Household Energy Prediction

Yein Jeon

Project Overview

saving energy is key to solving environmental problems

such as climate and energy crisis.

Therefore, it is necessary to improve Energy prediction

Goal:

Forecast Energy usage with weather features and time

Data

- Hourly **energy consumption** from January 1, 2017 - December 31, 2018
    - Source: Dataport
    - Variables are: localminute, dataid, and use (energy usage).

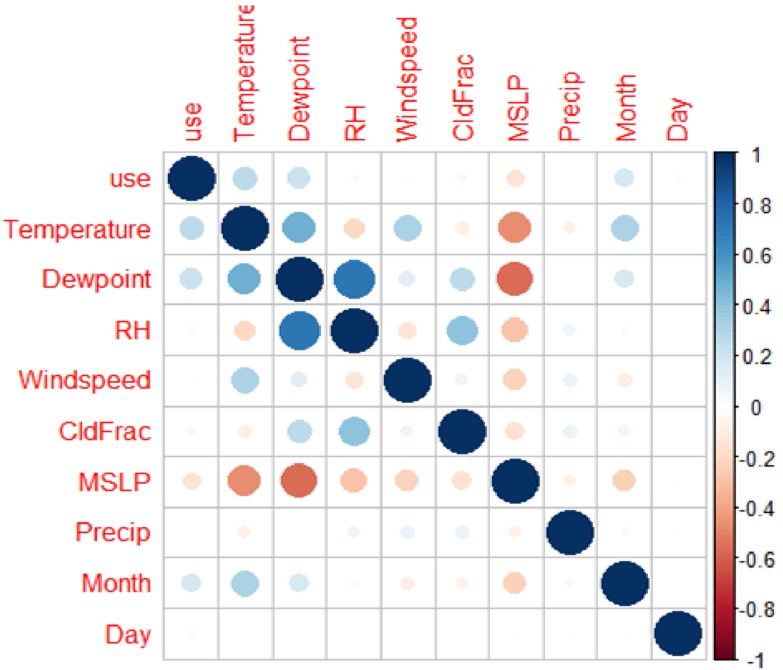
- **Hourly Climate data** for the same period.
    - Features include: temperature, dewpoint, relative humidity, windspeed, mean sea level pressure, precipitation, etc.
    - Source: http://www.frontierweather.com

