Names: Albert MANIRAHO
ANDREW ID: maniraha
Kigali, February 13, 2022

DATA ANALYTICS ASSIGNMENT 2 REPORT

All the libraries used:

```
from scipy import stats
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf
import math
import seaborn as sns
import statsmodels.api as sm
import warnings
import statistics
from statsmodels.graphics.tsaplots import plot_acf
import scipy.stats as stats
from arch.unitroot import VarianceRatio
from arch.unitroot import ADF
import itertools
from statsmodels.tsa.arima_model import ARMA
from statsmodels.tsa.arima_model import ARIMA
from sklearn.metrics import mean_absolute_error as mae
warnings.filterwarnings('ignore')
```

QUESTION 1.

The following are the ways gone through to solve the problem:

- Download the required dataset and load it into the IDE (Integrated Development Environment).
- The filled the missing values in the dataset.
- Generation of the dates and timestamps:

	Date	Time	Wind Generation
0	2014-01-01 02:00:00	1	1190.1
1	2014-01-01 04:00:00	2	1186.6
2	2014-01-01 06:00:00	3	1174.3
3	2014-01-01 08:00:00	4	1146.7
4	2014-01-01 10:00:00	5	1156.8
...	...	...	...
8755	2015-01-01 16:00:00	20	996.8
8756	2015-01-01 18:00:00	21	1031.8
8757	2015-01-01 20:00:00	22	1218.0
8758	2015-01-01 22:00:00	23	1312.1
8759	2015-01-02 00:00:00	24	1424.8

8760 rows × 3 columns

Figure 1: The dataset showing the dates and timestamps.

A graphic showing the time series of the wind generation over time on hourly basis:

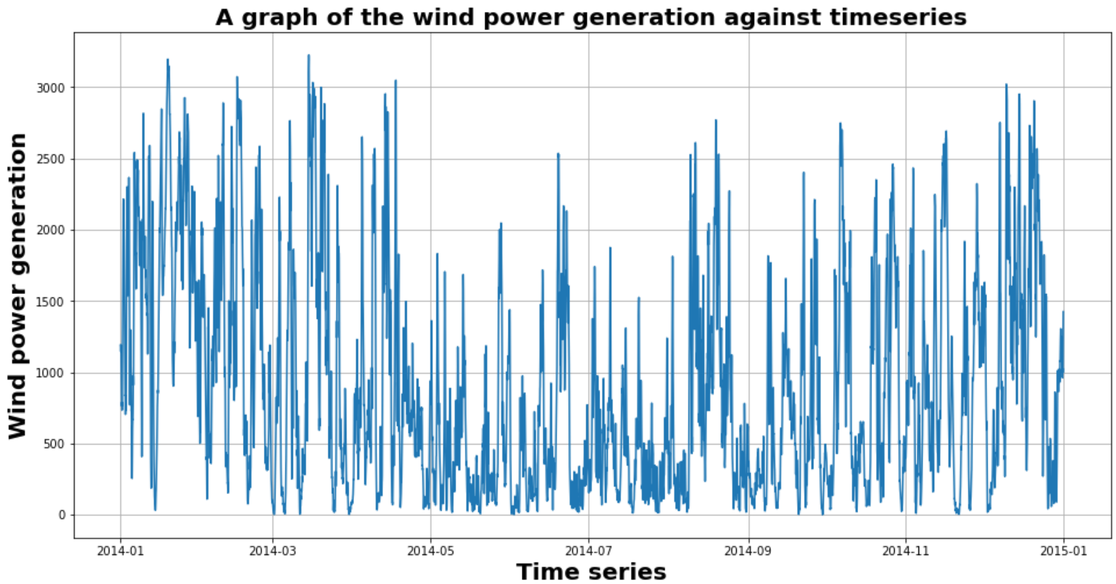


Figure 2: The graph of the wind power generation against timeseries.

- **Doing sampling for a better visualization:**
    - Daily sampling gives the graph below and is clearer compared to the first one:

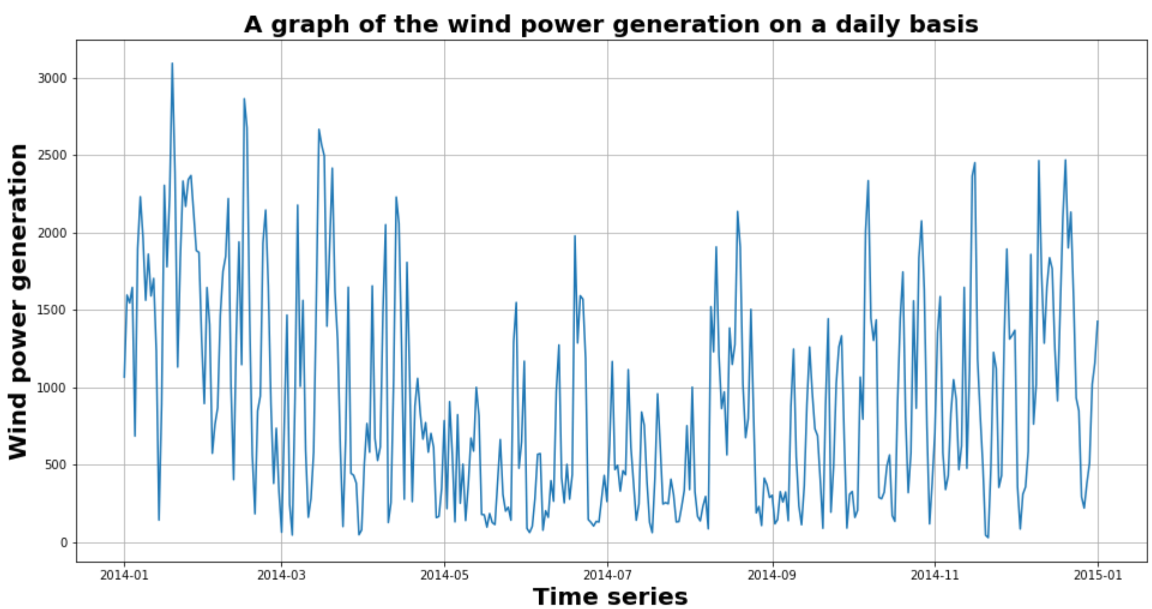


Figure 3: The graph of the wind power generation on a daily basis.

    - Weekly sampling gives a graph that is clear compared to the daily sampling:

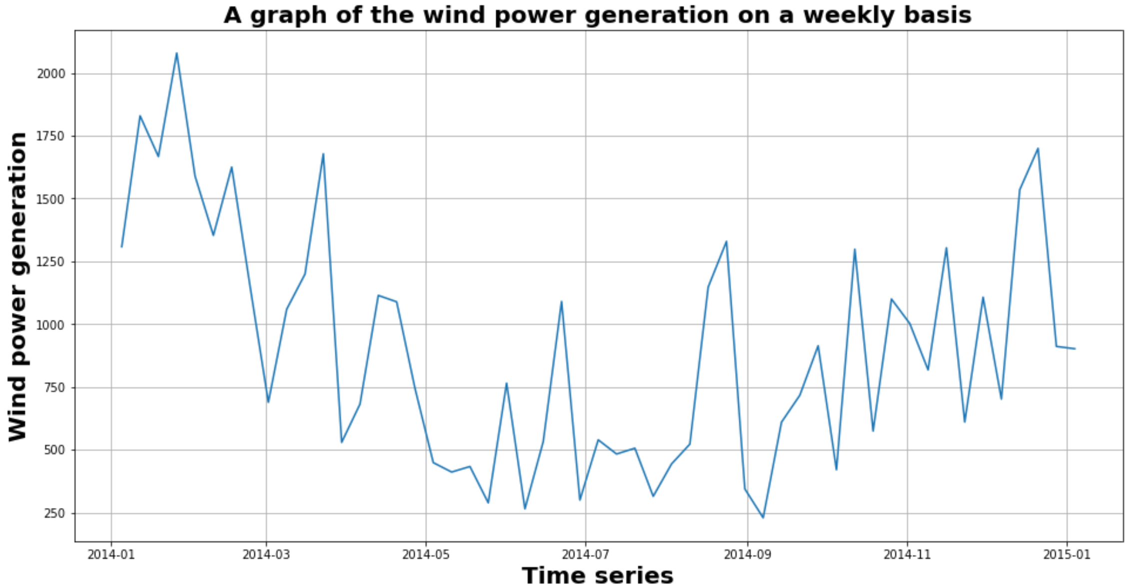


Figure 4: A graph of the wind power generation on a weekly basis.

