	5.666667	40.000000	4.633333	2016-01-11 18:30:00
4	1/11/2016	7:00:00 PM	780	20.273333	52.206667	19.717778	45.111111	19.937778	45.973333	19.000000	...	48.806667	16.914444	45.320556	6.000000	734.616667	90.500000	6.000000	40.000000	4.516667	2016-01-11 19:00:00

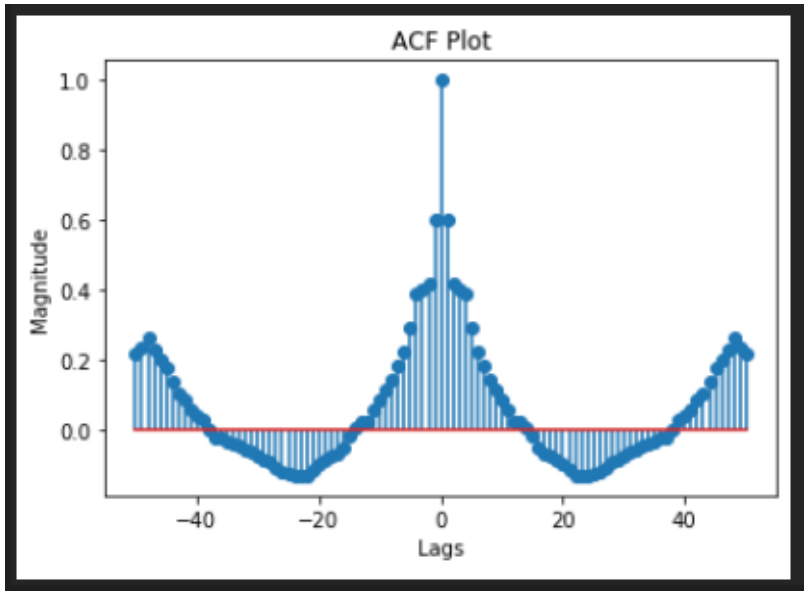
Correlation Matrix



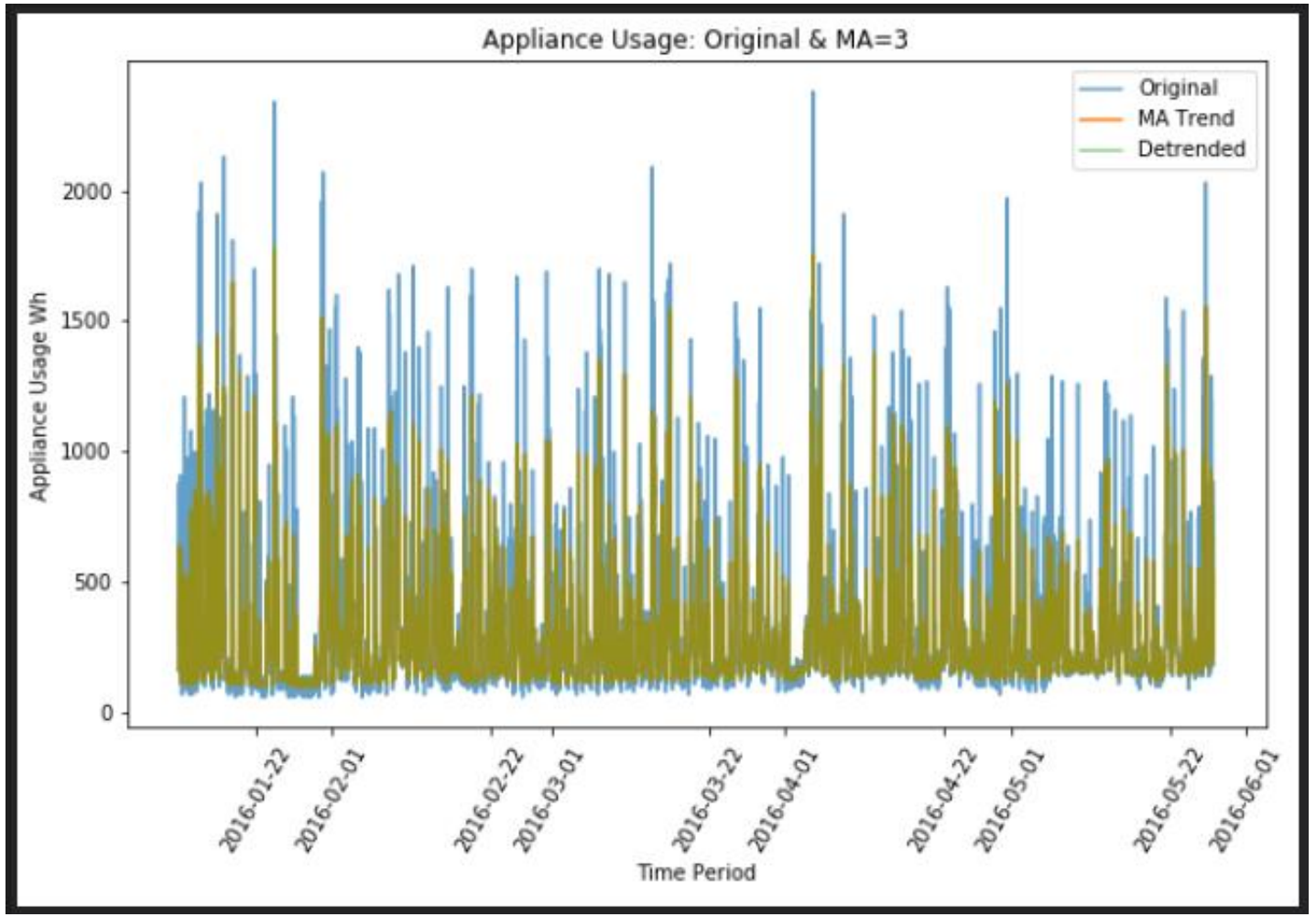
Stationary Analysis



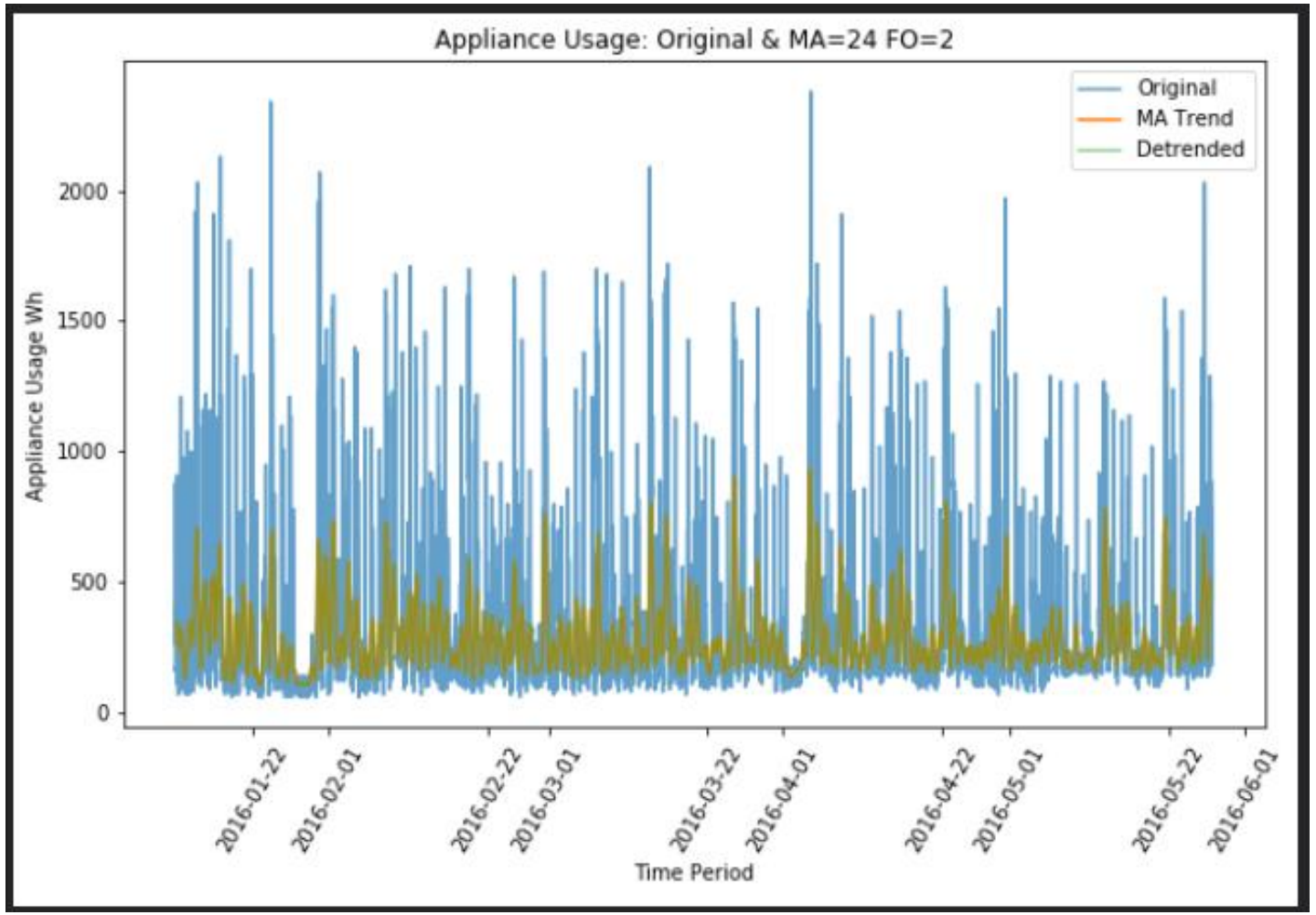

```
ADF Statistic: -18.473443
p-value: 0.000000
Critical Values:
        1%: -3.431
        5%: -2.862
        10%: -2.567
```



