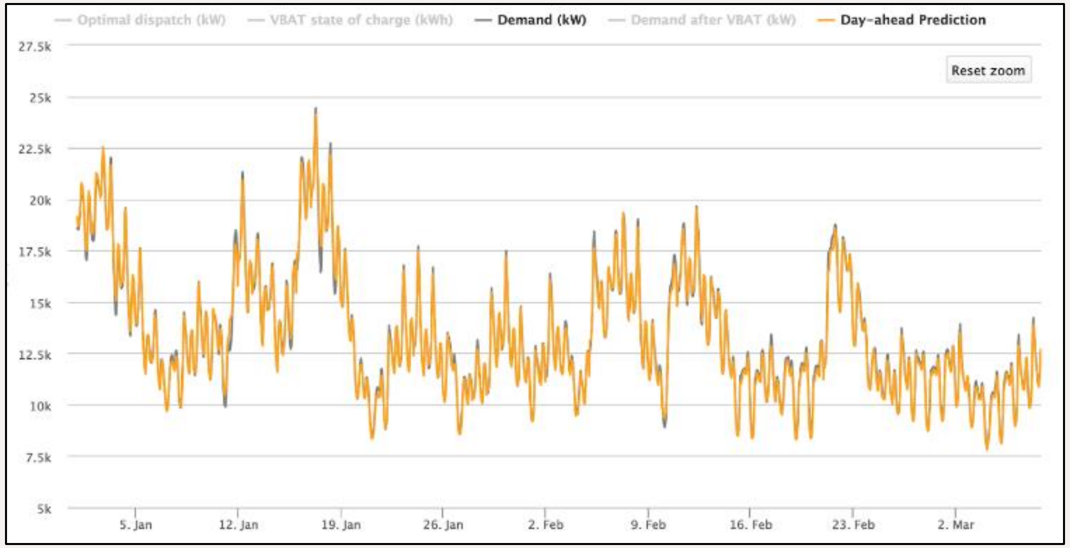
Indian Statistical Institute

PGDBA 2021-23

Load Prediction using Time Series



Submitted To
Prof. Smarajit Bose

Submitted By
Parag Agrawal (21BM6JP10)
Sanket Biswas (21BM6JP36)

TIME SERIES DATA

The 5216 rows dataset is about the Power Consumption of the three zones of the Tetuan City for the year 2017.

- The objective is to predict the total power consumption of all the zones of the Tetuan City. To make the required variable univariate, the Z1PC, Z2PC & Z3PC columns will be summed up for our calculations.

- For the given data, the sample time is 10 minutes, which makes the total number of 144 data points for a single day, hence, to make the data less complex, the sample time has been reduced to an hour.

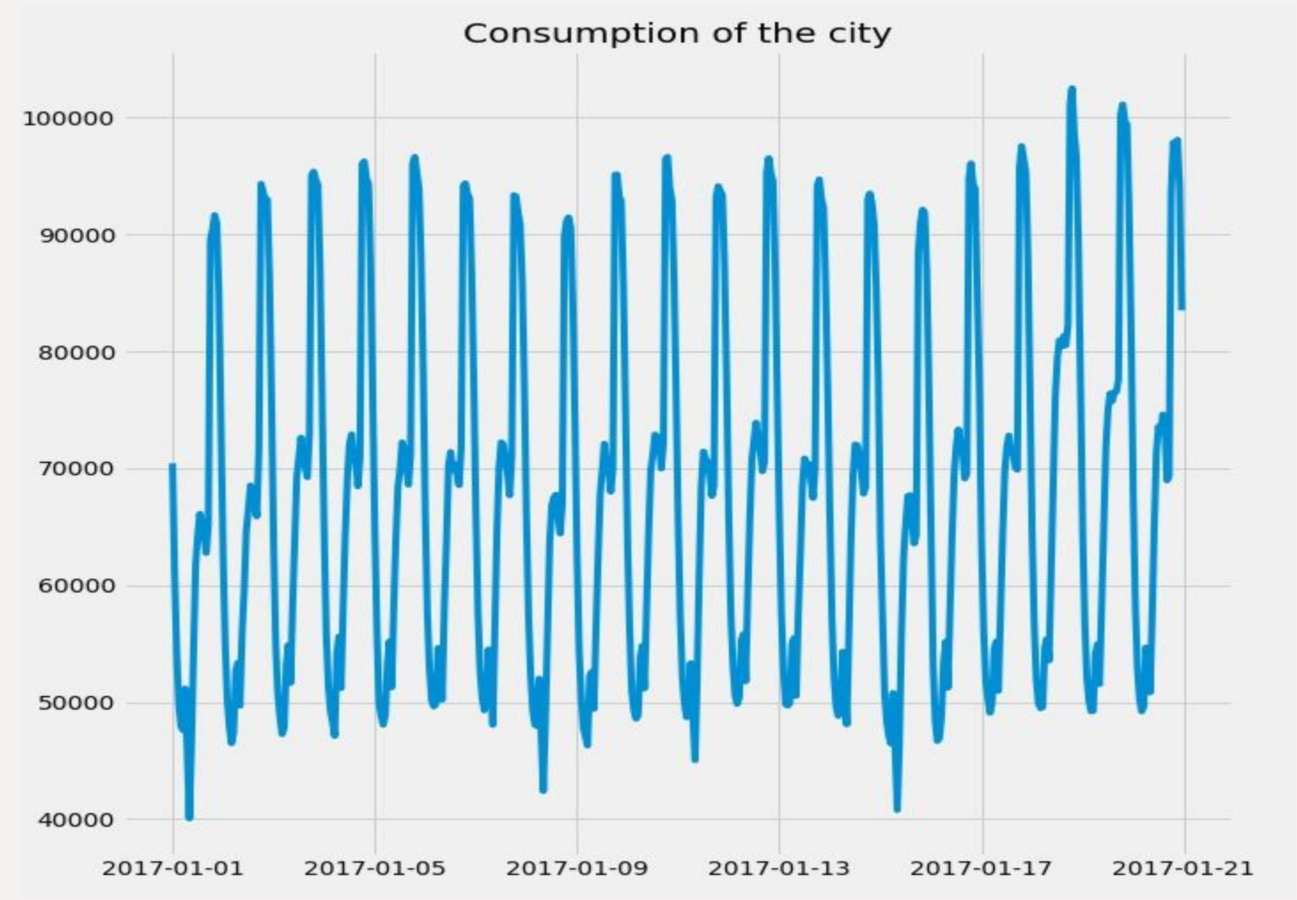
Wind Speed	general diffuse flows	diffuse flows	Z1PC	Z2PC	Z3PC
0.083	0.051	0.119	34055.69620	16128.87538	20240.96386
0.083	0.070	0.085	29814.68354	19375.07599	20131.08434
0.080	0.062	0.100	29128.10127	19006.68693	19668.43373
0.083	0.091	0.096	28228.86076	18361.09422	18899.27711
0.081	0.048	0.085	27335.69620	17872.34043	18442.40964