Data Preprocessing

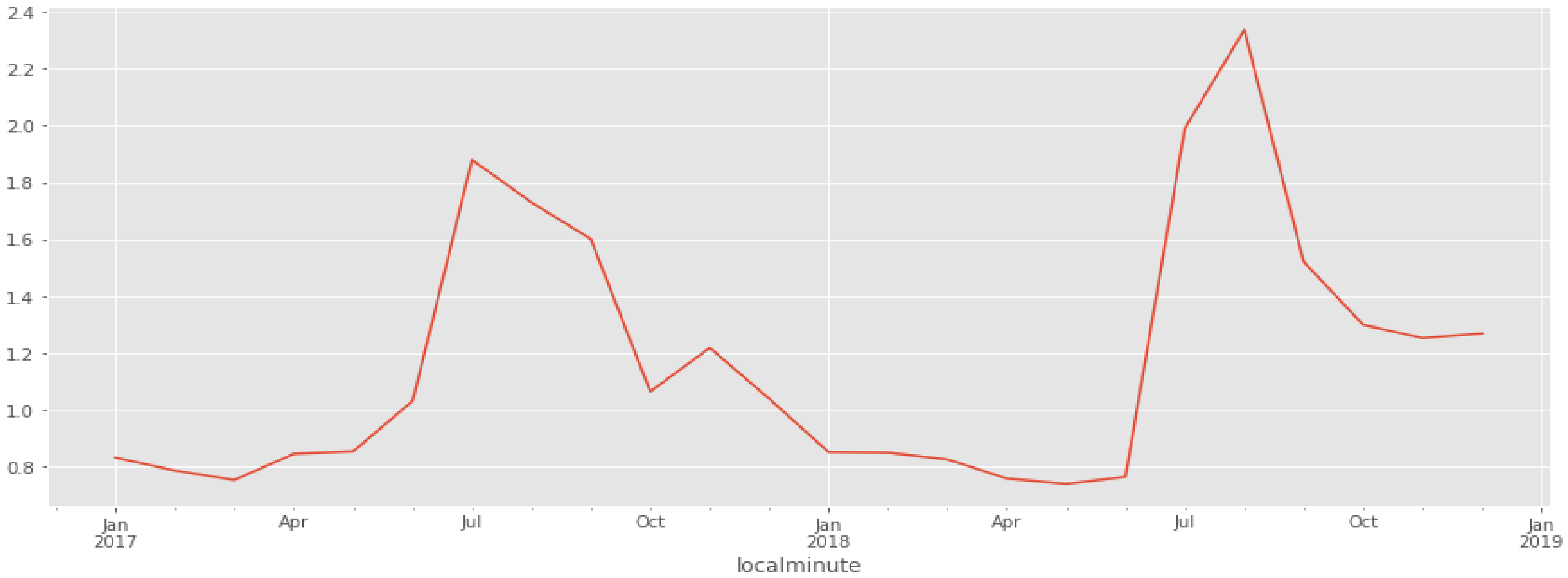
- Data Structure Check
- Dealing with Missing data

```
$ localminute: POSIXct, format: "2017-01-01 00:00:00" "2017-01-01 01:00:00" ...
$ dataid     : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
$ use        : num  0.388 0.593 0.591 0.464 0.295 ...
$ Temperature: num  52 52 51 51 50 50 50 50 50 53 ...
$ Dewpoint   : num  47 46 47 47 46 45 45 44 44 45 ...
$ RH         : num  83 80 86 86 86 83 83 80 80 74 ...
$ Windspeed  : num  0 5 0 3 0 0 3 6 3 6 ...
$ CldFrac    : num  1 0.7 0.7 0.7 0.7 0.4 0.7 0.4 0.4 0.4 ...
$ MSLP       : num  29.9 29.9 29.9 29.9 29.9 ...
$ Precip     : num  1e-02 1e-02 0e+00 0e+00 1e-04 0e+00 0e+00 0e+00 0e+00 0e+00 ...
$ Month      : Factor w/ 12 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
$ Day        : Factor w/ 7 levels "1","2","3","4",..: 7 7 7 7 7 7 7 7 7 7 ...
```