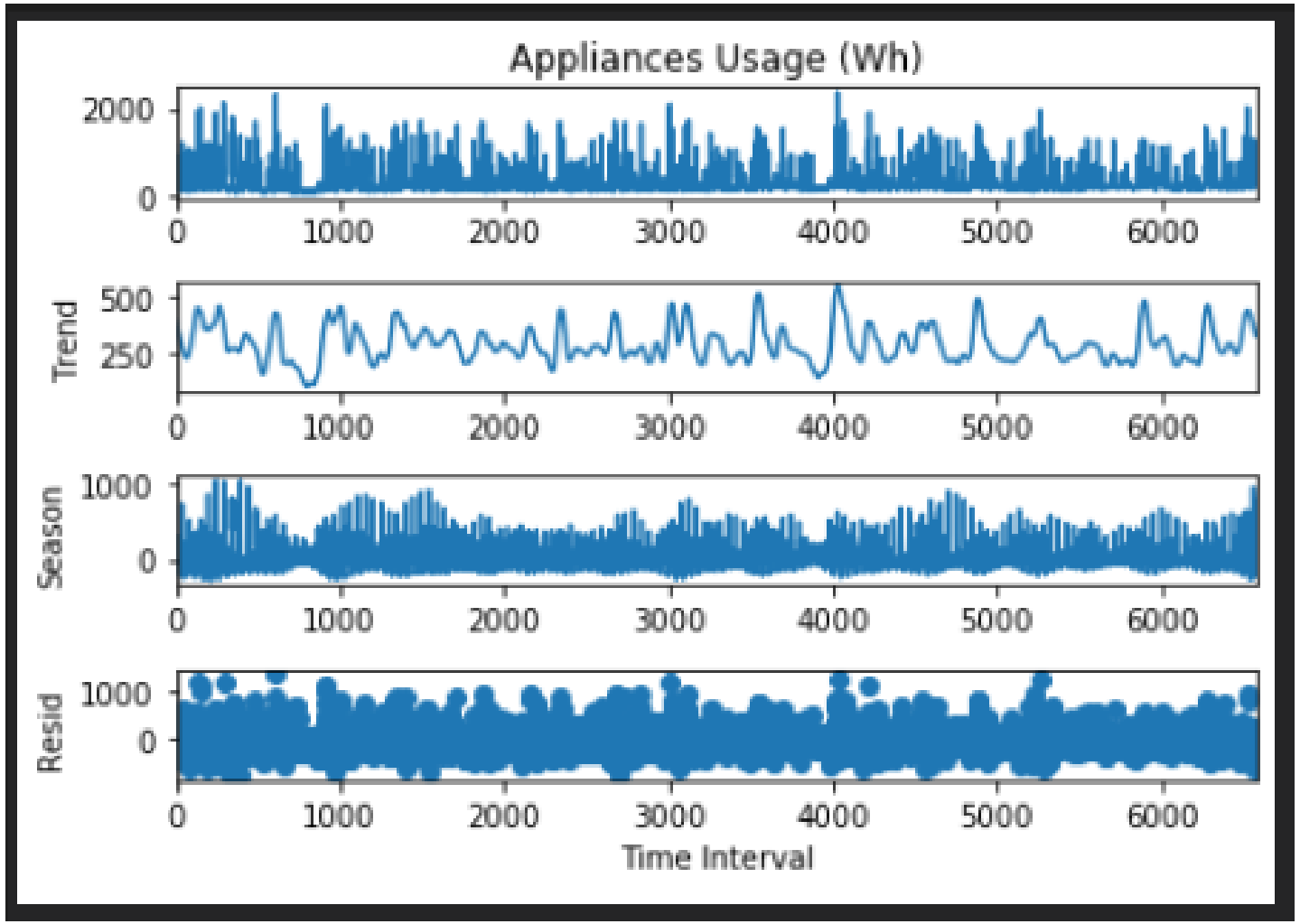
Time Series Decomposition



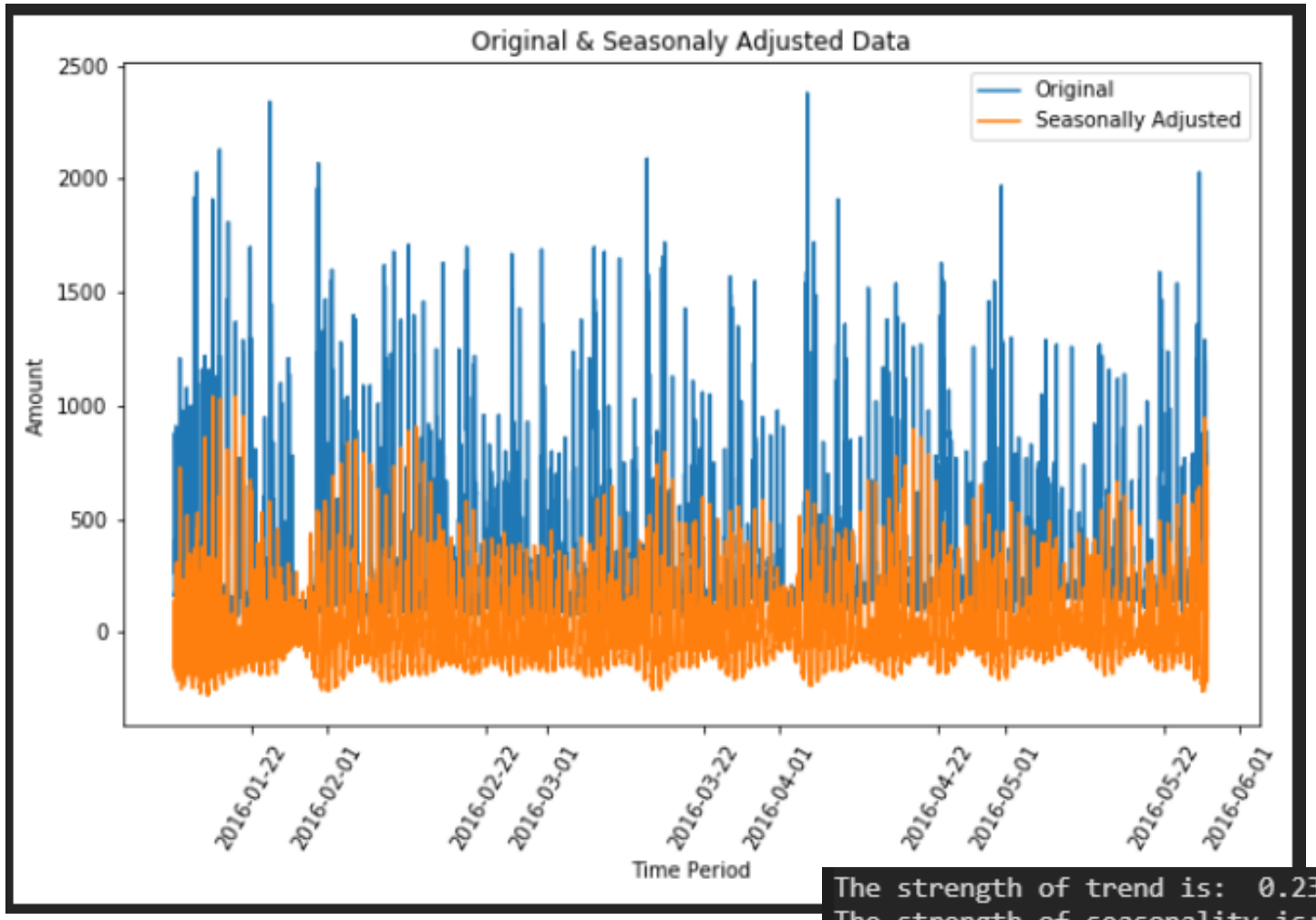
Time Series Decomposition



Time Series Decomposition



Time Series Decomposition

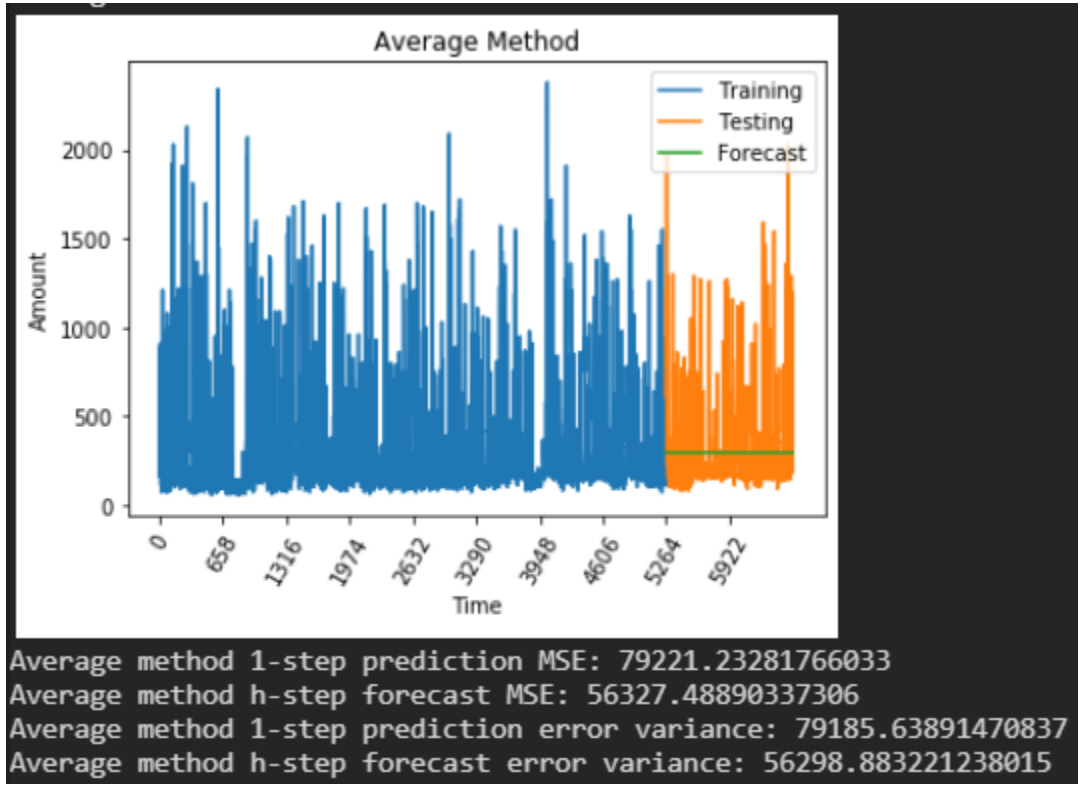



```
The strength of trend is:  0.2309627946897167
The strength of seasonality is:  0.5395603380024478
```

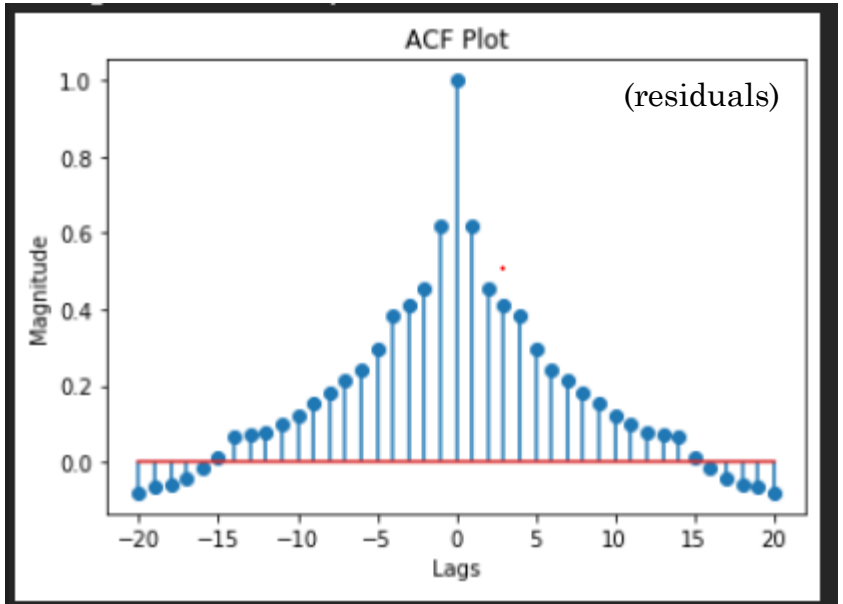
Model Creation

Base Models, Holt-Winters, Multiple Linear Regression, ARMA

Base Models: Average

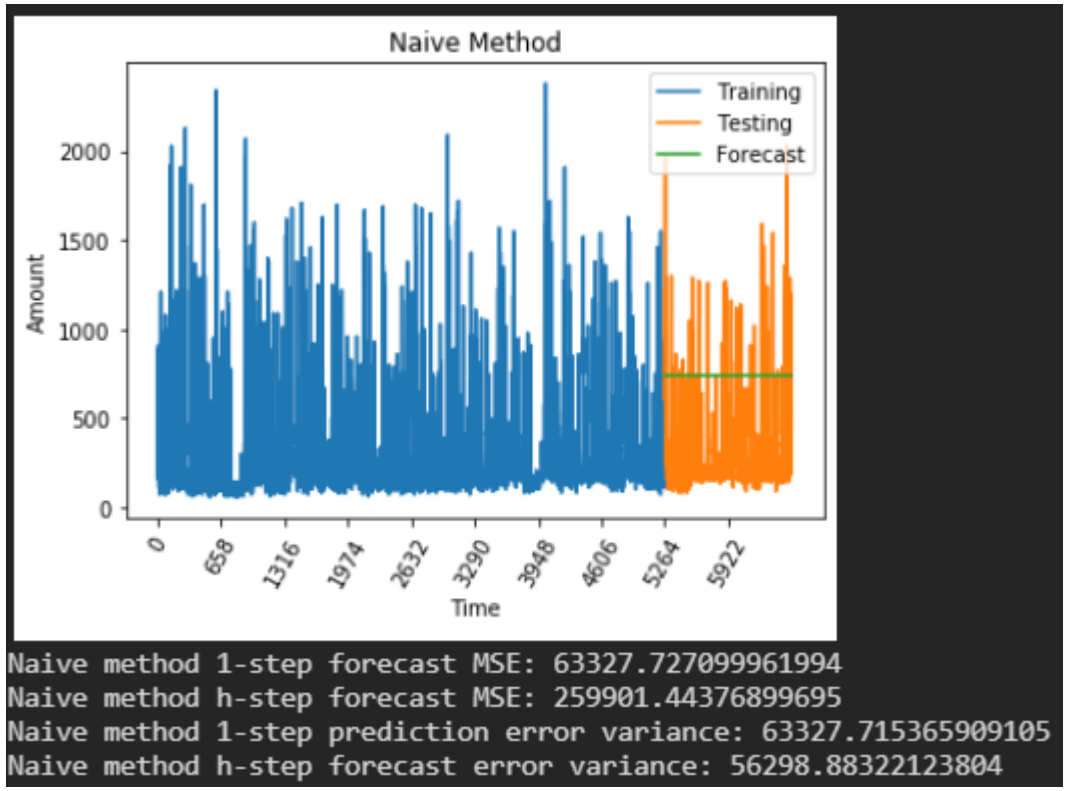


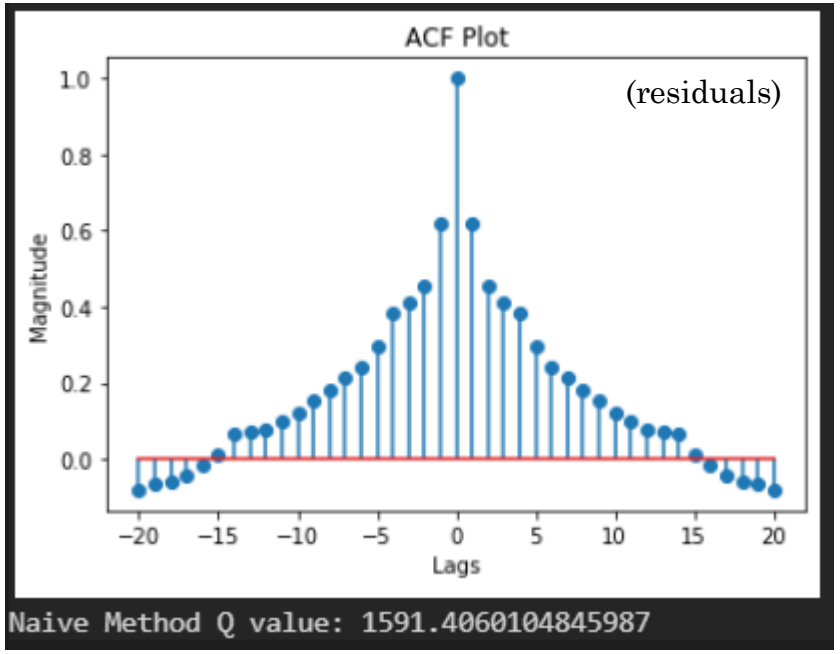

```
Average method 1-step prediction MSE: 79221.23281766033
Average method h-step forecast MSE: 56327.48890337306
Average method 1-step prediction error variance: 79185.63891470837
Average method h-step forecast error variance: 56298.883221238015
```



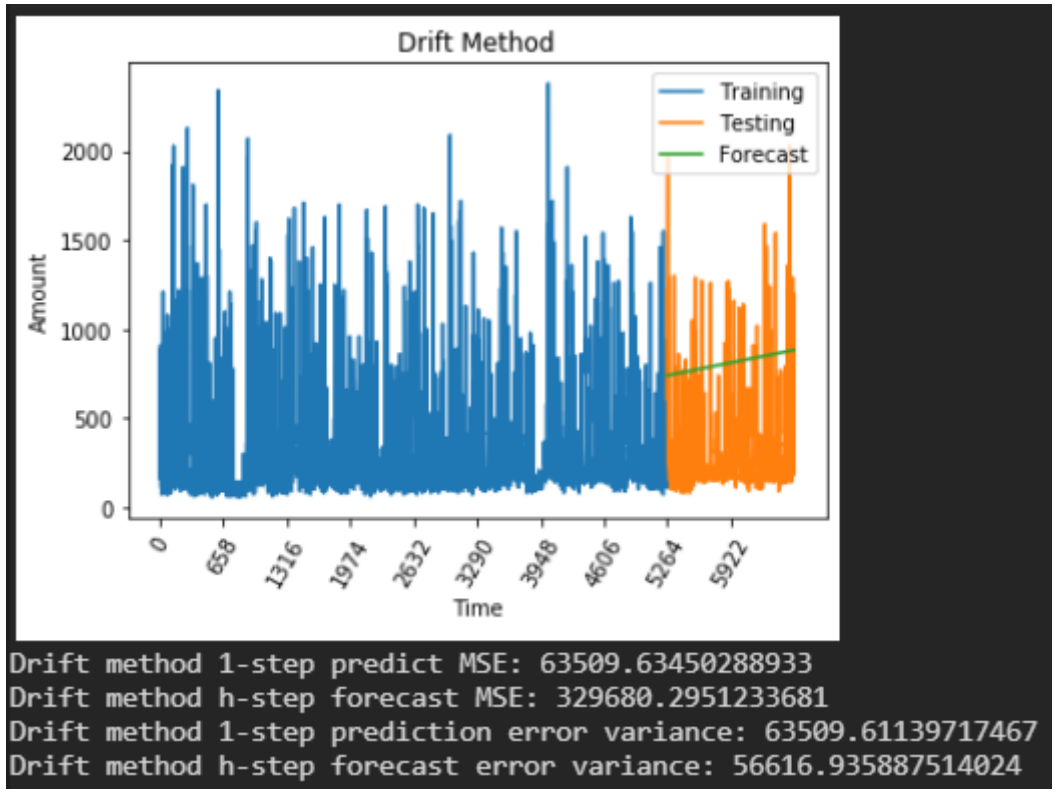

```
Average method Q value: 1591.4060104845985
```