- Monthly sampling produces the following graph:

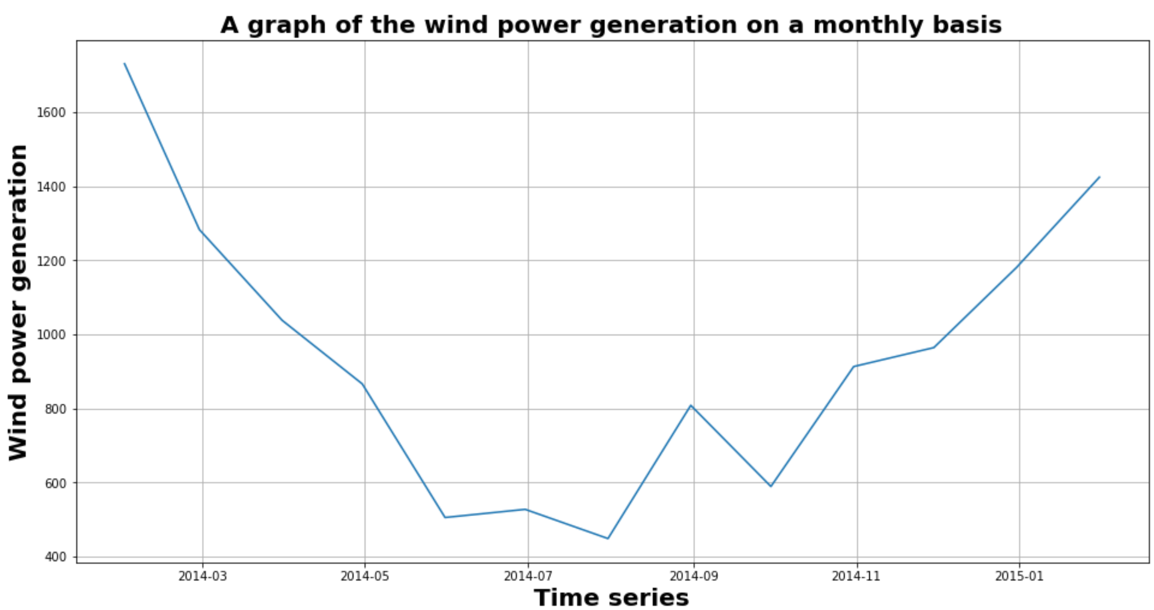


Figure 5: A graph of wind power generation on a monthly basis.

- Quarterly sampling produces the following graph:

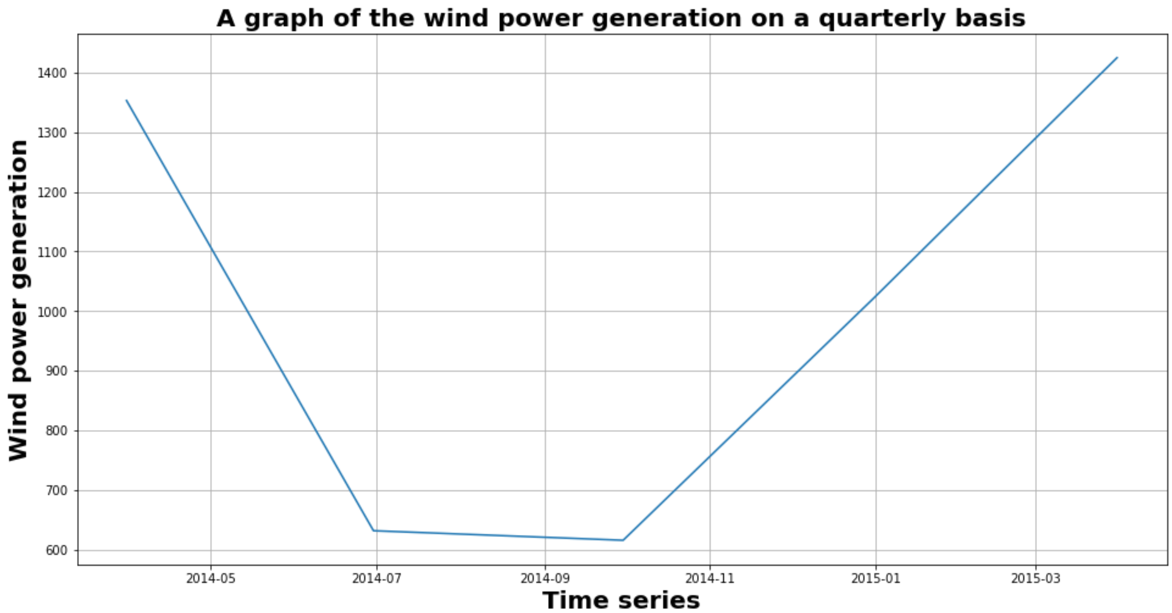


Figure 6: A graph of the wind power generation on a quarterly basis.

Yes, there is evidence of seasonality, because there is a regular pattern in the graph.

QUESTION 2.

The following steps have been taken to solve the question.

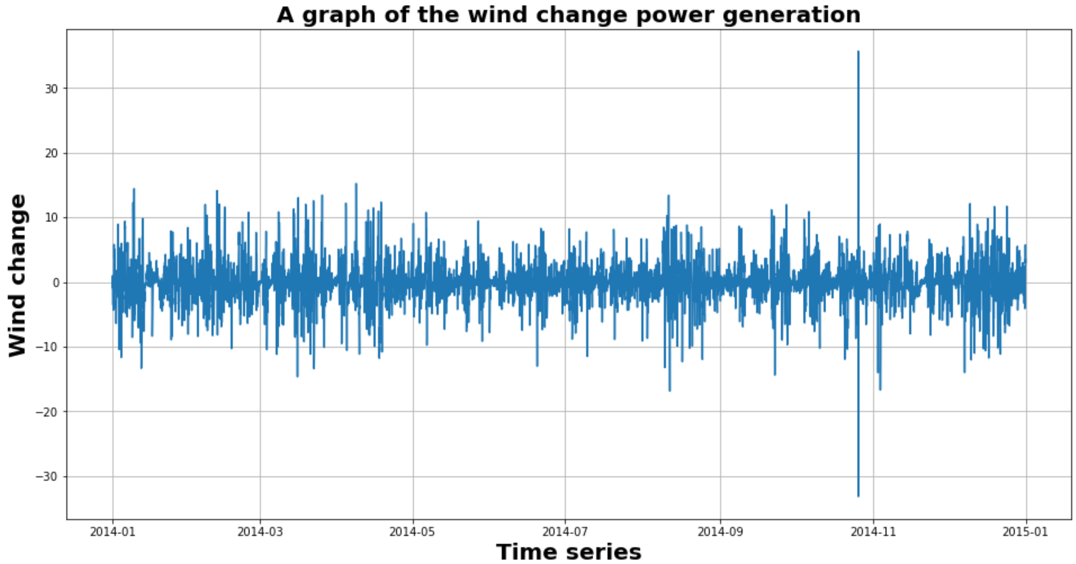
- Calculation of the change in wind generation using the appropriate formula and this is the results:

