Daily Forecast for Active Power generation for the Turbine

Load Libraries

```
library(tidyverse)
library(tidymodels)
library(janitor)
library(skimr)
library(kableExtra)
library(GGally)
library(vip)
library(fastshap)
library(MASS)
library(ISLR)
library(tree)
library(ggplot2)
library(dplyr)
library(lubridate)
library(imputeTS)
library(urca)
library(pracma)
library(fpp2)
library(astsa)
#install.packages("forecast")
library(forecast)
```

Load Turbine data

```
turbine <- read_csv("Turbine_Data.csv") %>%
  clean_names() %>%
# transfer date format to "date"
  mutate(date = as.Date(date, format="%m/%d/%Y"))
```

```
## Rows: 118224 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## dbl  (7): Year, Month, Day, ActivePower, AmbientTemperature, WindDirection, ...
## dttm (1): Date
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
turbine
```

year <dbl>	month <dbl>	day <dbl>	date <date>	active_power <dbl>	ambient_temperature <dbl>	wind_direction <dbl>	wind_speed <dbl>
2017	12	31	2017-12-31	NA	NA	NA	NA
2017	12	31	2017-12-31	NA	NA	NA	NA
2017	12	31	2017-12-31	NA	NA	NA	NA
2017	12	31	2017-12-31	NA	NA	NA	NA
2017	12	31	2017-12-31	NA	NA	NA	NA
2017	12	31	2017-12-31	NA	NA	NA	NA
2017	12	31	2017-12-31	NA	NA	NA	NA
2017	12	31	2017-12-31	NA	NA	NA	NA

year	month	day	date	active_power	ambient_temperature	wind_direction	wind_speed
<dbl>	<dbl>	<dbl>	<date>	<dbl>	<dbl>	<dbl>	<dbl>
2017	12	31	2017-12-31	NA	NA	NA	NA
2017	12	31	2017-12-31	NA	NA	NA	NA

1-10 of 10,000 rows Previous **1** 2 3 4 5 6 ... 1000 Next

```
# skim for exploring missing values
skim(turbine)
```

Data summary

Name	turbine
Number of rows	118224
Number of columns	8
_______________________	
Column type frequency:	
Date	1
numeric	7
________________________	
Group variables	None

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2017-12-31	2020-03-30	2019-02-14	821

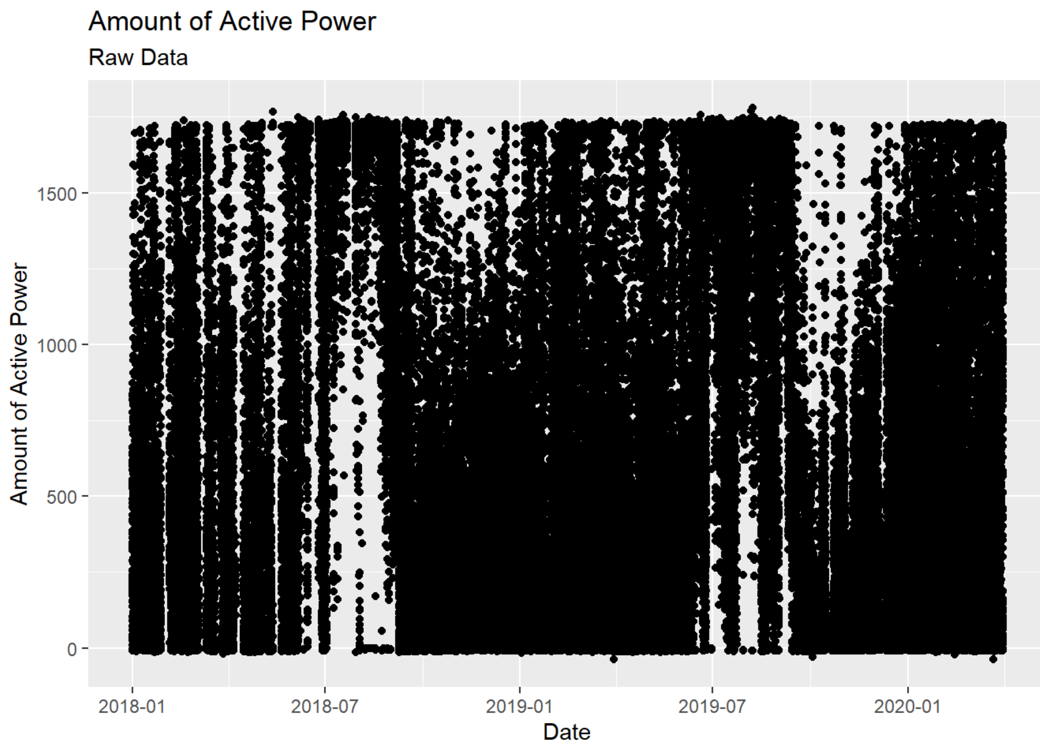
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
year	0	1.00	2018.66	0.67	2017.00	2018.00	2019.00	2019.00	2020.00	▁▇▁▇▂
month	0	1.00	6.04	3.56	1.00	3.00	6.00	9.00	12.00	▇▃▃▃▆
day	0	1.00	15.72	8.79	1.00	8.00	16.00	23.00	31.00	▇▇▇▇▆
active_power	23474	0.80	619.11	611.28	-38.52	79.64	402.65	1074.59	1779.03	▇▃▂▁▃
ambient_temperature	24407	0.79	28.77	4.37	0.00	25.63	28.34	31.66	42.41	▁▁▃▇▂
wind_direction	45946	0.61	196.29	88.30	0.00	145.00	182.00	271.00	357.00	▁▃▇▃▆
wind_speed	23629	0.80	5.88	2.62	0.00	3.82	5.56	7.51	22.97	▆▇▂▁▁

Plot raw data

```
# according to the plot, we should accumulate data to daily index to create time series object later
turbine %>%
ggplot(aes(x = date, y = active_power)) +
      geom_point() +
      labs(title = "Amount of Active Power",
           subtitle = "Raw Data",
           y = "Amount of Active Power",
           x = "Date")
```

```
## Warning: Removed 23474 rows containing missing values (geom_point).
```



Data Preparation - Accumulate Using a Daily Index

```
# 1. ACCUMULATE

turbine_day <- turbine %>%
  group_by(year, month, day) %>%
  summarize(sum_active_power = sum(active_power, na.rm=TRUE),
            avg_ambient_temperature = mean(ambient_temperature, na.rm=TRUE),
            avg_wind_direction = mean(wind_direction, na.rm=TRUE),
            avg_wind_speed = mean(wind_speed, na.rm=TRUE))
```

```
## `summarise()` has grouped output by 'year', 'month'. You can override using the
## `.groups` argument.
```

```
head(turbine_day)
```

year <dbl>	month <dbl>	d... <dbl>	sum_active_power <dbl>	avg_ambient_temperature <dbl>	avg_wind_direction <dbl>	avg_wind_speed <dbl>
2017	12	31	0.00	NaN	NaN	NaN
2018	1	1	26714.17	26.48230	282.4042	3.866226
2018	1	2	47865.31	25.64342	273.0716	4.677844
2018	1	3	43817.44	25.57071	275.2636	4.520639
2018	1	4	50659.94	24.96163	284.3388	5.063196
2018	1	5	30781.87	24.49187	280.7026	4.391144

6 rows

```
skim(turbine_day)
```

Data summary

Name	turbine_day

Number of rows	821
Number of columns	7
______________________	
Column type frequency:	
numeric	5
________________________	
Group variables	year, month

Variable type: numeric

skim_variable	year	month	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
day	2017	12	0	1.00	31.00	NA	31.00	31.00	31.00	31.00	31.00	
day	2018	1	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2018	2	0	1.00	14.50	8.23	1.00	7.75	14.50	21.25	28.00	
day	2018	3	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2018	4	0	1.00	15.50	8.80	1.00	8.25	15.50	22.75	30.00	
day	2018	5	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2018	6	0	1.00	15.50	8.80	1.00	8.25	15.50	22.75	30.00	
day	2018	7	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2018	8	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2018	9	0	1.00	15.50	8.80	1.00	8.25	15.50	22.75	30.00	
day	2018	10	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2018	11	0	1.00	15.50	8.80	1.00	8.25	15.50	22.75	30.00	
day	2018	12	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2019	1	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2019	2	0	1.00	14.50	8.23	1.00	7.75	14.50	21.25	28.00	
day	2019	3	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2019	4	0	1.00	15.50	8.80	1.00	8.25	15.50	22.75	30.00	
day	2019	5	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2019	6	0	1.00	15.50	8.80	1.00	8.25	15.50	22.75	30.00	
day	2019	7	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2019	8	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2019	9	0	1.00	15.50	8.80	1.00	8.25	15.50	22.75	30.00	
day	2019	10	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2019	11	0	1.00	15.50	8.80	1.00	8.25	15.50	22.75	30.00	
day	2019	12	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2020	1	0	1.00	16.00	9.09	1.00	8.50	16.00	23.50	31.00	
day	2020	2	0	1.00	15.00	8.51	1.00	8.00	15.00	22.00	29.00	
day	2020	3	0	1.00	15.50	8.80	1.00	8.25	15.50	22.75	30.00	
sum_active_power	2017	12	0	1.00	0.00	NA	0.00	0.00	0.00	0.00	0.00	
sum_active_power	2018	1	0	1.00	34741.33	23316.27	0.00	20689.73	31453.25	51168.94	86987.98	
sum_active_power	2018	2	0	1.00	50746.99	38932.26	0.00	21824.59	53186.52	74284.07	132483.33	
sum_active_power	2018	3	0	1.00	33193.52	32461.91	-34.03	0.00	37144.59	64548.81	101121.94	
sum_active_power	2018	4	0	1.00	27138.47	23373.33	0.00	1043.81	28615.12	41347.32	70295.16	
sum_active_power	2018	5	0	1.00	30443.78	32202.77	0.00	0.00	23507.68	46111.80	106672.02	
sum_active_power	2018	6	0	1.00	83452.86	81643.99	0.00	0.00	67492.93	155850.97	241795.68	
sum_active_power	2018	7	0	1.00	132448.04	100835.03	0.00	0.00	169936.81	221997.25	242396.49	
sum_active_power	2018	8	0	1.00	148095.68	58251.17	0.00	102385.70	165554.88	194994.85	225356.86	
sum_active_power	2018	9	0	1.00	71300.20	60324.46	1105.60	26310.78	48844.06	111352.93	220786.67	
sum_active_power	2018	10	0	1.00	36839.39	21270.73	-136.55	23500.01	32806.25	51955.27	78054.69	
sum_active_power	2018	11	0	1.00	39413.85	25873.19	0.00	23087.80	37592.81	57801.90	93757.08	
sum_active_power	2018	12	0	1.00	47888.61	20677.05	12218.84	35269.81	52787.31	60982.17	103859.12	
sum_active_power	2019	1	0	1.00	53305.42	26283.36	0.00	41862.13	54109.51	71489.89	98898.35	
sum_active_power	2019	2	0	1.00	71093.37	36437.92	14327.92	43602.63	76762.67	94498.74	166116.66	
sum_active_power	2019	3	0	1.00	60450.98	25140.07	26525.58	41846.44	56770.81	70367.44	132654.67	
sum_active_power	2019	4	0	1.00	44968.69	17823.34	13722.68	33642.27	39943.10	57504.94	99566.67	
sum_active_power	2019	5	0	1.00	64574.45	53597.91	0.00	28435.43	52282.87	84082.93	222377.16	
sum_active_power	2019	6	0	1.00	124609.61	68797.60	7110.35	67392.26	109360.44	186758.87	242131.73	