https://archive.ics.uci.edu/ml/datasets/Power+consumption+of+Tetouan+city

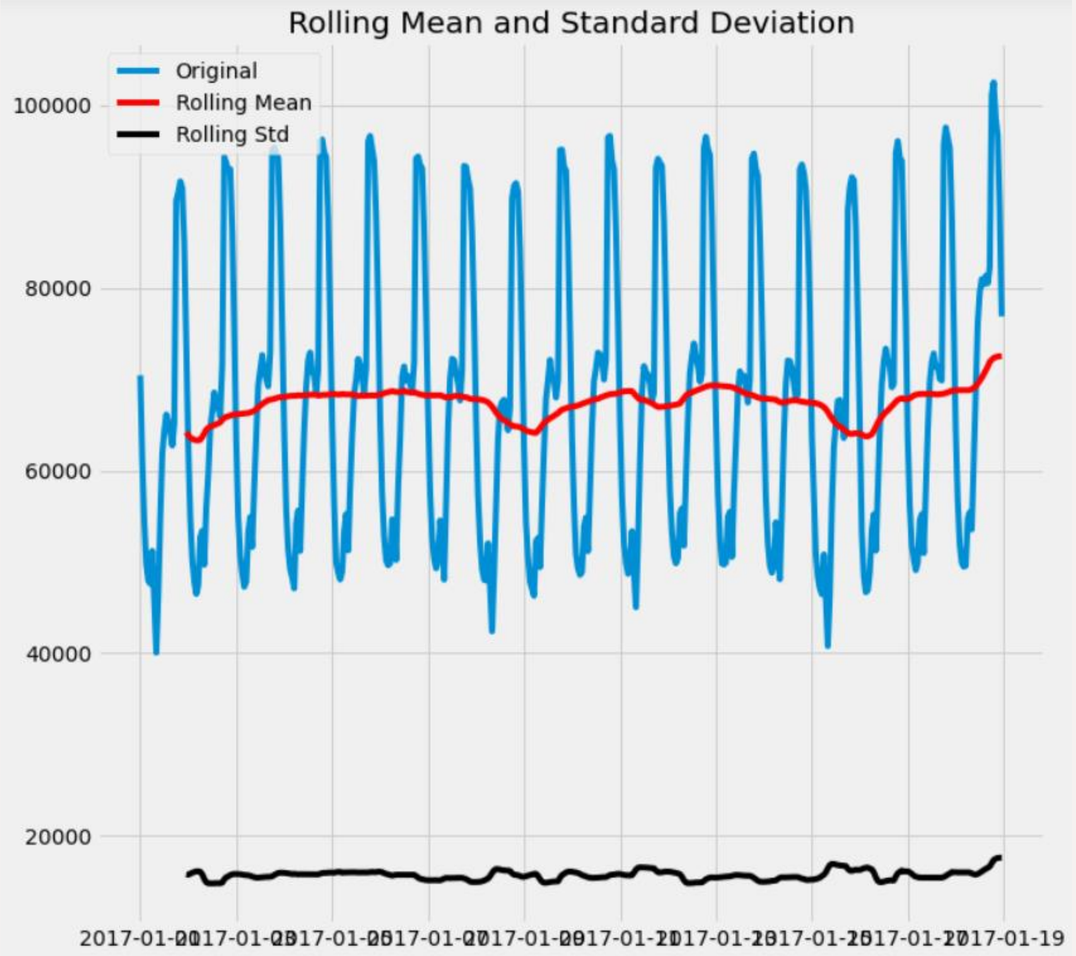
- **Training Dataset:** First 18 days of the year i.e January 1st 2017 to January 18th, 2017, constituting 432 rows.

- **Testing Dataset:** Next two days of the training data, i.e. January 19th & 20th 2017.