Base Models: Naïve

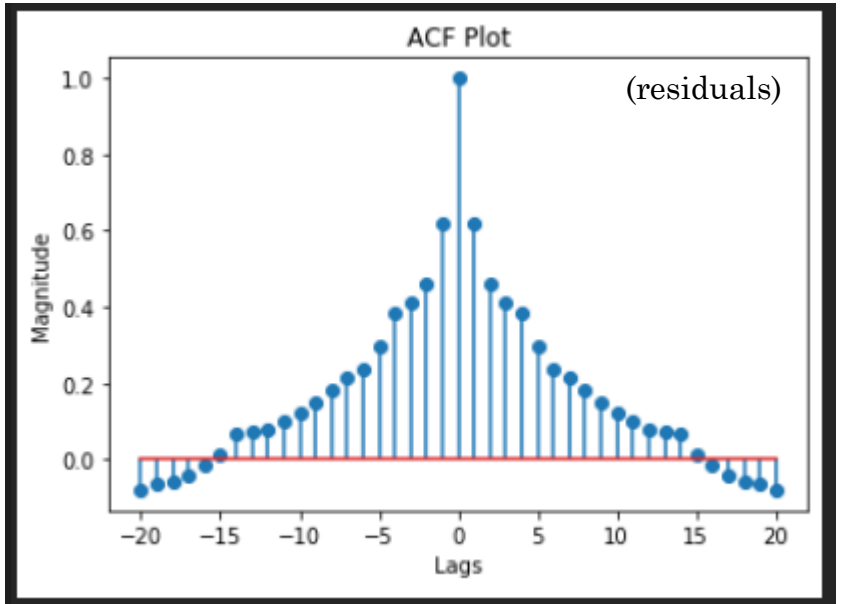




Base Models: Drift

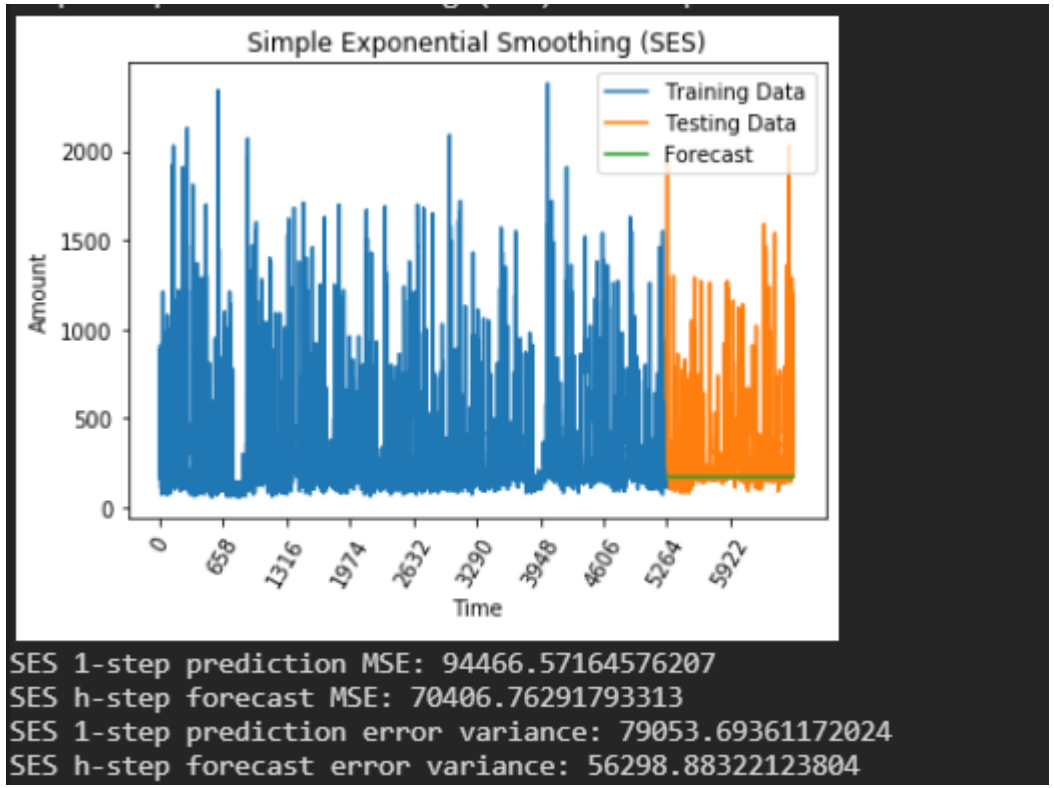


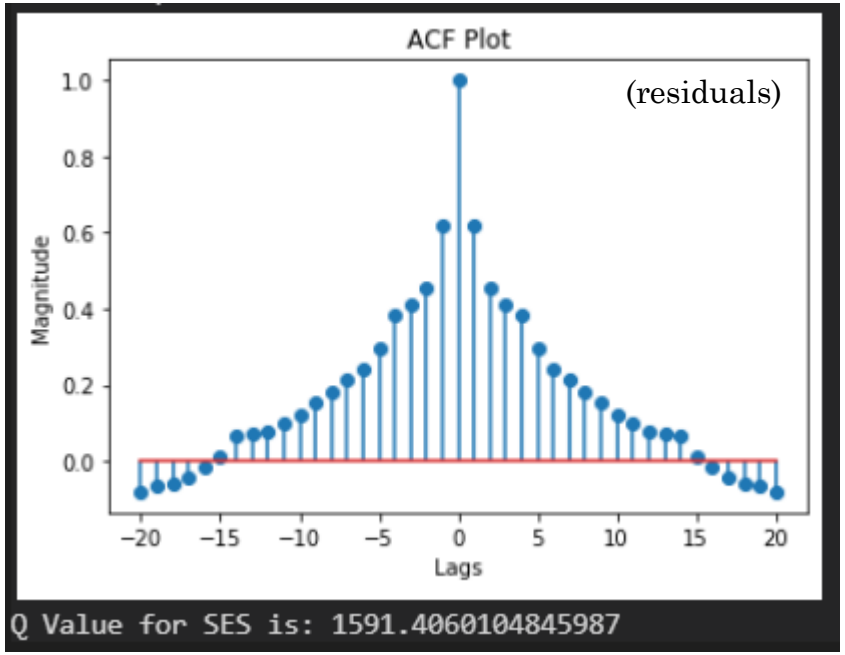

```
Drift method 1-step predict MSE: 63509.63450288933
Drift method h-step forecast MSE: 329680.2951233681
Drift method 1-step prediction error variance: 63509.61139717467
Drift method h-step forecast error variance: 56616.935887514024
```



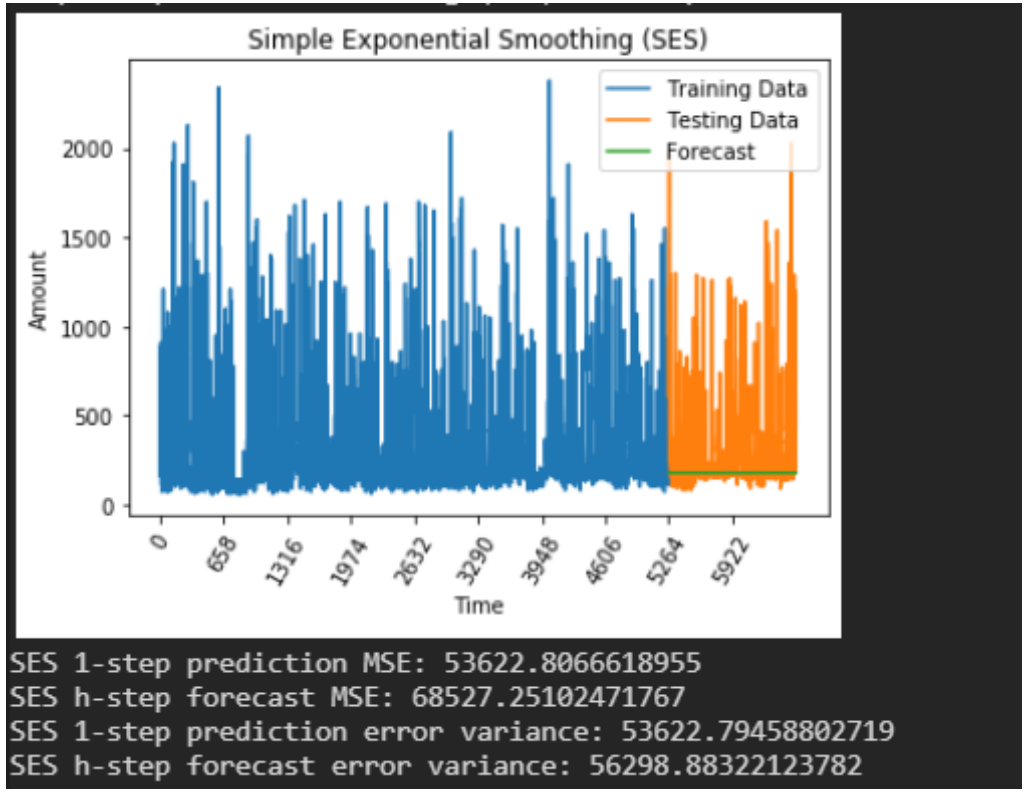

```
Drift method Q Value: 1598.7650305411544
```

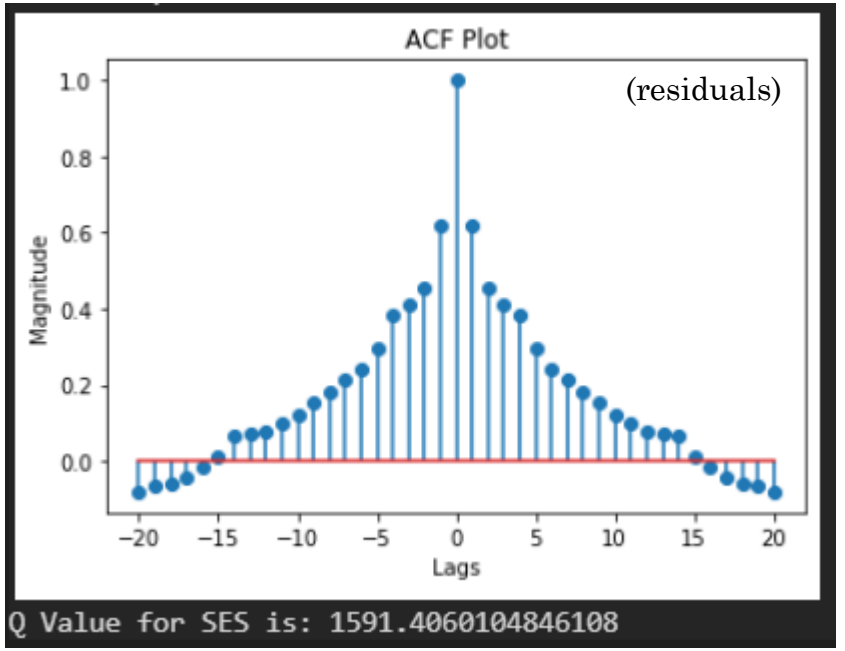
Base Models: SES (alpha=0)



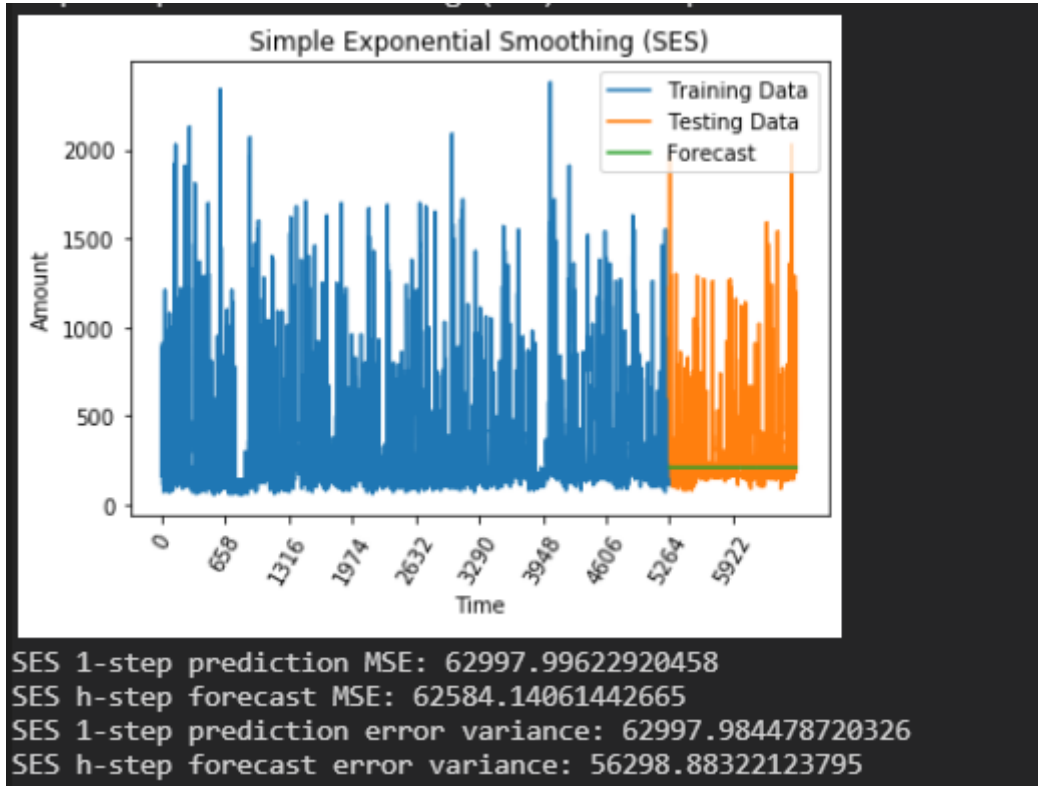


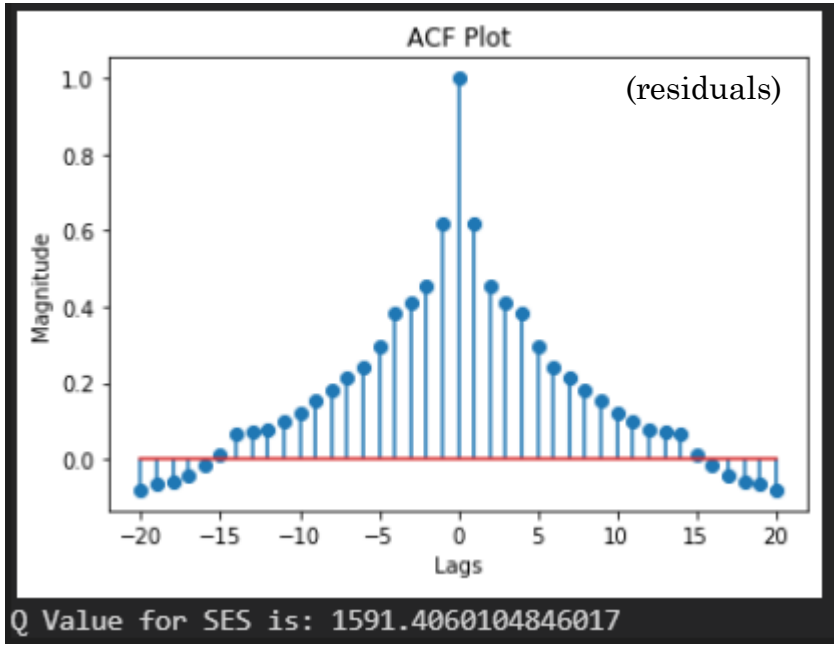
Base Models: SES (alpha=0.5)



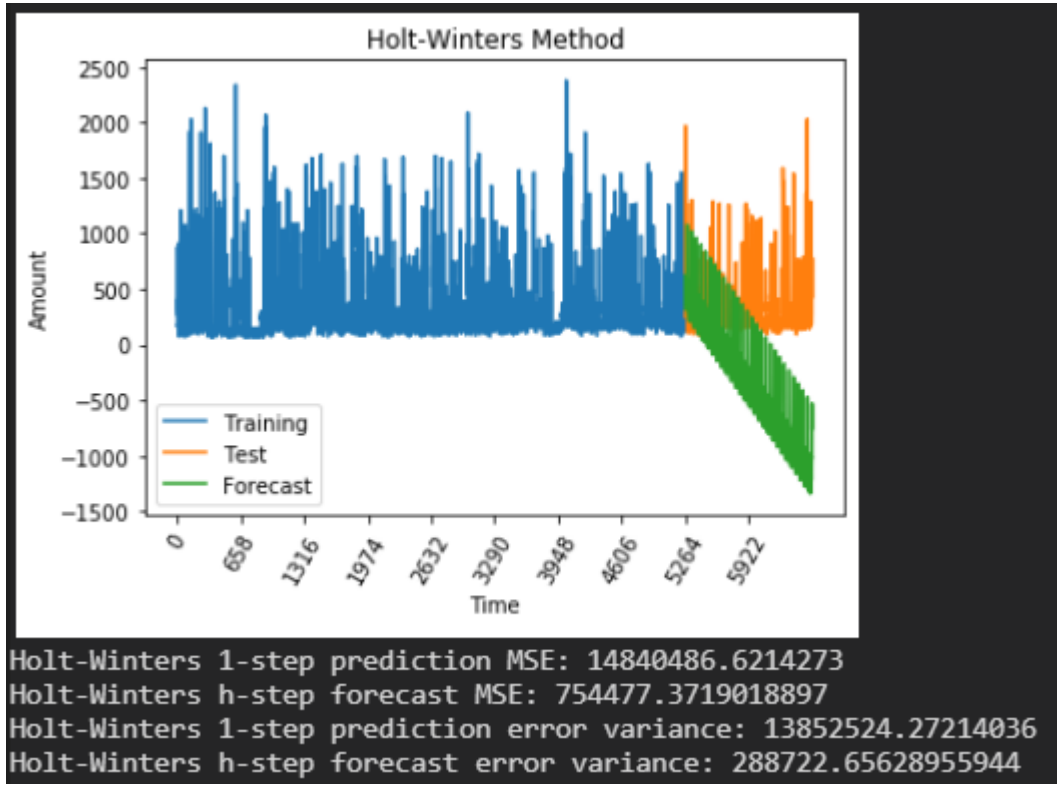


Base Models: SES (alpha=0.99)