skim_variable	year	month	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
sum_active_power	2019	7	0	1.00	179766.10	60683.96	32276.47	137214.18	213866.96	227181.02	245055.93	
sum_active_power	2019	8	0	1.00	174100.03	58504.02	66822.52	125396.51	189964.37	225286.54	246440.14	
sum_active_power	2019	9	0	1.00	94482.49	90577.16	-29.09	23669.08	44265.74	196808.87	244827.20	
sum_active_power	2019	10	0	1.00	16817.14	21057.02	-19.45	1600.08	8799.17	24548.53	90378.76	
sum_active_power	2019	11	0	1.00	36369.34	21585.77	3336.09	22249.24	35925.07	50573.43	79074.24	
sum_active_power	2019	12	0	1.00	52971.97	33905.07	-526.34	35927.44	54913.92	76540.42	112653.46	
sum_active_power	2020	1	0	1.00	72382.28	28569.14	19795.53	51140.68	70598.32	94330.37	120987.48	
sum_active_power	2020	2	0	1.00	78811.19	32126.33	12052.35	53536.73	69622.94	100331.71	138916.52	
sum_active_power	2020	3	0	1.00	67765.12	28320.36	7283.23	45384.91	69496.35	91295.42	114258.53	
avg_ambient_temperature	2017	12	1	0.00	NaN	NA	NA	NA	NA	NA	NA	
avg_ambient_temperature	2018	1	4	0.87	25.32	1.62	21.34	24.24	25.51	26.39	28.21	
avg_ambient_temperature	2018	2	5	0.82	27.15	1.68	24.19	25.73	27.20	28.38	30.07	
avg_ambient_temperature	2018	3	11	0.65	30.85	2.19	26.13	29.99	31.46	32.02	34.54	
avg_ambient_temperature	2018	4	8	0.73	33.23	2.41	25.38	32.37	33.93	34.49	36.39	
avg_ambient_temperature	2018	5	11	0.65	33.30	1.35	28.95	32.75	33.38	34.27	35.00	
avg_ambient_temperature	2018	6	9	0.70	29.47	1.46	27.32	27.92	29.40	30.45	32.30	
avg_ambient_temperature	2018	7	9	0.71	28.53	1.01	24.73	28.23	28.67	29.00	29.85	
avg_ambient_temperature	2018	8	1	0.97	27.79	1.10	25.85	26.83	27.91	28.48	29.89	
avg_ambient_temperature	2018	9	0	1.00	28.88	1.39	25.18	28.11	29.13	29.81	31.30	
avg_ambient_temperature	2018	10	0	1.00	29.03	1.52	25.34	28.00	29.00	30.14	31.87	
avg_ambient_temperature	2018	11	1	0.97	26.54	1.55	23.50	25.30	26.75	27.86	28.99	
avg_ambient_temperature	2018	12	0	1.00	24.96	1.36	22.23	23.92	25.06	26.07	27.13	
avg_ambient_temperature	2019	1	2	0.94	24.65	1.26	21.53	23.68	24.84	25.32	27.94	
avg_ambient_temperature	2019	2	0	1.00	28.37	2.38	24.90	26.15	28.20	30.40	32.46	
avg_ambient_temperature	2019	3	0	1.00	32.11	1.74	28.15	31.45	32.34	32.98	35.40	
avg_ambient_temperature	2019	4	0	1.00	34.13	1.17	31.88	33.45	34.34	34.88	35.85	
avg_ambient_temperature	2019	5	3	0.90	34.64	1.62	31.70	33.26	34.62	35.85	37.69	
avg_ambient_temperature	2019	6	0	1.00	31.46	1.63	27.58	30.27	31.54	32.49	34.77	
avg_ambient_temperature	2019	7	0	1.00	29.56	0.98	27.71	28.73	29.84	30.30	31.24	
avg_ambient_temperature	2019	8	0	1.00	28.31	1.11	25.54	27.68	28.67	29.04	29.88	
avg_ambient_temperature	2019	9	1	0.97	28.28	1.21	25.52	27.47	28.62	29.22	29.83	
avg_ambient_temperature	2019	10	5	0.84	27.90	1.87	23.63	26.49	28.35	29.06	31.80	
avg_ambient_temperature	2019	11	0	1.00	26.47	1.70	24.07	25.02	26.22	27.32	29.44	
avg_ambient_temperature	2019	12	0	1.00	24.17	0.64	22.61	23.90	24.19	24.63	25.14	
avg_ambient_temperature	2020	1	2	0.94	25.59	1.40	23.44	24.50	25.21	26.17	28.77	
avg_ambient_temperature	2020	2	0	1.00	26.98	1.07	25.28	25.97	27.18	27.63	29.21	
avg_ambient_temperature	2020	3	1	0.97	30.13	1.64	25.42	29.24	30.15	31.20	32.87	
avg_wind_direction	2017	12	1	0.00	NaN	NA	NA	NA	NA	NA	NA	
avg_wind_direction	2018	1	4	0.87	286.46	12.30	255.92	281.55	286.80	289.20	332.15	
avg_wind_direction	2018	2	5	0.82	282.70	17.06	235.37	277.89	284.69	293.19	305.16	
avg_wind_direction	2018	3	11	0.65	248.79	50.12	155.33	215.52	268.32	288.27	304.92	
avg_wind_direction	2018	4	8	0.73	220.32	53.96	131.86	174.02	221.24	276.27	293.73	
avg_wind_direction	2018	5	11	0.65	188.61	54.92	102.35	148.08	177.04	240.13	276.79	
avg_wind_direction	2018	6	9	0.70	97.19	20.22	75.22	87.58	89.99	95.77	150.90	
avg_wind_direction	2018	7	9	0.71	89.49	5.46	83.39	85.53	88.61	91.94	106.63	
avg_wind_direction	2018	8	2	0.94	91.60	7.15	74.76	89.23	90.65	96.27	105.17	
avg_wind_direction	2018	9	0	1.00	137.20	55.59	78.28	101.54	114.07	153.96	270.34	
avg_wind_direction	2018	10	0	1.00	232.58	51.27	111.18	198.34	253.39	271.07	281.39	
avg_wind_direction	2018	11	1	0.97	172.28	16.88	130.64	163.08	176.26	184.95	194.00	
avg_wind_direction	2018	12	0	1.00	170.98	35.34	54.14	172.55	181.26	189.93	213.34	
avg_wind_direction	2019	1	2	0.94	186.92	10.26	161.70	184.83	187.36	190.93	219.36	
avg_wind_direction	2019	2	0	1.00	181.83	24.93	126.32	171.61	186.86	195.95	237.62	
avg_wind_direction	2019	3	0	1.00	188.66	27.11	127.05	177.29	191.40	203.02	246.15	
avg_wind_direction	2019	4	0	1.00	197.43	20.72	156.88	180.92	197.92	214.39	241.36	
avg_wind_direction	2019	5	3	0.90	214.46	49.69	106.36	183.68	218.80	244.81	330.57	
avg_wind_direction	2019	6	0	1.00	256.78	63.08	105.53	217.81	262.55	310.31	346.00	
avg_wind_direction	2019	7	0	1.00	292.98	56.96	118.91	281.95	314.38	332.21	341.14	
avg_wind_direction	2019	8	0	1.00	281.95	59.22	110.09	250.97	298.51	332.32	342.91	
avg_wind_direction	2019	9	1	0.97	238.49	61.53	130.24	193.82	212.87	290.08	342.19	
avg_wind_direction	2019	10	4	0.87	180.96	55.45	105.60	145.64	164.99	197.88	320.15	
avg_wind_direction	2019	11	0	1.00	164.14	12.03	145.18	157.90	163.25	167.55	198.41	