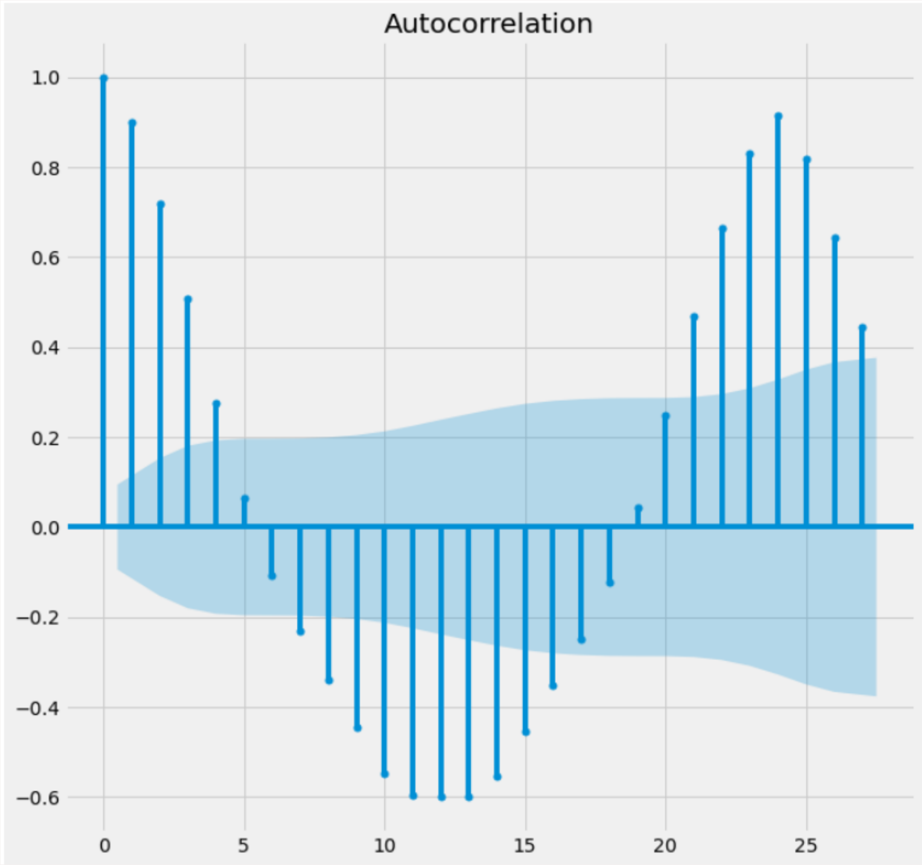
EXPLORATORY ANALYSIS



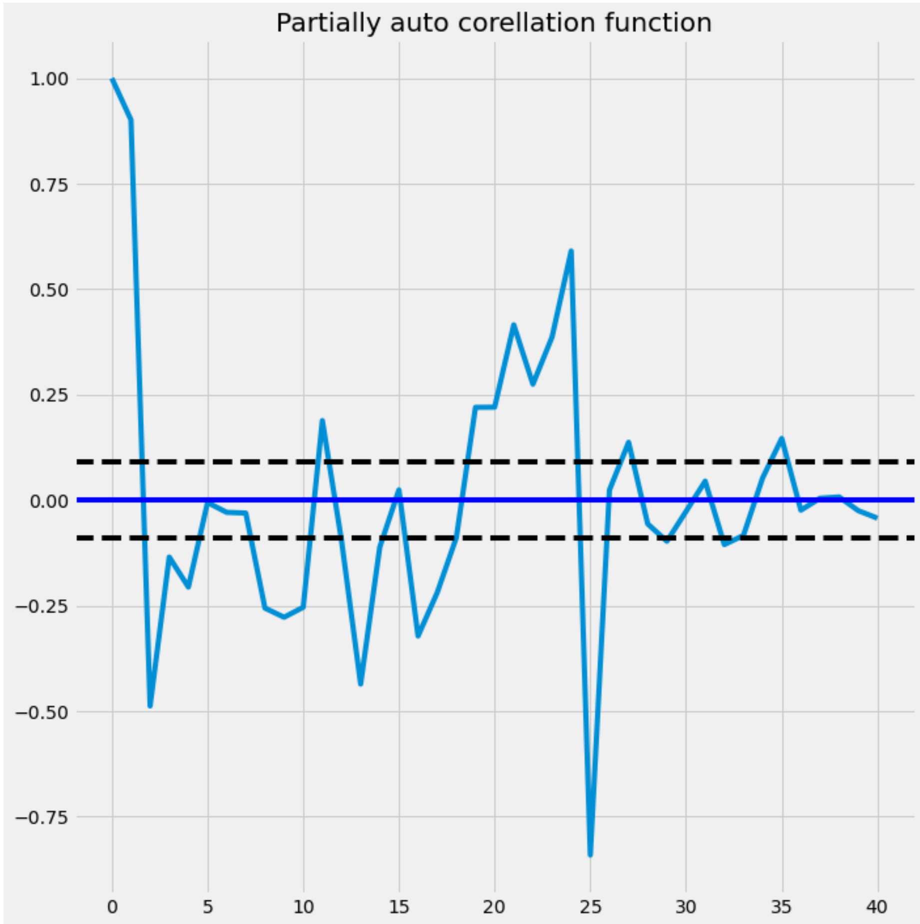
21 days electricity consumption plot for Tetuan city in January 2017.



EXPLORATORY ANALYSIS



ACF plot of training data without removing seasonality.



EXPLORATORY ANALYSIS

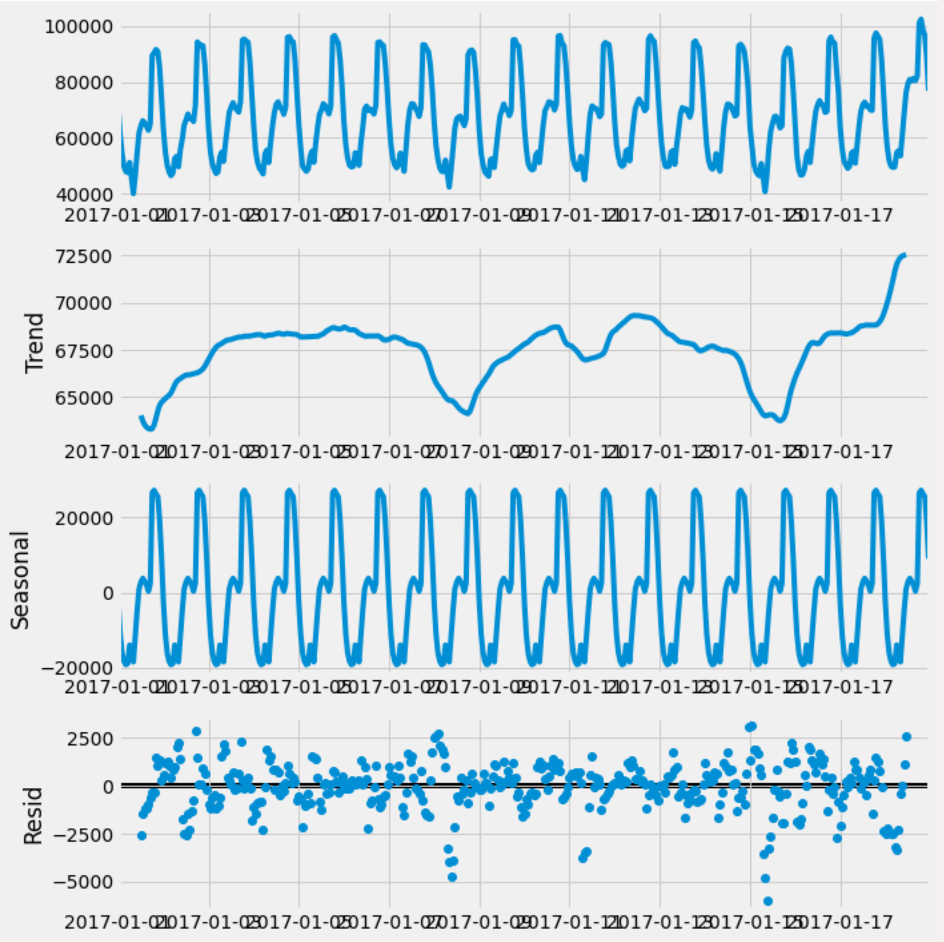
Trend: Looking at the data seems like there is a slight upward trend.