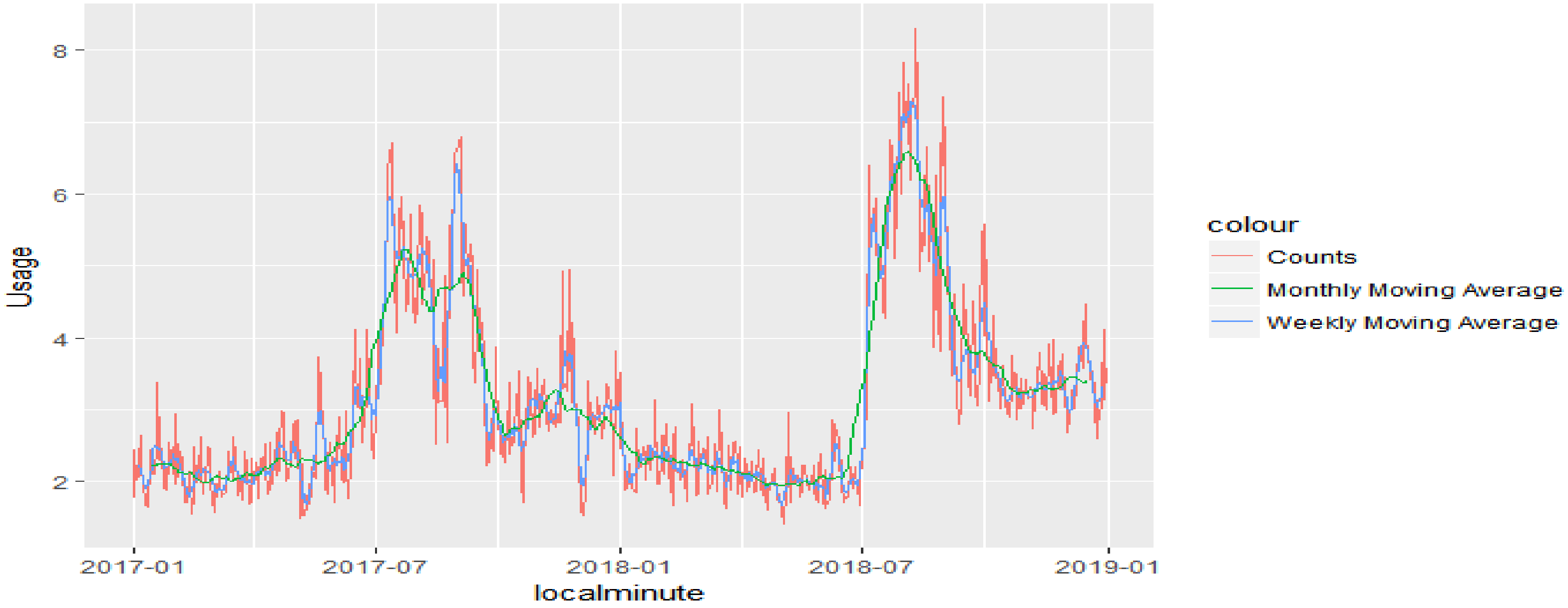
Correlation between Weather Variables and Energy Consumption