skim_variable	year	month	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
avg_wind_direction	2019	12	0	1.00	172.25	9.58	147.64	169.39	172.50	175.83	188.00	
avg_wind_direction	2020	1	0	1.00	180.79	12.22	157.69	175.90	178.31	182.20	229.22	
avg_wind_direction	2020	2	0	1.00	179.91	11.20	152.46	174.14	179.32	185.84	207.95	
avg_wind_direction	2020	3	0	1.00	179.43	12.92	148.27	171.16	178.30	188.83	204.73	
avg_wind_speed	2017	12	1	0.00	NaN	NA	NA	NA	NA	NA	NA	
avg_wind_speed	2018	1	4	0.87	4.81	0.78	3.26	4.29	4.93	5.23	6.89	
avg_wind_speed	2018	2	5	0.82	5.27	1.07	3.06	4.56	5.28	6.12	7.76	
avg_wind_speed	2018	3	11	0.65	5.03	1.02	2.40	4.47	5.29	5.72	6.43	
avg_wind_speed	2018	4	8	0.73	4.52	0.57	3.37	4.29	4.57	4.96	5.29	
avg_wind_speed	2018	5	11	0.65	4.78	1.33	2.95	4.05	4.49	5.15	9.07	
avg_wind_speed	2018	6	9	0.70	8.07	2.33	4.26	6.01	8.53	10.24	11.70	
avg_wind_speed	2018	7	9	0.71	9.90	1.73	5.29	8.98	9.86	11.07	12.71	
avg_wind_speed	2018	8	2	0.94	9.88	1.42	7.16	8.76	9.91	10.62	13.16	
avg_wind_speed	2018	9	0	1.00	5.36	1.73	2.77	3.99	5.11	6.60	8.88	
avg_wind_speed	2018	10	0	1.00	4.72	0.85	3.24	4.03	4.67	5.52	6.49	
avg_wind_speed	2018	11	1	0.97	5.06	0.79	3.34	4.65	5.16	5.52	6.64	
avg_wind_speed	2018	12	0	1.00	5.14	0.61	3.79	4.68	5.19	5.55	6.42	
avg_wind_speed	2019	1	2	0.94	5.43	0.67	4.42	4.95	5.28	5.97	6.93	
avg_wind_speed	2019	2	0	1.00	5.39	1.02	3.54	4.68	5.53	6.02	7.59	
avg_wind_speed	2019	3	0	1.00	5.07	0.76	4.06	4.50	5.01	5.38	7.12	
avg_wind_speed	2019	4	0	1.00	4.58	0.55	3.83	4.27	4.56	4.80	6.43	
avg_wind_speed	2019	5	3	0.90	5.43	1.40	3.59	4.50	4.98	5.95	9.11	
avg_wind_speed	2019	6	0	1.00	7.21	2.06	3.89	5.70	7.06	8.53	11.11	
avg_wind_speed	2019	7	0	1.00	8.63	1.94	4.08	7.09	9.60	10.13	11.28	
avg_wind_speed	2019	8	0	1.00	8.88	2.26	5.43	7.26	8.66	10.50	14.19	
avg_wind_speed	2019	9	1	0.97	6.40	2.75	2.92	4.05	6.00	8.60	11.33	
avg_wind_speed	2019	10	4	0.87	3.92	0.93	2.38	3.24	3.76	4.46	6.16	
avg_wind_speed	2019	11	0	1.00	4.48	0.92	2.69	3.94	4.80	5.17	5.82	
avg_wind_speed	2019	12	0	1.00	5.28	0.77	3.04	4.82	5.21	5.76	6.57	
avg_wind_speed	2020	1	0	1.00	5.41	0.84	3.65	4.77	5.48	6.11	6.81	
avg_wind_speed	2020	2	0	1.00	5.66	0.90	3.48	4.97	5.62	6.62	7.06	
avg_wind_speed	2020	3	0	1.00	5.15	0.81	3.31	4.51	5.29	5.72	6.54	

```
turbine_day <- subset(turbine_day, select=-c(year, month, day))
skim(turbine_day)
```

Data summary

Name	turbine_day
Number of rows	821
Number of columns	4
_______________________	
Column type frequency:	
numeric	4
________________________	
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
sum_active_power	0	1.00	71450.25	63976.28	-526.34	27565.48	54742.19	93542.88	246440.14	
avg_ambient_temperature	74	0.91	28.70	3.23	21.34	26.11	28.50	30.67	37.69	
avg_wind_direction	71	0.91	197.61	65.76	54.14	163.01	186.00	244.99	346.00	
avg_wind_speed	71	0.91	5.90	2.10	2.38	4.52	5.30	6.45	14.19	

```
#since the sum of null values returns 0, we change 0 value to null for impute
turbine_day[turbine_day == 0] <- NA
turbine_day
```

sum_active_power <dbl>	avg_ambient_temperature <dbl>	avg_wind_direction <dbl>	avg_wind_speed <dbl>
NA	NaN	NaN	NaN
26714.16695	26.48230	282.40415	3.866226
47865.31016	25.64342	273.07162	4.677844
43817.43605	25.57071	275.26357	4.520639
50659.93796	24.96163	284.33876	5.063196
30781.86943	24.49187	280.70259	4.391144
5057.91618	21.34054	332.15385	5.179815
NA	NaN	NaN	NaN
86987.97979	23.73776	280.55739	6.886447
71055.16284	23.54455	278.54561	5.644588

1-10 of 821 rows Previous **1** 2 3 4 5 6 ... 83 Next

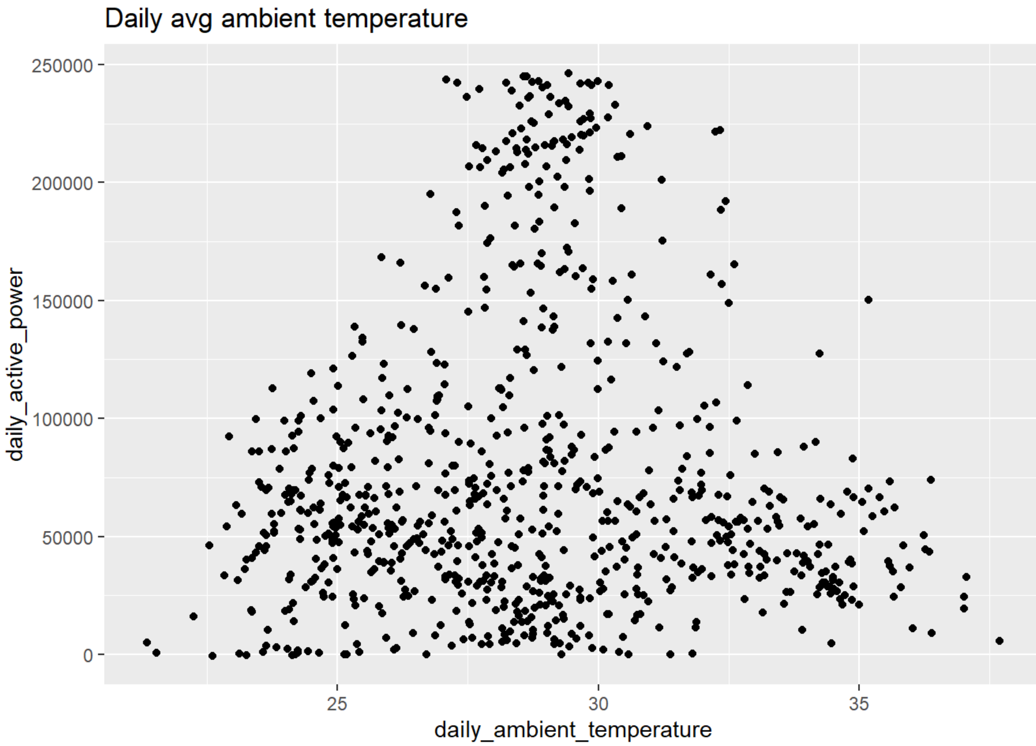
```
# 2. Impute Missing value

at_data <- subset(turbine_day, select=c(avg_ambient_temperature))
at_ts <- ts(at_data, start=c(2017,365), frequency = 365)
at_tsi <- na_interpolation(at_ts)

wd_data <- subset(turbine_day, select=c(avg_wind_direction))
wd_ts <- ts(wd_data, start=c(2017,365), frequency = 365)
wd_tsi <- na_interpolation(wd_ts)

ws_data <- subset(turbine_day, select=c(avg_wind_speed))
ws_ts <- ts(ws_data, start=c(2017,365), frequency = 365)
ws_tsi <- na_interpolation(ws_ts)

active_data <- subset(turbine_day, select=c(sum_active_power))
act_ts <- ts(active_data, start=c(2017,365), frequency = 365)
active_tsi <- na_interpolation(act_ts)

turbine_clean <- data.frame(daily_active_power= as.vector(active_tsi), daily_wind_direction= as.vector(wd_tsi), d
aily_ambient_temperature= as.vector(at_tsi),daily_wind_speed= as.vector(ws_tsi) )
turbine_clean
```

daily_active_power <dbl>	daily_wind_direction <dbl>	daily_ambient_temperature <dbl>	daily_wind_speed <dbl>
26714.16695	282.40415	26.48230	3.866226
26714.16695	282.40415	26.48230	3.866226
47865.31016	273.07162	25.64342	4.677844
43817.43605	275.26357	25.57071	4.520639
50659.93796	284.33876	24.96163	5.063196
30781.86943	280.70259	24.49187	4.391144
5057.91618	332.15385	21.34054	5.179815
46022.94799	306.35562	22.53915	6.033131
86987.97979	280.55739	23.73776	6.886447
71055.16284	278.54561	23.54455	5.644588

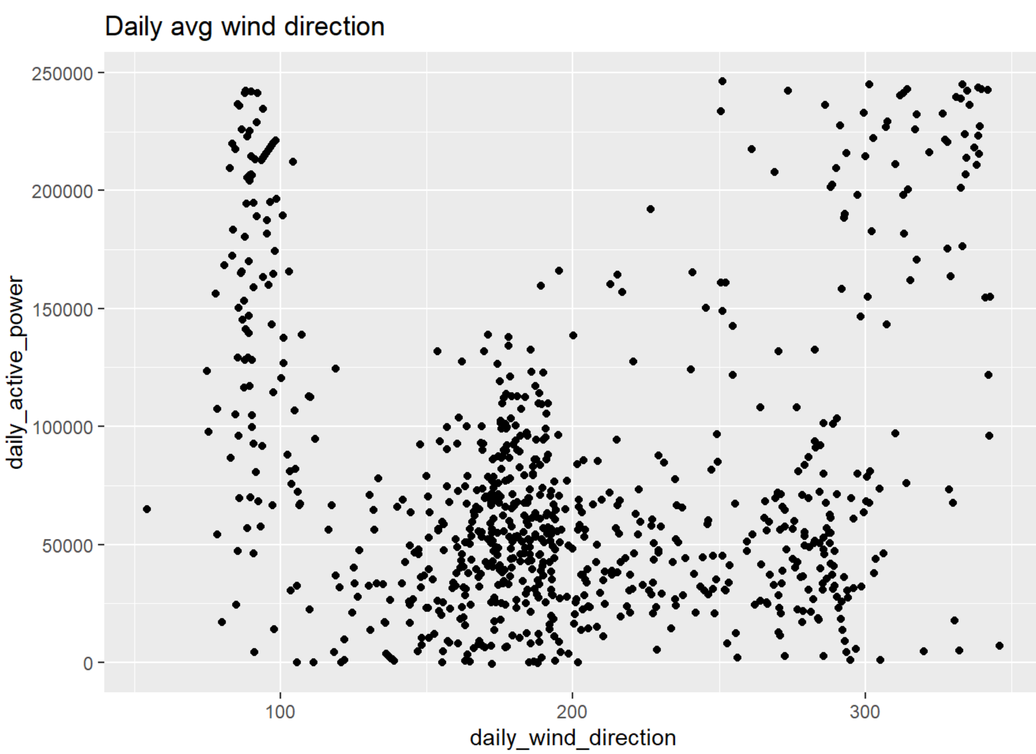
1-10 of 821 rows Previous **1** 2 3 4 5 6 ... 83 Next

