Configuration Manual

MSc Research Project
Data Analytics

Edmond Connolly
Student ID: 16126556

School of Computing
National College of Ireland

Supervisor: Majid Latifi

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Edmond Connolly
Student ID:	16126556
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Majid Latifi
Submission Due Date:	16/08/2021
Project Title:	Configuration Manual
Word Count:	1284
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Edmond Connolly
Date:	23rd September 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Edmond Connolly
16126556

1 Introduction

The following document will provide an illustration of the steps taken in order to model a time series forecast of Irelands Electricity consumption.

2 System Information

2.1 System Hardware

The implementation of this project was conducted in windows 10 home. The processor was an intel i7 and the environment operated with 16gb of ram. For a full system specification please refer to ***figure 1***.

Item	Value
OS Name	Microsoft Windows 10 Home
Version	10.0.19041 Build 19041
Other OS Description	Not Available
OS Manufacturer	Microsoft Corporation
System Name	DESKTOP-OC5C7O2
System Manufacturer	System manufacturer
System Model	System Product Name
System Type	x64-based PC
System SKU	ASUS_MB_CNL
Processor	Intel(R) Core(TM) i7-9700K CPU @ 3.60GHz, 3600 Mhz, 8 Core(s), 8 Logical Pr...
BIOS Version/Date	American Megatrends Inc. 1302, 02/09/2019
SMBIOS Version	3.2
Embedded Controller Version	255.255
BIOS Mode	UEFI
BaseBoard Manufacturer	ASUSTeK COMPUTER INC.
BaseBoard Product	PRIME Z390-A
BaseBoard Version	Rev 1.xx
Platform Role	Desktop
Secure Boot State	Off
PCR7 Configuration	Binding Not Possible
Windows Directory	C:\Windows
System Directory	C:\Windows\system32
Boot Device	\Device\HarddiskVolume1
Locale	United Kingdom
Hardware Abstraction Layer	Version = "10.0.19041.1151"
Username	DESKTOP-OC5C7O2\edoco
Time Zone	GMT Summer Time
Installed Physical Memory (RAM)	16.0 GB
Total Physical Memory	15.9 GB
Available Physical Memory	6.84 GB
Total Virtual Memory	32.9 GB
Available Virtual Memory	16.9 GB
Page File Space	17.0 GB

Figure 1: System Information

2.2 System Software

An Annaconda environment was created to install all the necessary Python 3 Packages for conducting the research.The preprocessing of data took place in exclusively in Jupyter notebooks and Microsoft Excel.Jupyter Notebooks was used for preliminary data exploration and to impute missing values and encode the days of the week. Output csvs' were aggregated and further processed in Excel. The Spyder programming editor was used for the implementation of the Sarimax and LSTM models. The Sarimax model was developed with pythons statsmodels package. The output was visualised with matplotlib. The LSTM was developed with the Keras library.

The figures below refer to the packages employed at each stage of the project. ***figure 2*** refers to the packages used in the preprocessing stage. This stage often included outputting data to csvs' for further preprocessing in Excel. ***Figure 24*** is a list of the packages employed in the implementation and Evaluation of the Sarimax model. ***Figure 4*** are the packages from which the LSTM was developed.

```
import csv
import os
import pandas as pd
import numpy as np
import sklearn
from pandas import read_csv
from numpy import isnan
from sklearn.impute import KNNImputer
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)
```

Figure 2: Preprocessing Packages

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.graphics.tsaplots as sgt
import statsmodels.tsa.stattools as sts
from statsmodels.tsa.seasonal import STL
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.metrics import r2_score, mean_absolute_error
import seaborn as sns
sns.set()
```

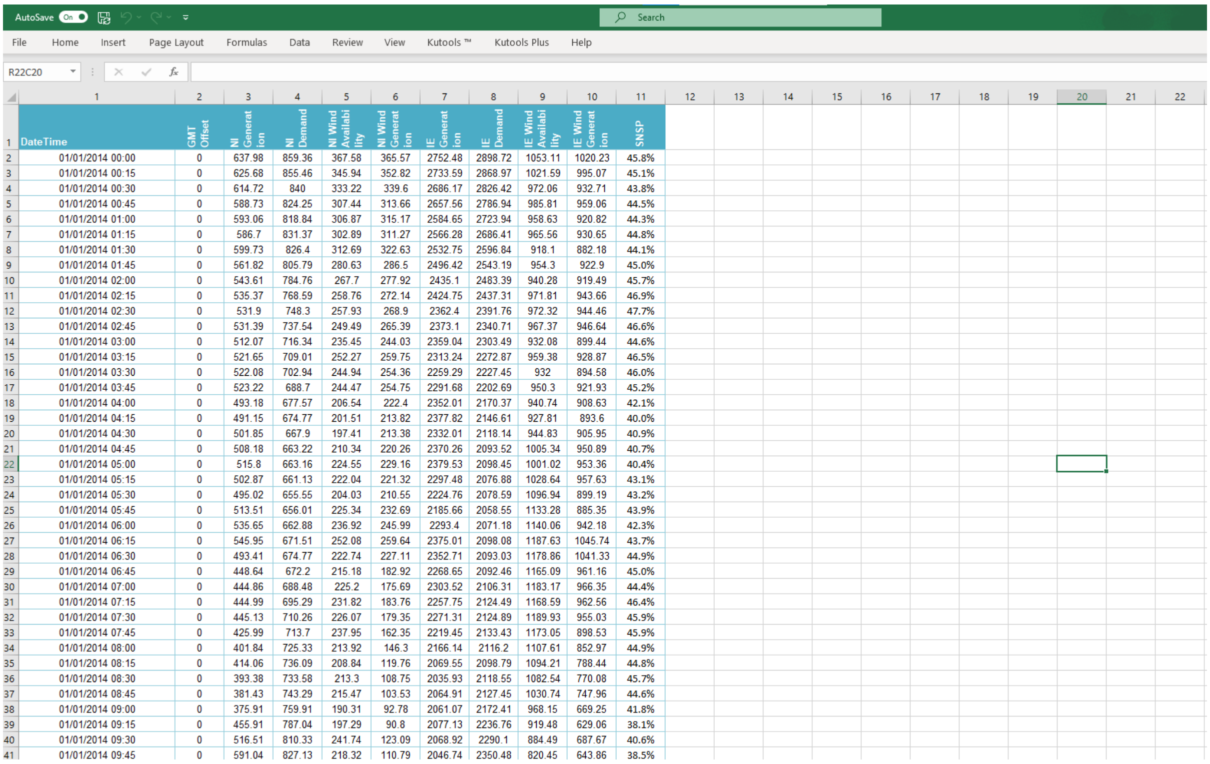
Figure 3: Sarimax Packages