	Date	Time	Change in generation
1	2014-01-01 02:00:00	2	-0.139518
2	2014-01-01 03:00:00	3	-0.412352
3	2014-01-01 04:00:00	4	-0.886712
4	2014-01-01 05:00:00	5	0.282136
5	2014-01-01 06:00:00	6	0.989025
...	...	...	...
8755	2014-12-31 20:00:00	20	-0.967322
8756	2014-12-31 21:00:00	21	1.054133
8757	2014-12-31 22:00:00	22	5.741923
8758	2014-12-31 23:00:00	23	2.886464
8759	2015-01-01 00:00:00	24	3.463136

8759 rows × 3 columns

Figure 7: A table showing the change in wind generation.

A graphic showing the plot of the change in wind generation against the timeseries on hourly



basis:

Figure 8: A graph of change in wind generation against timeseries.

- **Doing sampling for a better visualization:**
    - ➢ Daily sampling gives the graph below and is clearer compared to graph above:

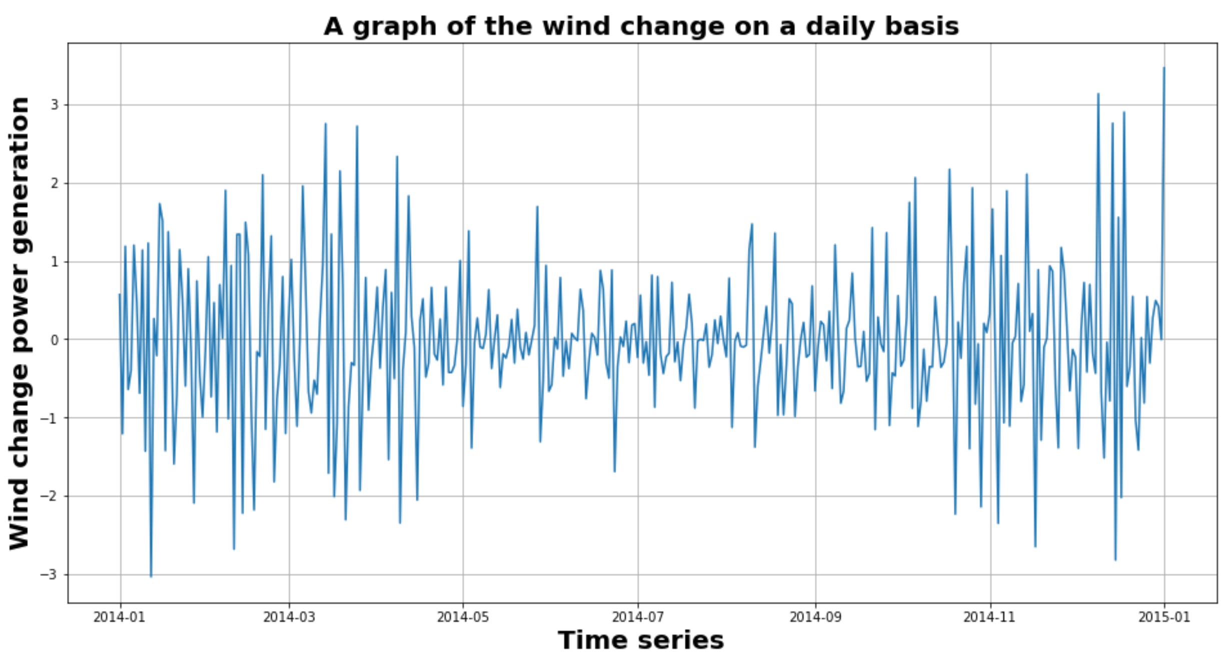


Figure 9: A graph of the change in wind generation on a daily basis.

- Weekly sampling gives a graph that is clear compared to the daily sampling:

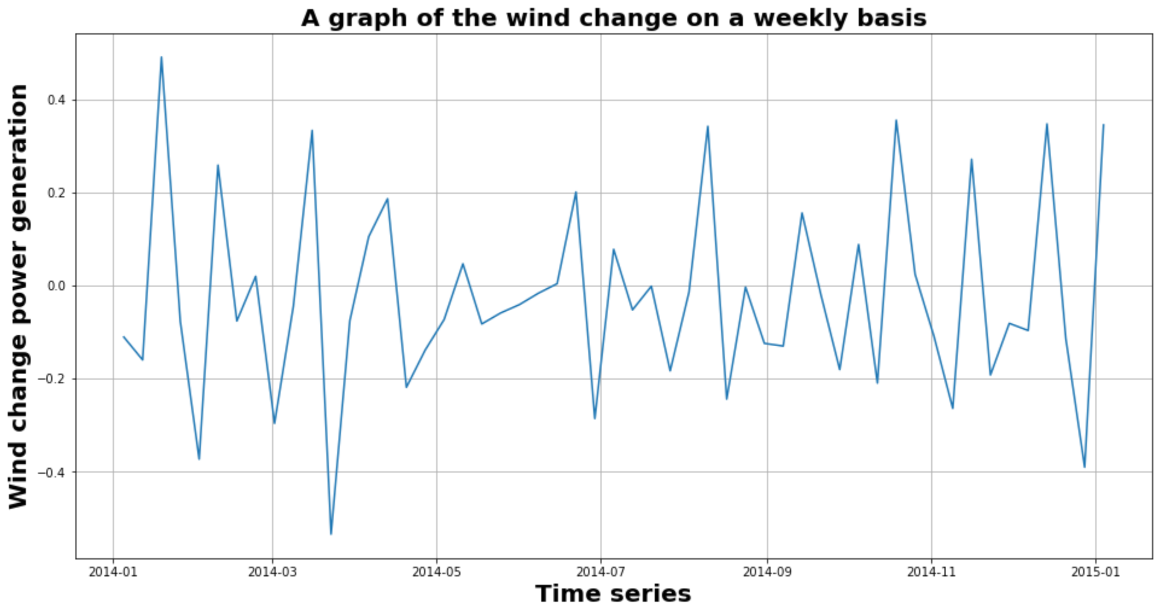


Figure 10: A graph of the change in wind generation on a weekly basis.