Exploratory analysis of three x variables

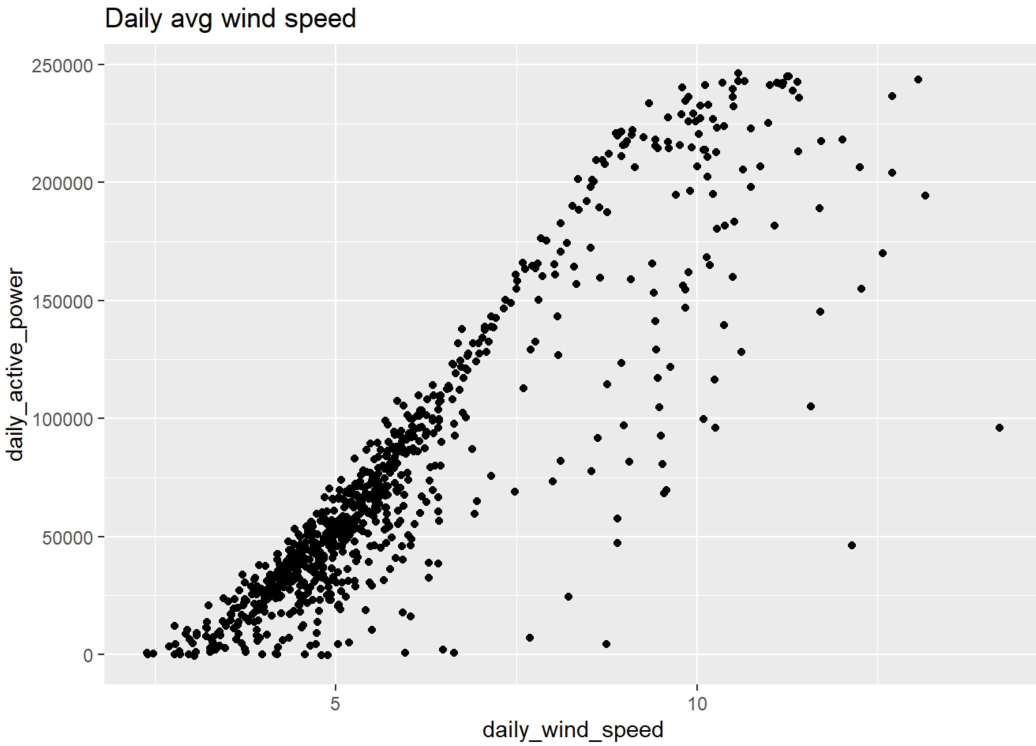
```
# according to the patterns of three plots below, we assume all three variables has some correlations with "active power;" we will do more test for them when we run "Arima" model later

ggplot(turbine_clean, aes(x=daily_ambient_temperature, y=daily_active_power)) + geom_point() + labs(title = "Daily avg ambient temperature")
```




```
ggplot(turbine_clean, aes(x=daily_wind_direction, y=daily_active_power)) + geom_point() + labs(title = "Daily avg wind direction")
```

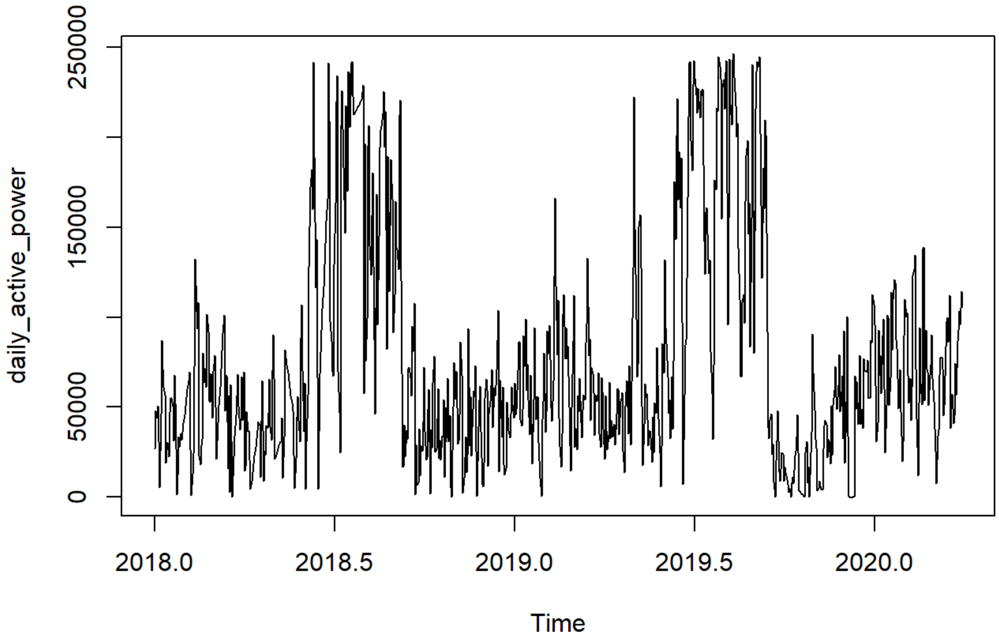


```
ggplot(turbine_clean, aes(x=daily_wind_speed, y=daily_active_power)) + geom_point() + labs(title = "Daily avg win
d speed")
```

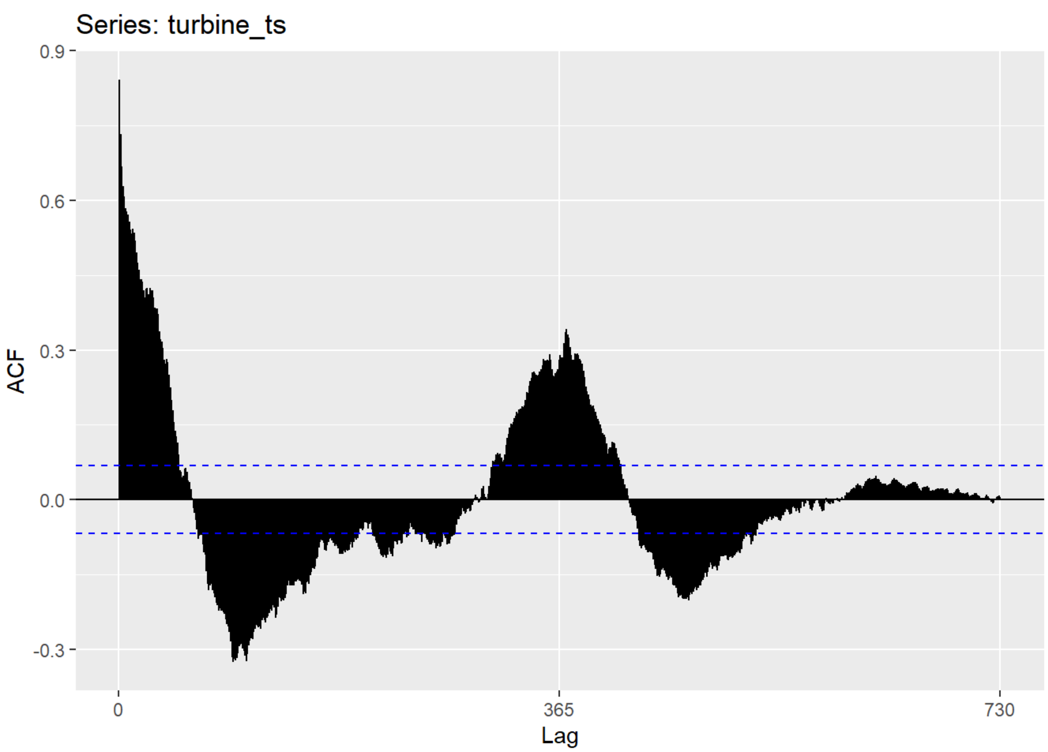


An overview/visual summary of the data that were used to generate the forecast - “Create a time series object for the data”

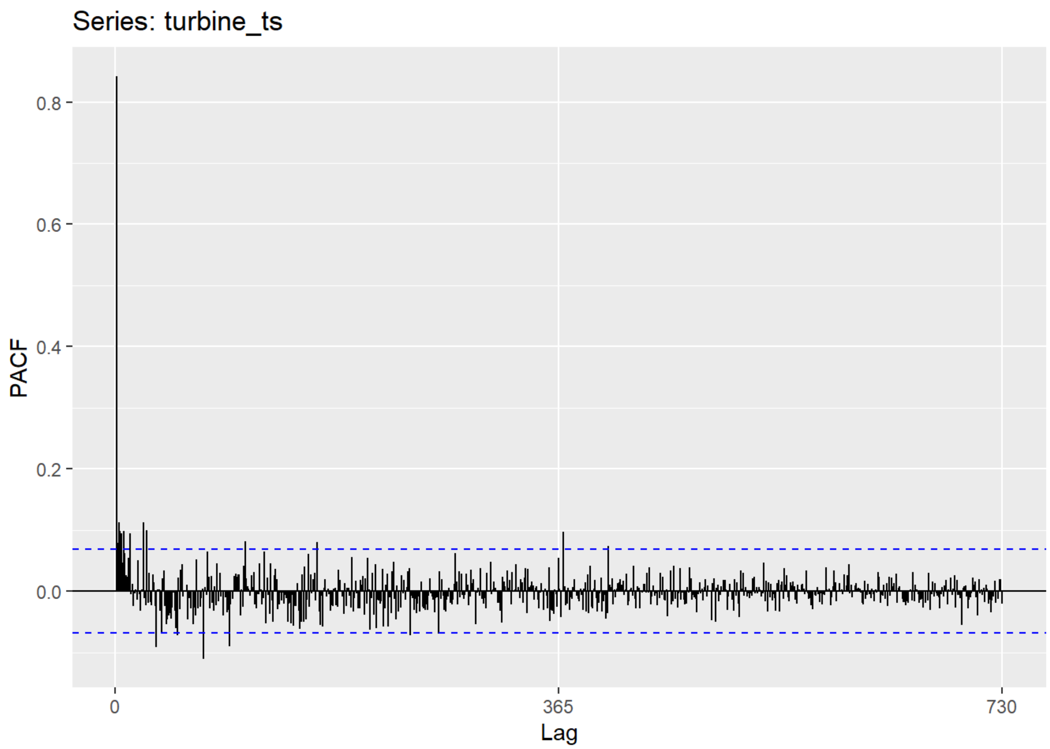
```
# Create time series object and plot time series

ts_prep <- subset(turbine_clean, select=c(daily_active_power))
turbine_ts <- ts(ts_prep, start=c(2017,365), frequency = 365)
plot(turbine_ts)
```




```
ggAcf(turbine_ts)
```




```
ggPacf(turbine_ts)
```