```
import numpy as np
import tensorflow as tf
from tensorflow import keras
import pandas as pd
import seaborn as sns
from pylab import rcParams
import matplotlib.pyplot as plt
from matplotlib import rc
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
from pandas.plotting import register_matplotlib_converters
```

Figure 4: LSTM Packages

3 Data Preparation

Data from three sources was used to model eletricity consumption in the Republic of Ireland. The first of these sources was Eirgrid. It is Eirgrid that are responsible for operating and maintaining the Electrical grid in Ireland. The raw dataset from Eirgrid contains date from 2014 through 2020 (***figure 5***) and is provided in 15 minute intervals. The only feature required form this dataset was the electricity consumption for the Republic of Ireland.

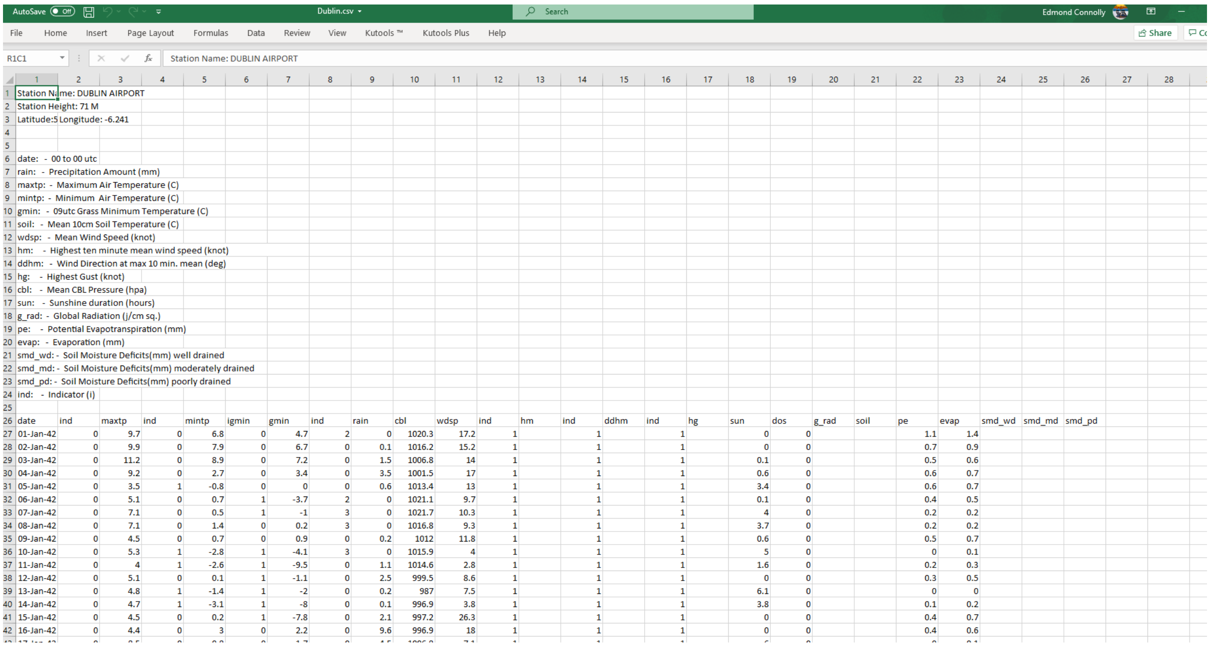


DateTime	GMT Offset	NI Generation	NI Demand	NI Wind Availability	NI Wind Generation	IE Generation	IE Demand	IE Wind Availability	IE Wind Generation	SNSP
01/01/2014 00:00	0	637.98	859.36	367.58	365.57	2752.48	2898.72	1053.11	1020.23	45.8%
01/01/2014 00:15	0	625.68	855.46	345.94	352.82	2733.59	2868.97	1021.59	995.07	45.1%
01/01/2014 00:30	0	614.72	840	333.22	339.6	2686.17	2826.42	972.06	932.71	43.8%
01/01/2014 00:45	0	588.73	824.25	307.44	313.66	2657.56	2786.94	985.81	959.06	44.5%
01/01/2014 01:00	0	593.06	818.84	306.87	315.17	2584.65	2723.94	958.63	920.82	44.3%
01/01/2014 01:15	0	586.7	831.37	302.89	311.27	2566.28	2686.41	965.56	930.65	44.8%
01/01/2014 01:30	0	599.73	826.4	312.69	322.63	2532.75	2596.84	918.1	882.18	44.1%
01/01/2014 01:45	0	561.82	805.79	280.63	286.5	2496.42	2543.19	954.3	922.9	45.0%
01/01/2014 02:00	0	543.61	784.76	267.7	277.92	2435.1	2483.39	940.28	919.49	45.7%
01/01/2014 02:15	0	535.37	768.59	258.76	272.14	2424.75	2437.31	971.81	943.66	46.9%
01/01/2014 02:30	0	531.9	748.3	257.93	268.9	2362.4	2391.76	972.32	944.46	47.7%
01/01/2014 02:45	0	531.39	737.54	249.49	265.39	2373.1	2340.71	967.37	946.64	46.6%
01/01/2014 03:00	0	512.07	716.34	235.45	244.03	2359.04	2303.49	932.08	899.44	44.6%
01/01/2014 03:15	0	521.65	709.01	252.27	259.75	2313.24	2272.87	959.38	928.87	46.5%
01/01/2014 03:30	0	522.08	702.94	244.94	254.36	2259.29	2227.45	932	894.58	46.0%
01/01/2014 03:45	0	523.22	688.7	244.47	254.75	2291.68	2202.69	950.3	921.93	45.2%
01/01/2014 04:00	0	493.18	677.57	206.54	222.4	2352.01	2170.37	940.74	908.63	42.1%
01/01/2014 04:15	0	491.15	674.77	201.51	213.82	2377.82	2146.61	927.81	893.6	40.0%
01/01/2014 04:30	0	501.85	667.9	197.41	213.38	2332.01	2118.14	944.83	905.95	40.9%
01/01/2014 04:45	0	508.18	663.22	210.34	220.26	2370.26	2093.52	1005.34	950.89	40.7%
01/01/2014 05:00	0	515.8	663.16	224.55	229.16	2379.53	2098.45	1001.02	953.36	40.4%
01/01/2014 05:15	0	502.87	661.13	222.04	221.32	2297.48	2076.88	1028.64	957.63	43.1%
01/01/2014 05:30	0	495.02	655.55	204.03	210.55	2224.76	2078.59	1096.94	899.19	43.2%
01/01/2014 05:45	0	513.51	656.01	225.34	232.69	2185.66	2058.55	1133.28	885.35	43.9%
01/01/2014 06:00	0	535.65	662.88	236.92	245.99	2293.4	2071.18	1140.06	942.18	42.3%
01/01/2014 06:15	0	545.95	671.51	252.08	259.64	2375.01	2098.08	1187.63	1045.74	43.7%
01/01/2014 06:30	0	493.41	674.77	222.74	227.11	2352.71	2093.03	1178.86	1041.33	44.9%
01/01/2014 06:45	0	448.64	672.2	215.18	182.92	2268.65	2092.46	1165.09	961.16	45.0%
01/01/2014 07:00	0	444.86	688.48	225.2	175.69	2303.52	2106.31	1183.17	966.35	44.4%
01/01/2014 07:15	0	444.99	695.29	231.82	183.76	2257.75	2124.49	1168.59	962.56	46.4%
01/01/2014 07:30	0	445.13	710.26	226.07	179.35	2271.31	2124.89	1189.93	955.03	45.9%
01/01/2014 07:45	0	425.99	713.7	237.95	162.35	2219.45	2133.43	1173.05	898.53	45.9%
01/01/2014 08:00	0	401.84	725.33	213.92	146.3	2166.14	2116.2	1107.61	852.97	44.9%
01/01/2014 08:15	0	414.06	736.09	208.84	119.76	2069.55	2098.79	1094.21	788.44	44.8%
01/01/2014 08:30	0	393.38	733.58	213.3	108.75	2035.93	2118.55	1082.54	770.08	45.7%
01/01/2014 08:45	0	381.43	743.29	215.47	103.53	2064.91	2127.45	1030.74	747.96	44.6%
01/01/2014 09:00	0	375.91	759.91	190.31	92.78	2061.07	2172.41	968.15	669.25	41.8%
01/01/2014 09:15	0	455.91	787.04	197.29	90.8	2077.13	2236.76	919.48	629.06	38.1%
01/01/2014 09:30	0	516.51	810.33	241.74	123.09	2068.92	2290.1	884.49	687.67	40.6%
01/01/2014 09:45	0	591.04	827.13	218.32	110.79	2046.74	2350.48	820.45	643.86	38.5%

Figure 5: Electricity Consumption 2014 -2020

Meteorological data was collected from a number of weather stations throughout Ireland to provide features for the LSTM and Sarimax independent variables. An example of this data is illustrated in ***figure 6*** and ***figure 7***.

The final source of data was from the National Oceanic and Atmospheric Administration. This is a sub-department of the U.S. Department of Commerce. This dataset contains daily information of sun-rise and sunset times and the length of daylight for each day. It also contains data relating to the aspect of the sun to geographic co-ordinates (***figure 8***). It contains climate data indicative of seasonal changes.



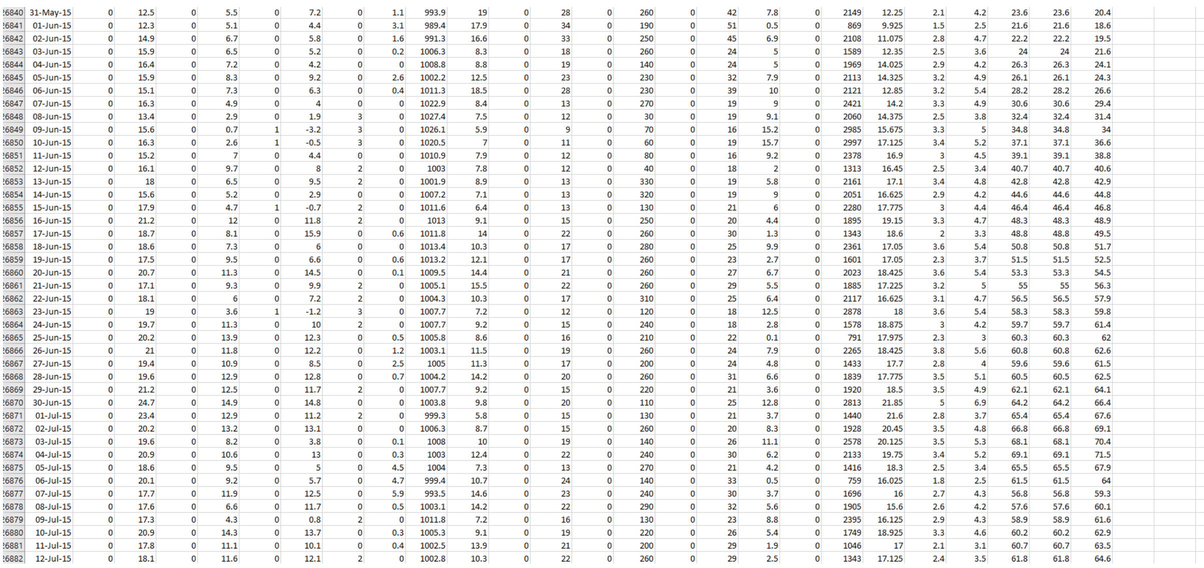
AutoSave Off Dublin.csv Search Edmond Connolly

File Home Insert Page Layout Formulas Data Review View Kutools ™ Kutools Plus Help Share

R1C1 Station Name: DUBLIN AIRPORT

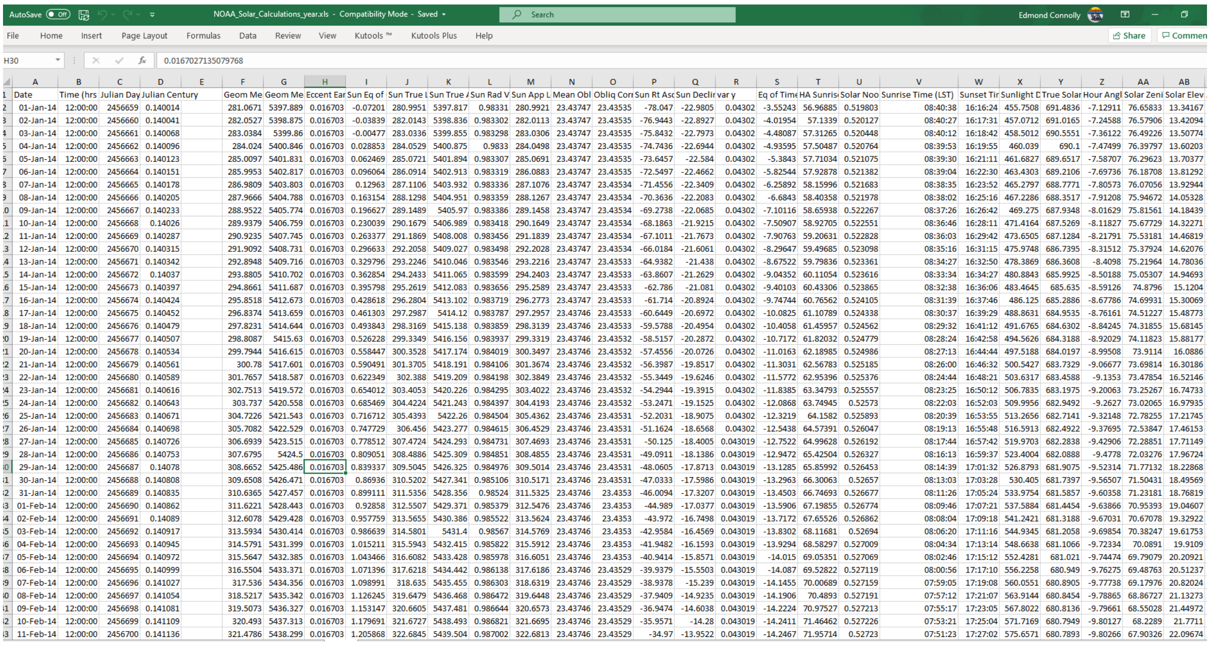
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	Station Name: DUBLIN AIRPORT																											
2	Station Height: 71 M																											
3	Latitude:5 Longitude: -6.241																											
4																												
5																												
6	date: - 00 to 00 utc																											
7	rain: - Precipitation Amount (mm)																											
8	maxtp: - Maximum Air Temperature (C)																											
9	mintp: - Minimum Air Temperature (C)																											
10	gmin: - 09utc Grass Minimum Temperature (C)																											
11	soil: - Mean 10cm Soil Temperature (C)																											
12	wdsp: - Mean Wind Speed (knot)																											
13	hm: - Highest ten minute mean wind speed (knot)																											
14	ddhm: - Wind Direction at max 10 min. mean (deg)																											
15	hg: - Highest Gust (knot)																											
16	cbl: - Mean CBL Pressure (hpa)																											
17	sun: - Sunshine duration (hours)																											
18	g_rad: - Global Radiation (j/cm sq.)																											
19	pe: - Potential Evapotranspiration (mm)																											
20	evap: - Evaporation (mm)																											
21	smd_wd: - Soil Moisture Deficits(mm) well drained																											
22	smd_md: - Soil Moisture Deficits(mm) moderately drained																											
23	smd_pd: - Soil Moisture Deficits(mm) poorly drained																											
24	ind: - Indicator (i)																											
25																												
26	date	ind	maxtp	ind	mintp	igmin	gmin	ind	rain	cbl	wdsp	ind	hm	ind	ddhm	ind	hg	sun	dos	g_rad	soil	pe	evap	smd_wd	smd_md	smd_pd		
27	01-Jan-42	0	9.7	0	6.8	0	4.7	2	0	1020.3	17.2	1		1		1		0	0			1.1	1.4					
28	02-Jan-42	0	9.9	0	7.9	0	6.7	0	0.1	1016.2	15.2	1		1		1		0	0			0.7	0.9					
29	03-Jan-42	0	11.2	0	8.9	0	7.2	0	1.5	1006.8	14	1		1		1		0.1	0			0.5	0.6					
30	04-Jan-42	0	9.2	0	2.7	0	3.4	0	3.5	1001.5	17	1		1		1		0.6	0			0.6	0.7					
31	05-Jan-42	0	3.5	1	-0.8	0	0	0	0.6	1013.4	13	1		1		1		3.4	0			0.6	0.7					
32	06-Jan-42	0	5.1	0	0.7	1	-3.7	2	0	1021.1	9.7	1		1		1		0.1	0			0.4	0.5					
33	07-Jan-42	0	7.1	0	0.5	1	-1	3	0	1021.7	10.3	1		1		1		4	0			0.2	0.2					
34	08-Jan-42	0	7.1	0	1.4	0	0.2	3	0	1016.8	9.3	1		1		1		3.7	0			0.2	0.2					
35	09-Jan-42	0	4.5	0	0.7	0	0.9	0	0.2	1012	11.8	1		1		1		0.6	0			0.5	0.7					
36	10-Jan-42	0	5.3	1	-2.8	1	-4.1	3	0	1015.9	4	1		1		1		5	0			0	0.1					
37	11-Jan-42	0	4	1	-2.6	1	-9.5	0	1.1	1014.6	2.8	1		1		1		1.6	0			0.2	0.3					
38	12-Jan-42	0	5.1	0	0.1	1	-1.1	0	2.5	999.5	8.6	1		1		1		0	0			0.3	0.5					
39	13-Jan-42	0	4.8	1	-1.4	1	-2	0	0.2	987	7.5	1		1		1		6.1	0			0	0					
40	14-Jan-42	0	4.7	1	-3.1	1	-8	0	0.1	996.9	3.8	1		1		1		3.8	0			0.1	0.2					
41	15-Jan-42	0	4.5	0	0.2	1	-7.8	0	2.1	997.2	26.3	1		1		1		0	0			0.4	0.7					
42	16-Jan-42	0	4.4	0	3	0	2.2	0	9.6	996.9	18	1		1		1		0	0			0.4	0.6					

Figure 6: Meteorological Data - Variables



26840	31-May-15	0	12.5	0	5.5	0	7.2	0	1.1	993.9	19	0	28	0	260	0	42	7.8	0	2149	12.25	2.1	4.2	23.6	23.6	20.4
26841	01-Jun-15	0	12.3	0	5.1	0	4.4	0	3.1	989.4	17.9	0	34	0	190	0	51	0.5	0	869	9.925	1.5	2.5	21.6	21.6	18.6
26842	02-Jun-15	0	14.9	0	6.7	0	5.8	0	1.6	991.3	16.6	0	33	0	250	0	45	6.9	0	2108	11.075	2.8	4.7	22.2	22.2	19.5
26843	03-Jun-15	0	15.9	0	6.5	0	5.2	0	0.2	1006.3	8.3	0	18	0	260	0	24	5	0	1589	12.35	2.5	3.6	24	24	21.6