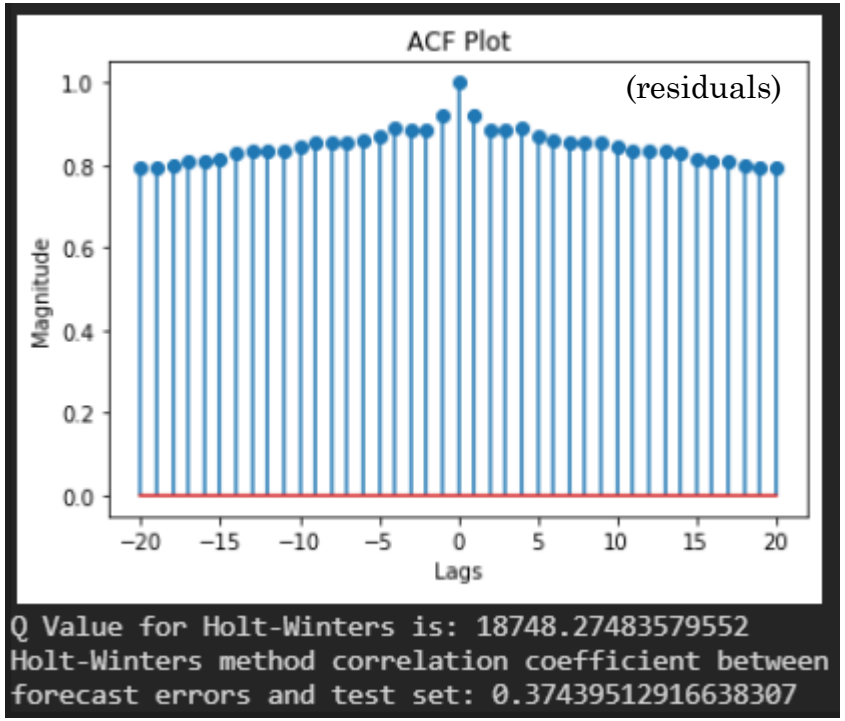




Holt-Winters (seasonal=48, damped=False)

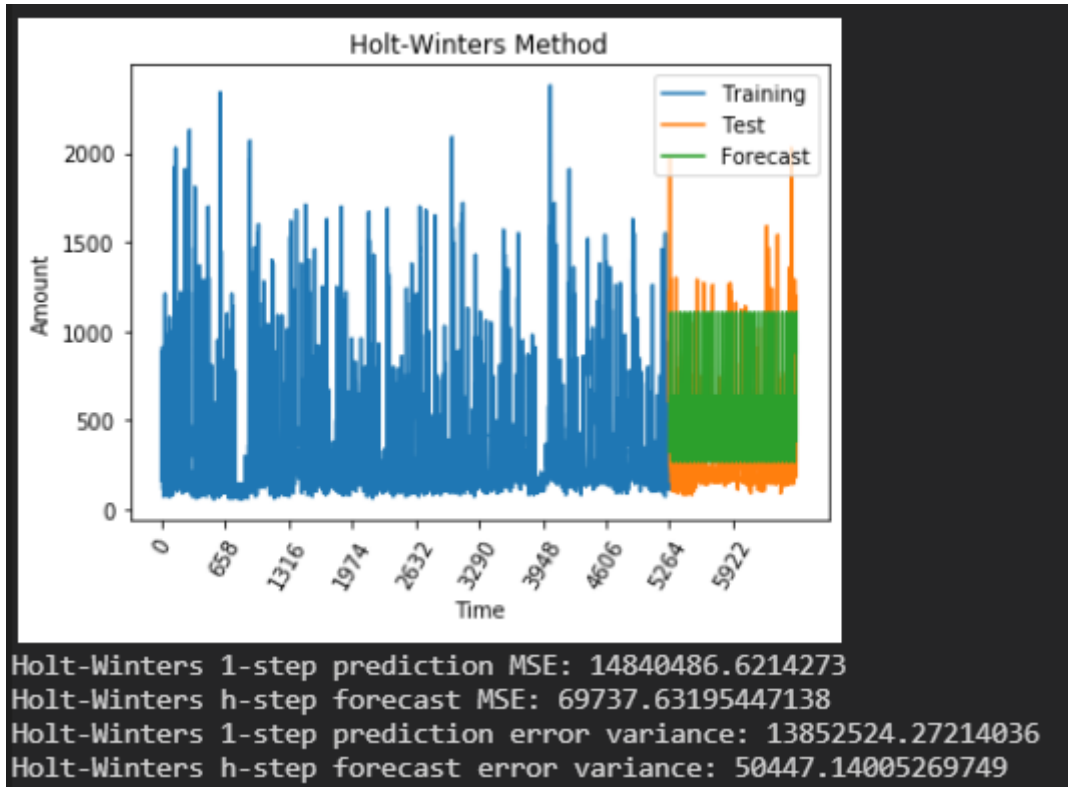



```
Holt-Winters 1-step prediction MSE: 14840486.6214273
Holt-Winters h-step forecast MSE: 754477.3719018897
Holt-Winters 1-step prediction error variance: 13852524.27214036
Holt-Winters h-step forecast error variance: 288722.65628955944
```

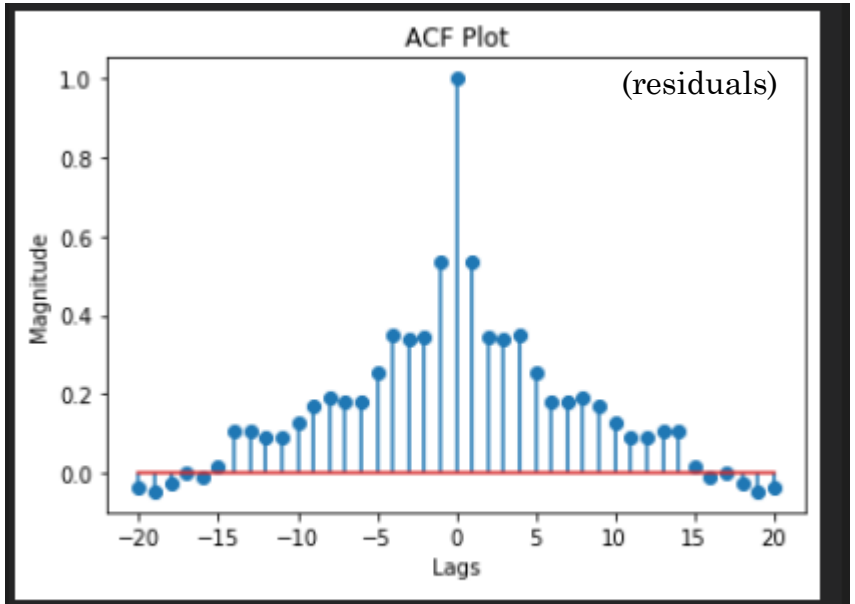



```
Q Value for Holt-Winters is: 18748.27483579552
Holt-Winters method correlation coefficient between
forecast errors and test set: 0.37439512916638307
```

Holt-Winters (seasonal=48, damped=True)

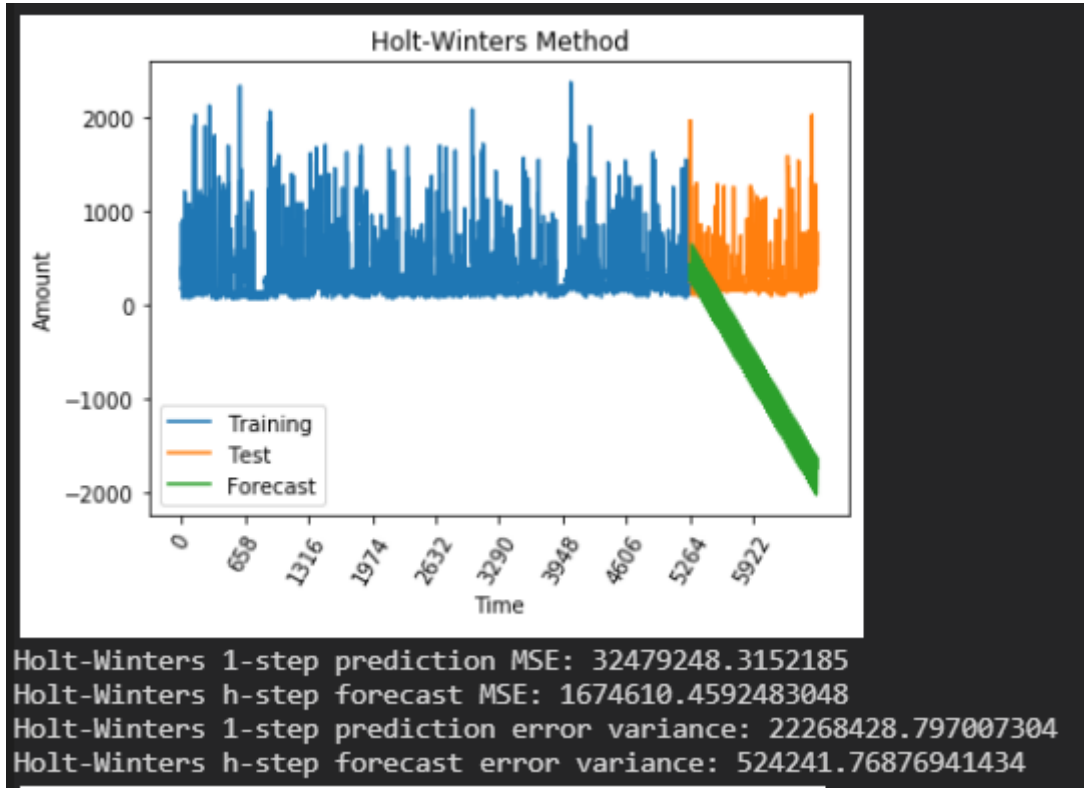



```
Holt-Winters 1-step prediction MSE: 14840486.6214273
Holt-Winters h-step forecast MSE: 69737.63195447138
Holt-Winters 1-step prediction error variance: 13852524.27214036
Holt-Winters h-step forecast error variance: 50447.14005269749
```

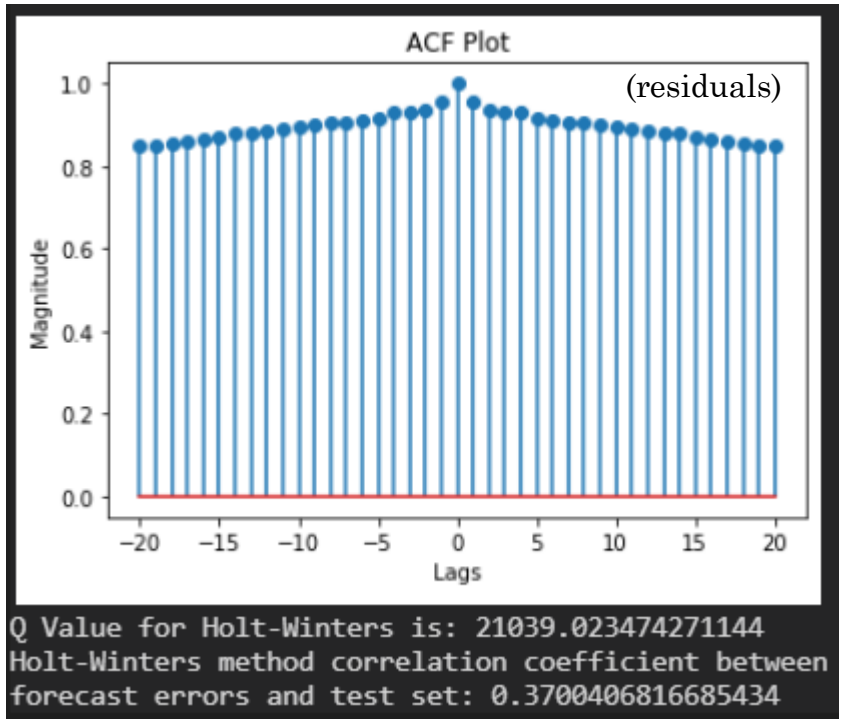



```
Q Value for Holt-Winters is: 1183.9802945365489
Holt-Winters method correlation coefficient between
forecast errors and test set: 0.7476723937687607
```

Holt-Winters (seasonal=24)

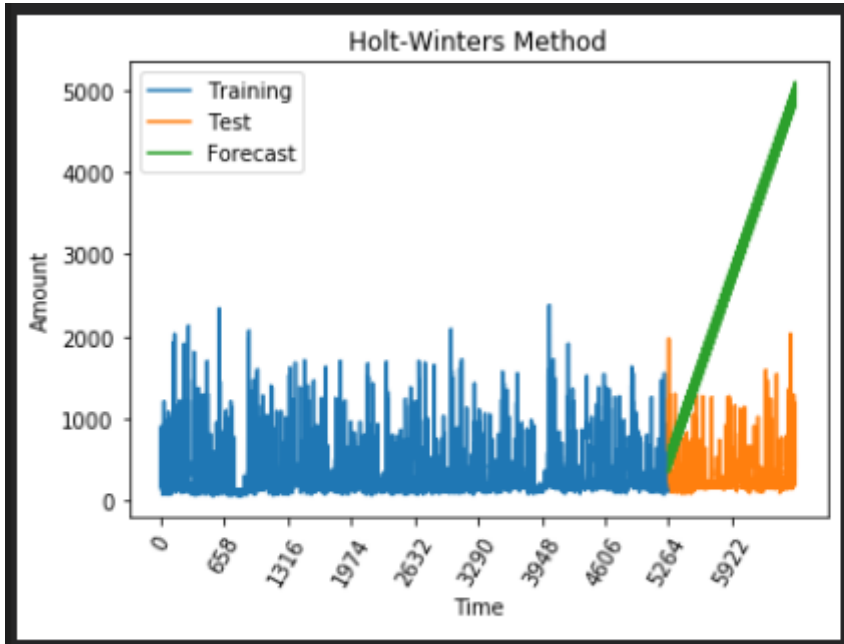



```
Holt-Winters 1-step prediction MSE: 32479248.3152185
Holt-Winters h-step forecast MSE: 1674610.4592483048
Holt-Winters 1-step prediction error variance: 22268428.797007304
Holt-Winters h-step forecast error variance: 524241.76876941434
```