Type 1. Seasonal Exponenetial Smoothing Model

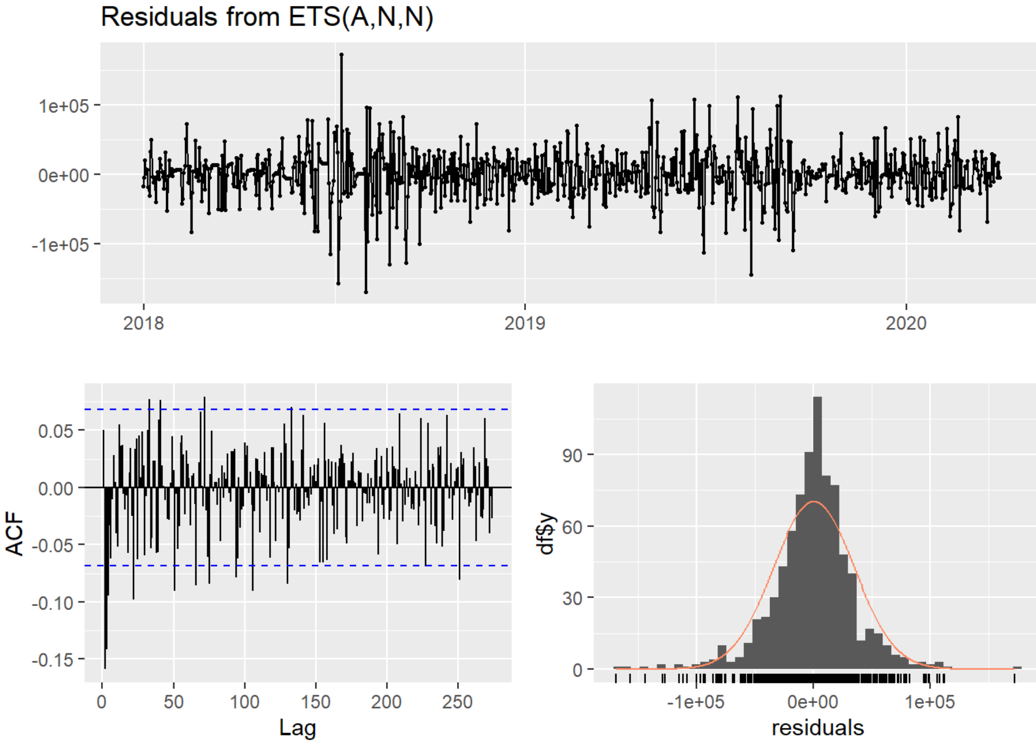
```
# due to "seasonality" from time series and ACF plot, we try to run seasonal expenential smoothing model first to
forecast

turbine_ets <- ets(turbine_ts, model="ZNZ")
```

```
## Warning in ets(turbine_ts, model = "ZNZ"): I can't handle data with frequency
## greater than 24. Seasonality will be ignored. Try stlf() if you need seasonal
## forecasts.
```

```
summary(turbine_ets)
```

```
## ETS(A,N,N)
##
## Call:
##  ets(y = turbine_ts, model = "ZNZ")
##
##   Smoothing parameters:
##     alpha = 0.7485
##
##   Initial states:
##     l = 43566.3515
##
##   sigma:  34659.04
##
##      AIC     AICc      BIC
## 22677.68 22677.71 22691.81
##
## Training set error measures:
##                    ME     RMSE      MAE      MPE    MAPE      MASE       ACF1
## Training set 102.6698 34616.79 24541.37 347.1835 548.622 0.5939651 0.05003906
```

```
checkresiduals(turbine_ets)
```

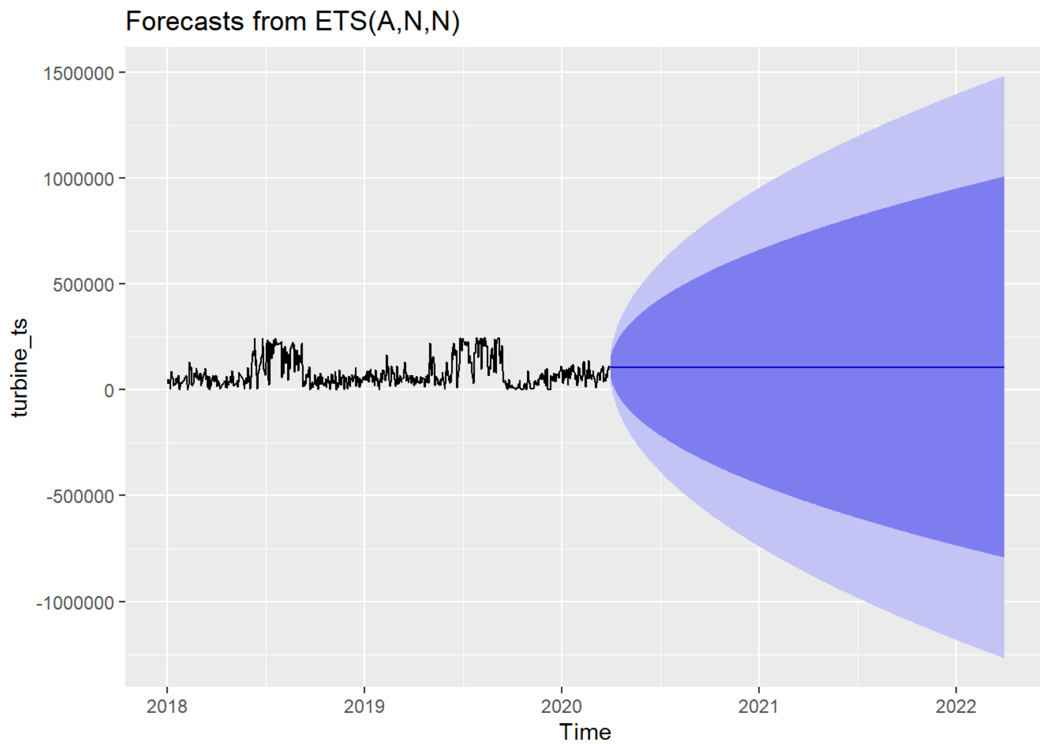



```
##
##  Ljung-Box test
##
## data:  Residuals from ETS(A,N,N)
## Q* = 257.48, df = 162, p-value = 2.629e-06
##
## Model df: 2.   Total lags used: 164
```

```
forecast(turbine_ets, h=5)
```

```
##           Point Forecast    Lo 80    Hi 80      Lo 95    Hi 95
## 2020.2466       106659.8 62242.50 151077.2  38729.383 174590.3
## 2020.2493       106659.8 51177.80 162141.9  21807.381 191512.3
## 2020.2521       106659.8 41978.98 171340.7   7739.001 205580.7
## 2020.2548       106659.8 33934.53 179385.2  -4563.922 217883.6
## 2020.2575       106659.8 26695.30 186624.4 -15635.368 228955.1
```

```
turbine_ets %>% forecast() %>% autoplot()
```




```
#relatively high forecast errors (RMSE)
#residuals are not white noises (According to p-values for Ljung-Box statistic of Residuals, residuals are white
noise. H0: white noise; Ha: not white noise. As p-value is smaller than 0.05, we reject null hypothesis, and stat
e residuals are not white noise.)
#flat pattern on forecast plot (forecast values) for all five forecast days
#this is not a good model for for Active Power generation for the turbine for the next 5 day period
```

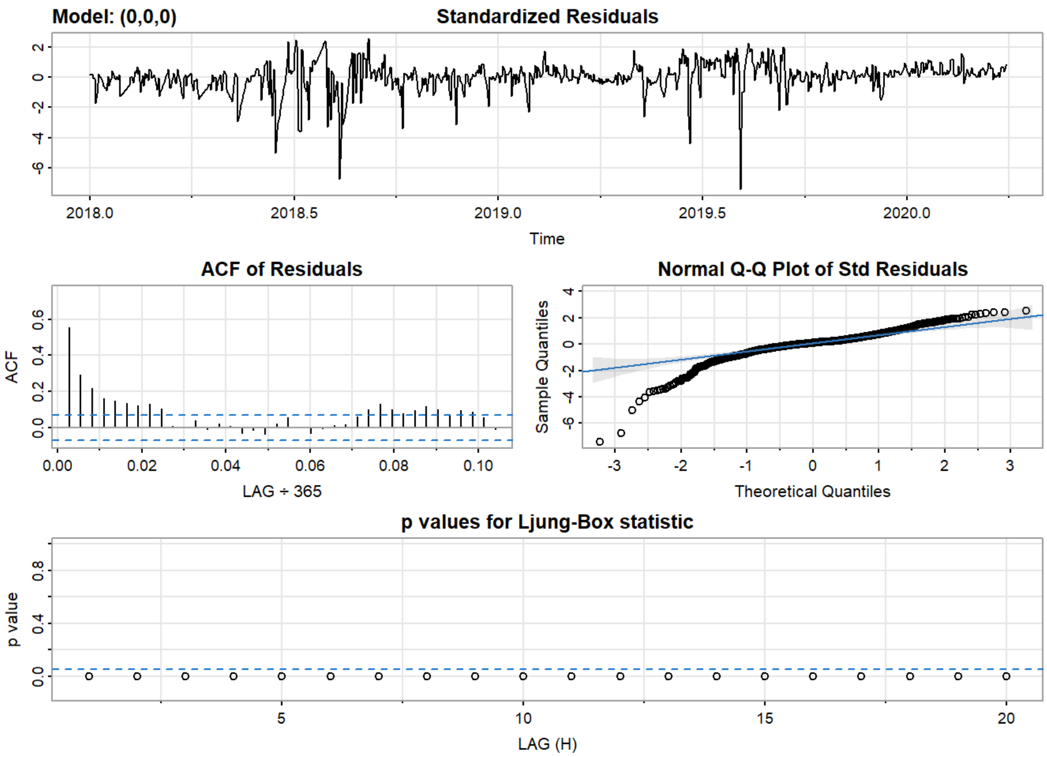
Type 2. “Arima” Method Models

Model 1 - Fit a regression model (including only the explanatory variables)