26844	04-Jun-15	0	16.4	0	7.2	0	4.2	0	0	1008.8	8.8	0	19	0	140	0	24	5	0	1969	14.025	2.9	4.2	26.3	26.3	24.1
26845	05-Jun-15	0	15.9	0	8.3	0	9.2	0	2.6	1002.2	12.5	0	23	0	230	0	32	7.9	0	2113	14.325	3.2	4.9	26.1	26.1	24.3
26846	06-Jun-15	0	15.1	0	7.3	0	6.3	0	0.4	1011.3	18.5	0	28	0	230	0	39	10	0	2121	12.85	3.2	5.4	28.2	28.2	26.6
26847	07-Jun-15	0	16.3	0	4.9	0	4	0	0	1022.9	8.4	0	13	0	270	0	19	9	0	2421	14.2	3.3	4.9	30.6	30.6	29.4
26848	08-Jun-15	0	13.4	0	2.9	0	1.9	3	0	1027.4	7.5	0	12	0	30	0	19	9.1	0	2060	14.375	2.5	3.8	32.4	32.4	31.4
26849	09-Jun-15	0	15.6	0	0.7	1	-3.2	3	0	1026.1	5.9	0	9	0	70	0	16	15.2	0	2985	15.675	3.3	5	34.8	34.8	34
26850	10-Jun-15	0	16.3	0	2.6	1	-0.5	3	0	1020.5	7	0	11	0	60	0	19	15.7	0	2997	17.125	3.4	5.2	37.1	37.1	36.6
26851	11-Jun-15	0	15.2	0	7	0	4.4	0	0	1010.9	7.9	0	12	0	80	0	16	9.2	0	2378	16.9	3	4.5	39.1	39.1	38.8
26852	12-Jun-15	0	16.1	0	9.7	0	8	2	0	1003	7.8	0	12	0	40	0	18	2	0	1313	16.45	2.5	3.4	40.7	40.7	40.6
26853	13-Jun-15	0	18	0	6.5	0	9.5	2	0	1001.9	8.9	0	13	0	330	0	19	5.8	0	2161	17.1	3.4	4.8	42.8	42.8	42.9
26854	14-Jun-15	0	15.6	0	5.2	0	2.9	0	0	1007.2	7.1	0	13	0	320	0	19	9	0	2051	16.625	2.9	4.2	44.6	44.6	44.8
26855	15-Jun-15	0	17.9	0	4.7	1	-0.7	2	0	1011.6	6.4	0	13	0	130	0	21	6	0	2280	17.775	3	4.4	46.4	46.4	46.8
26856	16-Jun-15	0	21.2	0	12	0	11.8	2	0	1013	9.1	0	15	0	250	0	20	4.4	0	1895	19.15	3.3	4.7	48.3	48.3	48.9
26857	17-Jun-15	0	18.7	0	8.1	0	15.9	0	0.6	1011.8	14	0	22	0	260	0	30	1.3	0	1343	18.6	2	3.3	48.8	48.8	49.5
26858	18-Jun-15	0	18.6	0	7.3	0	6	0	0	1013.4	10.3	0	17	0	280	0	25	9.9	0	2361	17.05	3.6	5.4	50.8	50.8	51.7
26859	19-Jun-15	0	17.5	0	9.5	0	6.6	0	0.6	1013.2	12.1	0	17	0	260	0	23	2.7	0	1601	17.05	2.3	3.7	51.5	51.5	52.5
26860	20-Jun-15	0	20.7	0	11.3	0	14.5	0	0.1	1009.5	14.4	0	21	0	260	0	27	6.7	0	2023	18.425	3.6	5.4	53.3	53.3	54.5
26861	21-Jun-15	0	17.1	0	9.3	0	9.9	2	0	1005.1	15.5	0	22	0	260	0	29	5.5	0	1885	17.225	3.2	5	55	55	56.3
26862	22-Jun-15	0	18.1	0	6	0	7.2	2	0	1004.3	10.3	0	17	0	310	0	25	6.4	0	2117	16.625	3.1	4.7	56.5	56.5	57.9
26863	23-Jun-15	0	19	0	3.6	1	-1.2	3	0	1007.7	7.2	0	12	0	120	0	18	12.5	0	2878	18	3.6	5.4	58.3	58.3	59.8
26864	24-Jun-15	0	19.7	0	11.3	0	10	2	0	1007.7	9.2	0	15	0	240	0	18	2.8	0	1578	18.875	3	4.2	59.7	59.7	61.4
26865	25-Jun-15	0	20.2	0	13.9	0	12.3	0	0.5	1005.8	8.6	0	16	0	210	0	22	0.1	0	791	17.975	2.3	3	60.3	60.3	62
26866	26-Jun-15	0	21	0	11.8	0	12.2	0	1.2	1003.1	11.5	0	19	0	260	0	24	7.9	0	2265	18.425	3.8	5.6	60.8	60.8	62.6
26867	27-Jun-15	0	19.4	0	10.9	0	8.5	0	2.5	1005	11.3	0	17	0	200	0	24	4.8	0	1433	17.7	2.8	4	59.6	59.6	61.5
26868	28-Jun-15	0	19.6	0	12.9	0	12.8	0	0.7	1004.2	14.2	0	20	0	260	0	31	6.6	0	1839	17.775	3.5	5.1	60.5	60.5	62.5
26869	29-Jun-15	0	21.2	0	12.5	0	11.7	2	0	1007.7	9.2	0	15	0	220	0	21	3.6	0	1920	18.5	3.5	4.9	62.1	62.1	64.1
26870	30-Jun-15	0	24.7	0	14.9	0	14.8	0	0	1003.8	9.8	0	20	0	110	0	25	12.8	0	2813	21.85	5	6.9	64.2	64.2	66.4
26871	01-Jul-15	0	23.4	0	12.9	0	11.2	2	0	999.3	5.8	0	15	0	130	0	21	3.7	0	1440	21.6	2.8	3.7	65.4	65.4	67.6
26872	02-Jul-15	0	20.2	0	13.2	0	13.1	0	0	1006.3	8.7	0	15	0	260	0	20	8.3	0	1928	20.45	3.5	4.8	66.8	66.8	69.1
26873	03-Jul-15	0	19.6	0	8.2	0	3.8	0	0.1	1008	10	0	19	0	140	0	26	11.1	0	2578	20.125	3.5	5.3	68.1	68.1	70.4
26874	04-Jul-15	0	20.9	0	10.6	0	13	0	0.3	1003	12.4	0	22	0	240	0	30	6.2	0	2133	19.75	3.4	5.2	69.1	69.1	71.5
26875	05-Jul-15	0	18.6	0	9.5	0	5	0	4.5	1004	7.3	0	13	0	270	0	21	4.2	0	1416	18.3	2.5	3.4	65.5	65.5	67.9
26876	06-Jul-15	0	20.1	0	9.2	0	5.7	0	4.7	999.4	10.7	0	24	0	140	0	33	0.5	0	759	16.025	1.8	2.5	61.5	61.5	64
26877	07-Jul-15	0	17.7	0	11.9	0	12.5	0	5.9	993.5	14.6	0	23	0	240	0	30	3.7	0	1696	16	2.7	4.3	56.8	56.8	59.3
26878	08-Jul-15	0	17.6	0	6.6	0	11.7	0	0.5	1003.1	14.2	0	22	0	290	0	32	5.6	0	1905	15.6	2.6	4.2	57.6	57.6	60.1
26879	09-Jul-15	0	17.3	0	4.3	0	0.8	2	0	1011.8	7.2	0	16	0	130	0	23	8.8	0	2395	16.125	2.9	4.3	58.9	58.9	61.6
26880	10-Jul-15	0	20.9	0	14.3	0	13.7	0	0.3	1005.3	9.1	0	19	0	220	0	26	5.4	0	1749	18.925	3.3	4.6	60.2	60.2	62.9
26881	11-Jul-15	0	17.8	0	11.1	0	10.1	0	0.4	1002.5	13.9	0	21	0	200	0	29	1.9	0	1046	17	2.1	3.1	60.7	60.7	63.5
26882	12-Jul-15	0	18.1	0	11.6	0	12.1	2	0	1002.8	10.3	0	22	0	260	0	29	2.5	0	1343	17.125	2.4	3.5	61.8	61.8	64.6

Figure 7: Meteorological Data



	A	B	C	D	E	F	G	H	I
1	Date	Time (hrs	Julian Day	Julian Century		Geom Me	Geom Me	Eccent Ear	Sun Eq of
2	01-Jan-14	12:00:00	2456659	0.140014		281.0671	5397.889	0.016703	-0.07201
3	02-Jan-14	12:00:00	2456660	0.140041		282.0527	5398.875	0.016703	-0.03839
4	03-Jan-14	12:00:00	2456661	0.140068		283.0384	5399.86	0.016703	-0.00477
5	04-Jan-14	12:00:00	2456662	0.140096		284.024	5400.846	0.016703	0.028853
6	05-Jan-14	12:00:00	2456663	0.140123		285.0097	5401.831	0.016703	0.062469
7	06-Jan-14	12:00:00	2456664	0.140151		285.9953	5402.817	0.016703	0.096064
8	07-Jan-14	12:00:00	2456665	0.140178		286.9809	5403.803	0.016703	0.12963
9	08-Jan-14	12:00:00	2456666	0.140205		287.9666	5404.788	0.016703	0.163154
10	09-Jan-14	12:00:00	2456667	0.140233		288.9522	5405.774	0.016703	0.196627
11	10-Jan-14	12:00:00	2456668	0.14026		289.9379	5406.759	0.016703	0.230039
12	11-Jan-14	12:00:00	2456669	0.140287		290.9235	5407.745	0.016703	0.263377
13	12-Jan-14	12:00:00	2456670	0.140315		291.9092	5408.731	0.016703	0.296633
14	13-Jan-14	12:00:00	2456671	0.140342		292.8948	5409.716	0.016703	0.329796
15	14-Jan-14	12:00:00	2456672	0.14037		293.8805	5410.702	0.016703	0.362854
16	15-Jan-14	12:00:00	2456673	0.140397		294.8661	5411.687	0.016703	0.395798
17	16-Jan-14	12:00:00	2456674	0.140424		295.8518	5412.673	0.016703	0.428618
18	17-Jan-14	12:00:00	2456675	0.140452		296.8374	5413.659	0.016703	0.461303
19	18-Jan-14	12:00:00	2456676	0.140479		297.8231	5414.644	0.016703	0.493843
20	19-Jan-14	12:00:00	2456677	0.140507		298.8087	5415.63	0.016703	0.526228
21	20-Jan-14	12:00:00	2456678	0.140534		299.7944	5416.615	0.016703	0.558447
22	21-Jan-14	12:00:00	2456679	0.140561		300.78	5417.601	0.016703	0.590491
23	22-Jan-14	12:00:00	2456680	0.140589		301.7657	5418.587	0.016703	0.622349
24	23-Jan-14	12:00:00	2456681	0.140616		302.7513	5419.572	0.016703	0.654012
25	24-Jan-14	12:00:00	2456682	0.140643		303.737	5420.558	0.016703	0.685469
26	25-Jan-14	12:00:00	2456683	0.140671		304.7226	5421.543	0.016703	0.716712
27	26-Jan-14	12:00:00	2456684	0.140698		305.7082	5422.529	0.016703	0.747729
28	27-Jan-14	12:00:00	2456685	0.140726		306.6939	5423.515	0.016703	0.778512
29	28-Jan-14	12:00:00	2456686	0.140753		307.6795	5424.5	0.016703	0.809051
30	29-Jan-14	12:00:00	2456687	0.14078		308.6652	5425.486	0.016703	0.839337
31	30-Jan-14	12:00:00	2456688	0.140808		309.6508	5426.471	0.016703	0.86936
32	31-Jan-14	12:00:00	2456689	0.140835		310.6365	5427.457	0.016703	0.899111
33	01-Feb-14	12:00:00	2456690	0.140862		311.6221	5428.443	0.016703	0.92858
34	02-Feb-14	12:00:00	2456691	0.14089		312.6078	5429.428	0.016703	0.957759
35	03-Feb-14	12:00:00	2456692	0.140917		313.5934	5430.414	0.016703	0.986639
36	04-Feb-14	12:00:00	2456693	0.140945		314.5791	5431.399	0.016703	1.015211
37	05-Feb-14	12:00:00	2456694	0.140972		315.5647	5432.385	0.016703	1.043466
38	06-Feb-14	12:00:00	2456695	0.140999		316.5504	5433.371	0.016703	1.071396
39	07-Feb-14	12:00:00	2456696	0.141027		317.536	5434.356	0.016703	1.098991
40	08-Feb-14	12:00:00	2456697	0.141054		318.5217	5435.342	0.016703	1.126245
41	09-Feb-14	12:00:00	2456698	0.141081		319.5073	5436.327	0.016703	1.153147
42	10-Feb-14	12:00:00	2456699	0.141109		320.493	5437.313	0.016703	1.179691
43	11-Feb-14	12:00:00	2456700	0.141136		321.4786	5438.299	0.016703	1.205868

	A	J	K	L	M	N	O	P	Q	R
1	Date	Sun True L	Sun True A	Sun Rad V	Sun App L	Mean Obl	Obliq Corr	Sun Rt Asc	Sun Declir	var y
2	01-Jan-14	280.9951	5397.817	0.98331	280.9921	23.43747	23.43535	-78.047	-22.9805	0.04302
3	02-Jan-14	282.0143	5398.836	0.983302	282.0113	23.43747	23.43535	-76.9443	-22.8927	0.04302
4	03-Jan-14	283.0336	5399.855	0.983298	283.0306	23.43747	23.43535	-75.8432	-22.7973	0.04302
5	04-Jan-14	284.0529	5400.875	0.9833	284.0498	23.43747	23.43535	-74.7436	-22.6944	0.04302
6	05-Jan-14	285.0721	5401.894	0.983307	285.0691	23.43747	23.43535	-73.6457	-22.584	0.04302
7	06-Jan-14	286.0914	5402.913	0.983319	286.0883	23.43747	23.43535	-72.5497	-22.4662	0.04302
8	07-Jan-14	287.1106	5403.932	0.983336	287.1076	23.43747	23.43534	-71.4556	-22.3409	0.04302
9	08-Jan-14	288.1298	5404.951	0.983359	288.1267	23.43747	23.43534	-70.3636	-22.2083	0.04302
10	09-Jan-14	289.1489	5405.97	0.983386	289.1458	23.43747	23.43534	-69.2738	-22.0685	0.04302
11	10-Jan-14	290.1679	5406.989	0.983418	290.1649	23.43747	23.43534	-68.1863	-21.9215	0.04302
12	11-Jan-14	291.1869	5408.008	0.983456	291.1839	23.43747	23.43534	-67.1011	-21.7673	0.04302
13	12-Jan-14	292.2058	5409.027	0.983498	292.2028	23.43747	23.43534	-66.0184	-21.6061	0.04302