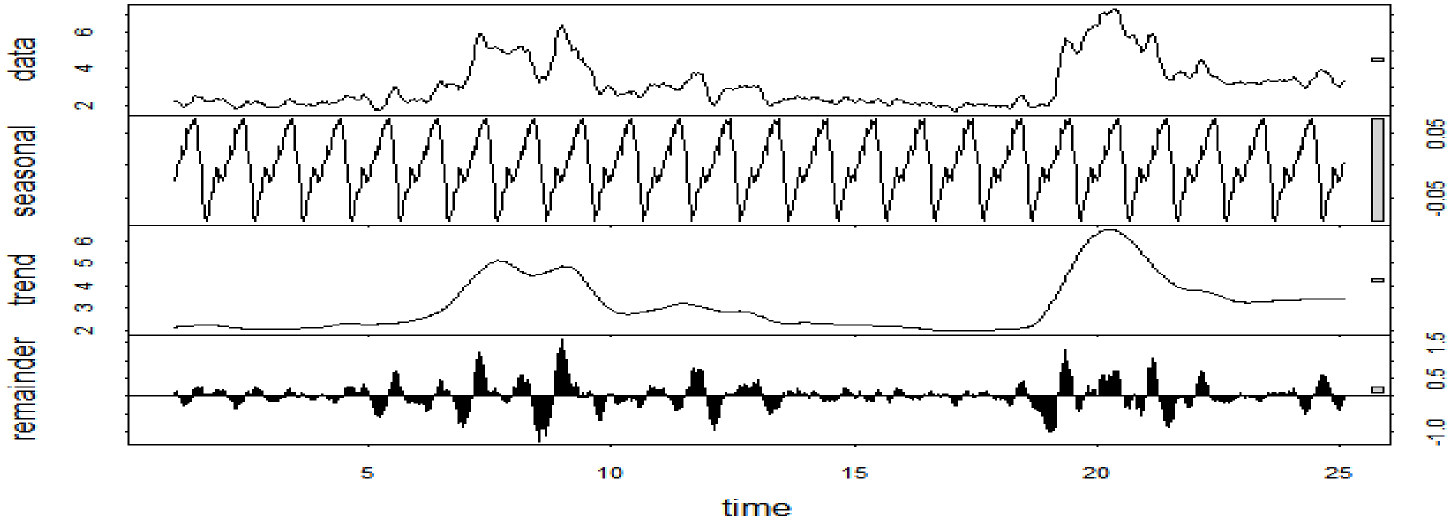
Data Visualization



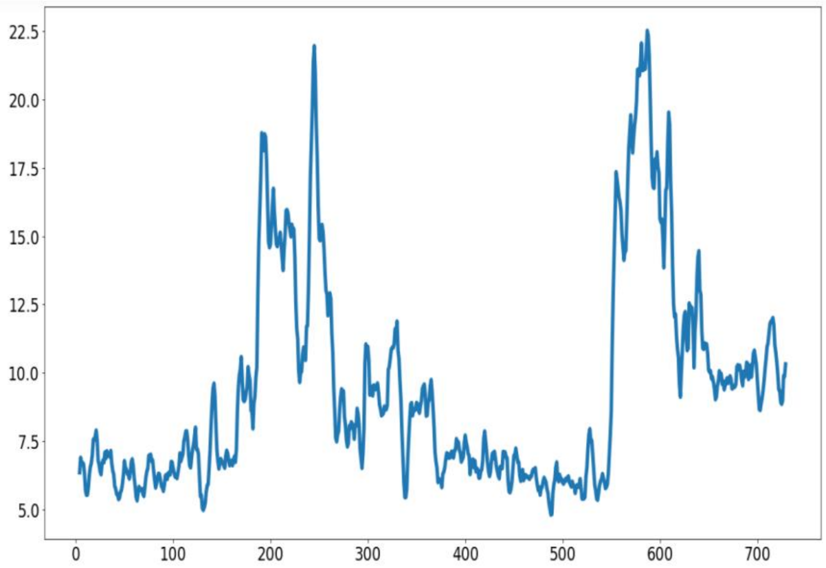
Time Series Visualization

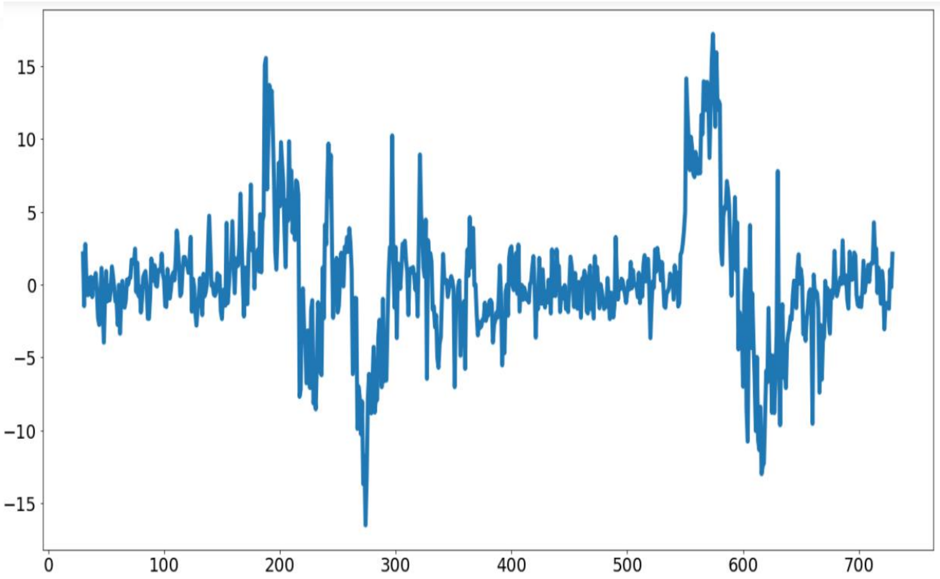