—-Examine the residuals to determine Which explanatory variables are important?

```
# set p, d, q=0, xreg= linear model
fit1_AR <- sarima(turbine_ts, 0, 0, 0, xreg=turbine_clean[,2:4]) #matrix of all variables
```

```
## initial  value 10.240925 
## iter   1 value 10.240925
## final  value 10.240925 
## converged
## initial  value 10.240925 
## iter   1 value 10.240925
## final  value 10.240925 
## converged
```




```
summary(fit1_AR)
```

```
##                    Length Class  Mode
## fit                14     Arima  list
## degrees_of_freedom  1     -none- numeric
## ttable             16     -none- numeric
## AIC                 1     -none- numeric
## AICc                1     -none- numeric
## BIC                 1     -none- numeric
```

```
fit1_AR
```

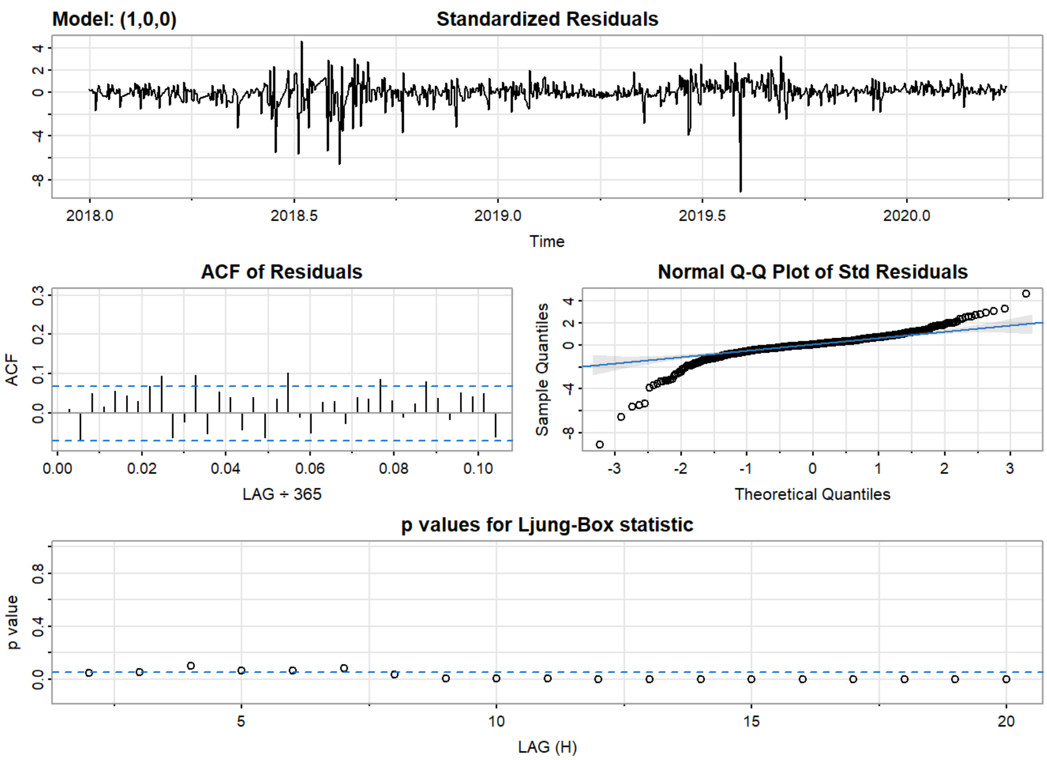
```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xreg, transform.pars = trans, fixed = fixed, optim.control = list(trace = trc,
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##         intercept  daily_wind_direction  daily_ambient_temperature
##       -134299.212               52.3642                  1492.8001
## s.e.     9668.273               14.6499                   305.0293
##       daily_wind_speed
##             26639.0337
## s.e.          462.6177
##
## sigma^2 estimated as 785515458:  log likelihood = -9572.75,  aic = 19155.5
##
## $degrees_of_freedom
## [1] 817
##
## $ttable
##                               Estimate        SE  t.value p.value
## intercept                 -134299.2116 9668.2732 -13.8907   0e+00
## daily_wind_direction           52.3642   14.6499   3.5744   4e-04
## daily_ambient_temperature    1492.8001  305.0293   4.8940   0e+00
## daily_wind_speed            26639.0337  462.6177  57.5833   0e+00
##
## $AIC
## [1] 23.33191
##
## $AICc
## [1] 23.33197
##
## $BIC
## [1] 23.3606
```

```
## ----Check p-value and drop insignificant, in this case, all explanatory variables are significant (Ho: Term is not needed in the model; Ha: Term is needed in the model. As p-value of all explanatory variables are smaller than 0.05, we reject null hypothesis, and state they are significant.)

#ACF some significant doesn't capture
#residuals are not white noise
#this is not a good model for forecasting
```

Model 2 - AR(1) with 3 Explanatory Variables (Ambient Termperature & Wind Direction & Wind Speed)

```
# According to ACF and PACF plots, ACF decays to zero, PACF drops quickly and have some significant values; start with AR(1) for ARIMA model, set p = 1
# to see significance of terms and whether residuals are white noises (residuals' plots)

fit2_AR3 <- sarima(turbine_ts, 1, 0, 0, xreg=turbine_clean[,2:4])
```

```
## initial  value 10.241524
## iter   2 value 10.058239
## iter   3 value 10.057796
## iter   4 value 10.054469
## iter   5 value 10.054400
## iter   6 value 10.054387
## iter   7 value 10.054379
## iter   8 value 10.054355
## iter   9 value 10.054340
## iter  10 value 10.054335
## iter  11 value 10.054334
## iter  11 value 10.054334
## iter  11 value 10.054334
## final  value 10.054334
## converged
## initial  value 10.054009
## iter   2 value 10.054008
## iter   2 value 10.054008
## iter   2 value 10.054008
## final  value 10.054008
## converged
```




```
summary(fit2_AR3)
```

```
##                    Length Class  Mode
## fit                14     Arima  list
## degrees_of_freedom  1     -none- numeric
## ttable             20     -none- numeric
## AIC                 1     -none- numeric
## AICc                1     -none- numeric
## BIC                 1     -none- numeric
```

```
fit2_AR3
```

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xreg, transform.pars = trans, fixed = fixed, optim.control = list(trace = trc,
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1   intercept  daily_wind_direction  daily_ambient_temperature
##       0.5683  -150182.37               17.7704                  2372.2506
## s.e.  0.0298    15902.18               21.0675                   501.7667
##       daily_wind_speed
##              26219.152
## s.e.           708.979
##
## sigma^2 estimated as 540249321:  log likelihood = -9419.29,  aic = 18850.58
##
## $degrees_of_freedom
## [1] 816
##
## $ttable
##                              Estimate         SE t.value p.value
## ar1                            0.5683     0.0298 19.1001  0.0000
## intercept                 -150182.3726 15902.1755 -9.4441  0.0000
## daily_wind_direction           17.7704    21.0675  0.8435  0.3992
## daily_ambient_temperature    2372.2506   501.7667  4.7278  0.0000
## daily_wind_speed            26219.1519   708.9790 36.9816  0.0000
##
## $AIC
## [1] 22.96051
##
## $AICc
## [1] 22.9606
##
## $BIC
## [1] 22.99493
```

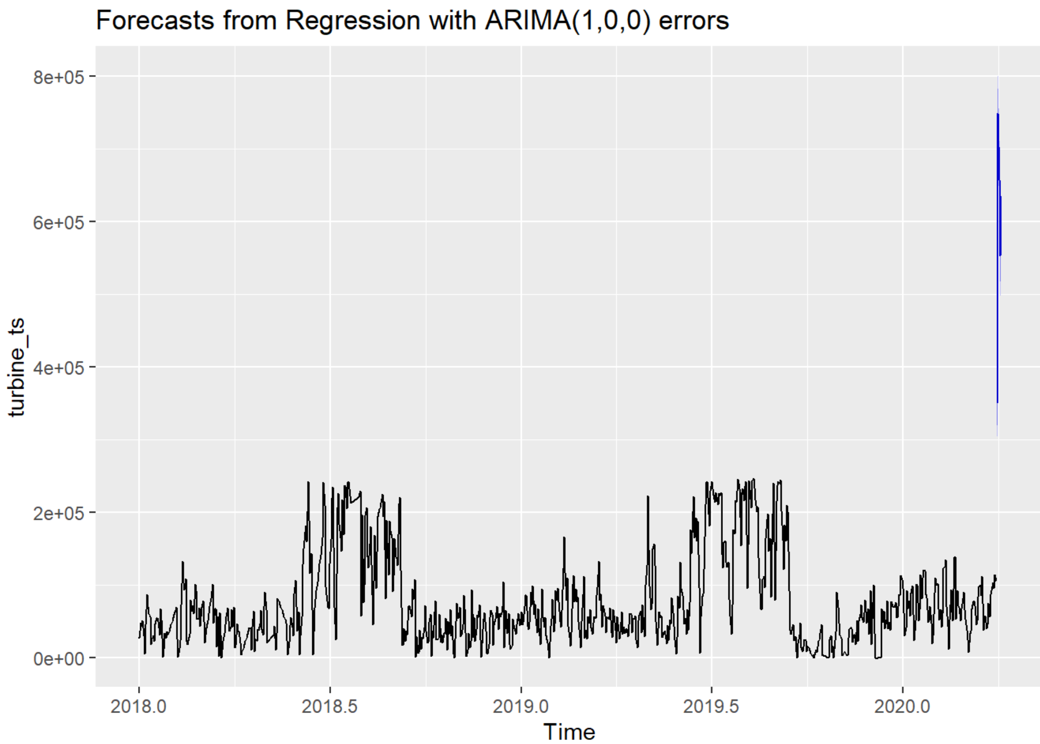
```
# Create matrix of covariates for next 5 time periods
xdat <- c(32.25, 142.25, 5.73, 35.59, 331.22, 4.03, 34.68,295.51, 3.88, 33.44,239.83, 5.01, 34.06,279.92, 4.51)
xdat1 <- matrix(xdat,nrow=5,ncol=3,byrow=TRUE)

xregmat = as.matrix(turbine_clean[,2:4])