14	13-Jan-14	293.2246	5410.046	0.983546	293.2216	23.43747	23.43533	-64.9382	-21.438	0.04302
15	14-Jan-14	294.2433	5411.065	0.983599	294.2403	23.43747	23.43533	-63.8607	-21.2629	0.04302
16	15-Jan-14	295.2619	5412.083	0.983656	295.2589	23.43747	23.43533	-62.786	-21.081	0.04302
17	16-Jan-14	296.2804	5413.102	0.983719	296.2773	23.43747	23.43533	-61.714	-20.8924	0.04302
18	17-Jan-14	297.2987	5414.12	0.983787	297.2957	23.43746	23.43533	-60.6449	-20.6972	0.04302
19	18-Jan-14	298.3169	5415.138	0.983859	298.3139	23.43746	23.43533	-59.5788	-20.4954	0.04302
20	19-Jan-14	299.3349	5416.156	0.983937	299.3319	23.43746	23.43532	-58.5157	-20.2872	0.04302
21	20-Jan-14	300.3528	5417.174	0.984019	300.3497	23.43746	23.43532	-57.4556	-20.0726	0.04302
22	21-Jan-14	301.3705	5418.191	0.984106	301.3674	23.43746	23.43532	-56.3987	-19.8517	0.04302
23	22-Jan-14	302.388	5419.209	0.984198	302.3849	23.43746	23.43532	-55.3449	-19.6246	0.04302
24	23-Jan-14	303.4053	5420.226	0.984295	303.4022	23.43746	23.43532	-54.2944	-19.3915	0.04302
25	24-Jan-14	304.4224	5421.243	0.984397	304.4193	23.43746	23.43532	-53.2471	-19.1525	0.04302
26	25-Jan-14	305.4393	5422.26	0.984504	305.4362	23.43746	23.43531	-52.2031	-18.9075	0.04302
27	26-Jan-14	306.456	5423.277	0.984615	306.4529	23.43746	23.43531	-51.1624	-18.6568	0.04302
28	27-Jan-14	307.4724	5424.293	0.984731	307.4693	23.43746	23.43531	-50.125	-18.4005	0.043019
29	28-Jan-14	308.4886	5425.309	0.984851	308.4855	23.43746	23.43531	-49.0911	-18.1386	0.043019
30	29-Jan-14	309.5045	5426.325	0.984976	309.5014	23.43746	23.43531	-48.0605	-17.8713	0.043019
31	30-Jan-14	310.5202	5427.341	0.985106	310.5171	23.43746	23.43531	-47.0333	-17.5986	0.043019
32	31-Jan-14	311.5356	5428.356	0.98524	311.5325	23.43746	23.4353	-46.0094	-17.3207	0.043019
33	01-Feb-14	312.5507	5429.371	0.985379	312.5476	23.43746	23.4353	-44.989	-17.0377	0.043019
34	02-Feb-14	313.5655	5430.386	0.985522	313.5624	23.43746	23.4353	-43.972	-16.7498	0.043019
35	03-Feb-14	314.5801	5431.4	0.98567	314.5769	23.43746	23.4353	-42.9584	-16.4569	0.043019
36	04-Feb-14	315.5943	5432.415	0.985822	315.5912	23.43746	23.4353	-41.9482	-16.1593	0.043019
37	05-Feb-14	316.6082	5433.428	0.985978	316.6051	23.43746	23.4353	-40.9414	-15.8571	0.043019
38	06-Feb-14	317.6218	5434.442	0.986138	317.6186	23.43746	23.43529	-39.9379	-15.5503	0.043019
39	07-Feb-14	318.635	5435.455	0.986303	318.6319	23.43746	23.43529	-38.9378	-15.239	0.043019
40	08-Feb-14	319.6479	5436.468	0.986472	319.6448	23.43746	23.43529	-37.9409	-14.9235	0.043019
41	09-Feb-14	320.6605	5437.481	0.986644	320.6573	23.43746	23.43529	-36.9474	-14.6038	0.043019
42	10-Feb-14	321.6727	5438.493	0.986821	321.6695	23.43746	23.43529	-35.9571	-14.28	0.043019
43	11-Feb-14	322.6845	5439.504	0.987002	322.6813	23.43746	23.43529	-34.97	-13.9522	0.043019

	A	S	T	U	V	W	X	Y	Z	AA	AB
1	Date	Eq of Time	HA Sunris	Solar Noo	Sunrise Time (LST)	Sunset Tir	Sunlight D	True Solar	Hour Angl	Solar Zeni	Solar Elev
2	01-Jan-14	-3.55243	56.96885	0.519803	08:40:38	16:16:24	455.7508	691.4836	-7.12911	76.65833	13.34167
3	02-Jan-14	-4.01954	57.1339	0.520127	08:40:27	16:17:31	457.0712	691.0165	-7.24588	76.57906	13.42094
4	03-Jan-14	-4.48087	57.31265	0.520448	08:40:12	16:18:42	458.5012	690.5551	-7.36122	76.49226	13.50774
5	04-Jan-14	-4.93595	57.50487	0.520764	08:39:53	16:19:55	460.039	690.1	-7.47499	76.39797	13.60203
6	05-Jan-14	-5.3843	57.71034	0.521075	08:39:30	16:21:11	461.6827	689.6517	-7.58707	76.29623	13.70377
7	06-Jan-14	-5.82544	57.92878	0.521382	08:39:04	16:22:30	463.4303	689.2106	-7.69736	76.18708	13.81292
8	07-Jan-14	-6.25892	58.15996	0.521683	08:38:35	16:23:52	465.2797	688.7771	-7.80573	76.07056	13.92944
9	08-Jan-14	-6.6843	58.40358	0.521978	08:38:02	16:25:16	467.2286	688.3517	-7.91208	75.94672	14.05328
10	09-Jan-14	-7.10116	58.65938	0.522267	08:37:26	16:26:42	469.275	687.9348	-8.01629	75.81561	14.18439
11	10-Jan-14	-7.50907	58.92705	0.522551	08:36:46	16:28:11	471.4164	687.5269	-8.11827	75.67729	14.32271
12	11-Jan-14	-7.90763	59.20631	0.522828	08:36:03	16:29:42	473.6505	687.1284	-8.21791	75.53181	14.46819
13	12-Jan-14	-8.29647	59.49685	0.523098	08:35:16	16:31:15	475.9748	686.7395	-8.31512	75.37924	14.62076
14	13-Jan-14	-8.67522	59.79836	0.523361	08:34:27	16:32:50	478.3869	686.3608	-8.4098	75.21964	14.78036
15	14-Jan-14	-9.04352	60.11054	0.523616	08:33:34	16:34:27	480.8843	685.9925	-8.50188	75.05307	14.94693
16	15-Jan-14	-9.40103	60.43306	0.523865	08:32:38	16:36:06	483.4645	685.635	-8.59126	74.8796	15.1204
17	16-Jan-14	-9.74744	60.76562	0.524105	08:31:39	16:37:46	486.125	685.2886	-8.67786	74.69931	15.30069
18	17-Jan-14	-10.0825	61.10789	0.524338	08:30:37	16:39:29	488.8631	684.9535	-8.76161	74.51227	15.48773
19	18-Jan-14	-10.4058	61.45957	0.524562	08:29:32	16:41:12	491.6765	684.6302	-8.84245	74.31855	15.68145
20	19-Jan-14	-10.7172	61.82032	0.524779	08:28:24	16:42:58	494.5626	684.3188	-8.92029	74.11823	15.88177
21	20-Jan-14	-11.0163	62.18985	0.524986	08:27:13	16:44:44	497.5188	684.0197	-8.99508	73.9114	16.0886
22	21-Jan-14	-11.3031	62.56783	0.525185	08:26:00	16:46:32	500.5427	683.7329	-9.06677	73.69814	16.30186
23	22-Jan-14	-11.5772	62.95396	0.525376	08:24:44	16:48:21	503.6317	683.4588	-9.1353	73.47854	16.52146
24	23-Jan-14	-11.8385	63.34793	0.525557	08:23:25	16:50:12	506.7835	683.1975	-9.20063	73.25267	16.74733
25	24-Jan-14	-12.0868	63.74945	0.52573	08:22:03	16:52:03	509.9956	682.9492	-9.2627	73.02065	16.97935
26	25-Jan-14	-12.3219	64.1582	0.525893	08:20:39	16:53:55	513.2656	682.7141	-9.32148	72.78255	17.21745
27	26-Jan-14	-12.5438	64.57391	0.526047	08:19:13	16:55:48	516.5913	682.4922	-9.37695	72.53847	17.46153
28	27-Jan-14	-12.7522	64.99628	0.526192	08:17:44	16:57:42	519.9703	682.2838	-9.42906	72.28851	17.71149
29	28-Jan-14	-12.9472	65.42504	0.526327	08:16:13	16:59:37	523.4004	682.0888	-9.4778	72.03276	17.96724
30	29-Jan-14	-13.1285	65.85992	0.526453	08:14:39	17:01:32	526.8793	681.9075	-9.52314	71.77132	18.22868
31	30-Jan-14	-13.2963	66.30063	0.52657	08:13:03	17:03:28	530.405	681.7397	-9.56507	71.50431	18.49569
32	31-Jan-14	-13.4503	66.74693	0.526677	08:11:26	17:05:24	533.9754	681.5857	-9.60358	71.23181	18.76819
33	01-Feb-14	-13.5906	67.19855	0.526774	08:09:46	17:07:21	537.5884	681.4454	-9.63866	70.95393	19.04607
34	02-Feb-14	-13.7172	67.65526	0.526862	08:08:04	17:09:18	541.2421	681.3188	-9.67031	70.67078	19.32922
35	03-Feb-14	-13.8302	68.11681	0.52694	08:06:20	17:11:16	544.9345	681.2058	-9.69854	70.38247	19.61753
36	04-Feb-14	-13.9294	68.58297	0.527009	08:04:34	17:13:14	548.6638	681.1066	-9.72334	70.0891	19.9109
37	05-Feb-14	-14.015	69.05351	0.527069	08:02:46	17:15:12	552.4281	681.021	-9.74474	69.79079	20.20921
38	06-Feb-14	-14.087	69.52822	0.527119	08:00:56	17:17:10	556.2258	680.949	-9.76275	69.48763	20.51237
39	07-Feb-14	-14.1455	70.00689	0.527159	07:59:05	17:19:08	560.0551	680.8905	-9.77738	69.17976	20.82024
40	08-Feb-14	-14.1906	70.4893	0.527191	07:57:12	17:21:07	563.9144	680.8454	-9.78865	68.86727	21.13273
41	09-Feb-14	-14.2224	70.97527	0.527213	07:55:17	17:23:05	567.8022	680.8136	-9.79661	68.55028	21.44972
42	10-Feb-14	-14.2411	71.46462	0.527226	07:53:21	17:25:04	571.7169	680.7949	-9.80127	68.2289	21.7711
43	11-Feb-14	-14.2467	71.95714	0.52723	07:51:23	17:27:02	575.6571	680.7893	-9.80266	67.90326	22.09674

Figure 8: NOAA Data

4 Feature Selection

In order to determine the appropriate features to use in modelling the data a correlation matrix was used. The features which had a correlation above .5 were used in the final dataset. A seaborn heatmap was used to plot the results of the correlation matrix function (***figure 9***). Examples of these correlation matrices can be seen in ***figure 10*** , ***figure 11*** and ***figure 12***.

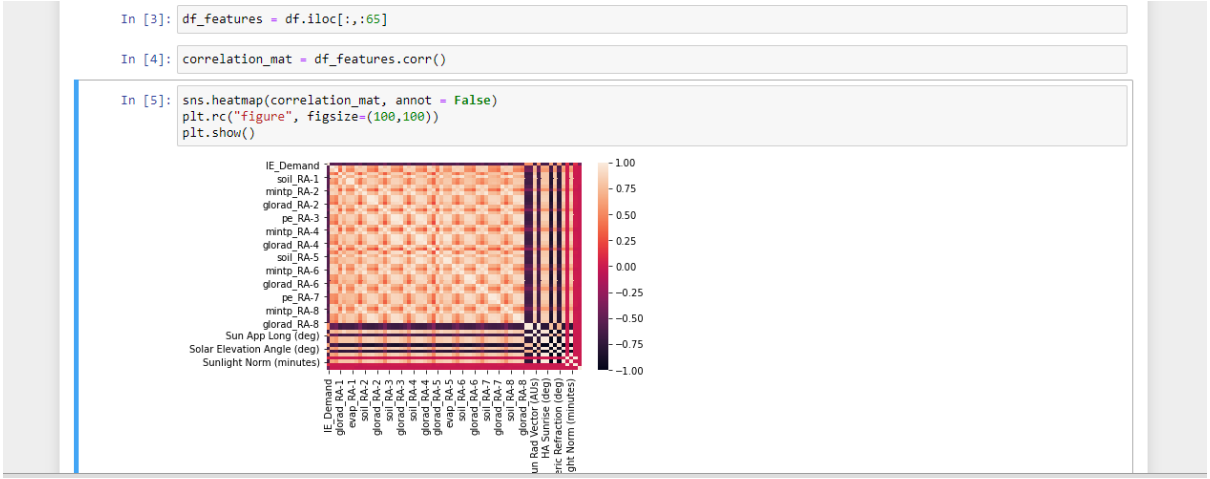


Figure 9: Correlation Matrix Syntax

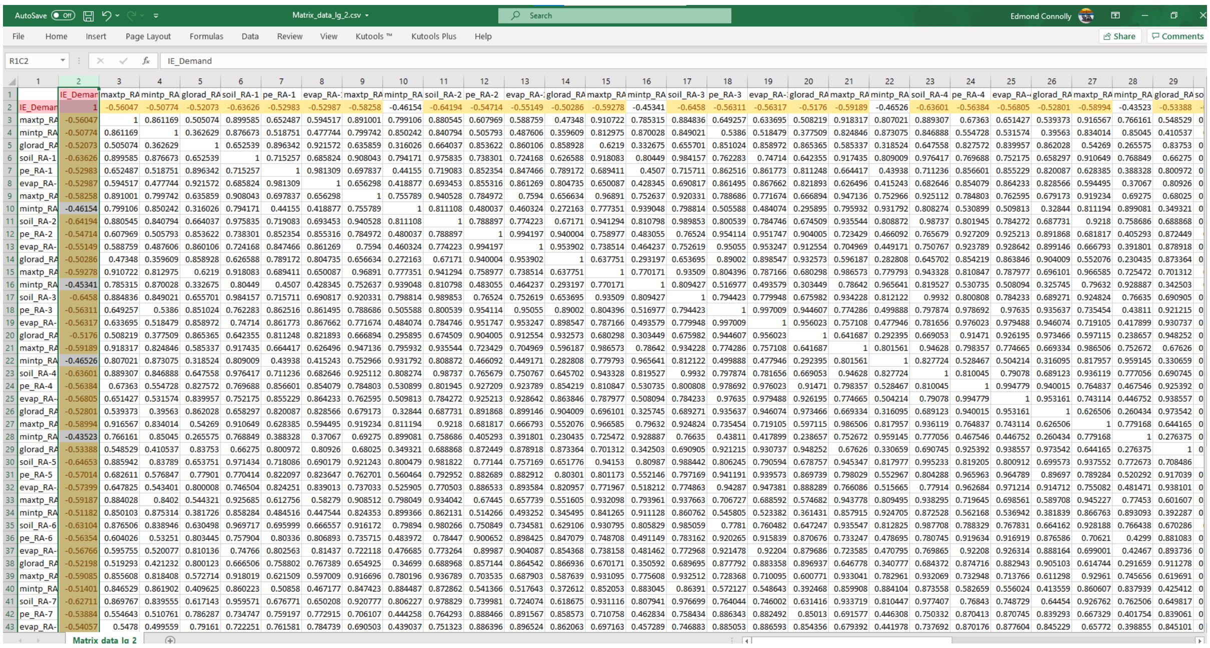


Figure 10: Correlation Matrix Values

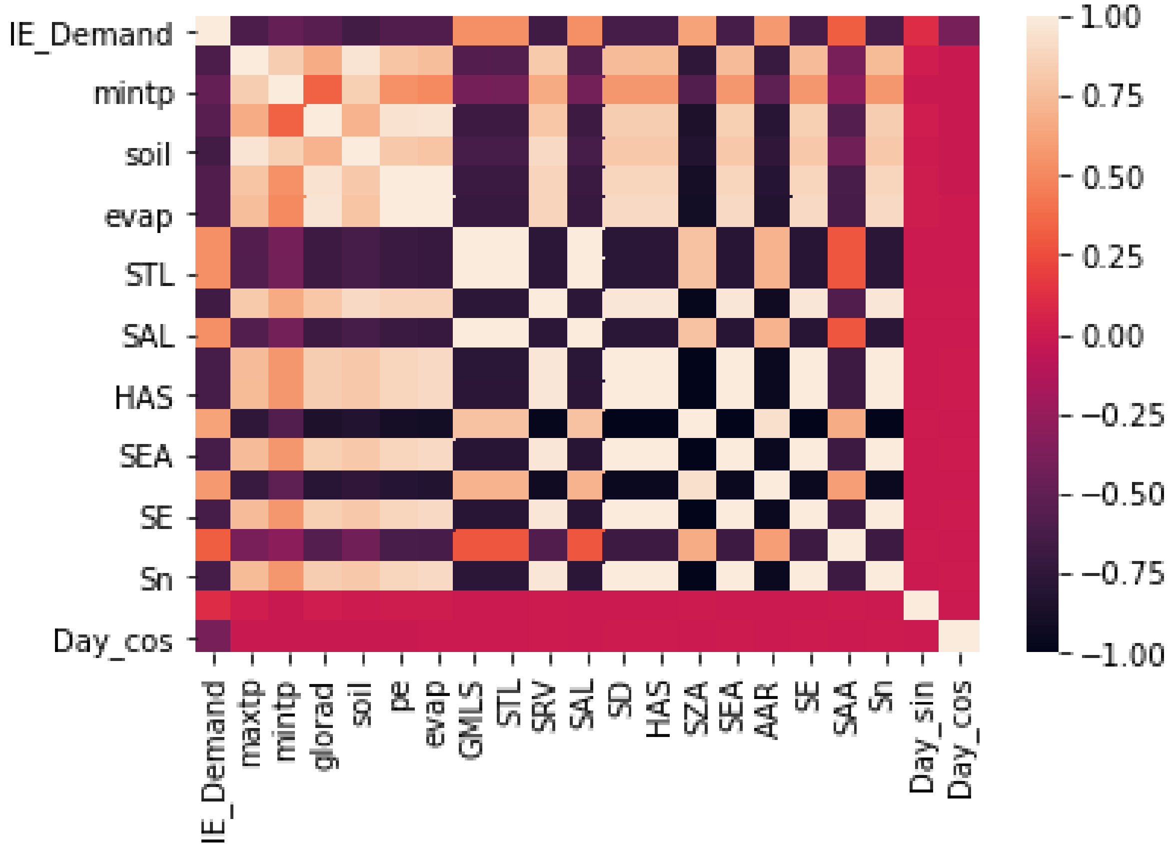


Figure 11: Correlation Matrix - Electricity Demand and Solar Data

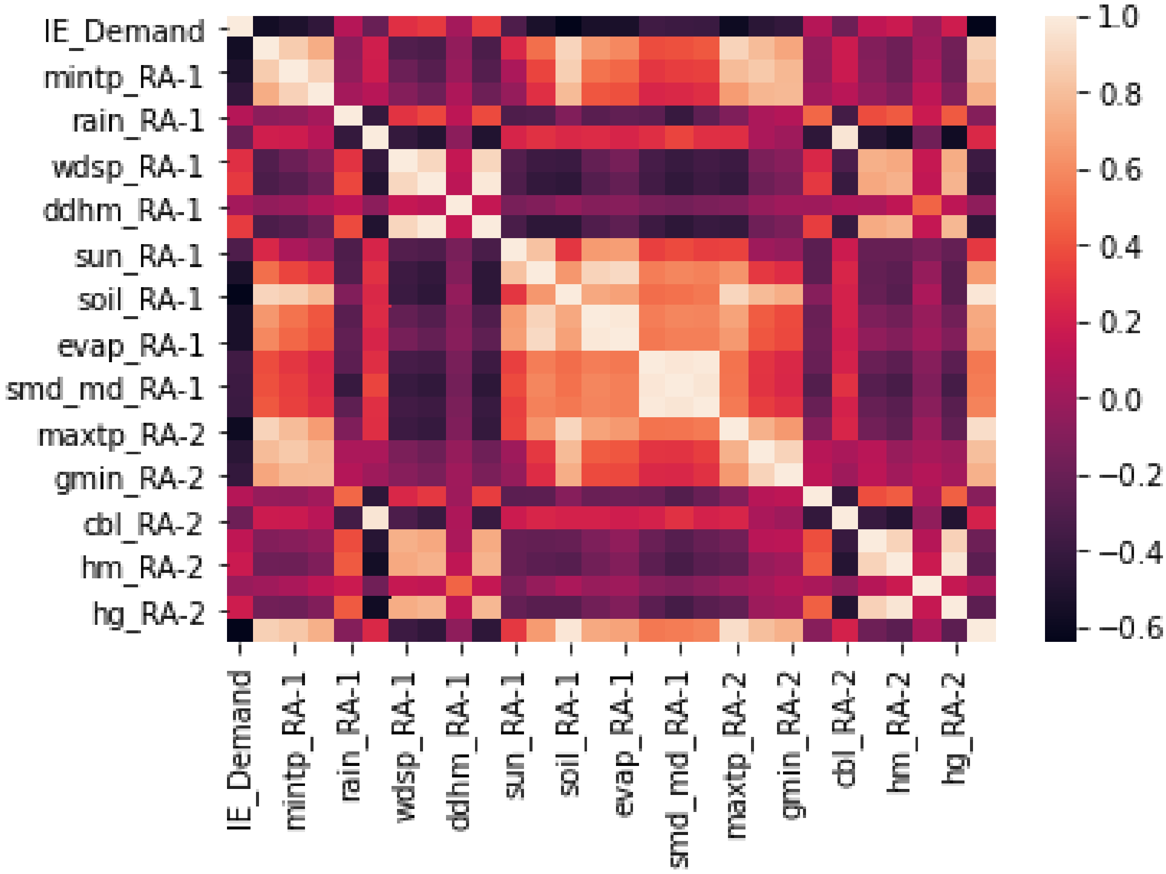


Figure 12: Correlation Matrix - Electricity Demand and Meteorological Data

5 Missing Values

There were not a lot of missing values in the dataset. Any missing values were limited to a few features on days throughout the data range in the weather station data. The knn algoithm was used to impute these features. The function used to complete this task can be seen in ***figure 13*** . The 85 values imputed refer to the original dataset of 2558 rows containing 144 features. Missing values were were imputed on the original dataset to provide as best information as possible for the algorithm. Ultimately this means that some of the values imputed were for features that were not required. This function was adapted from a tutorial from (Jason$_{B}rownlee$ (2020)).