Seasonality: There is a seasonality with the period of one day that is 24 data points.

Mean: Figure-2 shows that the rolling mean is constant.

Variance: The variance is almost constant.

Conclusion: Pertaining to the seasonality with period of a day, the time series does not have stationarity.



AUGMENTED DICKEY FULLER (ADF) TEST

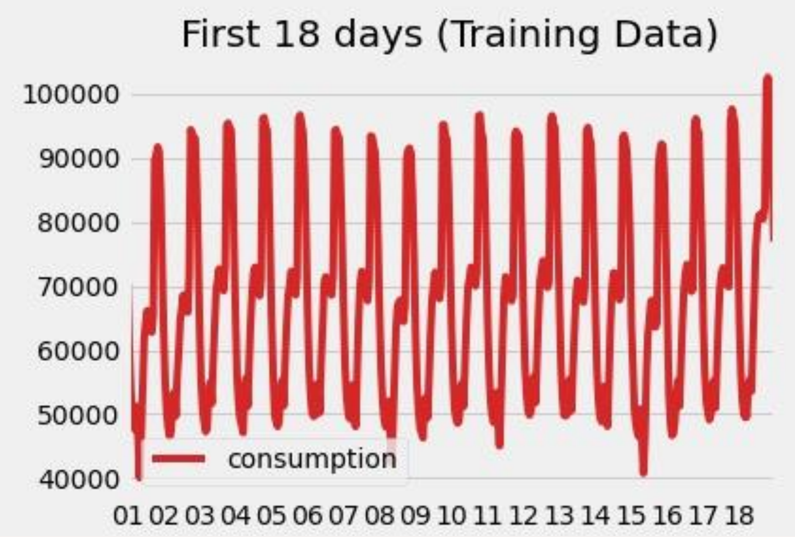
```
Results of dickey fuller test
Test Statistics                 -7.904235e+00
p-value                          4.107309e-12
No. of lags used                 1.800000e+01
Number of observations used      4.130000e+02
critical value (1%)             -3.446283e+00
critical value (5%)             -2.868564e+00
critical value (10%)            -2.570511e+00
dtype: float64
```

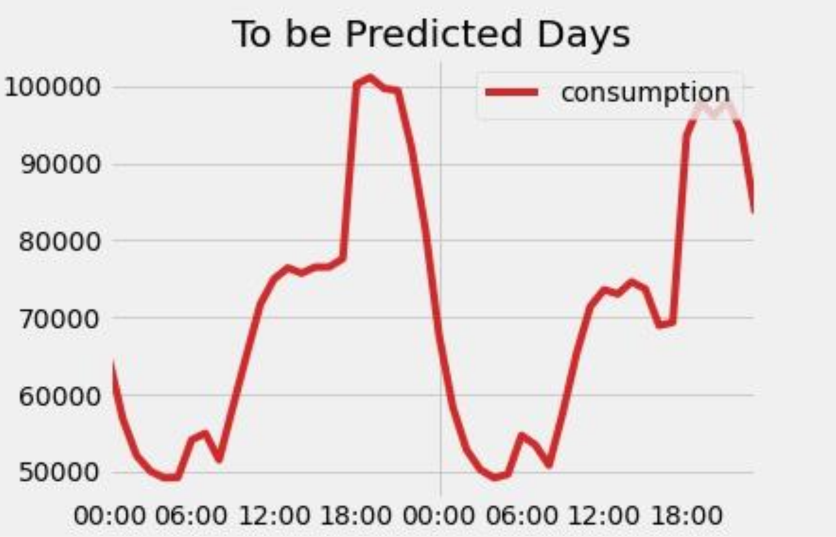
- Applying the Augmented Dickey Fuller (ADF) Test on the training data to check whether the data is stationary.
- We get a highly significant p value, i.e. 4.1+e-12 and test statistic as -7.90.
- Hence, we need to accept the alternative hypothesis i.e. data is stationary.
- However, this test doesn't seem to consider the seasonality in finding the stationarity.