- Monthly sampling produces the following graph:

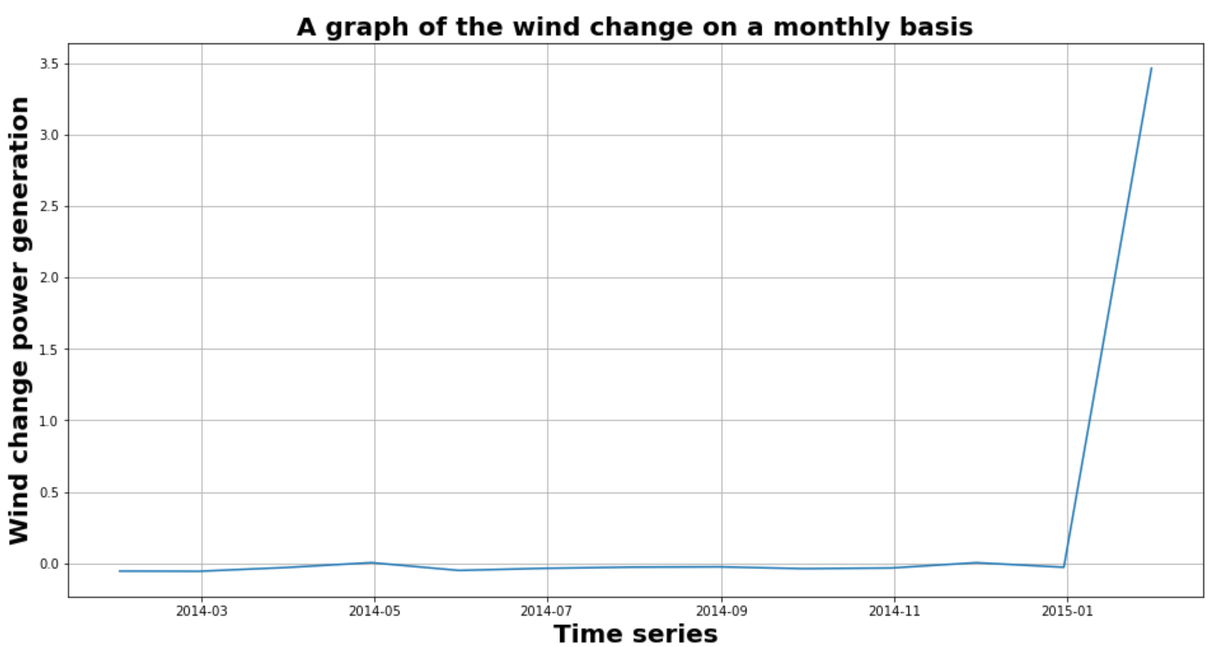


Figure 11: A graph of the change in wind generation on a monthly basis

.

- Quarterly sampling produces the following graph:

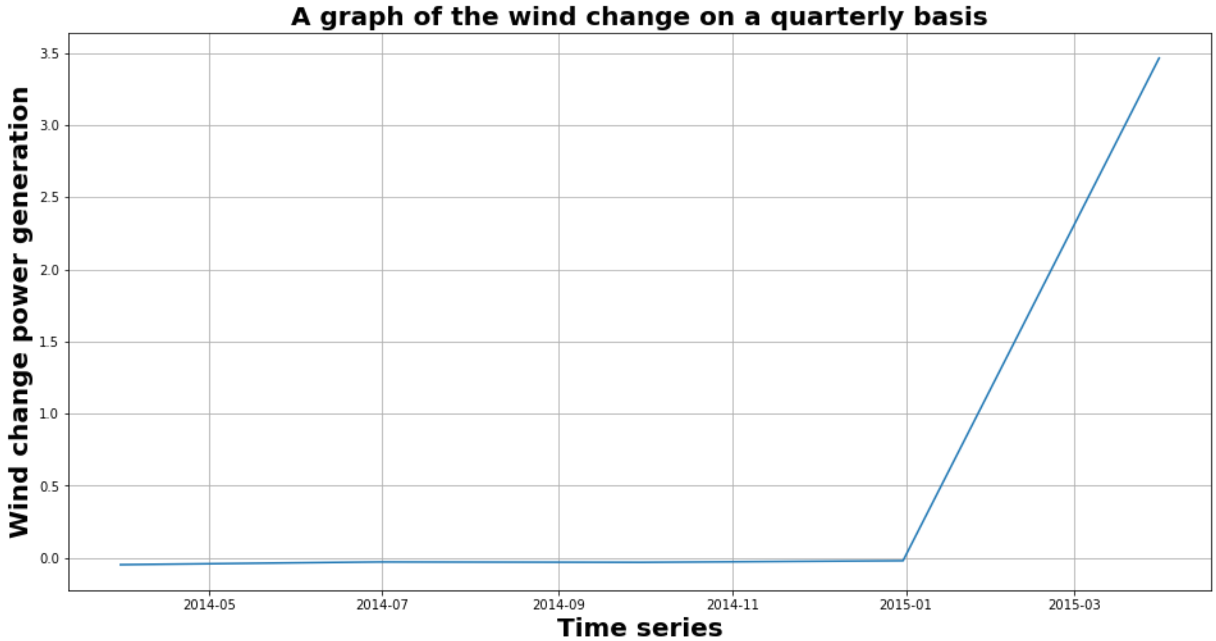


Figure 12: A graph of the change in wind generation on a quarterly basis.

Yes, there is evidence of seasonality, because there is a regular pattern in the graph.

QUESTION 3.

The following are the steps taken to solve the problem:

- The ramps calculations using the formula: **r(t,d) =100 * [x (t + d) - x(t)] / max(x)**, where d=1 on hourly sampling.
- The positive and negative ramps were separated using the appropriate codes.
- Both the positive and negative ramps were sorted on an ascending order.
- For the negative ramps, we had to find the absolute values before sorting them.
- Finding of the y-positive and y-negative.
- Plot of the ramps using a semilogy plot.

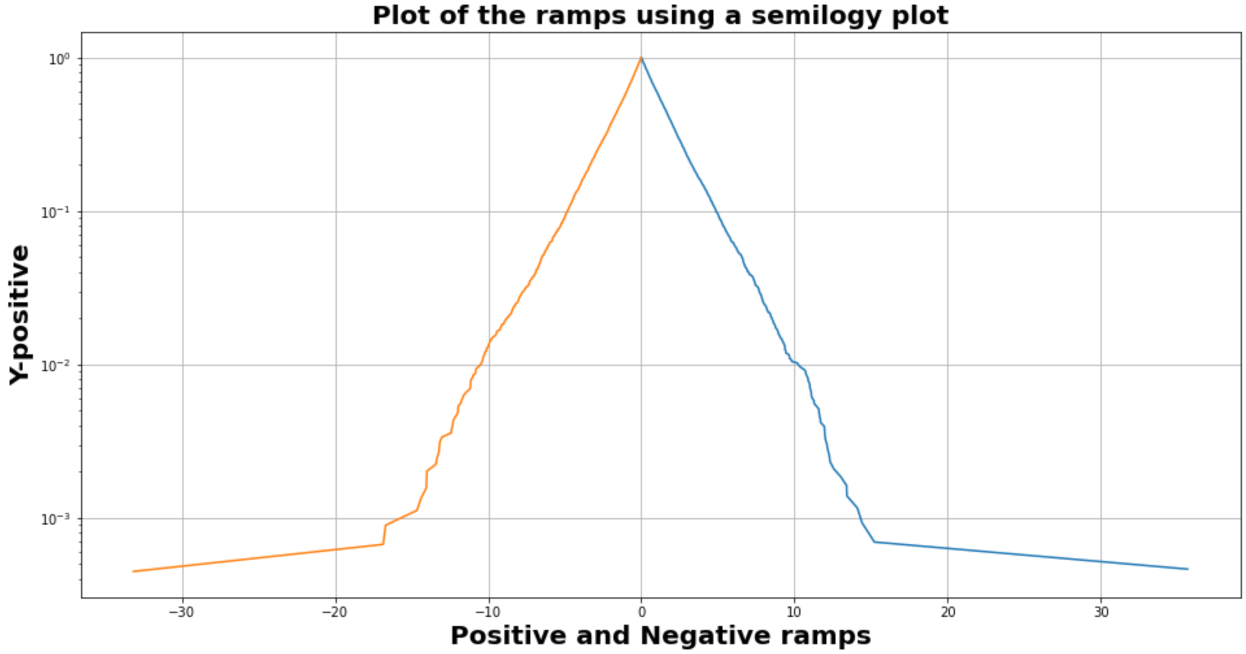


Figure 13: Plot of the ramps using a semilogy plot

- Plot of a normal distribution CDF graph:

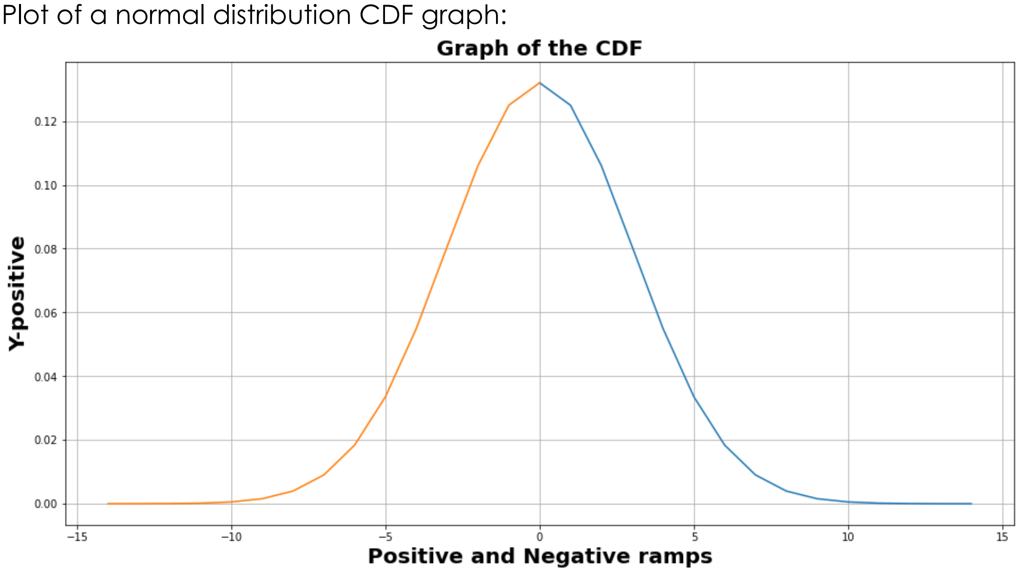


Figure 14: Plotting of the normal distribution CDF.

- Graph of the combined semiology and normal distribution CDF.

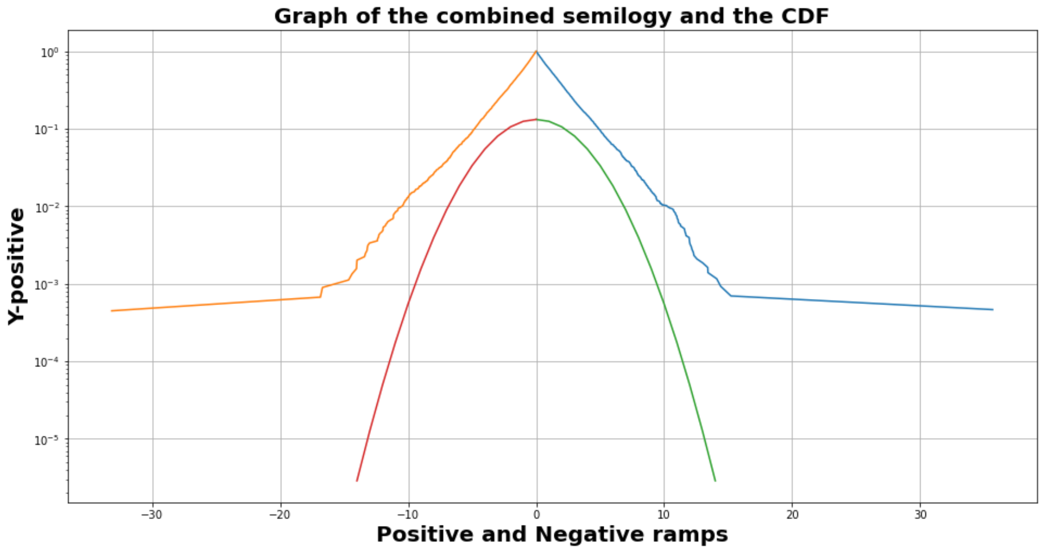


Figure 15: Graph of the combined semilogy and the CDF.

Is the normal distribution a good model for wind power extremes?

Answer:

No

QUESTION 4.

These are the steps taken to solve this problem:

- Finding of the wind power generations over a time scale of 24hours(from 1h up to 24h).
- Calculations of the 1%, 5%, 95%, 99% percentile and the table of the percentiles is here below:

	0.01	0.05	0.95	0.99
h1	-8.716841	-4.815062	4.880015	8.479010
h2	-16.060179	-9.059496	9.192968	16.316798
h3	-22.552397	-12.898090	12.910182	22.088826
h4	-27.915421	-16.268060	16.290538	27.687760
h5	-32.825107	-19.332486	19.254356	32.462981
h6	-36.398183	-22.313667	22.045018	36.205959
h7	-39.360730	-24.445030	24.267533	39.665995
h8	-42.256247	-27.089663	26.147145	41.597538
h9	-44.676846	-29.287840	27.863366	44.366745
h10	-46.464687	-31.066689	29.746388	46.685310
h11	-48.123458	-32.863831	31.411143	48.319960
h12	-49.138463	-34.074998	32.485893	51.162677
h13	-50.511781	-35.232529	33.481274	53.170087
h14	-51.630216	-36.212253	34.659422	54.191604
h15	-52.600329	-37.182985	35.664104	55.483847
h16	-53.816736	-37.763068	36.621194	56.060613
h17	-54.713865	-38.699541	37.468996	56.617722
h18	-55.060551	-39.260557	38.351832	57.769083
h19	-55.676815	-39.494481	39.206920	58.427885
h20	-55.510665	-39.534941	39.807156	58.312922
h21	-55.698084	-39.456967	40.442426	59.380015
h22	-56.383301	-39.480375	40.814163	59.125690
h23	-56.663453	-39.785453	41.467105	59.350902
h24	-57.091462	-39.971321	41.890153	59.804737

- Plotting of the percentiles against timescales.