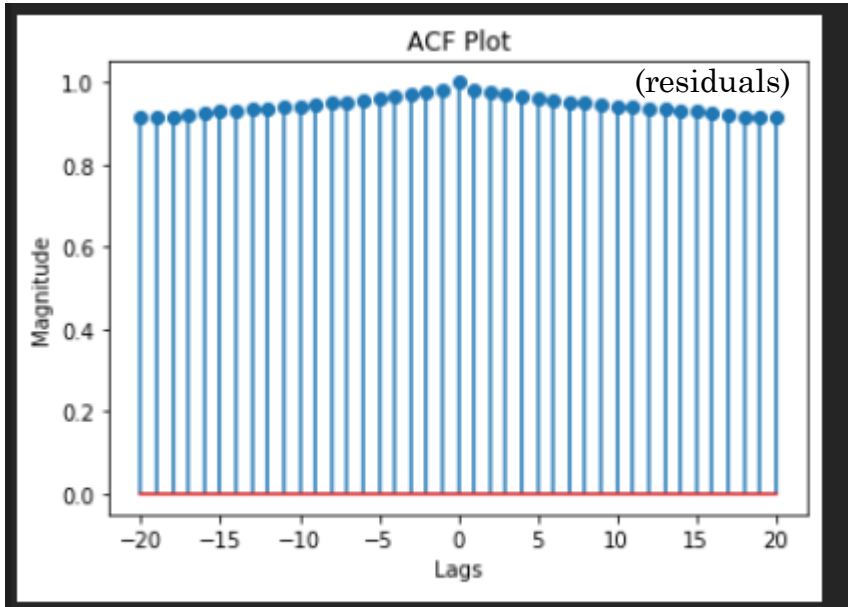



```
Q Value for Holt-Winters is: 21039.023474271144
Holt-Winters method correlation coefficient between
forecast errors and test set: 0.3700406816685434
```

Holt-Winters (seasonal=12)

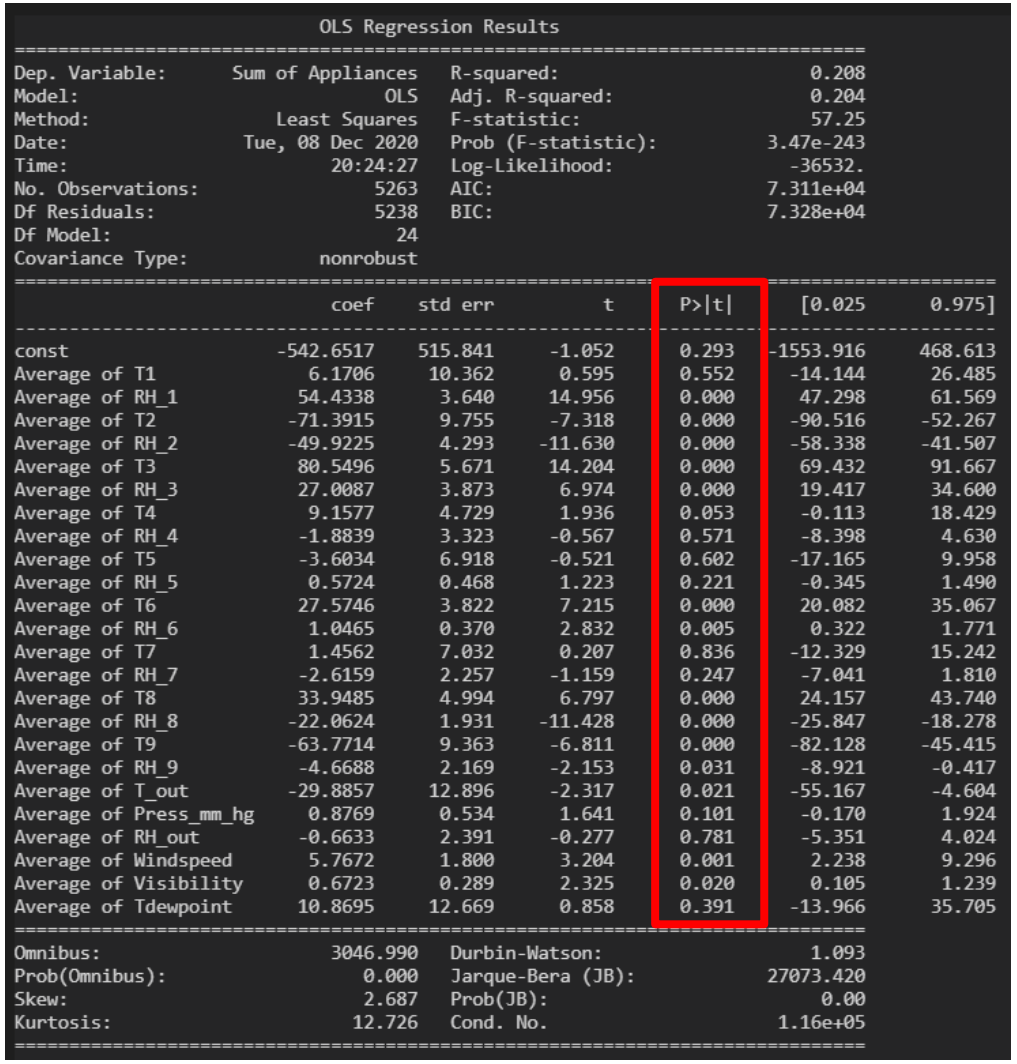



```
Holt-Winters 1-step prediction MSE: 370109233.15393287
Holt-Winters h-step forecast MSE: 7284517.902727775
Holt-Winters 1-step prediction error variance: 319814380.29831034
Holt-Winters h-step forecast error variance: 1682696.470105548
```




```
Q Value for Holt-Winters is: 23414.111719314675
Holt-Winters method correlation coefficient between
forecast errors and test set: 0.10669524622815553
```

Multiple Linear Regression



OLS Regression Results

Dep. Variable:	Sum of Appliances	R-squared:	0.208
Model:	OLS	Adj. R-squared:	0.204
Method:	Least Squares	F-statistic:	57.25
Date:	Tue, 08 Dec 2020	Prob (F-statistic):	3.47e-243
Time:	20:24:27	Log-Likelihood:	-36532.
No. Observations:	5263	AIC:	7.311e+04
Df Residuals:	5238	BIC:	7.328e+04
Df Model:	24		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-542.6517	515.841	-1.052	0.293	-1553.916	468.613
Average of T1	6.1706	10.362	0.595	0.552	-14.144	26.485
Average of RH_1	54.4338	3.640	14.956	0.000	47.298	61.569
Average of T2	-71.3915	9.755	-7.318	0.000	-90.516	-52.267
Average of RH_2	-49.9225	4.293	-11.630	0.000	-58.338	-41.507
Average of T3	80.5496	5.671	14.204	0.000	69.432	91.667
Average of RH_3	27.0087	3.873	6.974	0.000	19.417	34.600
Average of T4	9.1577	4.729	1.936	0.053	-0.113	18.429
Average of RH_4	-1.8839	3.323	-0.567	0.571	-8.398	4.630
Average of T5	-3.6034	6.918	-0.521	0.602	-17.165	9.958
Average of RH_5	0.5724	0.468	1.223	0.221	-0.345	1.490
Average of T6	27.5746	3.822	7.215	0.000	20.082	35.067
Average of RH_6	1.0465	0.370	2.832	0.005	0.322	1.771
Average of T7	1.4562	7.032	0.207	0.836	-12.329	15.242
Average of RH_7	-2.6159	2.257	-1.159	0.247	-7.041	1.810
Average of T8	33.9485	4.994	6.797	0.000	24.157	43.740
Average of RH_8	-22.0624	1.931	-11.428	0.000	-25.847	-18.278
Average of T9	-63.7714	9.363	-6.811	0.000	-82.128	-45.415
Average of RH_9	-4.6688	2.169	-2.153	0.031	-8.921	-0.417
Average of T_out	-29.8857	12.896	-2.317	0.021	-55.167	-4.604
Average of Press_mm_hg	0.8769	0.534	1.641	0.101	-0.170	1.924
Average of RH_out	-0.6633	2.391	-0.277	0.781	-5.351	4.024
Average of Windspeed	5.7672	1.800	3.204	0.001	2.238	9.296
Average of Visibility	0.6723	0.289	2.325	0.020	0.105	1.239
Average of Tdewpoint	10.8695	12.669	0.858	0.391	-13.966	35.705

Omnibus:	3046.990	Durbin-Watson:	1.093
Prob(Omnibus):	0.000	Jarque-Bera (JB):	27073.420
Skew:	2.687	Prob(JB):	0.00
Kurtosis:	12.726	Cond. No.	1.16e+05

Multiple Linear Regression

	Variables	AIC	BIC	Adj. R-squared	R-squared
65520	[[Average of RH_1, Average of T2, Average of R...	73106.836113	73205.362960	0.619794	0.620878
65534	[[Average of RH_1, Average of T2, Average of R...	73108.443250	73213.538554	0.619750	0.620906
65406	[[Average of RH_1, Average of T2, Average of R...	73111.095824	73203.054215	0.619414	0.620427
65445	[[Average of RH_1, Average of T2, Average of R...	73111.303587	73203.261977	0.619399	0.620412
65528	[[Average of RH_1, Average of T2, Average of R...	73112.347851	73210.874699	0.619396	0.620481
...	...	...	...	...	...
14	[[Average of Windspeed]]	75272.058834	75278.627291	0.424768	0.424877
12	[[Average of T_out]]	75418.568722	75425.137179	0.408530	0.408642
96	[[Average of T6, Average of T_out]]	75420.082641	75433.219554	0.408472	0.408697
7	[[Average of RH_6]]	75445.825555	75452.394011	0.405459	0.405572
6	[[Average of T6]]	75482.731192	75489.299649	0.401275	0.401389

65535 rows × 5 columns

```
The best model includes these variables: ['Average of RH_1', 'Average of T2', 'Average of RH_2', 'Average of T3',
'Average of RH_3', 'Average of T4', 'Average of T6', 'Average of RH_6', 'Average of T8', 'Average of RH_8', 'Aver
age of T9', 'Average of RH_9', 'Average of T_out', 'Average of Windspeed', 'Average of Visibility']
```

```
The F-statistic for the best model is: 572.9671
The p-value of the F-statistic for the best model is: 0.0
```

Multiple Linear Regression

Best model summary stats:

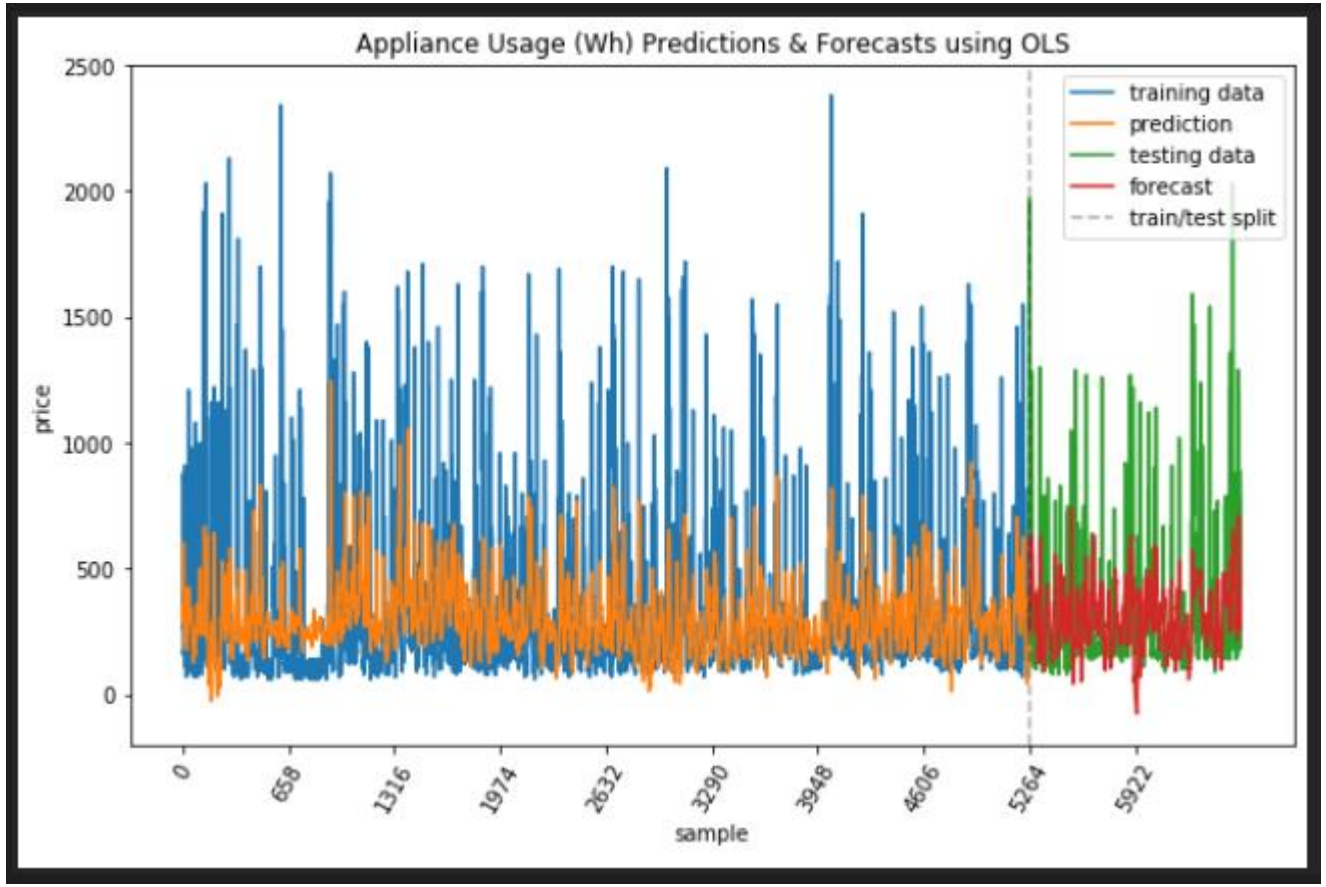
OLS Regression Results

Dep. Variable:	Sum of Appliances	R-squared (uncentered):	0.621
Model:	OLS	Adj. R-squared (uncentered):	0.620
Method:	Least Squares	F-statistic:	573.0
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	12:49:16	Log-Likelihood:	-36538.
No. Observations:	5263	AIC:	7.311e+04
Df Residuals:	5248	BIC:	7.321e+04
Df Model:	15		
Covariance Type:	nonrobust		