TRAINING DATA & TESTING DATA

	DateTime	consumption
0	01-01-2017 00:00	70425.53544
6	01-01-2017 01:00	60937.36065
12	01-01-2017 02:00	54290.73952
18	01-01-2017 03:00	49783.65137
24	01-01-2017 04:00	47930.28589
...	...	...
2850	1/20/2017 19:00	97881.98310
2856	1/20/2017 20:00	96194.23179
2862	1/20/2017 21:00	98114.90754
2868	1/20/2017 22:00	94066.76339
2874	1/20/2017 23:00	83507.12040

480 rows × 2 columns

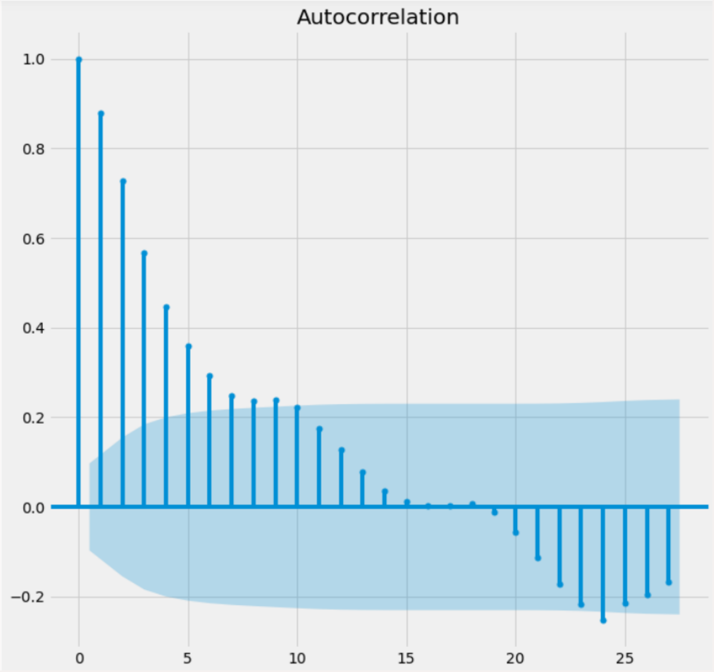


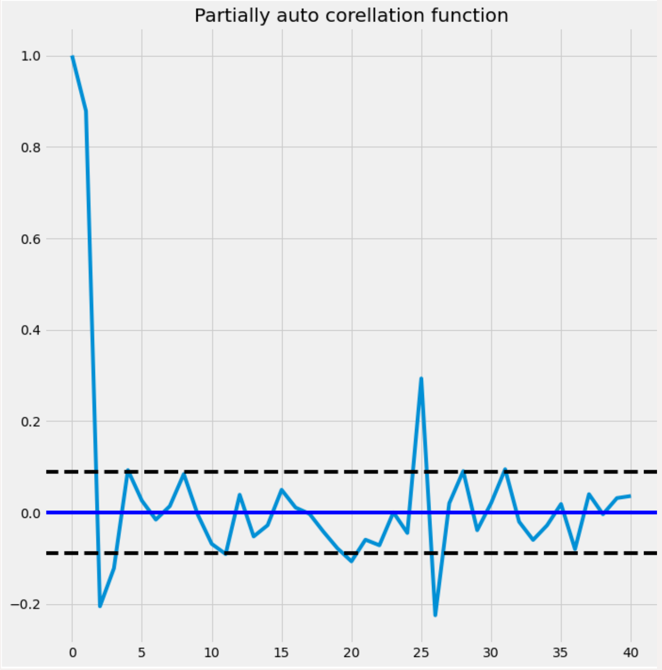


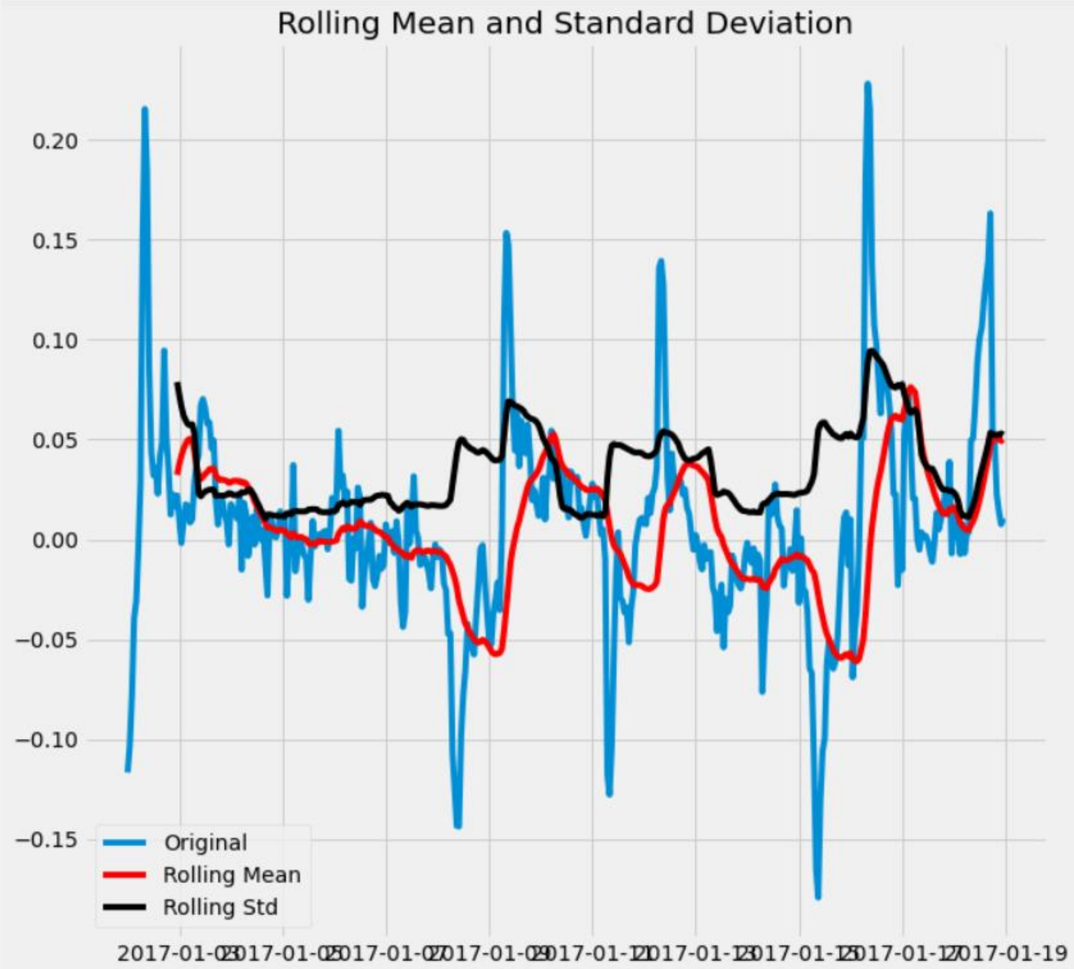
FITTING MODEL AFTER REMOVING SEASONALITY

Plot after log differencing.

```
data['consumption']=data['consumption'].apply(lambda x:np.log(x))
data_shift=pd.DataFrame(data['consumption']-data['consumption'].shift(24))
```