```
In [11]: # split into input and output elements
         data = dataframe.values
         ix = [i for i in range(data.shape[1]) if i !=143]
         X, y = data[:, ix], data[:, 143]
         # print total missing
         print('Missing: %d' % sum(isnan(X).flatten()))
         # define imputer
         imputer = KNNImputer()
         # fit on the dataset
         imputer.fit(X)
         # transform the dataset
         Xtrans = imputer.transform(X)
         # print total missing
         print('Missing: %d' % sum(isnan(Xtrans).flatten()))

         Missing: 85
         Missing: 0
```

Figure 13: Application of the KNN Algorithm to impute missing values.

6 Encoding Days of the Week

The encoding of the days of the week used a Periodic Cyclic transformation. The purpose of this is to transform the name of the day into variables that still retain a degree of proximity to each other. This is to allow the algorithm using these features to determine that the days follow one another and repeat in a cycle. This is not achieved through one-hot-encoding or by simply giving each day a number. Two variables are created for each day. The formula uses cosine and sin to produce each feature respectively. For an illustration of this formula and the values it returns please see ***figure 14***. A scatterplot illustrates how a machine learning algorithm views this Periodic Cyclic encoding.

```
In [1]: %matplotlib inline
        import matplotlib.pyplot as plt
        import pandas as pd
        import numpy as np
```