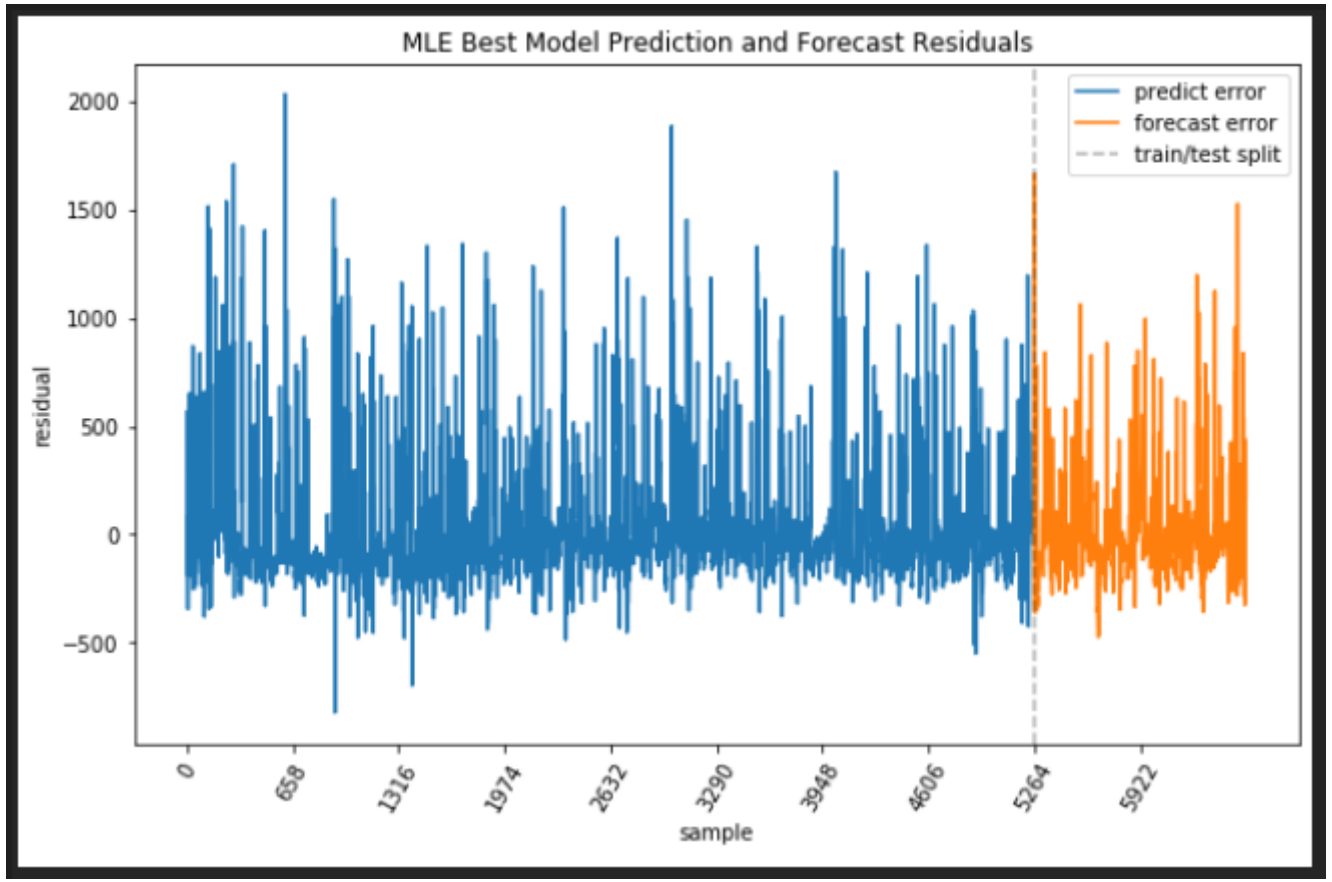
	coef	std err	t	P>|t|	[0.025	0.975]
Average of RH_1	52.4506	2.996	17.506	0.000	46.577	58.324
Average of T2	-68.4863	5.710	-11.995	0.000	-79.679	-57.293
Average of RH_2	-47.3194	3.132	-15.109	0.000	-53.459	-41.180
Average of T3	81.9814	4.875	16.818	0.000	72.425	91.538
Average of RH_3	23.7243	3.302	7.184	0.000	17.250	30.198
Average of T4	10.7720	4.240	2.540	0.011	2.459	19.085
Average of T6	27.4949	3.450	7.969	0.000	20.731	34.259
Average of RH_6	1.2964	0.313	4.143	0.000	0.683	1.910
Average of T8	35.3418	4.010	8.813	0.000	27.480	43.203
Average of RH_8	-22.5086	1.465	-15.363	0.000	-25.381	-19.636
Average of T9	-65.2364	6.969	-9.361	0.000	-78.898	-51.575
Average of RH_9	-4.9097	1.965	-2.499	0.012	-8.761	-1.058
Average of T_out	-21.8921	3.753	-5.833	0.000	-29.249	-14.535
Average of Windspeed	4.6664	1.636	2.852	0.004	1.459	7.874
Average of Visibility	0.7276	0.283	2.568	0.010	0.172	1.283

Omnibus:	3044.480	Durbin-Watson:	1.087
Prob(Omnibus):	0.000	Jarque-Bera (JB):	26993.766
Skew:	2.684	Prob(JB):	0.00
Kurtosis:	12.709	Cond. No.	290.

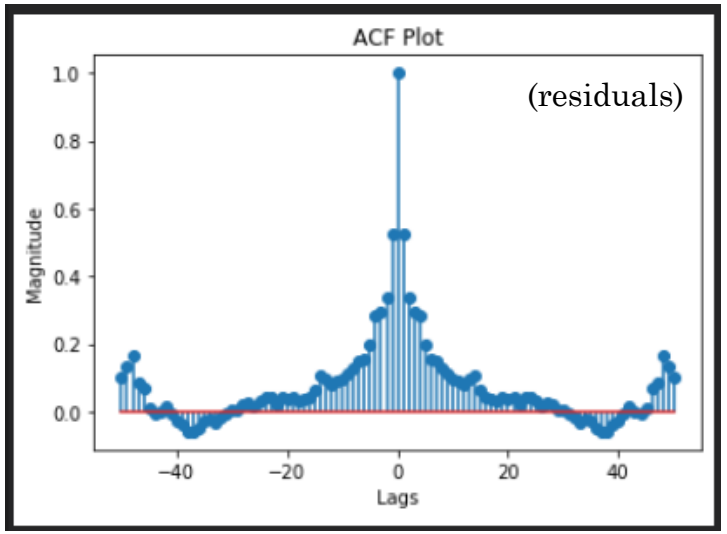
Multiple Linear Regression



Multiple Linear Regression

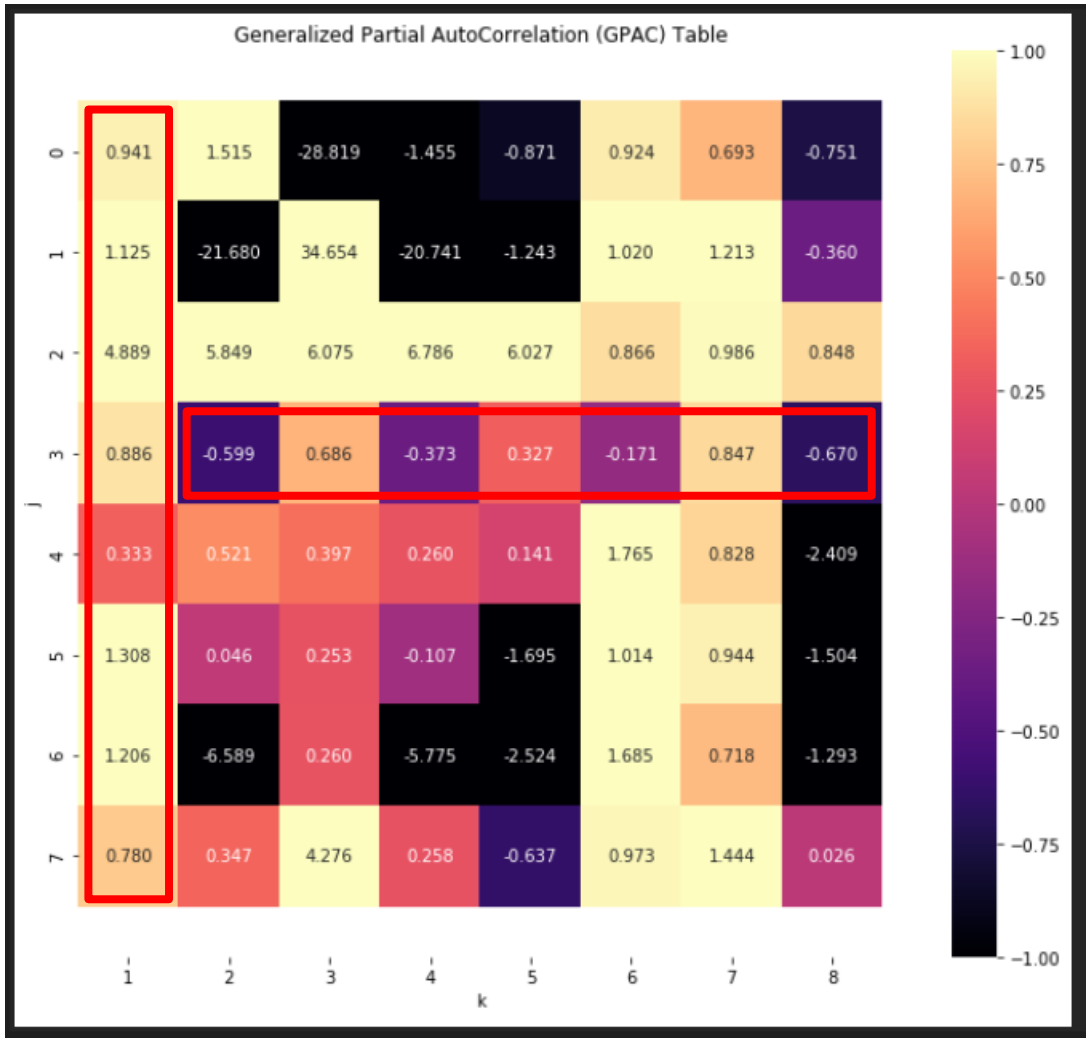



```
MLE Q value: 957.3814657604295
MLE h-step MSE:  48225.69938450677
```




```
Best MLE 1-step prediction error variance: 62764.08780218717
Best MLE 1-step prediction error standard deviation: 250.5276188410914
Best MLE h-step forecast error variance: 47877.52842560792
Best MLE h-step forecast error standard deviation: 218.8093426378497
```

Generalized Partial Autocorrelation



ARMA(1,3)

ARMA Model Results

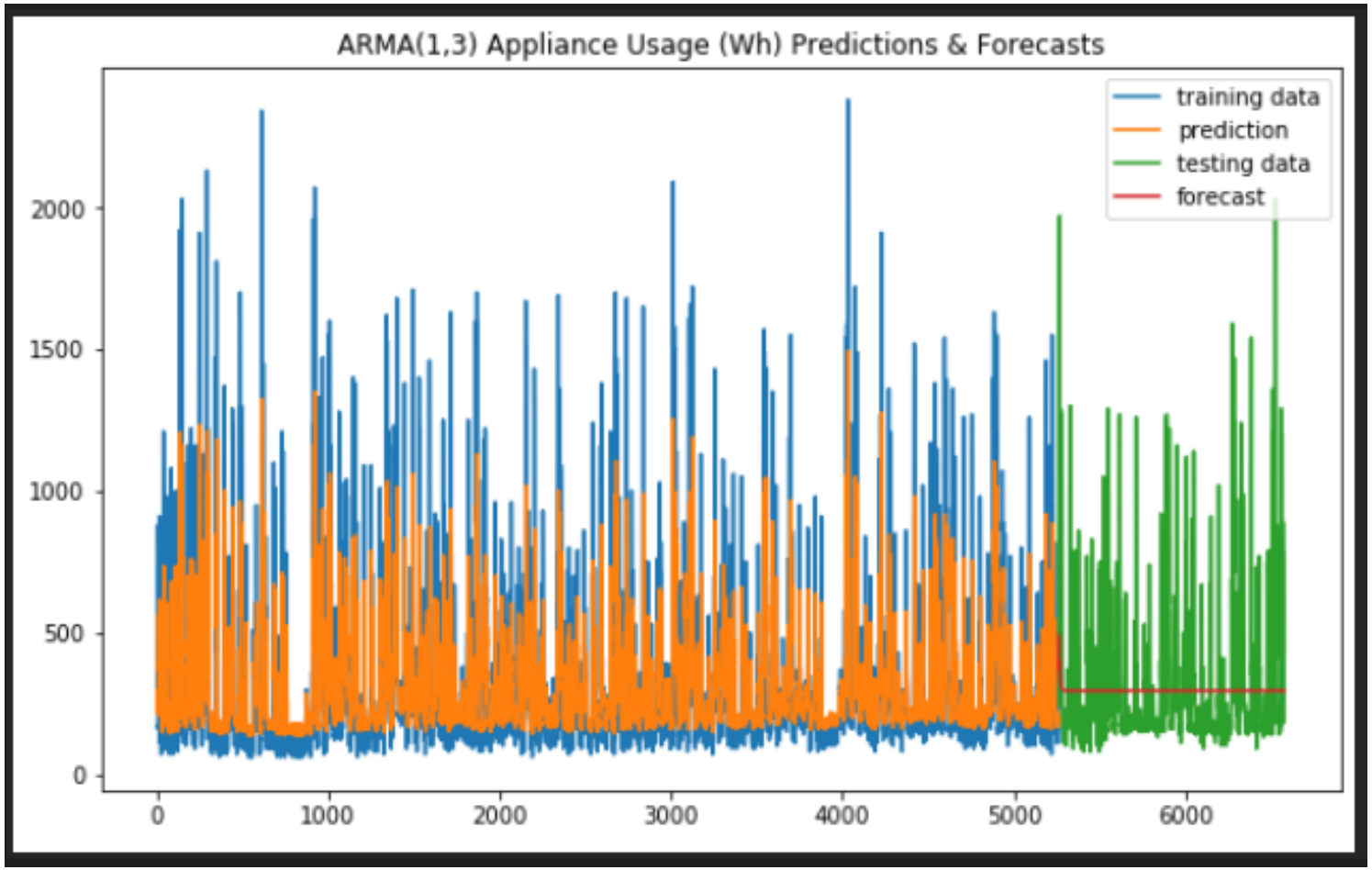
Dep. Variable:	Sum of Appliances	No. Observations:	5263
Model:	ARMA(1, 3)	Log Likelihood	-35849.662
Method:	css-mle	S.D. of innovations	219.784
Date:	Tue, 08 Dec 2020	AIC	71711.324
Time:	20:45:56	BIC	71750.735
Sample:	0	HQIC	71725.101

	coef	std err	z	P>|z|	[0.025	0.975]
const	294.3816	11.221	26.236	0.000	272.390	316.374
ar.L1.Sum of Appliances	0.8486	0.014	62.479	0.000	0.822	0.875
ma.L1.Sum of Appliances	-0.3425	0.018	-18.553	0.000	-0.379	-0.306
ma.L2.Sum of Appliances	-0.1721	0.015	-11.663	0.000	-0.201	-0.143
ma.L3.Sum of Appliances	0.0760	0.016	4.851	0.000	0.045	0.107

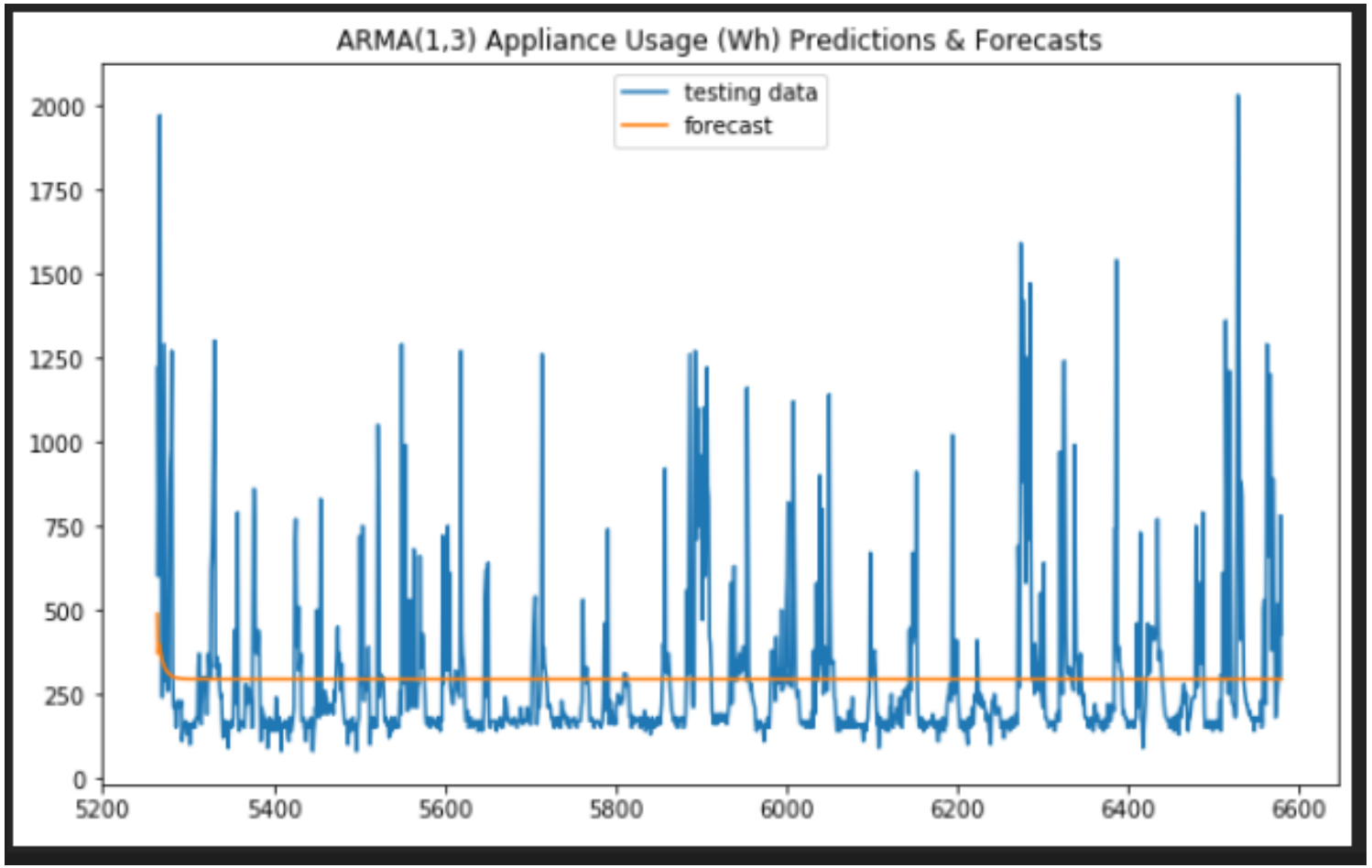
Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.1785	+0.0000j	1.1785	0.0000
MA.1	-2.2714	-0.0000j	2.2714	-0.5000
MA.2	2.2681	-0.8067j	2.4073	-0.0544
MA.3	2.2681	+0.8067j	2.4073	0.0544