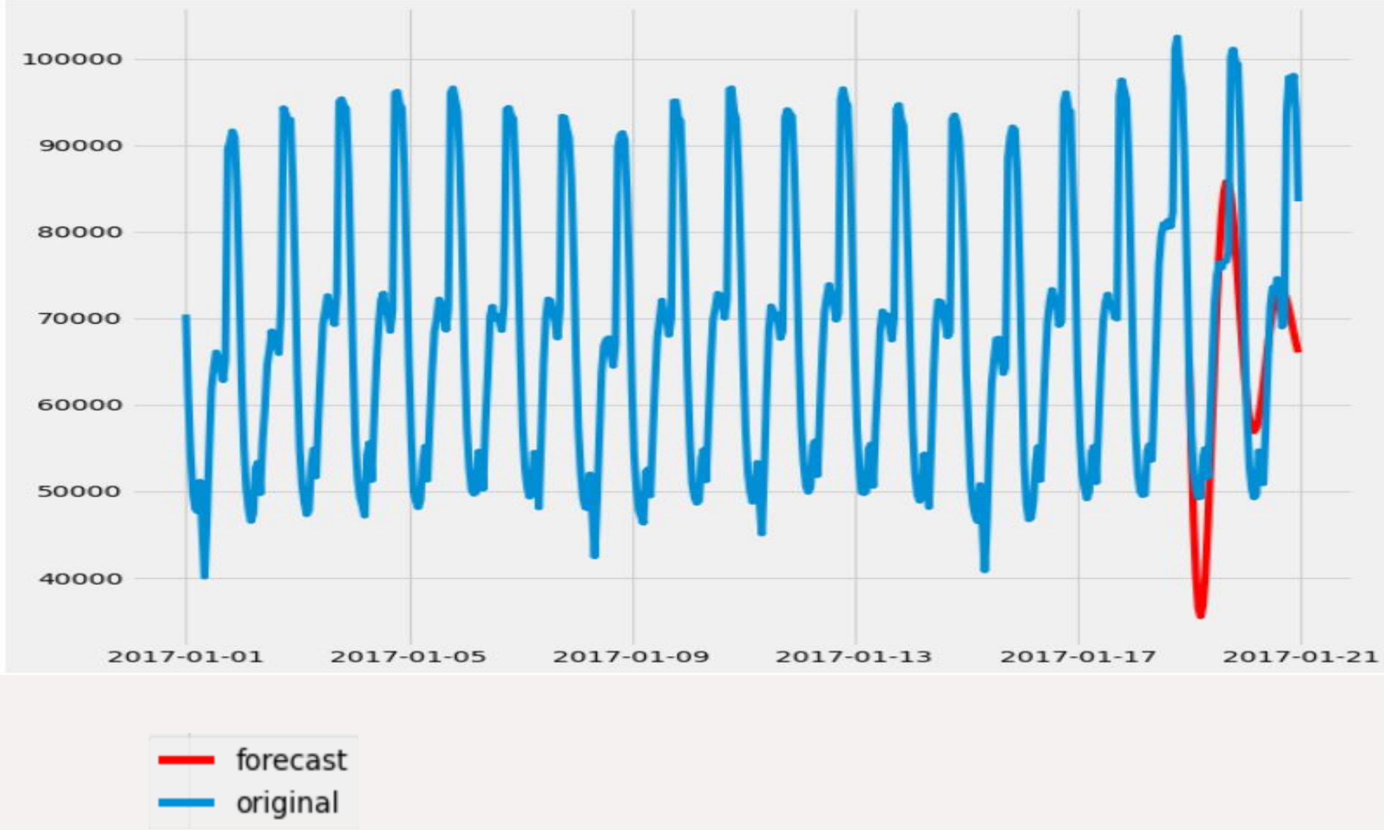




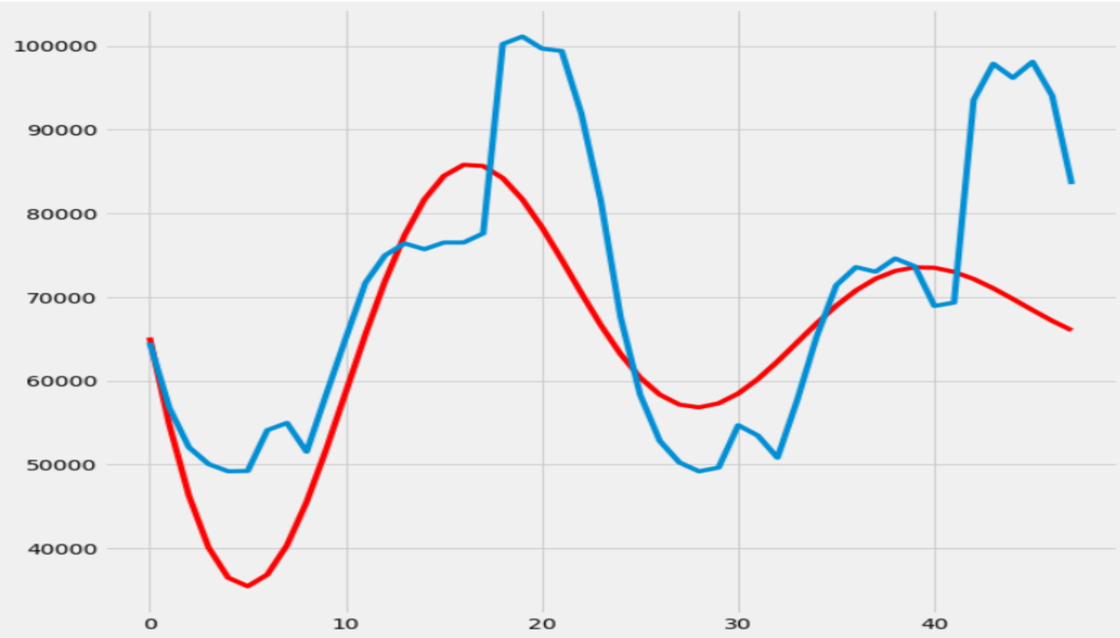


FITTING MODEL WITHOUT REMOVING SEASONALITY

Fitting ARMA (2,4) Model




```python
from statsmodels.tsa.arima_model import ARIMA
model = ARIMA(traindata, order=(2,0,4))
result_AR = model.fit(disp = 0)
plt.plot(data)
plt.plot(result_AR.fittedvalues, color='red')
```



FITTING MODEL WITHOUT REMOVING SEASONALITY

Fitting ARMA (2,2) Model

Dep. Variable:	consumption	**No. Observations:**	432
Model:	ARMA(2, 2)	**Log Likelihood**	-4335.330
Method:	css-mle	**S.D. of innovations**	5500.316
Date:	Tue, 01 Mar 2022	**AIC**	8682.661
Time:	15:33:51	**BIC**	8707.071
Sample:	01-01-2017	**HQIC**	8692.298
	- 01-18-2017		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	6.747e+04	341.918	197.330	0.000	6.68e+04	6.81e+04
ar.L1.consumption	1.8356	0.023	81.115	0.000	1.791	1.880
ar.L2.consumption	-0.9086	0.022	-40.521	0.000	-0.953	-0.865
ma.L1.consumption	-0.6777	0.054	-12.593	0.000	-0.783	-0.572
ma.L2.consumption	-0.2307	0.053	-4.374	0.000	-0.334	-0.127

- The predicted curve is a simple sinusoidal curve, which only predicts the overall seasonality of the of the data.
- We are getting high RSS (Sum of Residual Squares) in this case.