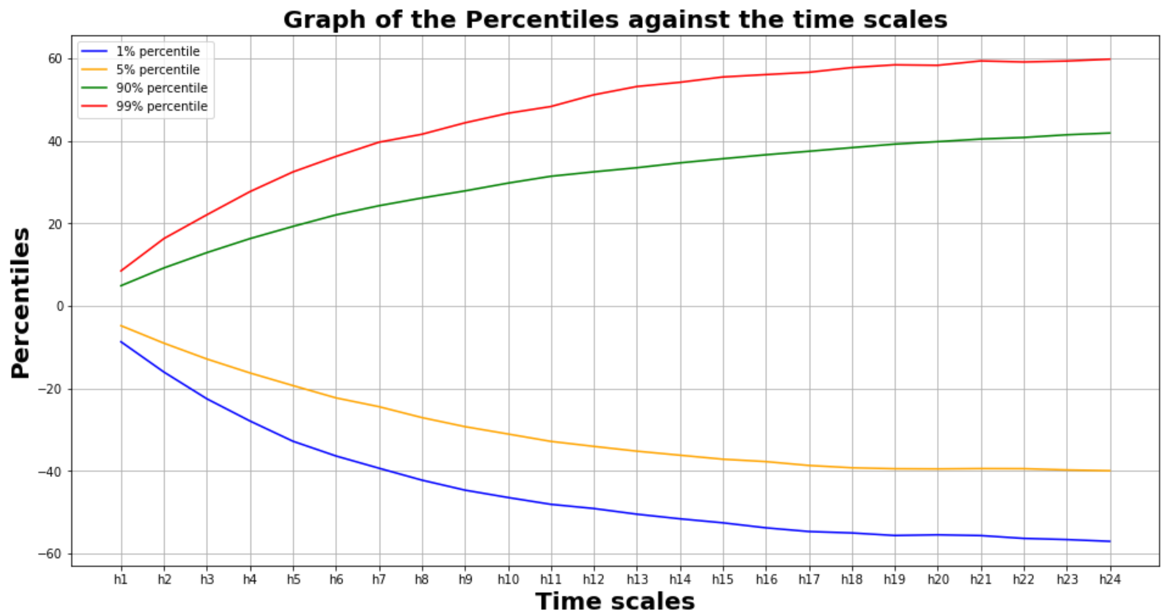


Figure 16: Graph of the percentiles against the timescales.

95% and 99% percentiles increase the generation while 1% and 5% percentiles decrease the generation.

QUESTION 5.

The steps taken to solve the problem:

- Calculated the autocorrelation of the wind generation (actual) for 10 days lags. The number of lags in **10 days lags = 1hr * 24hr * 10 = 240 lags.**
- The autocorrelation is as follows:

```
array([1.        , 0.99194168, 0.97259635, 0.94626189, 0.91599517,
       0.88370139, 0.85076521, 0.81829992, 0.78702714, 0.75737762,
       0.72965525, 0.70392066, 0.68004856, 0.65796888, 0.63768012,
       0.61926857, 0.60286594, 0.58848166, 0.57613357, 0.56566213,
       0.55678102, 0.54897868, 0.5413343 , 0.53276825, 0.52224839,
       0.50919157, 0.49420076, 0.47828309, 0.46224332, 0.44650854,
       0.43138   , 0.4170529 , 0.4036398 , 0.39130513, 0.38002356,
       0.36961585, 0.35979554, 0.35056   , 0.34198318, 0.33422265,
       0.32748535, 0.32183807, 0.31742221, 0.31432072, 0.3123028 ,
       0.31082684, 0.3093631 , 0.30736068, 0.30398574, 0.29879197,
       0.2919623 , 0.28424961, 0.27632077, 0.26870707, 0.26172618,
       0.25557006, 0.25044336, 0.24634119, 0.24308739, 0.24038062,
       0.23789009, 0.23551838, 0.23331617, 0.23151362, 0.23028232,
       0.22968379, 0.22979245, 0.23063303, 0.23206178, 0.23379058,
```

```
0.23538313, 0.2362559 , 0.23568915, 0.23323731, 0.22921278,
0.22434811, 0.21914068, 0.21395152, 0.20905333, 0.20460451,
0.20085169, 0.19796213, 0.19594404, 0.19466872, 0.19398825,
0.19370871, 0.19373366, 0.19408805, 0.19489318, 0.19618568,
0.19788193, 0.19984082, 0.20179256, 0.20342359, 0.20429849,
0.20381548, 0.20132172, 0.1966148 , 0.19026045, 0.18311677,
0.17585447, 0.16888802, 0.16251876, 0.15695867, 0.15242113,
0.14896329, 0.14646409, 0.14472588, 0.14357177, 0.14300774,
0.14308035, 0.1437869 , 0.14515649, 0.14717797, 0.14976546,
0.15281734, 0.15633136, 0.16008932, 0.16369226, 0.16641126,
0.16750006, 0.16659355, 0.16398024, 0.16039221, 0.15642525,
0.15262619, 0.1495408 , 0.14743637, 0.14640381, 0.14662393,
0.14801935, 0.15040569, 0.15344713, 0.15701725, 0.1609494 ,
0.16502785, 0.16923936, 0.17351185, 0.17787733, 0.18240017,
0.18706312, 0.19157777, 0.19562787, 0.19857512, 0.19987703,
0.19942568, 0.19736634, 0.19432019, 0.19088167, 0.1874845 ,
0.18443633, 0.18192362, 0.1801681 , 0.1792333 , 0.17899522,
0.17919911, 0.17957532, 0.18022563, 0.18126304, 0.18286567,
0.18513467, 0.18818392, 0.19197664, 0.19642785, 0.20135239,
0.20636535, 0.21097461, 0.21446087, 0.2161955 , 0.21596717,
0.21400525, 0.21074211, 0.20672493, 0.20243916, 0.19833552,
0.19475252, 0.191858  , 0.18974013, 0.18836813, 0.18759747,
0.18724612, 0.18718941, 0.18754457, 0.18849482, 0.19007072,
0.19226616, 0.19512003, 0.1985351 , 0.20230757, 0.20598536,
0.20916728, 0.21145353, 0.2123571 , 0.21159188, 0.20933863,
0.2061714 , 0.20273152, 0.19942123, 0.19637636, 0.19370244,
0.19141776, 0.18959466, 0.1881757 , 0.18700301, 0.18606553,
0.18547669, 0.18544477, 0.18597839, 0.18702991, 0.18870477,
0.19087976, 0.19337949, 0.19593551, 0.19835207, 0.20023825,
0.20110634, 0.20029232, 0.19755108, 0.19324104, 0.18810128,
0.18282727, 0.17793122, 0.17364795, 0.17010012, 0.16736541,
0.16543272, 0.16410929, 0.16312718, 0.16225739, 0.16133974,
0.16041219, 0.15947715, 0.15865228, 0.15799645, 0.15751915,
0.15727247, 0.15721578, 0.15720152, 0.15680528, 0.15547998,
0.15264941])
```

- The graph of the autocorrelation using wing generation.

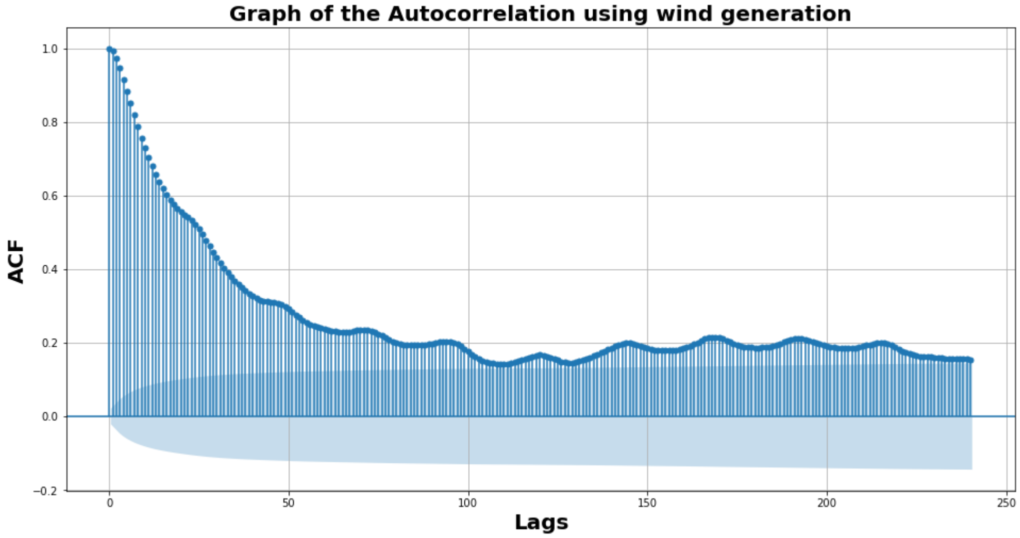


Figure 17: The graph of the autocorrelation using the wind generation.