ARMA(1,3)

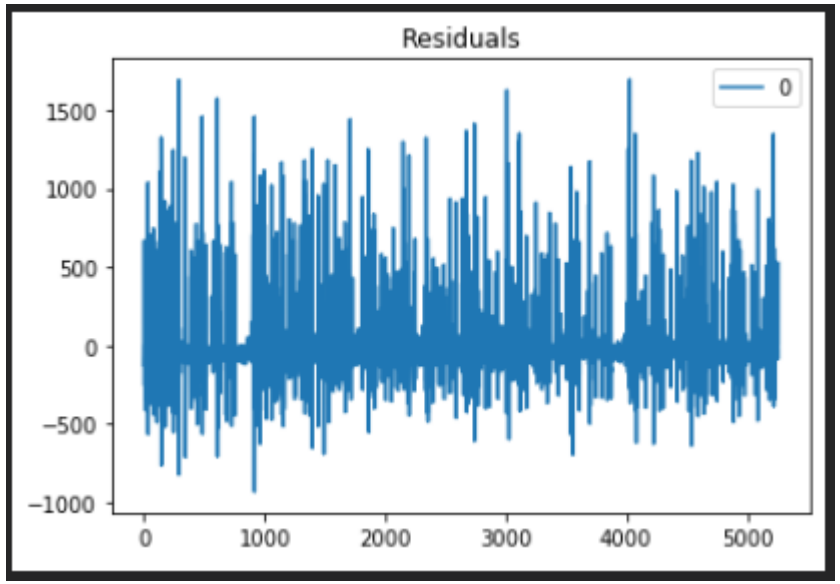

ARMA(1,3) Appliance Usage (Wh) Predictions & Forecasts
training data
prediction
testing data
forecast
2000
1500
1000
500
0
0
1000
2000
3000
4000
5000
6000

ARMA(1,3)


ARMA(1,3) Appliance Usage (Wh) Predictions & Forecasts
testing data
forecast
2000
1750
1500
1250
1000
750
500
250
0
5200
5400
5600
5800
6000
6200
6400
6600

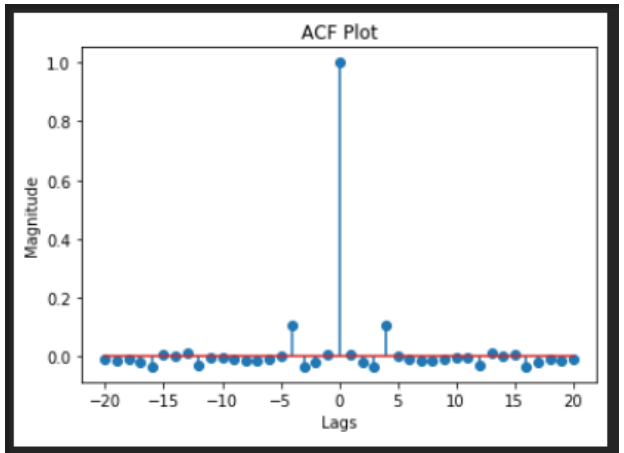
ARMA(1,3) - Diagnostics

(plot of residuals)

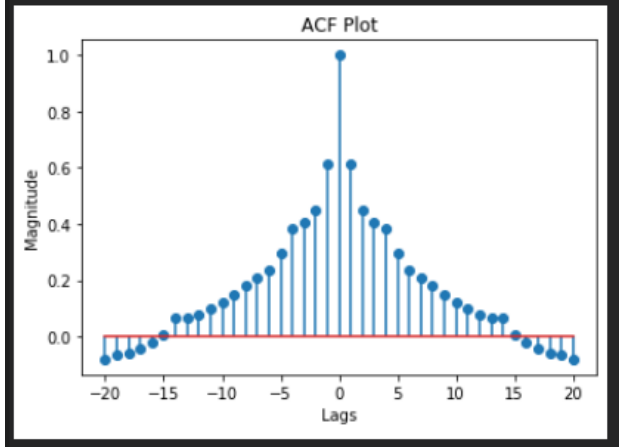



```
AR roots:  [1.17846872]
MA roots:  [-2.27135577-0.j          2.26812744-0.80669424j  2.26812744+0.80669424j]
```

Unbiased because the confidence intervals of the coefficients do not contain 0 so there is no zero/pole cancelation.




```
One-step Q value: 21.347087976402953
```




```
H-step Q value: 45.950434066381206
```

ARMA(1,3) - Diagnostics

Covariance Matrix

	const	ar.L1.Sum of Appliances	ma.L1.Sum of Appliances	ma.L2.Sum of Appliances	ma.L3.Sum of Appliances
const	125.901951	-0.000019	0.000034	-0.000010	-0.000061
ar.L1.Sum of Appliances	-0.000019	0.000184	-0.000160	-0.000094	-0.000079
ma.L1.Sum of Appliances	0.000034	-0.000160	0.000341	0.000026	-0.000002
ma.L2.Sum of Appliances	-0.000010	-0.000094	0.000026	0.000218	0.000003
ma.L3.Sum of Appliances	-0.000061	-0.000079	-0.000002	0.000003	0.000245

```
Variance of prediction error:  48315.728072226164
Variance of forecast error:  55510.23026050083
ARMA(1,3) h-step MSE:  55504.94998339806
```

ARMA(0,3)

ARMA Model Results

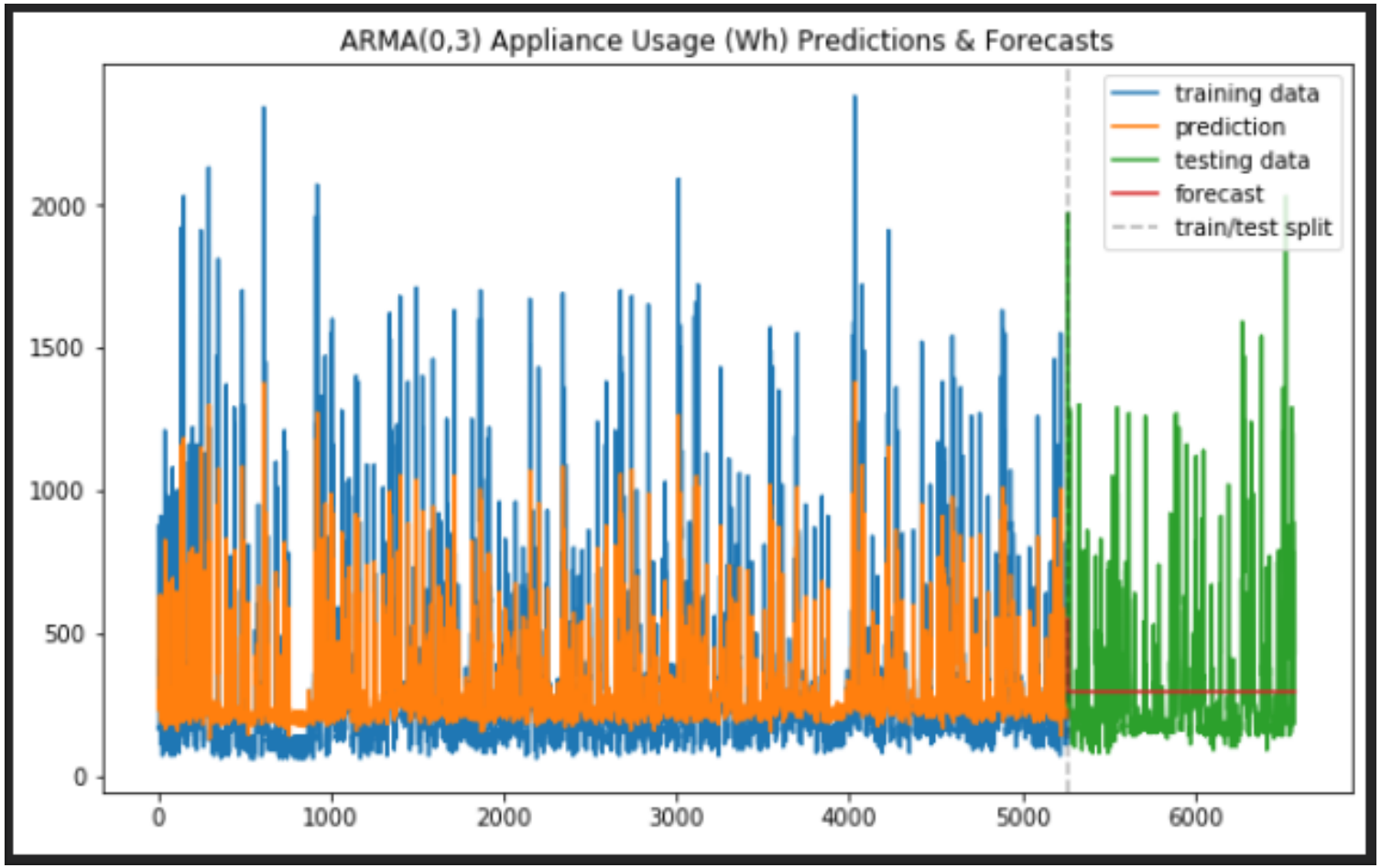
Dep. Variable:	Sum of Appliances	No. Observations:	5263
Model:	ARMA(0, 3)	Log Likelihood	-36084.151
Method:	css-mle	S.D. of innovations	229.803
Date:	Tue, 08 Dec 2020	AIC	72178.302
Time:	21:36:16	BIC	72211.144
Sample:	0	HQIC	72189.783

	coef	std err	z	P>|z|	[0.025	0.975]
const	294.1861	6.091	48.301	0.000	282.249	306.124
ma.L1.Sum of Appliances	0.5588	0.015	38.409	0.000	0.530	0.587
ma.L2.Sum of Appliances	0.2445	0.013	18.995	0.000	0.219	0.270
ma.L3.Sum of Appliances	0.1197	0.013	9.242	0.000	0.094	0.145

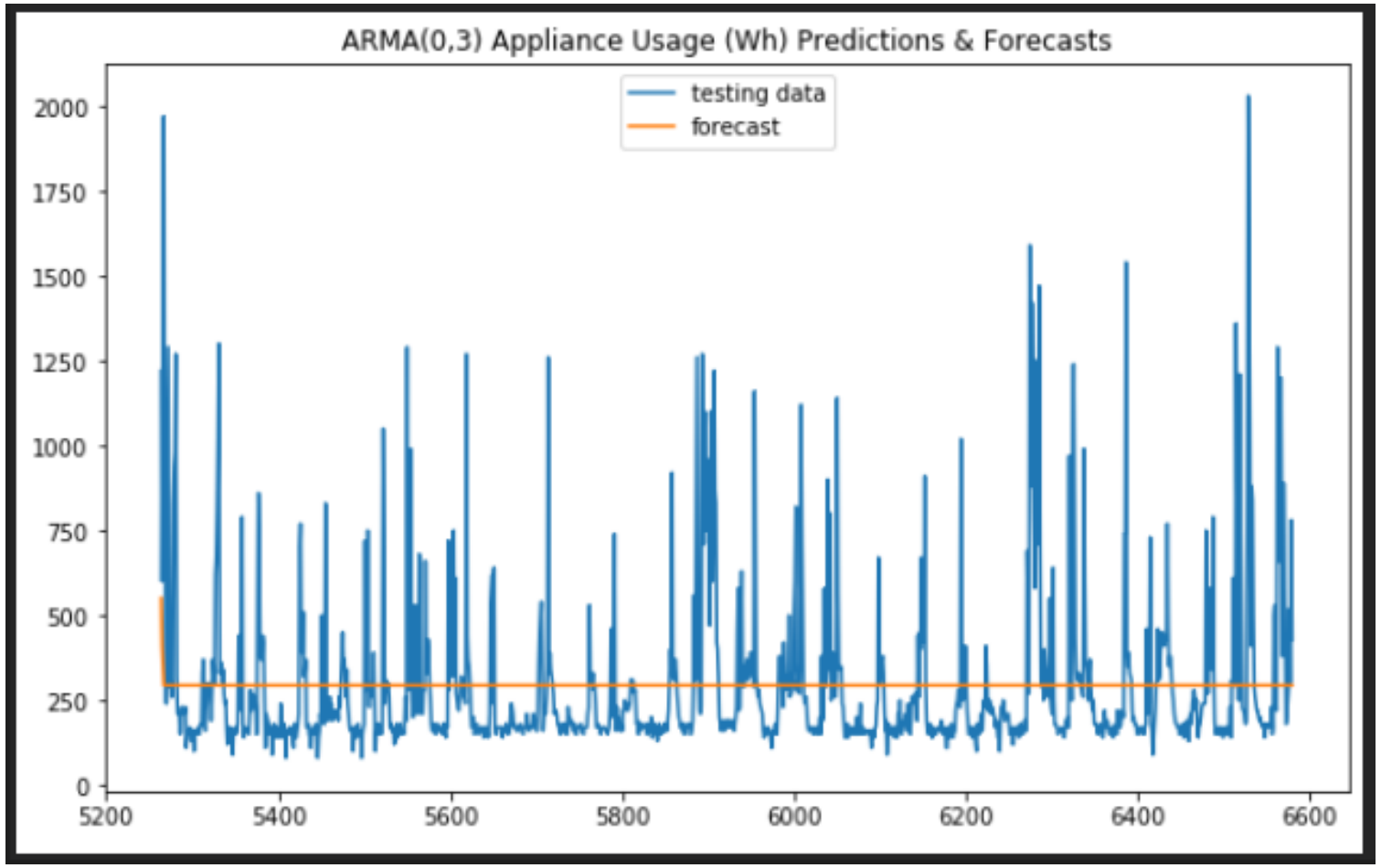
Roots

	Real	Imaginary	Modulus	Frequency
MA.1	-1.8997	-0.0000j	1.8997	-0.5000
MA.2	-0.0713	-2.0957j	2.0969	-0.2554
MA.3	-0.0713	+2.0957j	2.0969	0.2554

ARMA(0,3)