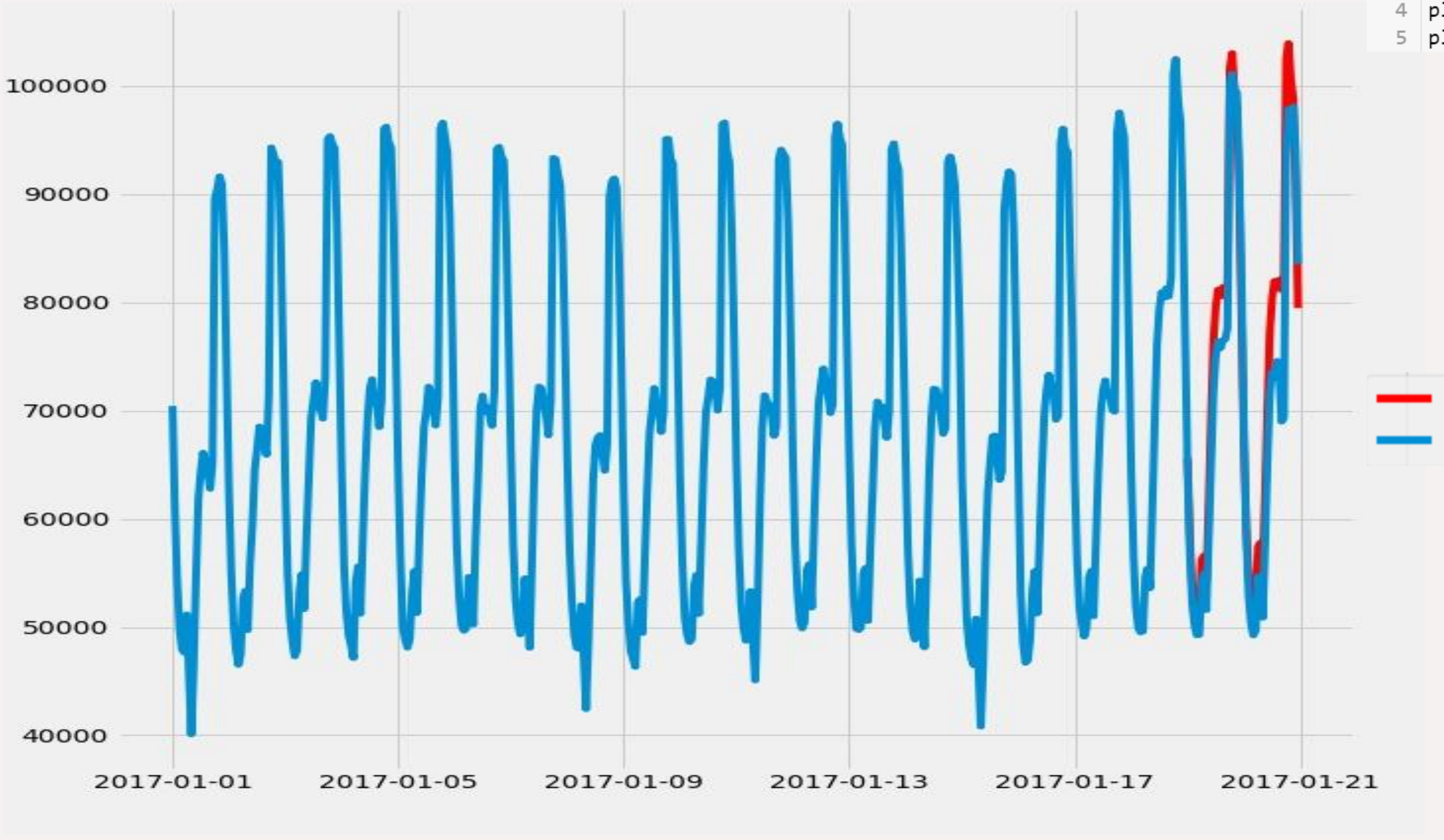
```
RSS : 13122326065.528435
```

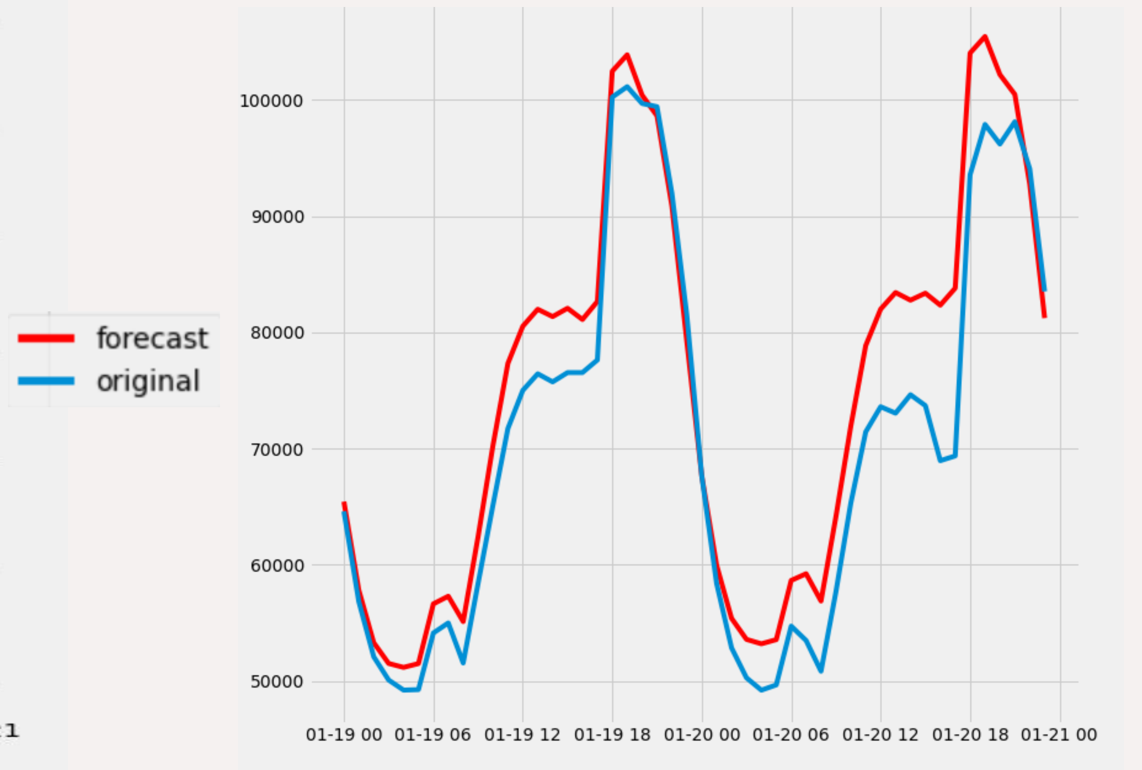
- Hence, a better fitting model has be looked for.

FITTING MODEL WITHOUT REMOVING SEASONALITY

Fitting SARIMA (2,1,4) (1,1,1,24) Model

```python
import statsmodels.api as sm
model = sm.tsa.statespace.SARIMAX(traindata, order=(2,1,3), seasonal_order=(1,1,1,24))
result_AR = model.fit(disp = 0)
plt.plot(data)
plt.plot(result_AR.fittedvalues, color='red')
```





FITTING MODEL WITHOUT REMOVING SEASONALITY

Fitting SARIMA (2,1,4) (1,1,1,24) Model

SARIMAX Results

Dep. Variable:	consumption	**No. Observations:**	432
Model:	SARIMAX(2, 1, 4)x(1, 1, [1], 24)	**Log Likelihood**	-3508.849
Date:	Tue, 01 Mar 2022	**AIC**	7035.697
Time:	16:21:27	**BIC**	7071.777
Sample:	01-01-2017	**HQIC**	7049.976
	- 01-18-2017		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.2939	7.270	0.040	0.968	-13.954	14.542
ar.L2	0.3700	5.549	0.067	0.947	-10.506	11.246
ma.L1	-0.3191	7.264	-0.044	0.965	-14.556	13.918