```
In [2]: data = [['Sunday', 0], ['monday', 1], ['tuesday', 2],['wednesday', 3],['thursday',4],['friday',5],['saturday',6]]
        df_day = pd.DataFrame(data, columns=['day_name','day_number'])
```

```
In [3]: df_day['sin_time']= np.sin(2*np.pi*df_day.day_number/7)
        df_day['cos_time']= np.cos(2*np.pi*df_day.day_number/7)
        df_day
```

Out[3]:

	day_name	day_number	sin_time	cos_time
0	Sunday	0	0.000000	1.000000
1	monday	1	0.781831	0.623490
2	tuesday	2	0.974928	-0.222521
3	wednesday	3	0.433884	-0.900969
4	thursday	4	-0.433884	-0.900969
5	friday	5	-0.974928	-0.222521
6	saturday	6	-0.781831	0.623490

```
In [6]: df_day.plot.scatter('sin_time','cos_time').set_aspect('equal');
```

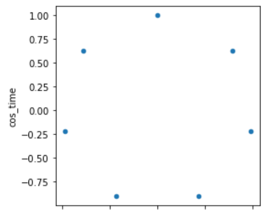


Figure 14: Encoding the Days of the Week

7 Finalised Data

The completed datasets after the optimal features were selected are illustrated below in ***figure 15*** and ***figure 16***. This data set contains electricity consumption and weather variables from 8 weather stations throughout Ireland for the years 2014 through 2020. Variables from the NOAA are also included and days of the week have been encoded. In total there are 2558 rows of data and 64 features. This dataset contains no missing values and is the finalised data set to be used with both the Sarimax model and the LSTM.

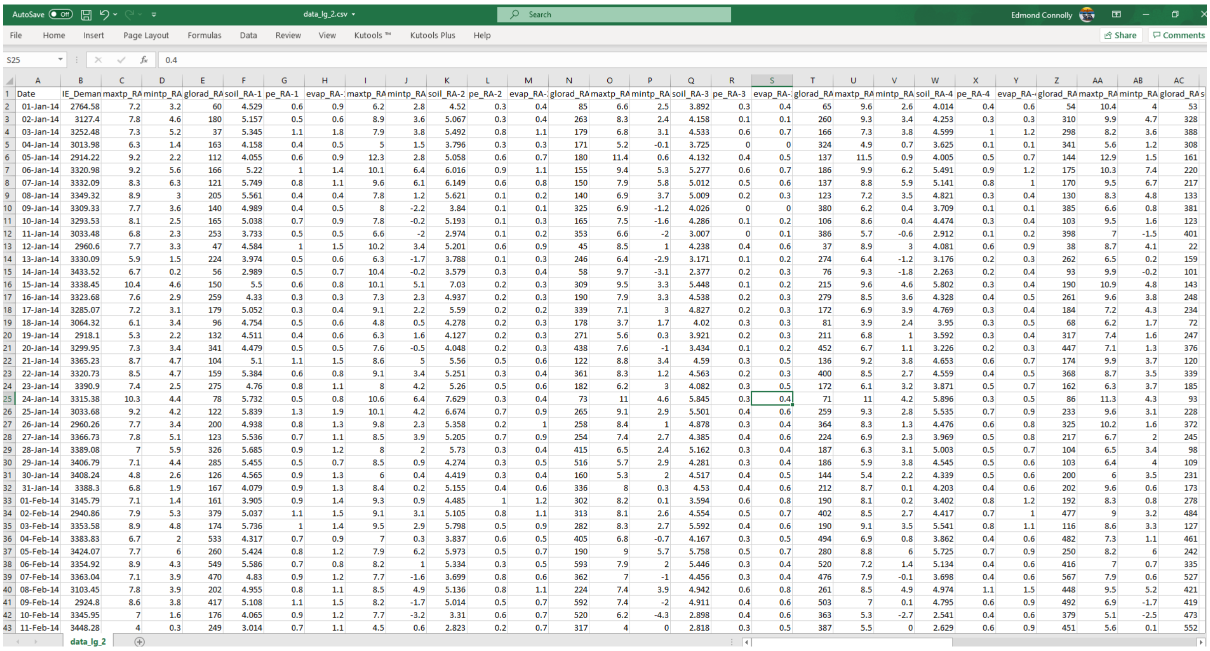


Figure 15: Finalised Data

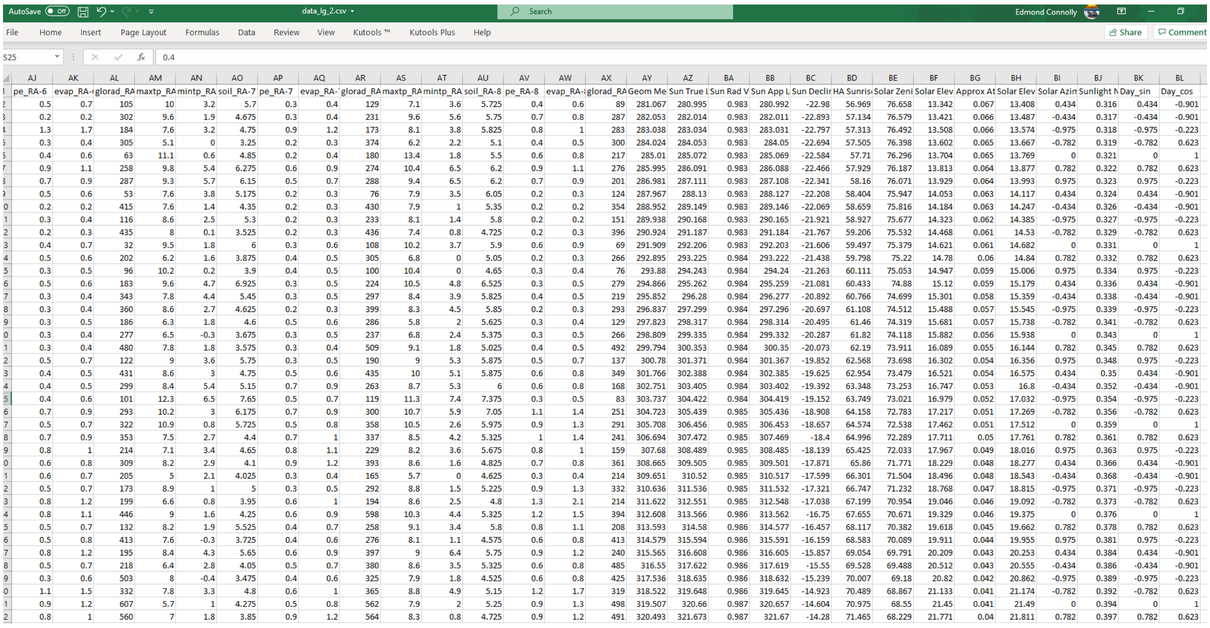


Figure 16: Finalised Data

8 Modelling

The implementation of the models took place in an Annaconda environment using python 3. The datasets had been prepared and only needed to be divided into training and testing sets for each model. The Eirgrid data that is to be modelled is plotted in ***figure 17***

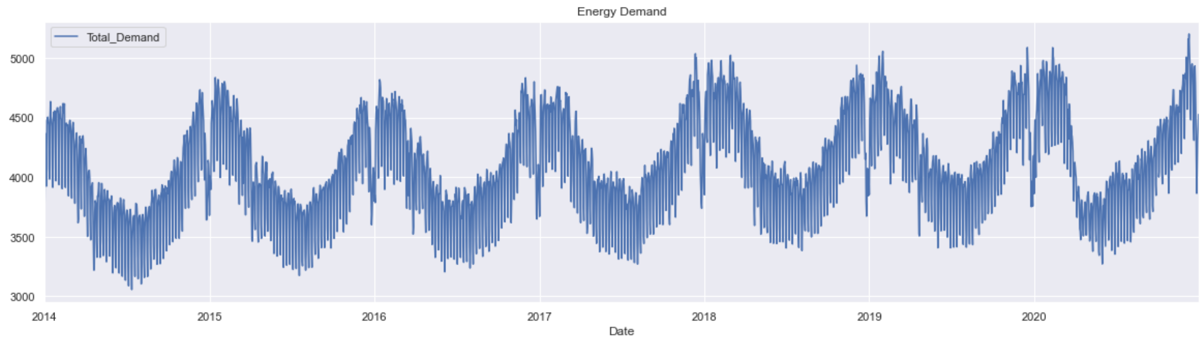


Figure 17: Electricity Consumption in The Republic of Ireland 2014 - 2020

8.1 Implementation of the Sarimax Model

The data was first loaded into the Annaconda environment in spyder. The index was set to the date field of the dataset (***figure 18***).

The first step in modelling this data was to look at the Auto Correlation Function Graph

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.graphics.tsaplots as sgt
import statsmodels.tsa.stattools as sts
from statsmodels.tsa.seasonal import STL
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.metrics import r2_score, mean_absolute_error
import seaborn as sns
sns.set()

raw_daily_data = pd.read_csv("data_lg_2.csv")
df_daily = raw_daily_data.copy()

df_daily.Date = pd.to_datetime(df_daily.Date, dayfirst = True)
df_daily.set_index("Date")
```

Figure 18: Loading the Data and Setting the Index

and the Partial Auto-Correlation Function Graph(***figure 19***). As mentioned previously this gives an indication of how many lags to use for the Auto-Regressive components (***figure 20***) and the Moving Averages components(***figure 21***). The Dickey-Fuller test had already established that the data was not stationary to the critical value for the 5% level of significance (***figure 22***). This indicated that the integration step was required.