Time Series Visualization



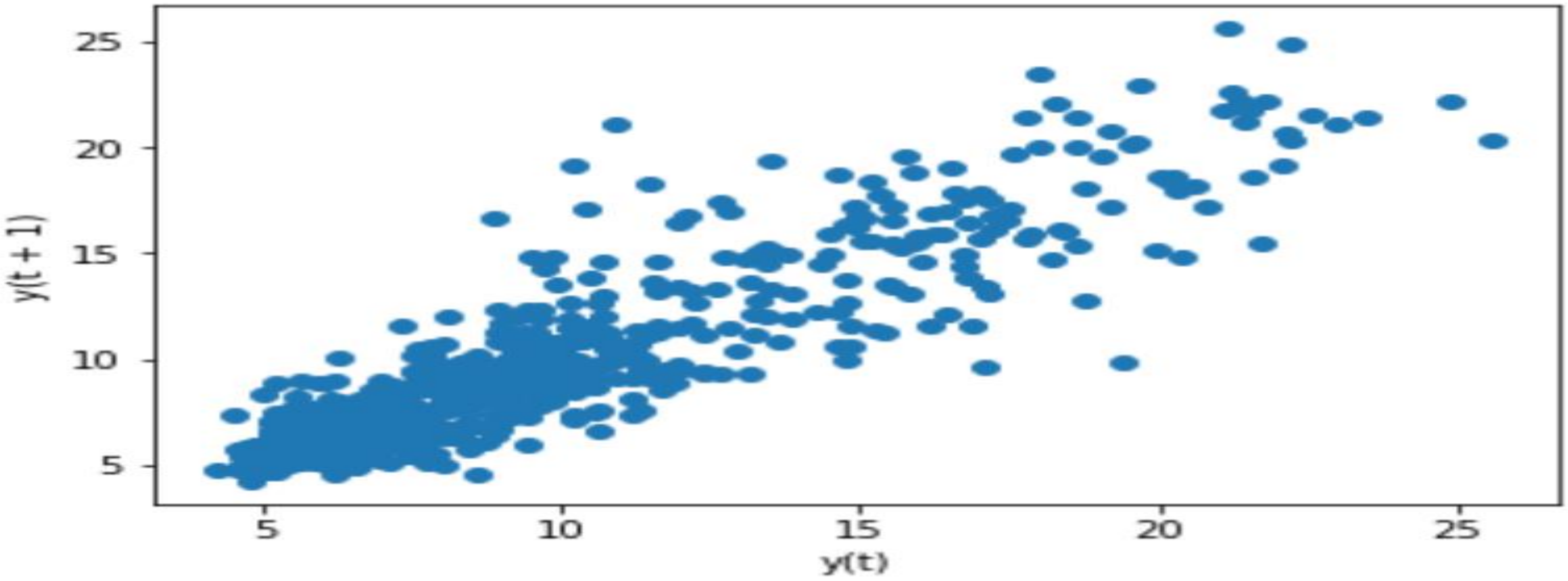
Time Series Visualization



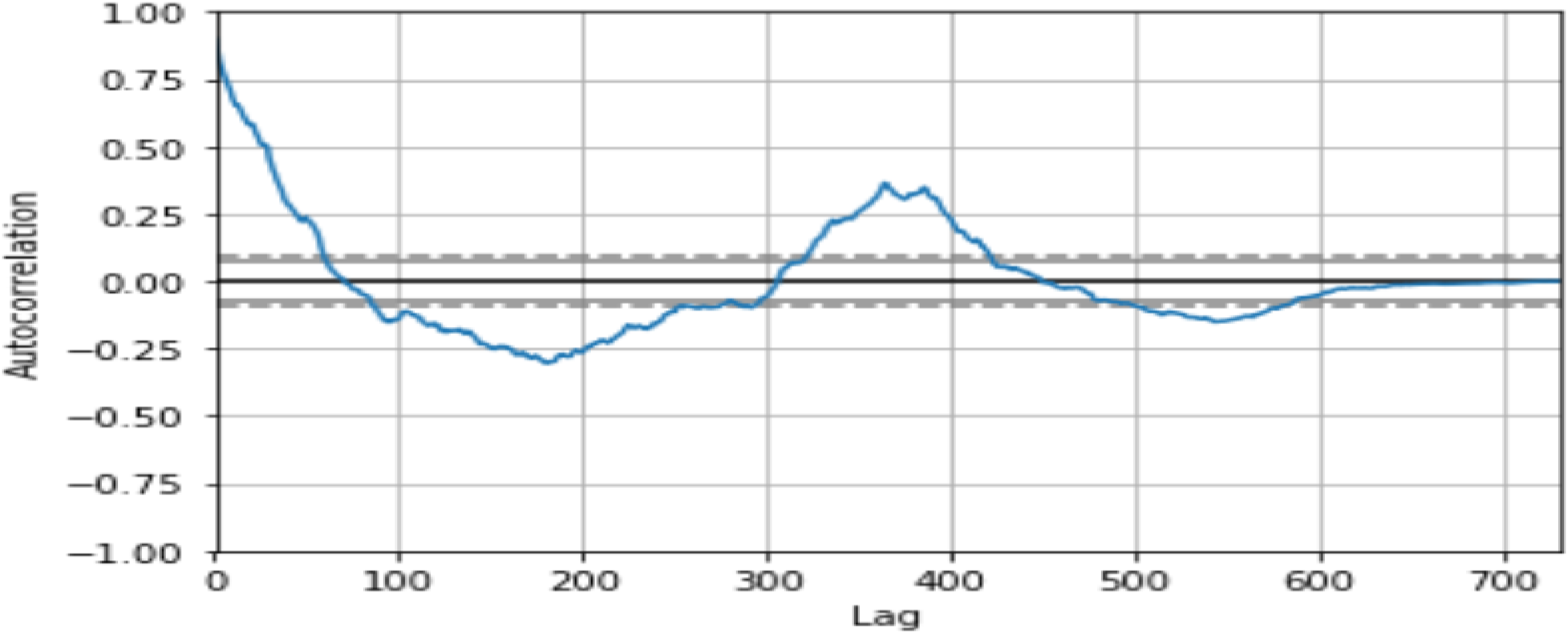


Trend analysis / seasonal variation