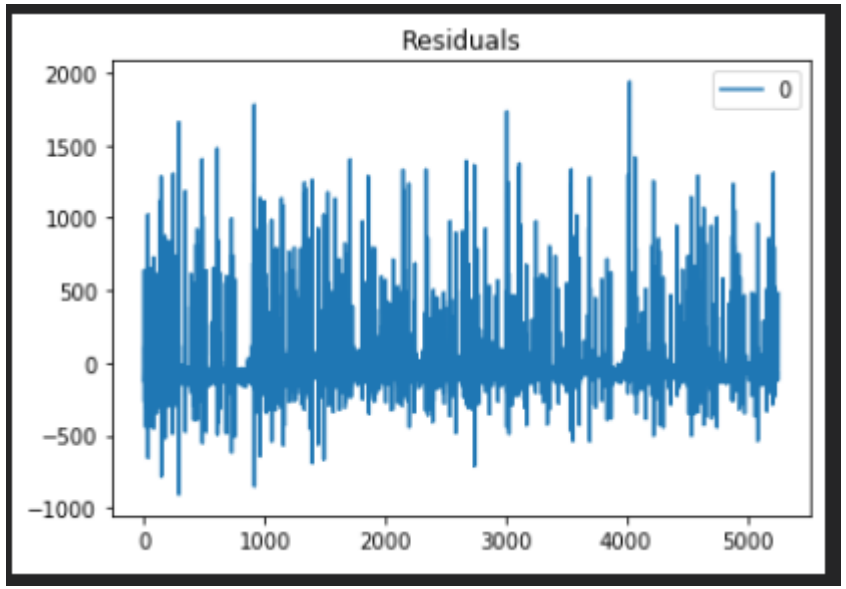

ARMA(0,3) Appliance Usage (Wh) Predictions & Forecasts
training data
prediction
testing data
forecast
train/test split
2000
1500
1000
500
0
0
1000
2000
3000
4000
5000
6000

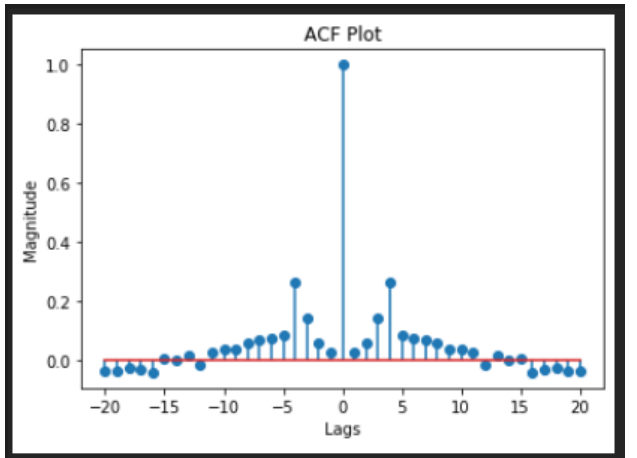
ARMA(0,3)


ARMA(0,3) Appliance Usage (Wh) Predictions & Forecasts
testing data
forecast
2000
1750
1500
1250
1000
750
500
250
0
5200
5400
5600
5800
6000
6200
6400
6600

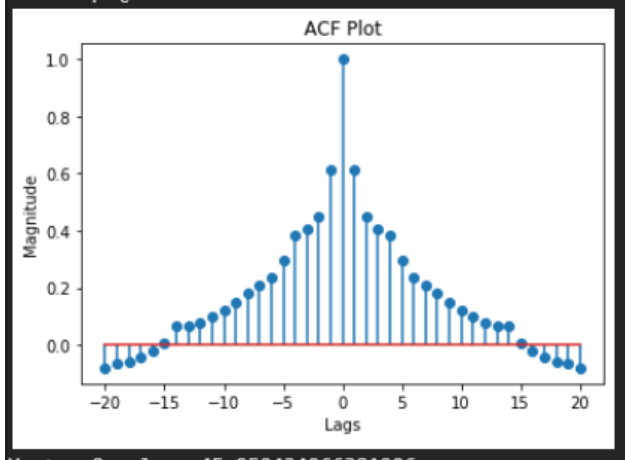
ARMA(0,3) - Diagnostics

(plot of residuals)





One-step Q value: 23.62950974107968



H-step Q value: 45.950434066381206

```
AR roots:  []
MA roots:  [-1.89967919-0.j         -0.07126426-2.09570971j -0.07126426+2.09570971j]
```

Unbiased because the confidence intervals of the coefficients do not contain 0 so there is no zero/pole cancelation.

ARMA(0,3) - Diagnostics

Covariance Matrix

	const	ma.L1.Sum of Appliances	ma.L2.Sum of Appliances	ma.L3.Sum of Appliances
const	37.096003	0.000035	0.000033	-0.000007
ma.L1.Sum of Appliances	0.000035	0.000212	0.000091	-0.000039
ma.L2.Sum of Appliances	0.000033	0.000091	0.000166	0.000033
ma.L3.Sum of Appliances	-0.000007	-0.000039	0.000033	0.000168

```
Variance of prediction error:  52820.62205516074
Variance of forecast error:  55510.23026050083
ARMA(0,3) h-step MSE:  55922.817983088746
```

Final Model Selection

Metrics Comparison

Forecast MSE Comparison

Model	Forecast MSE
MLE Best	48,225.7
ARMA(1,3)	55,509.3
ARMA(0,3)	55,922.8
Average	56,327.5
SES (alpha=0.99)	62,584.1
SES (alpha=0.5)	68,527.3
Holt-Winters (seasonal=48, damped=True)	69,737.6
SES (alpha=0)	70,406.8
Naïve	259,901.4
Drift	329,680.3
Holt-Winters (seasonal=48, damped=False)	754,477.4
Holt-Winters (seasonal=24, damped=False)	1,674,610.5
Holt-Winters (seasonal=12, damped=False)	7,284,517.9

Multiple Linear Regression

Y = Average of RH_1*(52.4506) + Average of T2*(-68.4863) + Average of RH_2*(-47.3194) + Average of T3*(81.9814) + Average of RH_3*(23.7243) + Average of T4*(10.772) + Average of T6*(27.4949) + Average of RH_6*(1.2964) + Average of T8*(35.3418) + Average of RH_8*(-22.5086) + Average of T9*(-65.2364) + Average of RH_9*(-4.9097) + Average of T_out*(-21.8921) + Average of Windspeed*(4.6664) + Average of Visibility*(.7276)

Multiple Linear Regression

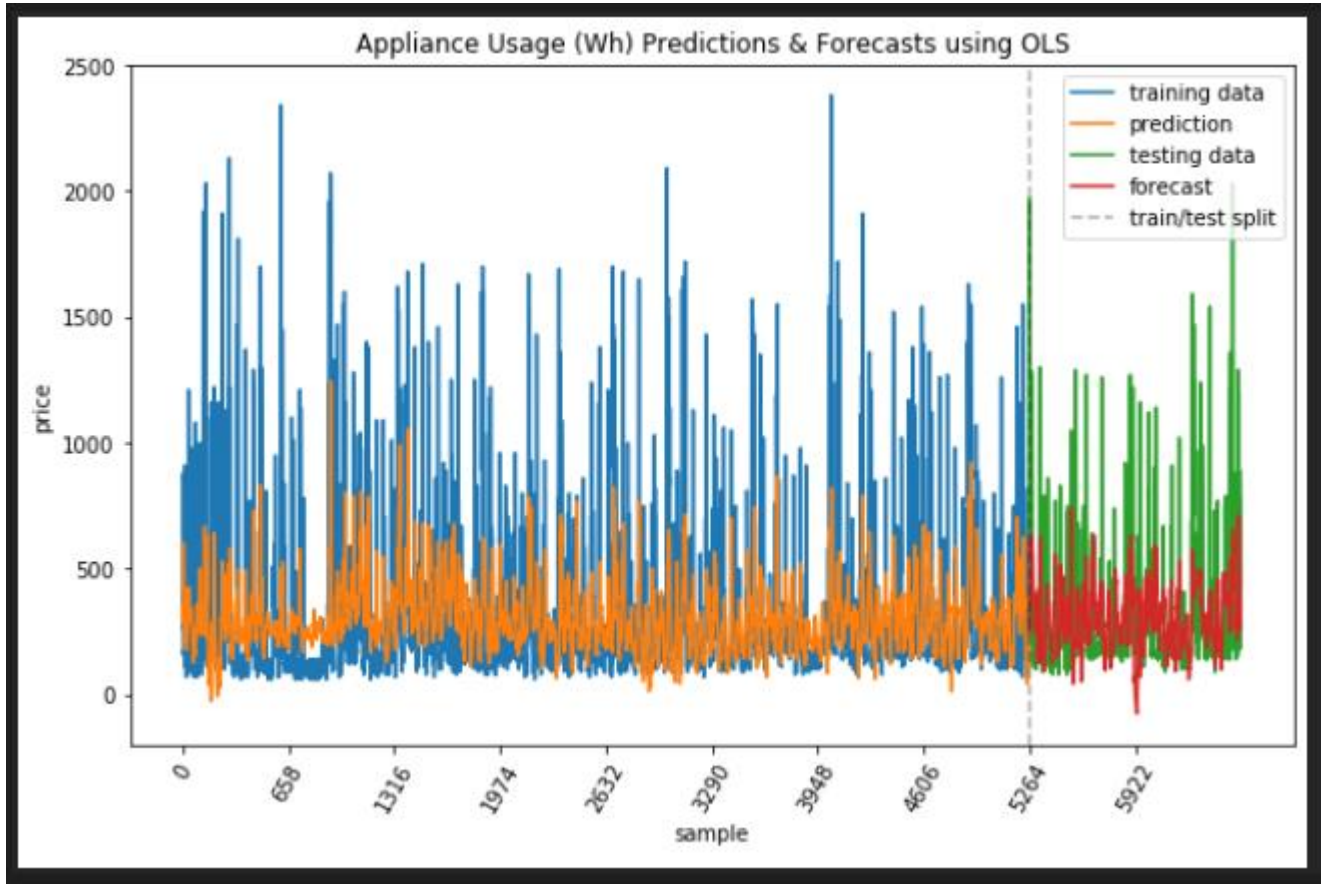
OLS Regression Results

Dep. Variable:	Sum of Appliances	R-squared (uncentered):	0.621
Model:	OLS	Adj. R-squared (uncentered):	0.620
Method:	Least Squares	F-statistic:	573.0
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	12:49:16	Log-Likelihood:	-36538.
No. Observations:	5263	AIC:	7.311e+04
Df Residuals:	5248	BIC:	7.321e+04
Df Model:	15		
Covariance Type:	nonrobust		

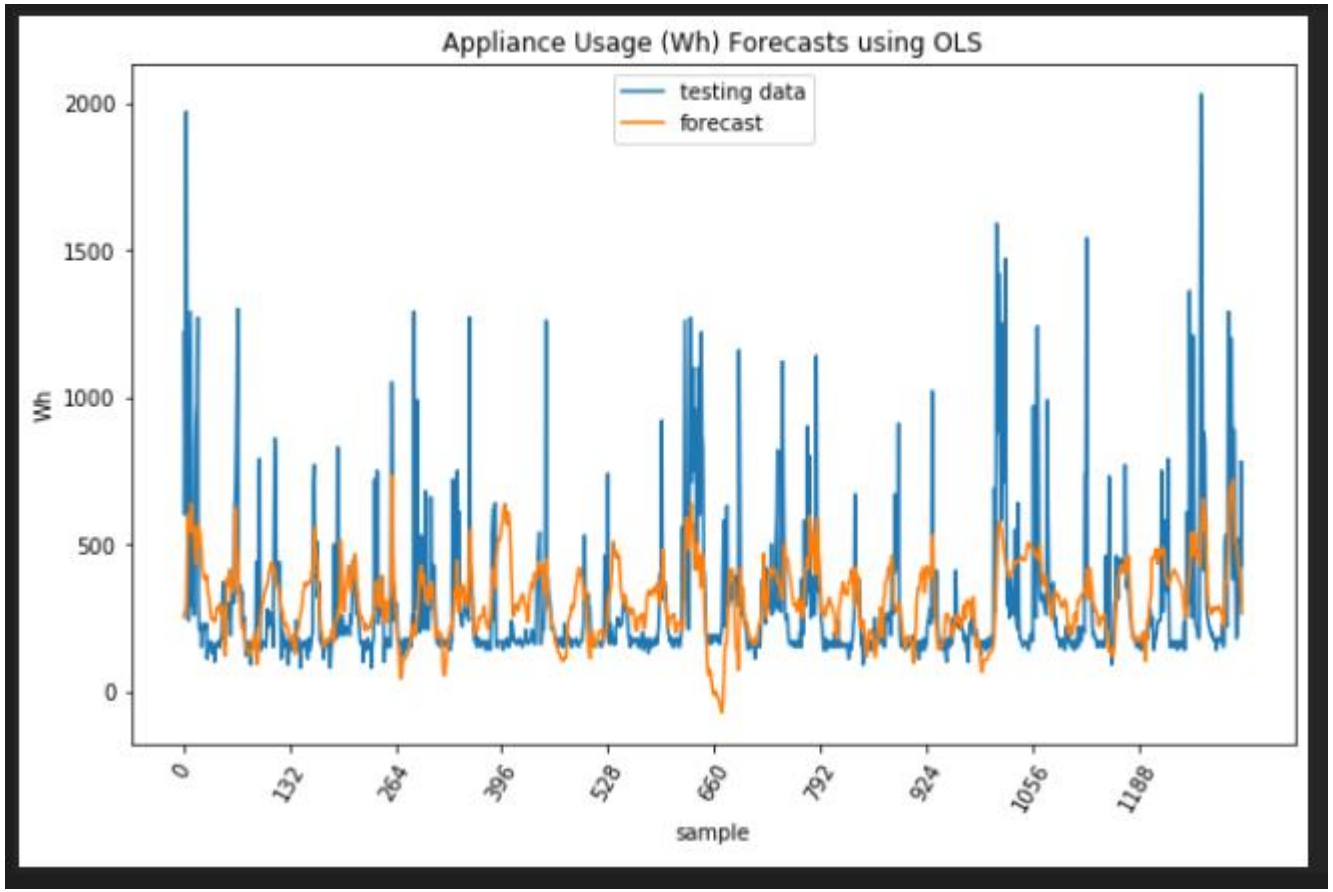
	coef	std err	t	P>|t|	[0.025	0.975]
Average of RH_1	52.4506	2.996	17.506	0.000	46.577	58.324
Average of T2	-68.4863	5.710	-11.995	0.000	-79.679	-57.293
Average of RH_2	-47.3194	3.132	-15.109	0.000	-53.459	-41.180
Average of T3	81.9814	4.875	16.818	0.000	72.425	91.538
Average of RH_3	23.7243	3.302	7.184	0.000	17.250	30.198
Average of T4	10.7720	4.240	2.540	0.011	2.459	19.085
Average of T6	27.4949	3.450	7.969	0.000	20.731	34.259
Average of RH_6	1.2964	0.313	4.143	0.000	0.683	1.910
Average of T8	35.3418	4.010	8.813	0.000	27.480	43.203
Average of RH_8	-22.5086	1.465	-15.363	0.000	-25.381	-19.636
Average of T9	-65.2364	6.969	-9.361	0.000	-78.898	-51.575
Average of RH_9	-4.9097	1.965	-2.499	0.012	-8.761	-1.058
Average of T_out	-21.8921	3.753	-5.833	0.000	-29.249	-14.535
Average of Windspeed	4.6664	1.636	2.852	0.004	1.459	7.874
Average of Visibility	0.7276	0.283	2.568	0.010	0.172	1.283

Omnibus:	3044.480	Durbin-Watson:	1.087
Prob(Omnibus):	0.000	Jarque-Bera (JB):	26993.766
Skew:	2.684	Prob(JB):	0.00
Kurtosis:	12.709	Cond. No.	290.

Multiple Linear Regression



Multiple Linear Regression



Thank You

Questions?