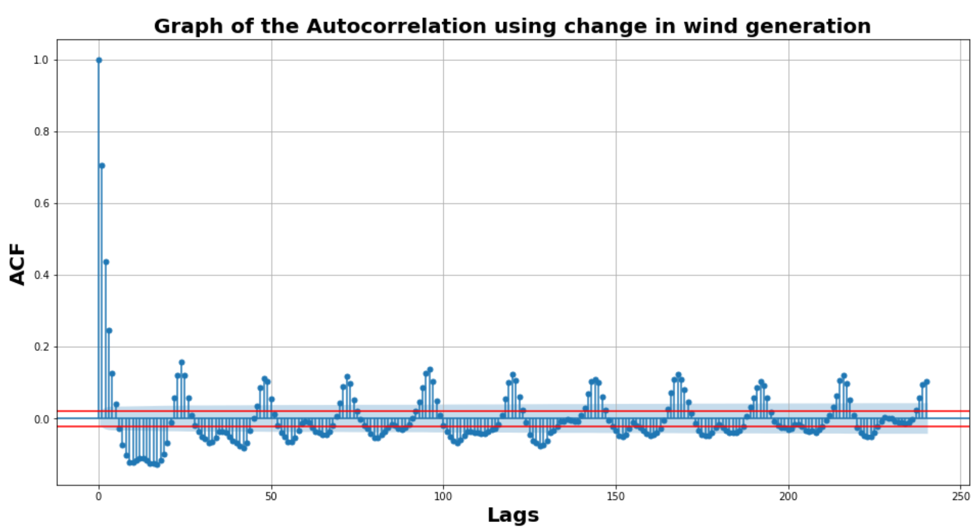
QUESTION 6.

The autocorrelation of change in wind generation for lags over 10 days.

```
array([ 1.00000000e+00,  7.04045607e-01,  4.36179997e-01,  2.46152725e-01,
        1.26411811e-01,  3.98890544e-02, -2.90144122e-02, -7.35407728e-02,
       -1.01639421e-01, -1.20740732e-01, -1.23252559e-01, -1.15504223e-01,
       -1.11521616e-01, -1.11763654e-01, -1.17585571e-01, -1.25511950e-01,
       -1.26091744e-01, -1.27831557e-01, -1.17226686e-01, -9.95542263e-02,
       -6.77612993e-02, -1.02642999e-02,  5.69326741e-02,  1.21242920e-01,
        1.57294589e-01,  1.20040012e-01,  5.79841197e-02,  8.25254791e-03,
       -1.88755709e-02, -3.74276701e-02, -4.97104756e-02, -5.68415922e-02,
       -6.69179078e-02, -6.48468912e-02, -5.35719969e-02, -3.64015897e-02,
       -3.59899603e-02, -4.06410656e-02, -5.03478701e-02, -6.28681474e-02,
       -6.77481176e-02, -7.66737079e-02, -8.17890372e-02, -6.73262692e-02,
       -3.34811447e-02, -2.03079578e-04,  3.38067342e-02,  8.52294305e-02,
        1.13025690e-01,  1.01998812e-01,  5.50971489e-02,  1.29435361e-02,
       -2.00795930e-02, -3.97410560e-02, -5.14921078e-02, -6.43706725e-02,
       -6.40290702e-02, -5.30413121e-02, -3.36019208e-02, -1.31883761e-02,
       -7.35026297e-03, -1.02061214e-02, -2.43111866e-02, -3.52120163e-02,
       -3.96290678e-02, -4.41788124e-02, -4.48024980e-02, -3.60839504e-02,
       -1.88612627e-02,  7.59003468e-03,  4.40248355e-02,  8.89703922e-02,
        1.16779363e-01,  9.76152682e-02,  5.27776402e-02,  2.05650218e-02,
       -2.19007417e-03, -1.93124042e-02, -2.78968795e-02, -4.19029033e-02,
       -5.24545851e-02, -5.27493747e-02, -4.48055201e-02, -3.60846484e-02,
       -2.38753608e-02, -1.75303830e-02, -2.00032630e-02, -2.82569401e-02,
       -3.09537281e-02, -2.58472050e-02, -1.69679777e-02,  4.84285738e-05,
        1.97071353e-02,  4.69782803e-02,  8.48476881e-02,  1.25363418e-01,
        1.38044945e-01,  1.03125812e-01,  4.96665006e-02,  7.95279520e-03,
       -1.81864784e-02, -3.73548041e-02, -5.01062406e-02, -6.31210242e-02,
       -6.67907215e-02, -5.90554818e-02, -4.70668224e-02, -3.62213280e-02,
```

```
-3.68501732e-02, -4.01203863e-02, -3.97579303e-02, -4.13733590e-02,
-4.07324606e-02, -3.55864003e-02, -2.93280189e-02, -2.92595406e-02,
-1.57779982e-02,  8.82925708e-03,  5.42527766e-02,  1.01294593e-01,
 1.24328993e-01,  1.05487063e-01,  5.95706695e-02,  2.29633772e-02,
-1.11930932e-02, -4.48305544e-02, -6.13913240e-02, -6.67358115e-02,
-7.76737264e-02, -7.30408334e-02, -6.14622981e-02, -4.06868359e-02,
-3.34732281e-02, -2.28685216e-02, -9.05621154e-03, -7.69002493e-03,
-3.27311053e-03, -5.32982703e-03, -9.29053883e-03, -8.27777736e-03,
 9.53576878e-03,  2.91669769e-02,  6.92021448e-02,  1.03221383e-01,
 1.09256144e-01,  9.95920536e-02,  6.06342311e-02,  2.32774485e-02,
-3.87498342e-03, -2.26889174e-02, -3.42522884e-02, -4.79041654e-02,
-5.17323758e-02, -4.43100800e-02, -2.88101647e-02, -1.21677137e-02,
-1.79065117e-02, -2.46124388e-02, -3.47922803e-02, -4.18337896e-02,
-4.72495910e-02, -4.57640902e-02, -4.05581361e-02, -2.84353486e-02,
-4.17423230e-03,  2.60795294e-02,  7.05402082e-02,  1.08419405e-01,
 1.22557604e-01,  1.08316251e-01,  8.02357213e-02,  4.69480910e-02,
 1.60642106e-02, -1.23262890e-02, -3.35069150e-02, -4.36420565e-02,
-4.86722334e-02, -4.68656863e-02, -3.85989019e-02, -2.53378901e-02,
-1.73913379e-02, -2.37303483e-02, -3.45088505e-02, -3.85789216e-02,
-3.88427281e-02, -4.06182851e-02, -3.44402140e-02, -2.13827483e-02,
 6.91408686e-03,  3.17414743e-02,  5.62706956e-02,  8.66418315e-02,
 1.04033346e-01,  9.23143406e-02,  5.65316633e-02,  1.66056475e-02,
-8.90147209e-03, -1.80341686e-02, -2.45888479e-02, -2.61482107e-02,
-3.08195927e-02, -2.65989199e-02, -1.62059747e-02, -1.53257318e-02,
-2.12981973e-02, -3.43007375e-02, -3.52953292e-02, -3.24163007e-02,
-3.88803672e-02, -3.09706078e-02, -2.19190706e-02, -4.42894834e-03,
 8.51880812e-03,  3.33393269e-02,  6.44771267e-02,  1.05580476e-01,
 1.20832086e-01,  9.76081354e-02,  5.18543235e-02,  7.80386402e-03,
-2.39660447e-02, -3.80610943e-02, -4.65317999e-02, -5.12629294e-02,
-4.99331741e-02, -3.80202673e-02, -2.09705169e-02, -7.27043953e-03,
 3.09796349e-03,  9.86919113e-04, -1.03015810e-04, -6.76968543e-03,
-1.14134326e-02, -1.12466655e-02, -1.26196204e-02, -1.20196061e-02,
-2.78112302e-03,  2.29092470e-02,  5.82668844e-02,  9.38363404e-02,
 1.04489331e-01])
```

Plotting of horizontal lines to detect statistically significance values ($p<0.05$).