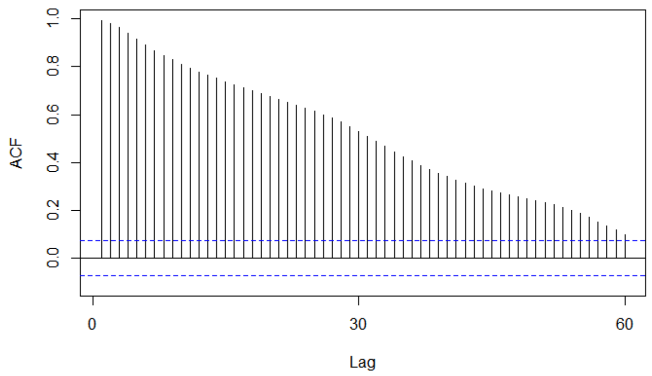
Lag plot

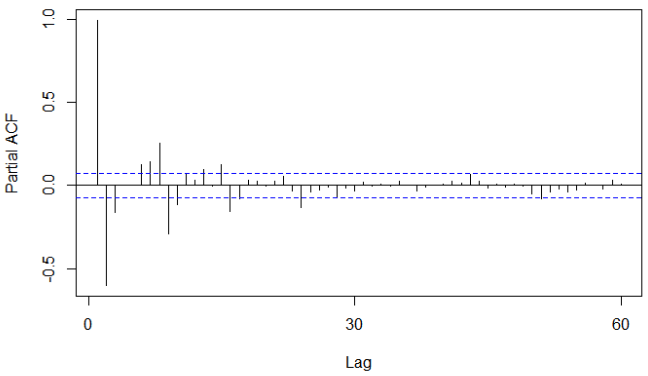


Autocorrelation



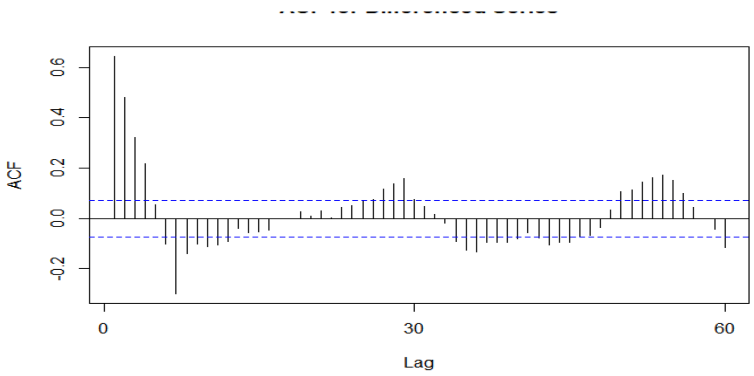
ACF & PACF & Stationery

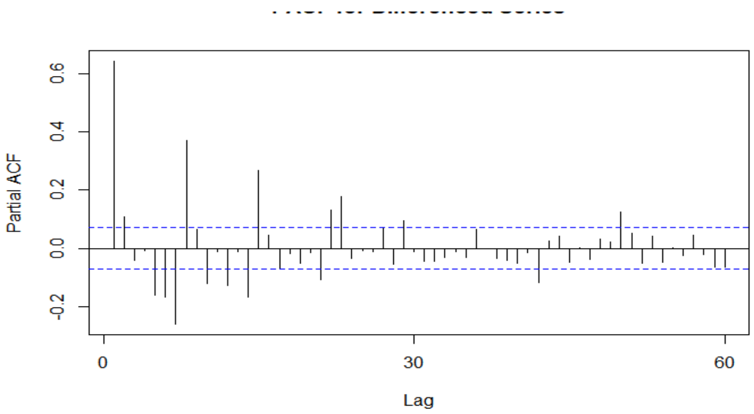





```
        Augmented Dickey-Fuller Test

data:  count_ma
Dickey-Fuller = -2.8292, Lag order = 8, p-value = 0.2273
alternative hypothesis: stationary
```