ma.L2	-0.3658	5.724	-0.064	0.949	-11.585	10.853
ma.L3	-0.0634	0.130	-0.487	0.626	-0.319	0.192
ma.L4	-0.0299	0.517	-0.058	0.954	-1.042	0.983
ar.S.L24	0.4539	0.082	5.508	0.000	0.292	0.615
ma.S.L24	-0.6008	0.091	-6.622	0.000	-0.779	-0.423
sigma2	1.792e+06	7.91e+04	22.640	0.000	1.64e+06	1.95e+06

- The model seems to be a very close and best prediction to the testing data.
- We are getting comparatively low RSS (Sum of Residual Squares) than the ARMA(2,2) model.

```
RSS : 10260343156.844927
```

- Hence, we consider this as a fitting model.

CONCLUSION

MODEL	RSS Value
ARMA (2,2)	13122326065.5284
SARIMA (2,1,2) (1,1,1,24)	10475061221.7473
SARIMA (2,1,3) (1,1,1,24)	10301488099.4280
SARIMA (2,1,4) (1,1,1,24)	10260343156.8449
SARIMA (2,1,5) (1,1,1,24)	10276348078.6051

- Tabel shows the RSS values for different models.
- It shows that the minimum value of RSS is when we select the SARIMA model with ordering (2,1,4)(1,1,1,24).
- Hence the model is fitting the best values.

THANK YOU