There is a diurnal seasonality.

Yes, I might be an appropriate model the change in wind generation than the actual wind generation.

QUESTION 7.

The variance ratio test:

Variance-Ratio Test Results

Test Statistic	-2.558
P-value	0.011
Lags	240

Computed with overlapping blocks (de-biased)

The pvalue of the variance ratio test:

```
0.010536253517002159
```

Yes, Is there evidence of either mean-reversion because the test statistic values is less than critical values.

QUESTION 8.

The mean absolute error (MAE) between the SMA and the actual wind power.

```
[0.0,
 33.382471461187215,
 62.86293378995434,
 89.39071347031964,
 113.50287671232876,
 135.5138926940639,
 155.78285225048924,
 174.58115724885846,
 191.8615474378488,
 207.69734132420092,
 222.12505188875053,
 235.40405917047184,
 247.84681067790657,
 259.39654843444225,
 270.06914992389653,
 279.9369941495434,
 289.12357238785927,
 297.7328570522577,
 305.64494592645997,
 312.8542305936074,
 319.4559795607741,
 325.48077314238276,
 331.0720274965257,
 336.4262685502283]
```

The minimum MAE is on number 1.

QUESTION 9.

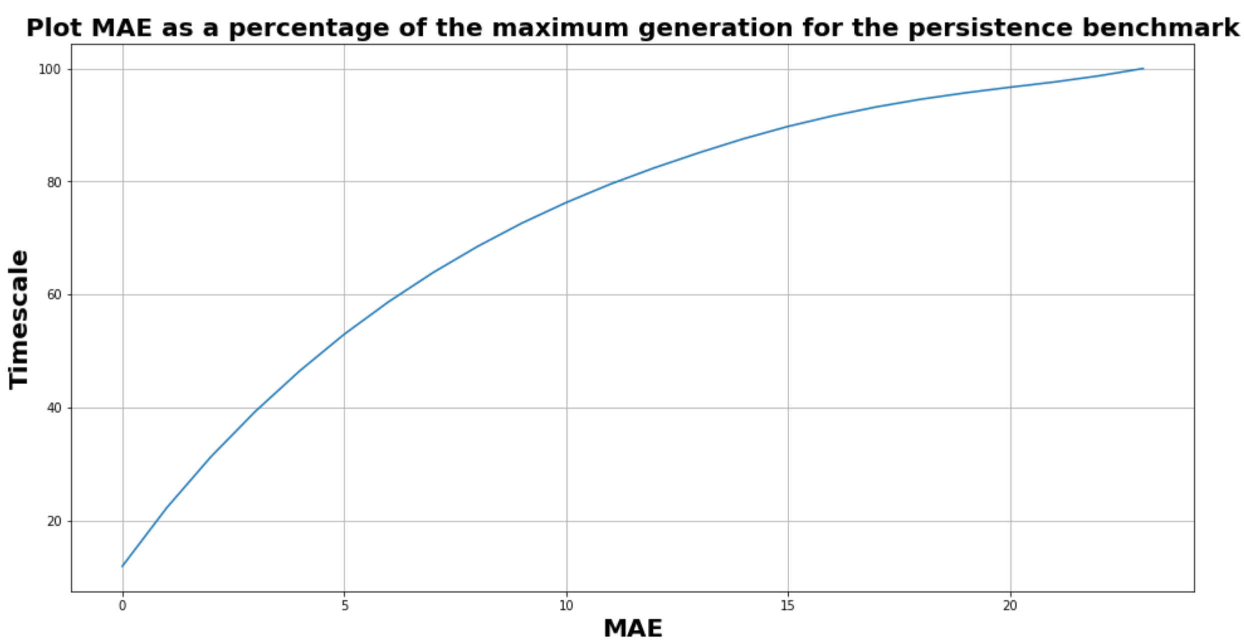
The mean absolute error (MAE) between the predicted wind power and the actual wind power.

```
[66.62908675799086,
 124.43453196347032,
 175.14756849315071,
 219.9664497716895,
 260.05316210045663,
 296.1855365296804,
 328.33710045662104,
 357.1499429223744,
 382.9671461187214,
 406.09070776255714,
 426.5038356164383,
 444.8217351598174,
 461.0433447488585,
 475.89309360730596,
 489.6733219178082,
 501.832397260274,
 512.2439155251142,
 521.2381506849315,
 528.7853881278539,
 535.1028310502282,
 540.6013470319634,
 545.7994520547945,
 551.8389269406392,
 559.1393835616439]
```

MAE as a percentage of the maximum generation for the persistence benchmark.

```
[11.916364455240553,
 22.254653423058635,
 31.32449146713363,
 39.340181757638376,
 46.50954122457846,
 52.9716820594925З,
 58.72186973579954,
 63.874939491361985,
 68.4922503006802,
 72.62781333266369,
 76.27862535807536,
 79.5547135897246,
 82.45588815655829,
 85.11171053198399,
 87.57625313363798,
 89.7508585540278,
 91.6129198880945,
 93.22150540795633,
 94.57130076575201,
 95.70115194563725,
 96.68454108676882,
 97.61420284475835,
 98.69434047473071,
 100.0]
```

Plot MAE as a percentage of the maximum generation for the persistence benchmark.



QUESTION 10.

AutoRegressive Integrated Moving Average (ARIMA) model.

```
p:1 q:1
- AIC:98979.61517258997 - BIC:99007.92652067862
p:1 q:2
- AIC:98966.01291330518 - BIC:99001.402098416
p:1 q:3
- AIC:98967.70461801608 - BIC:99010.17164014906
p:1 q:4
- AIC:98967.26843680524 - BIC:99016.81329596038
p:2 q:1
- AIC:98855.04467372017 - BIC:98890.43385883099
p:2 q:2
- AIC:98785.83633762335 - BIC:98828.30335975632
p:2 q:3
- AIC:98784.35927360562 - BIC:98833.90413276076
p:2 q:4
- AIC:98784.87158467267 - BIC:98841.49428084998
p:3 q:1
- AIC:98783.3892256722 - BIC:98825.85624780518
p:3 q:2
- AIC:98785.23069604157 - BIC:98834.77555519671
p:3 q:3
- AIC:98785.29545288862 - BIC:98841.91814906592
p:3 q:4
- AIC:98788.05260494823 - BIC:98851.7531381477
p:4 q:1
- AIC:98785.18769071798 - BIC:98834.73254987312
p:4 q:2
- AIC:98786.31940830333 - BIC:98842.94210448064
p:4 q:3
- AIC:98785.65715552024 - BIC:98849.35768871971
p:4 q:4
- AIC:98776.47210715484 - BIC:98847.25047737647
```

Optimal ARIMA model parameters are 3, 1, 1.