Following this the test and train datasets were created and then tested for Seasonality (***figure 23***). The Sarimax model was then trained and fitted (***figure 24*** and ***figure 25***). The actual values(blue) of the test set were then plotted against the predicted values(red) (***figure 26***) and the evaluation metrics calculated (***figure 27***.

```python
sgt.plot_acf(df_daily.IE_Demand, lags=20, zero = False)
plt.title("ACF Consumption/Demand", size=24)
plt.show()

sgt.plot_pacf(df_daily.IE_Demand, lags = 20, zero = False, method = ('ols'))
plt.title("PACF Energy Demand", size = 24)
plt.show()

df_train = df_daily.iloc[:2496]
df_test = df_daily.iloc[2496:]

res_test = STL(df_test.IE_Demand,seasonal=9, robust = True).fit()
plt.rc("figure", figsize=(20,6))
res_test.plot()
plt.show()

start_date = "2020-11-01"
end_date = "2020-12-31"
```

Figure 19: Computing the ACF and the PACF

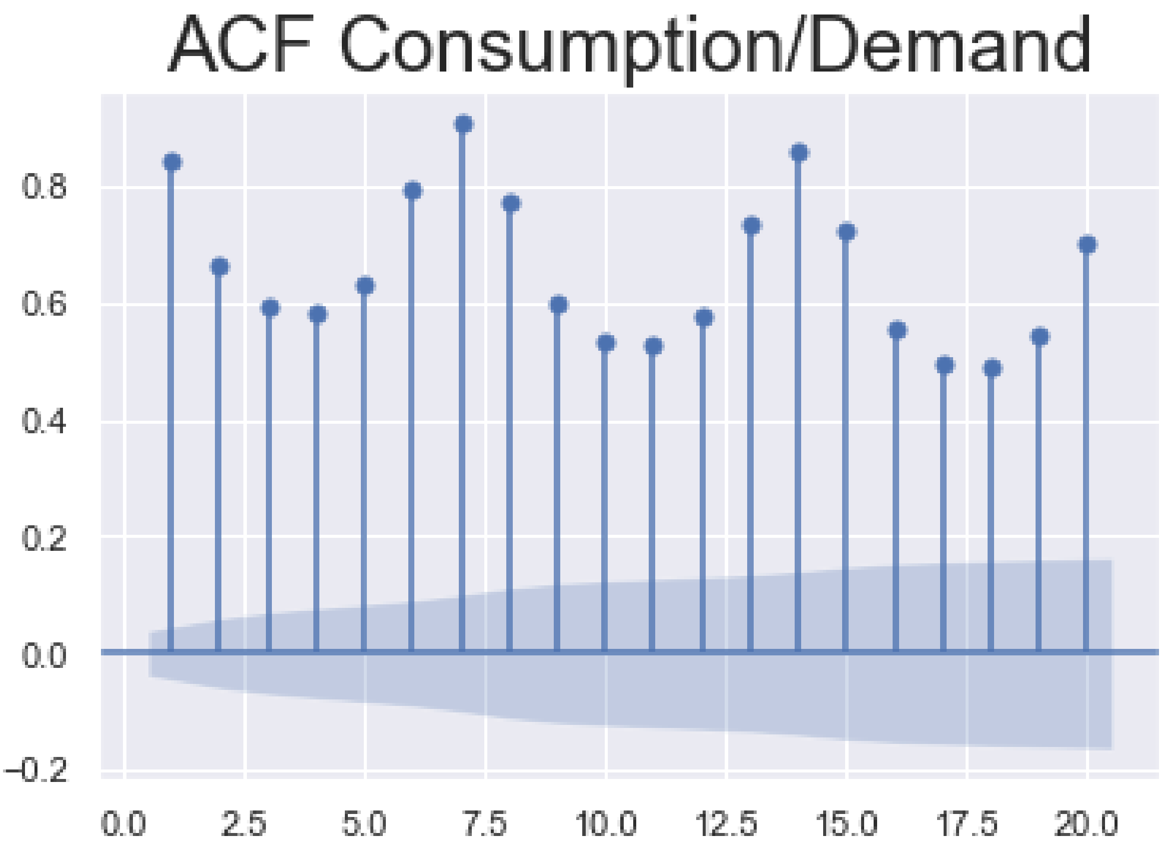


Figure 20: Auto Correlation Function

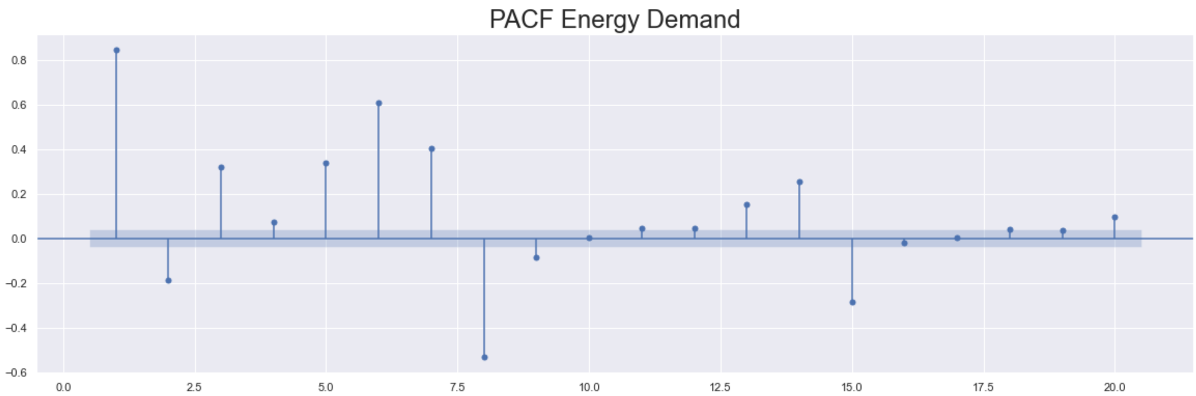


Figure 21: Partial Auto Correlation Function

```
In [11]: sts.adfuller(df_daily.IE_Demand)
Out[11]: (-2.3071780595395177,
          0.16966335475741245,
          27,
          2529,
          {'1%': -3.432938355012086,
           '5%': -2.8626835272597217,
           '10%': -2.567378742868999},
          29521.768484352164)
```

Figure 22: Augmented Dickey-Fuller Test

Figure 23: Test for Seasonality

```
start_date = "2020-11-01"
end_date = "2020-12-31"

#end_date = "2020-12-01"
model_demand_sarimax = SARIMAX(df_train.IE_Demand, exog = df_train[['maxtp_RA-1','mintp_RA-1','glorad_RA-1','soil_RA-1',
                                                    'pe_RA-1','evap_RA-1','maxtp_RA-2','mintp_RA-2','soil_RA-2',
                                                    'pe_RA-2','evap_RA-2','glorad_RA-2','maxtp_RA-3','mintp_RA-3',
                                                    'soil_RA-3','pe_RA-3','evap_RA-3','glorad_RA-3','maxtp_RA-4',
                                                    'mintp_RA-4','soil_RA-4','pe_RA-4','evap_RA-4','glorad_RA-4',
                                                    'maxtp_RA-5','mintp_RA-5','glorad_RA-5','soil_RA-5','pe_RA-5',
                                                    'evap_RA-5','maxtp_RA-6','mintp_RA-6','soil_RA-6','pe_RA-6','evap_RA-6',
                                                    'glorad_RA-6','maxtp_RA-7','mintp_RA-7','soil_RA-7','pe_RA-7','evap_RA-7',
                                                    'glorad_RA-7','maxtp_RA-8','mintp_RA-8','soil_RA-8','pe_RA-8','evap_RA-8',
                                                    'glorad_RA-8','Geom Mean Long Sun (deg)','Sun True Long (deg)',
                                                    'Sun Rad Vector (AUs)','Sun App Long (deg)','Sun Declin (deg)',
                                                    'HA Sunrise (deg)','Solar Zenith Angle (deg)','Solar Elevation Angle (deg)',
                                                    'Approx Atmospheric Refraction (deg)','Solar Elevation corrected for atm refraction (deg)',
                                                    'Solar Azimuth Angle (deg cw from N)','Sunlight Norm (minutes)','Day_sin','Day_cos']],
                        order = (1,1,2), seasonal_order = (1,0,2,7))
results_demand_sarimax = model_demand_sarimax.fit()

df_daily_pred_sarimax = results_demand_sarimax.predict(start = start_date, end = end_date,
                                    exog = df_test[['maxtp_RA-1','mintp_RA-1','glorad_RA-1','soil_RA-1',
                                                    'pe_RA-1','evap_RA-1','maxtp_RA-2','mintp_RA-2','soil_RA-2',
                                                    'pe_RA-2','evap_RA-2','glorad_RA-2','maxtp_RA-3','mintp_RA-3',
                                                    'soil_RA-3','pe_RA-3','evap_RA-3','glorad_RA-3','maxtp_RA-4',
                                                    'mintp_RA-4','soil_RA-4','pe_RA-4','evap_RA-4','glorad_RA-4',
                                                    'maxtp_RA-5','mintp_RA-5','glorad_RA-5','soil_RA-5','pe_RA-5',
                                                    'evap_RA-5','maxtp_RA-6','mintp_RA-6','soil_RA-6','pe_RA-6','evap_RA-6',
                                                    'glorad_RA-6','maxtp_RA-7','mintp_RA-7','soil_RA-7','pe_RA-7','evap_RA-7',
                                                    'glorad_RA-7','maxtp_RA-8','mintp_RA-8','soil_RA-8','pe_RA-8','evap_RA-8',
                                                    'glorad_RA-8','Geom Mean Long Sun (deg)','Sun True Long (deg)',
                                                    'Sun Rad Vector (AUs)','Sun App Long (deg)','Sun Declin (deg)',
                                                    'HA Sunrise (deg)','Solar Zenith Angle (deg)','Solar Elevation Angle (deg)',
                                                    'Approx Atmospheric Refraction (deg)','Solar Elevation corrected for atm refraction (deg)',
                                                    'Solar Azimuth Angle (deg cw from N)','Sunlight Norm (minutes)','Day_sin','Day_cos']][start_date:end_dat

df_daily_pred_sarimax[start_date:end_date].plot(figsize = (20,5), color = "red")
df_test.IE_Demand[start_date:end_date].plot(color = "blue")
plt.title("Predictions vs Actual", size = 24)
plt.show()
```

Figure 24: Sarimax Model Implementation

```python
results_demand_sarimax.mse

results_demand_sarimax.mae

ans = pd.DataFrame(df_daily_pred_sarimax)

df_results = pd.concat([df_test['IE_Demand'],ans['predicted_mean']],axis=1)


df_Results_comp = df_results.iloc[0:60].round(2)

df_Results_comp['error'] = df_Results_comp['IE_Demand'] - df_Results_comp['predicted_mean']



rmse = np.sqrt(np.mean(df_Results_comp['error']**2)).round(2)
mape = np.round(np.mean(np.abs(100*df_Results_comp['error']/df_Results_comp['IE_Demand'])), 2)

print('RMSE =', rmse, 'MW')
print('MAPE =', mape, '%')
print('R2 =', r2_score(df_Results_comp['IE_Demand'], df_Results_comp['predicted_mean']).round(2))
print('MAE =', mean_absolute_error(df_Results_comp['IE_Demand'], df_Results_comp['predicted_mean']).round(2))
```

Figure 25: Sarimax Model Implementation Plotting Results

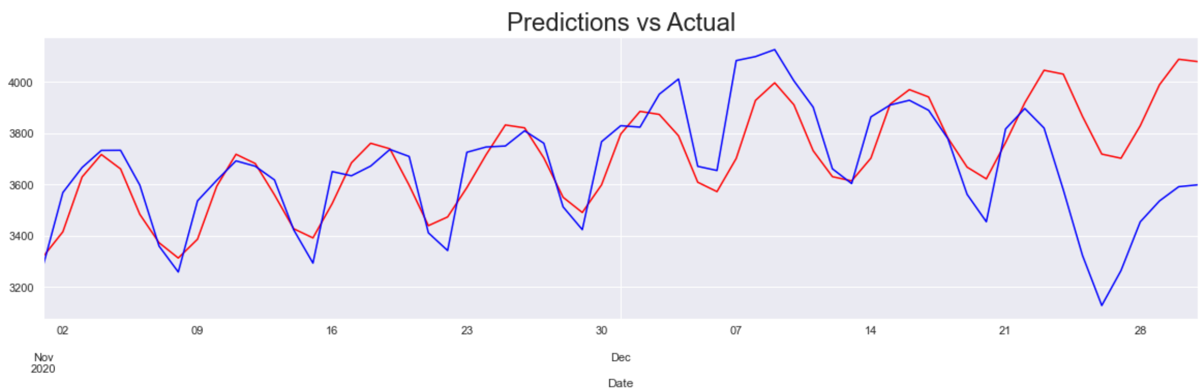


Figure 26: Sarimax Model Implementation Results

```
RMSE = 194.13 MW
MAPE = 3.58 %
R2 = 0.24
MAE = 128.28
```