# re-run model with Arima and produce forecast for next 5 time periods

fit2_v2 <- Arima(turbine_ts, order=c(1, 0, 0), xreg=xregmat) #only works for matrix

#Arima could forecast variable

forecast(fit2_v2, xreg = xdat1,#value for the next 5 time period
         h = 5)
```

```
## Warning in forecast.forecast_ARIMA(fit2_v2, xreg = xdat1, h = 5): xreg contains
## different column names from the xreg used in training. Please check that the
## regressors are in the same order.
```

```
##           Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2020.2466       349811.2 319932.7 379689.8 304115.9 395506.6
## 2020.2493       748517.2 714151.1 782883.3 695958.8 801075.6
## 2020.2521       656976.8 621281.8 692671.7 602386.0 711567.5
## 2020.2548       552859.8 516746.1 588973.4 497628.7 608090.8
## 2020.2575       633935.2 597687.3 670183.0 578498.9 689371.5
```

```
autoplot(forecast(fit2_v2, xreg=xdat1, h=5))
```

```
## Warning in forecast.forecast_ARIMA(fit2_v2, xreg = xdat1, h = 5): xreg contains
## different column names from the xreg used in training. Please check that the
## regressors are in the same order.
```




```
summary(fit2_v2)
```

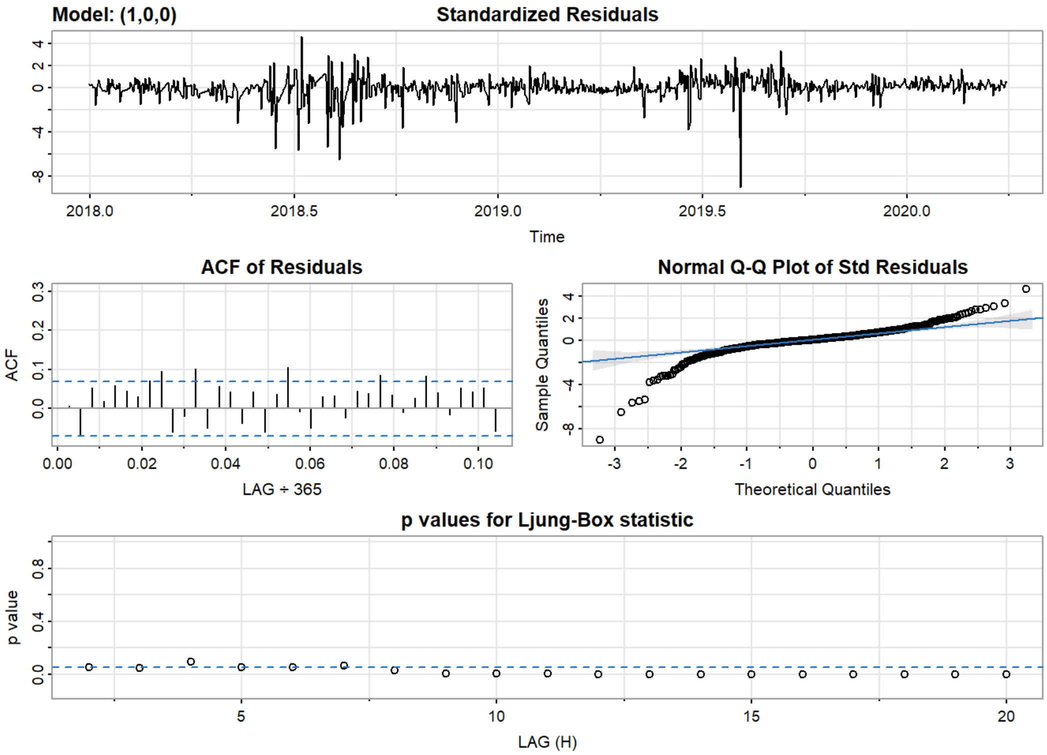
```
## Series: turbine_ts
## Regression with ARIMA(1,0,0) errors
##
## Coefficients:
##          ar1   intercept  daily_wind_direction  daily_ambient_temperature
##       0.5683  -150182.37               17.7704                  2372.2506
## s.e.  0.0298    15902.18               21.0675                   501.7667
##       daily_wind_speed
##              26219.152
## s.e.           708.979
##
## sigma^2 = 543559672:  log likelihood = -9419.29
## AIC=18850.58   AICc=18850.68   BIC=18878.84
##
## Training set error measures:
##                    ME     RMSE     MAE       MPE     MAPE      MASE        ACF1
## Training set 5.178305 23243.26 14476.7 -205.9281 365.3024 0.3503738 0.009068147
```

```
#daily_wind_direction is not significant (Ho: Term is not needed in the model; Ha: Term is needed in the model. A
s p-value of dailt wind direction is bigger than 0.05, we fail to reject null hypothesis, and state it is insigni
ficant.)
#ar term is significant (Ho: Term is not needed in the model; Ha: Term is needed in the model. As p-value of ar1
is smaller than 0.05, we reject null hypothesis, and state ar term is insignificant.)
#not too much significant in acf
#but not quite white noise (According to p-values for Ljung-Box statistic and ACF of Residuals, the residuals are
not white noise. H0: white noise; Ha: not white noise. i. Not all lags within the confidence bands, supporting th
at not all residuals are white noise. ii. Not all p-value points are above the alpha = 0.05 line, the residuals a
re not all white noises.)
#forecast plot pattern are too high the previous data, not reasonable pattern
#not the lowest forecast error
#this is not a good model for for Active Power generation for the turbine for the next 5 day period
```

Model 3 - AR(1) with 2 Explanatory Variables (Ambient Termperature & Wind Direction)

```
# Based on Model 2 above, daily_wind_direction is not significant, drop it
# According to ACF and PACF plots, ACF decays to zero, PACF drops quickly and have some significant values; start
with AR(1) for ARIMA model, set p = 1
# to see significance of terms and whether residuals are white noises (residuals' plots)

xreg_2 <- subset(turbine_clean, select=-c(daily_wind_direction))

fit3_AR3 <- sarima(turbine_ts, 1, 0, 0, xreg=xreg_2[,2:3])
```

```
## initial  value 10.249204
## iter   2 value 10.056662
## iter   3 value 10.056479
## iter   4 value 10.054813
## iter   5 value 10.054770
## iter   6 value 10.054765
## iter   7 value 10.054764
## iter   8 value 10.054758
## iter   9 value 10.054757
## iter  10 value 10.054756
## iter  10 value 10.054756
## final  value 10.054756
## converged
## initial  value 10.054455
## iter   2 value 10.054455
## iter   2 value 10.054455
## iter   2 value 10.054455
## final  value 10.054455
## converged
```




```
summary(fit3_AR3)
```

```
##                    Length Class  Mode
## fit                14     Arima  list
## degrees_of_freedom  1     -none- numeric
## ttable             16     -none- numeric
## AIC                 1     -none- numeric
## AICc                1     -none- numeric
## BIC                 1     -none- numeric
```

```
fit3_AR3
```

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xreg, transform.pars = trans, fixed = fixed, optim.control = list(trace = trc,
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1   intercept  daily_ambient_temperature  daily_wind_speed
##       0.5726  -147469.24                  2394.1466        26246.8905
## s.e.  0.0293    15645.32                   503.9683          712.0449
##
## sigma^2 estimated as 540727291:  log likelihood = -9419.66,  aic = 18849.31
##
## $degrees_of_freedom
## [1] 817
##
## $ttable
##                              Estimate         SE t.value p.value
## ar1                            0.5726     0.0293 19.5730       0
## intercept                 -147469.2442 15645.3158 -9.4258       0
## daily_ambient_temperature    2394.1466   503.9683  4.7506       0
## daily_wind_speed            26246.8905   712.0449 36.8613       0
##
## $AIC
## [1] 22.95897
##
## $AICc
## [1] 22.95903
##
## $BIC
## [1] 22.98765
```