ACF & PACF & Stationery for Differenced Series

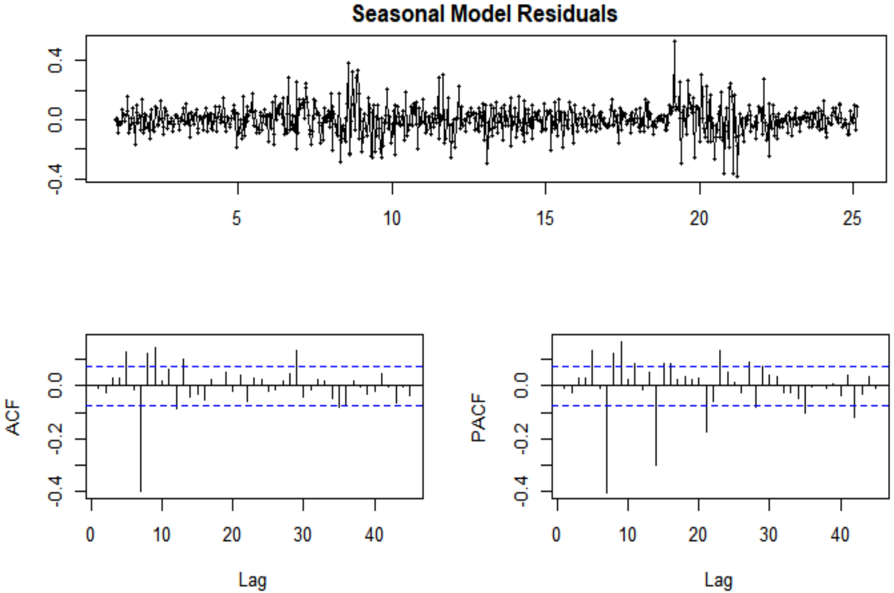




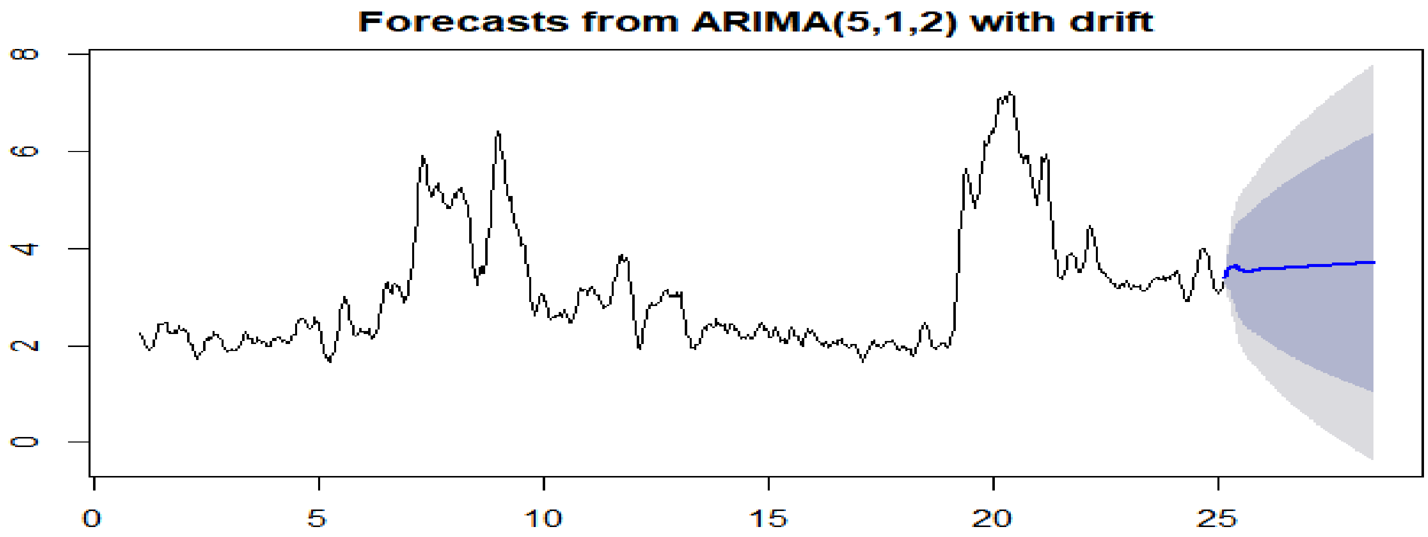

```
p-value smaller than printed p-value
        Augmented Dickey-Fuller Test

data:  count_d1
Dickey-Fuller = -7.3477, Lag order = 8, p-value = 0.01
alternative hypothesis: stationary
```

ARIMA with Seasonality

```
Series: deseasonal_use
ARIMA(5,1,2) with drift

Coefficients:
         ar1     ar2      ar3     ar4      ar5     ma1      ma2   drift
      0.5144  0.6098  -0.2465  0.0321  -0.1908  0.0585  -0.4719  0.0019
s.e.  0.0970  0.0853   0.0701  0.0445   0.0385  0.0948   0.0775  0.0076

sigma^2 estimated as 0.00965:  log likelihood=655.32
AIC=-1292.64   AICc=-1292.39   BIC=-1251.39
```



ARIMA Forecasting



ARIMA Prediction & Accuracy

	Predicted value	SE	Actual Value
2019/01/01	3.392099	0.09769617	3.590859
2019/01/02	3.484767	0.18209222	3.251024
2019/01/03	3.530264	0.26748686	3.215609
2019/01/04	3.592768	0.35129686	3.585437
2019/01/05	3.606636	0.43657021	3.847791
2019/01/06	3.627345	0.51201976	3.265519
2019/01/07	3.614837	0.58065545	3.153457

```
                        ME       RMSE        MAE           MPE       MAPE
Training set -3.606931e-05 0.09762437 0.07076282  1.743486e-03   2.335626
Test set     -5.614305e-01 1.27007409 0.66902933 -3.652872e+02 368.175579
                  MASE        ACF1
Training set 0.7597808 -0.00727974
Test set     7.1833713          NA
```

THANK YOU……

DO YOU HAVE ANY QUESTIONS ?