Figure 27: Sarimax Model Evaluation Metrics

8.2 Implementation of the Bi-Directional Long Short Term Memory Neural Network

The data was first loaded into the Annaconda environment in spyder. The index was set to the date field of the dataset. Training and test sets were created and the a selection of columns scaled. RobustScaler was used as this function allows the rescaling of data. MinmaxScaler was also considered. The encoded day values were not scaled.Following this the training set was reshaped into a 3 dimensional numpy array for input to the LSTM (***figure 28***). The reshaping of data into a numpy array was created from a tutorial by Hristo$_{M}avrodief$ (2019). After this the model was trained with the parameters as discussed in the accompanying document (***figure 29***). A plot to view the validation loss form the validation set was produced. This helped with viewing the convergence of the loss between the datasets and assisted in determining the optimal number of epochs for the training of the LSTM (***figure 30***). Once the model was fitted the results were plotted and the evaluation metrics computed for the results. This involved flattening the the output array and rescaling the values (***figure 31***). The final graph was a comparison between the models predicted values and actual values from the test set (***figure 32***). The metrics for the model can be viewed in ***figure 33***

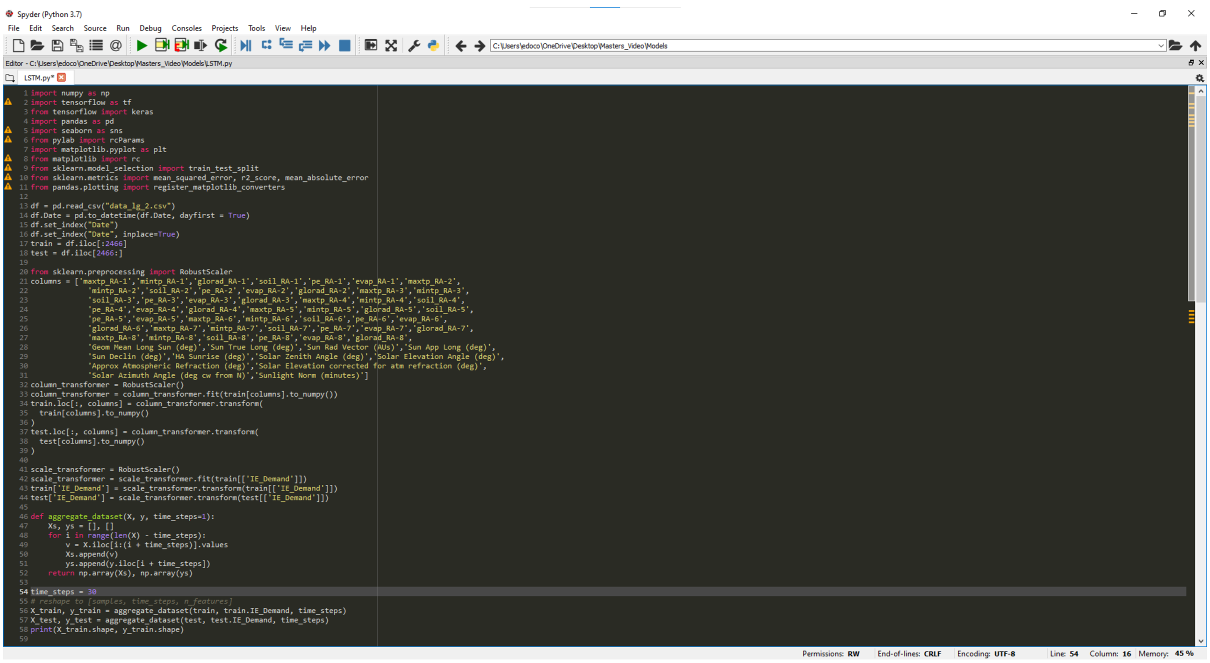


Figure 28: LSTM Preparation

```python

from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers import Dropout


model = keras.Sequential()
model.add(
  keras.layers.Bidirectional(
    keras.layers.LSTM(
      units=128,
        return_sequences = True,
      input_shape=(X_train.shape[1], X_train.shape[2])
    )
  )
)
model.add(keras.layers.Dropout(rate=0.2))


model.add(keras.layers.Bidirectional(keras.layers.LSTM(units=128, return_sequences = True,)))
model.add(keras.layers.Dropout(rate=0.2))


model.add(keras.layers.Bidirectional(keras.layers.LSTM(units=128)))
model.add(keras.layers.Dropout(rate=0.2))

model.add(keras.layers.Dense(units=1))
model.compile(loss='mean_squared_error', optimizer='adam')


history = model.fit(
    X_train, y_train,
    epochs=30,
    batch_size=32,
    validation_split=0.1,
    shuffle=False
)

```

Figure 29: LSTM Training and Fitting

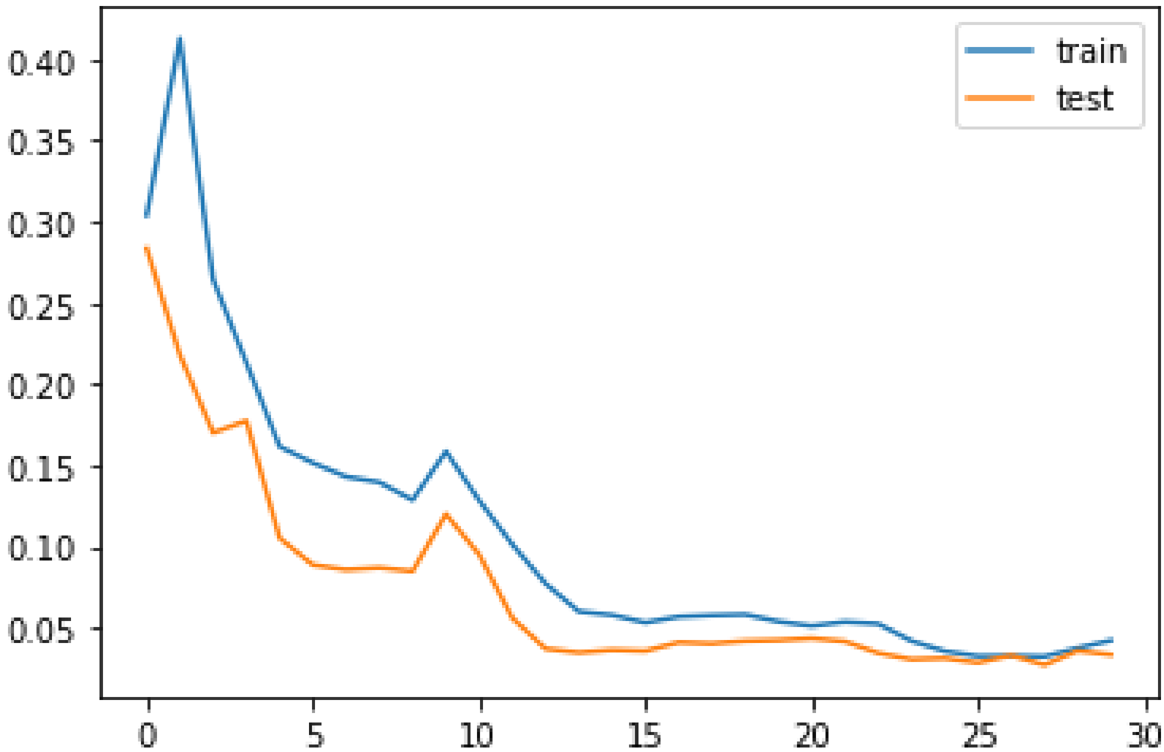


Figure 30: Validation Loss

```python
plt.plot(history.history['loss'], label='train')
plt.plot(history.history['val_loss'], label='test')
plt.legend();

y_pred = model.predict(X_test)


y_train_inv = scale_transformer.inverse_transform(y_train.reshape(1, -1))
y_test_inv = scale_transformer.inverse_transform(y_test.reshape(1, -1))
y_pred_inv = scale_transformer.inverse_transform(y_pred)



plt.plot(y_test_inv.flatten(), marker='.', label="true")
plt.plot(y_pred_inv.flatten(), 'r', label="prediction")
plt.ylabel('Electricity Demand')
plt.xlabel('Time Period')
plt.legend()
plt.show();

y_test_sh = y_test_inv.reshape(-1,1)

y_pred_sh = y_pred_inv.reshape(-1,1)

df4 = pd.DataFrame(y_test_sh)

df5 = pd.DataFrame(y_pred_sh)

df_results = pd.concat([df4,df5],axis=1)

df_results.columns=['IE_Demand','Predicted_Demand']

df_results['error'] = df_results['IE_Demand'] - df_results['Predicted_Demand']

rmse = np.sqrt(np.mean(df_results['error']**2)).round(2)
mape = np.round(np.mean(np.abs(100*df_results['error']/df_results['IE_Demand'])), 2)
print('RMSE =', rmse, 'MW')
print('MAPE =', mape, '%')
print('R2 =', r2_score(y_test, y_pred).round(2))
print('MAE =', mean_absolute_error(y_test, y_pred).round(2))

print(df_results)
```

Figure 31: LSTM Evaluation

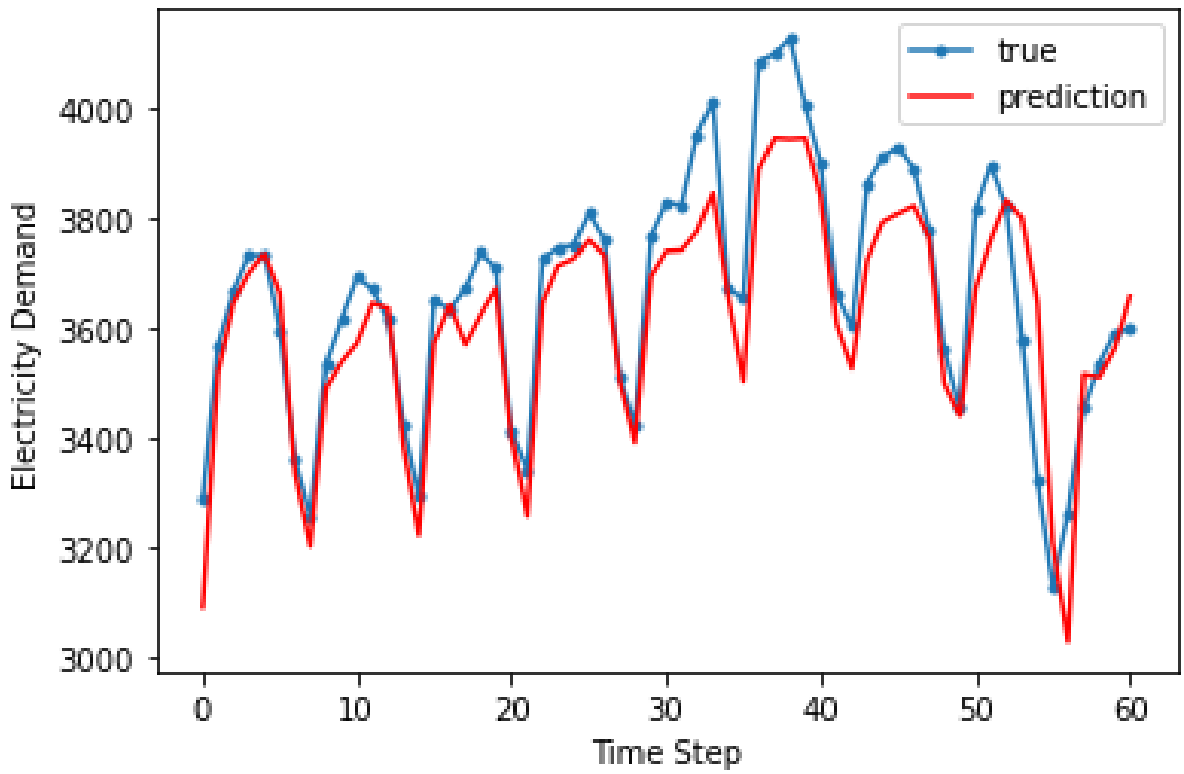


Figure 32: Predicted and Actual Values

```
RMSE = 82.67 MW
MAPE = 1.85 %
R2 = 0.74
MAE = 66.34
```

Figure 33: Metrics

References

Hristo$_M$*avrodief*(2019).*Demandpredictionwithlstmsusingtensorflow2andkerasinpython.*
URL:*https://curiousily.com/posts/demand-prediction-with-lstms-using-tensorflow-*

Jason$_B$*rownlee*(2020).*knnimputationformissingvaluesinmachinelearning.*
URL:*https://machinelearningmastery.com/knn-imputation-for-missing-values-in-mach-*