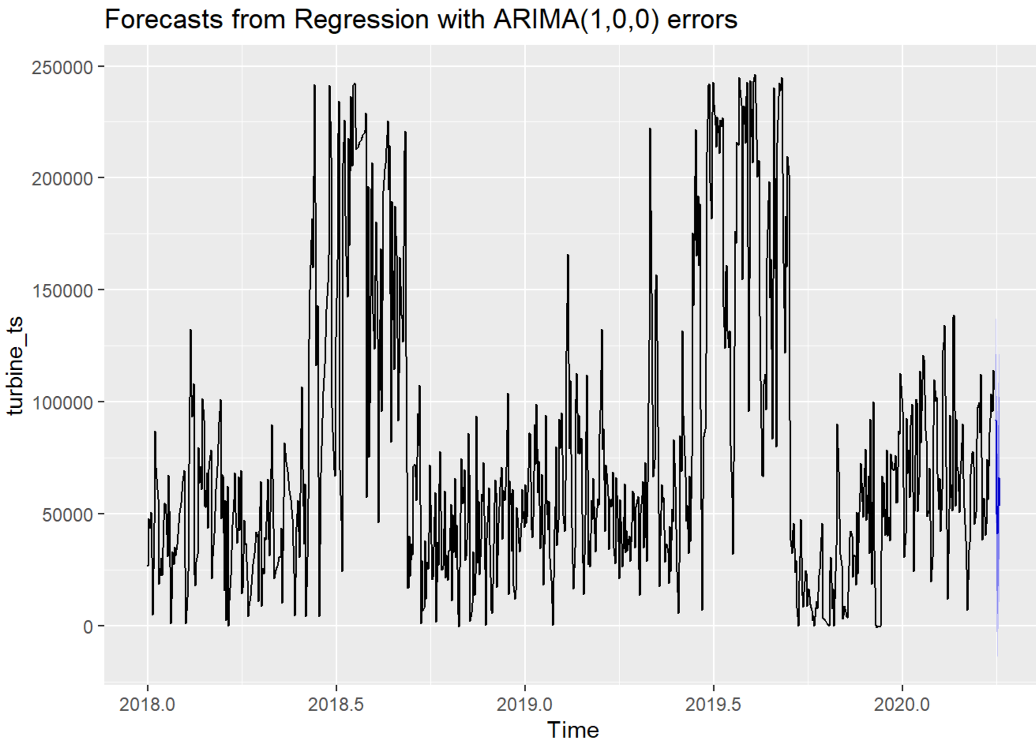
```
# Create matrix of covariates for next 5 time periods
xdat <- c(32.25, 5.73, 35.59, 4.03, 34.68, 3.88, 33.44, 5.01, 34.06, 4.51)
xdat1 <- matrix(xdat,nrow=5,ncol=2,byrow=TRUE)

xregmat = as.matrix(xreg_2[,2:3])

# re-run model with Arima and produce forecast for next 5 time period

fit3_v2 <- Arima(turbine_ts, order=c(1, 0, 0), xreg=xregmat) #only works for matrix
#Arima could forecast variable

forecast(fit3_v2, xreg = xdat1,#value for the next 5 time period
         h = 5)
```

```
## Warning in forecast.forecast_ARIMA(fit3_v2, xreg = xdat1, h = 5): xreg contains
## different column names from the xreg used in training. Please check that the
## regressors are in the same order.
```

```
##           Point Forecast     Lo 80     Hi 80      Lo 95     Hi 95
## 2020.2466       91851.29 61977.809 121724.76  46163.745 137538.83
## 2020.2493       50221.12 15797.027  84645.21  -2425.985 102868.22
## 2020.2521       41238.49  5448.147  77028.83 -13498.115  95975.09
## 2020.2548       66287.15 30060.082 102514.22  10882.629 121691.67
## 2020.2575       53708.12 17339.003  90077.24  -1913.644 109329.88
```

```
autoplot(forecast(fit3_v2, xreg=xdat1, h=5))
```

```
## Warning in forecast.forecast_ARIMA(fit3_v2, xreg = xdat1, h = 5): xreg contains
## different column names from the xreg used in training. Please check that the
## regressors are in the same order.
```




```
summary(fit3_v2)
```

```
## Series: turbine_ts
## Regression with ARIMA(1,0,0) errors
##
## Coefficients:
##          ar1   intercept  daily_ambient_temperature  daily_wind_speed
##       0.5726  -147469.24                  2394.1466        26246.8905
## s.e.  0.0293    15645.32                   503.9683          712.0449
##
## sigma^2 = 543374670:  log likelihood = -9419.66
## AIC=18849.31   AICc=18849.39   BIC=18872.86
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 6.685737 23253.54 14467.12 -179.5402 347.5104 0.3501419
##                     ACF1
## Training set 0.006574936
```

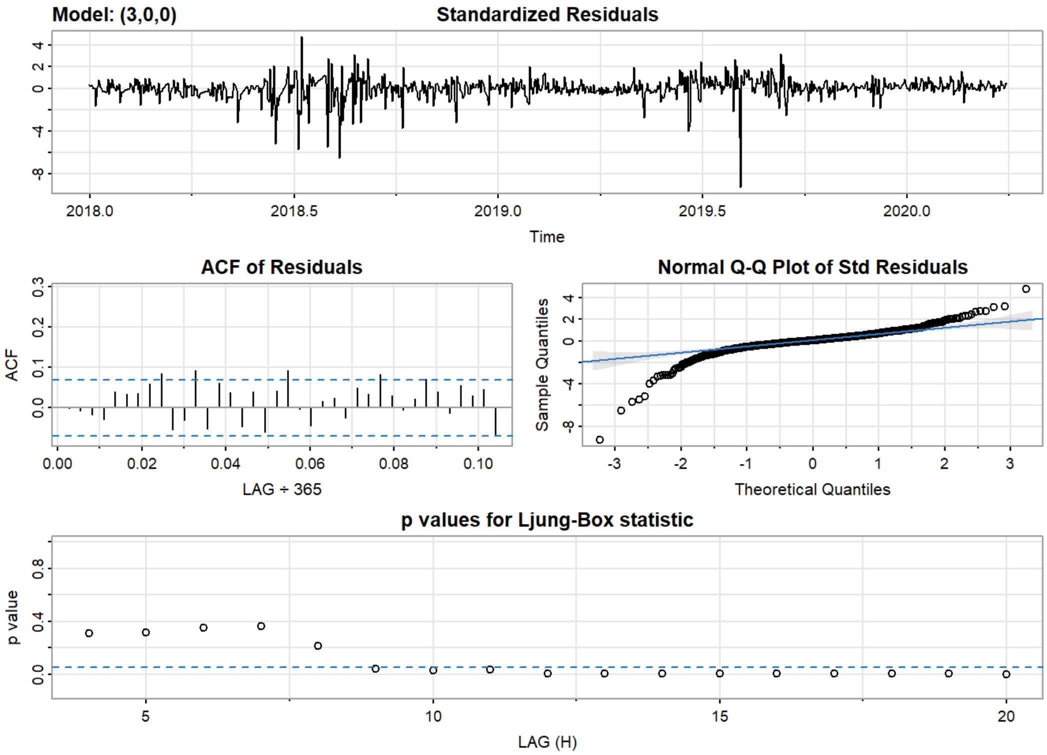
```
#all explanatory variables are significant (Ho: Term is not needed in the model; Ha: Term is needed in the model.
As p-value of all explanatory variables are smaller than 0.05, we reject null hypothesis, and state they are sign
ificant.)
#ar term is significant (Ho: Term is not needed in the model; Ha: Term is needed in the model. As p-value of ar1
is smaller than 0.05, we reject null hypothesis, and state ar term is insignificant.)
#not too much significant in acf
#but not quite white noise (According to p-values for Ljung-Box statistic and ACF of Residuals, the residuals are
not white noise. H0: white noise; Ha: not white noise. i. Not all lags within the confidence bands, supporting th
at not all residuals are white noise. ii. Not all p-value points are above the alpha = 0.05 line, the residuals a
re not all white noises.)
#not the lowest forecast error
#this is not a good model for for Active Power generation for the turbine for the next 5 day period
```

Model 4 - AR(3) with 2 Explanatory Variables (Ambient Termperature & Wind Direction)

—-This is our final model. This model has almost all AR terms are significant and all explanotry variables are significant. It also has the lowest forecast error (RMSE). Even if its residuals may not quite be white noise, those residuals are white noise at lower lag. In addition, its forecast pattern on forecast plot (forecast values) for all five forecast days is reasonable. This model is a simpler model. This is the best model for for Active Power generation for the turbine for the next 5 day period.

```
# According to ACF and PACF plots, ACF decays to zero, PACF drops quickly and have some significant values; try A
R(3) for ARIMA model, set p = 3
# to see significance of terms and whether residuals are white noises (residuals' plots)

fit4_AR1 <- sarima(turbine_ts, 3, 0, 0, xreg=xreg_2[,2:3])
```

```
## initial  value 10.250311
## iter   2 value 10.193930
## iter   3 value 10.061073
## iter   4 value 10.053602
## iter   5 value 10.051653
## iter   6 value 10.051458
## iter   7 value 10.051092
## iter   8 value 10.050555
## iter   9 value 10.050168
## iter  10 value 10.050055
## iter  11 value 10.050044
## iter  12 value 10.050043
## iter  13 value 10.050041
## iter  14 value 10.050037
## iter  15 value 10.050036
## iter  16 value 10.050036
## iter  17 value 10.050035
## iter  18 value 10.050035
## iter  19 value 10.050034
## iter  20 value 10.050033
## iter  20 value 10.050033
## iter  20 value 10.050033
## final  value 10.050033
## converged
## initial  value 10.048600
## iter   2 value 10.048599
## iter   3 value 10.048599
## iter   4 value 10.048599
## iter   5 value 10.048598
## iter   6 value 10.048598
## iter   7 value 10.048597
## iter   7 value 10.048597
## iter   7 value 10.048597
## final  value 10.048597
## converged
```




```
summary(fit4_AR1)
```

```
##                     Length Class  Mode
## fit                 14     Arima  list
## degrees_of_freedom   1     -none- numeric
## ttable              24     -none- numeric
## AIC                  1     -none- numeric
## AICc                 1     -none- numeric
## BIC                  1     -none- numeric
```

```
fit4_AR1
```

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xreg, transform.pars = trans, fixed = fixed, optim.control = list(trace = trc,
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ar2     ar3   intercept  daily_ambient_temperature
##       0.5803  -0.0719  0.1076  -150016.30                  2477.4677
## s.e.  0.0351   0.0401  0.0347    16525.82                   530.4599
##       daily_wind_speed
##             26279.5400
## s.e.          735.6432
##
## sigma^2 estimated as 534405647:  log likelihood = -9414.85,  aic = 18843.69
##
## $degrees_of_freedom
## [1] 815
##
## $ttable
##                              Estimate         SE t.value p.value
## ar1                            0.5803     0.0351 16.5496  0.0000
## ar2                           -0.0719     0.0401 -1.7945  0.0731
## ar3                            0.1076     0.0347  3.0965  0.0020
## intercept                 -150016.3018 16525.8209 -9.0777  0.0000
## daily_ambient_temperature    2477.4677   530.4599  4.6704  0.0000
## daily_wind_speed            26279.5400   735.6432 35.7232  0.0000
##
## $AIC
## [1] 22.95212
##
## $AICc
## [1] 22.95225
##
## $BIC
## [1] 22.99229
```

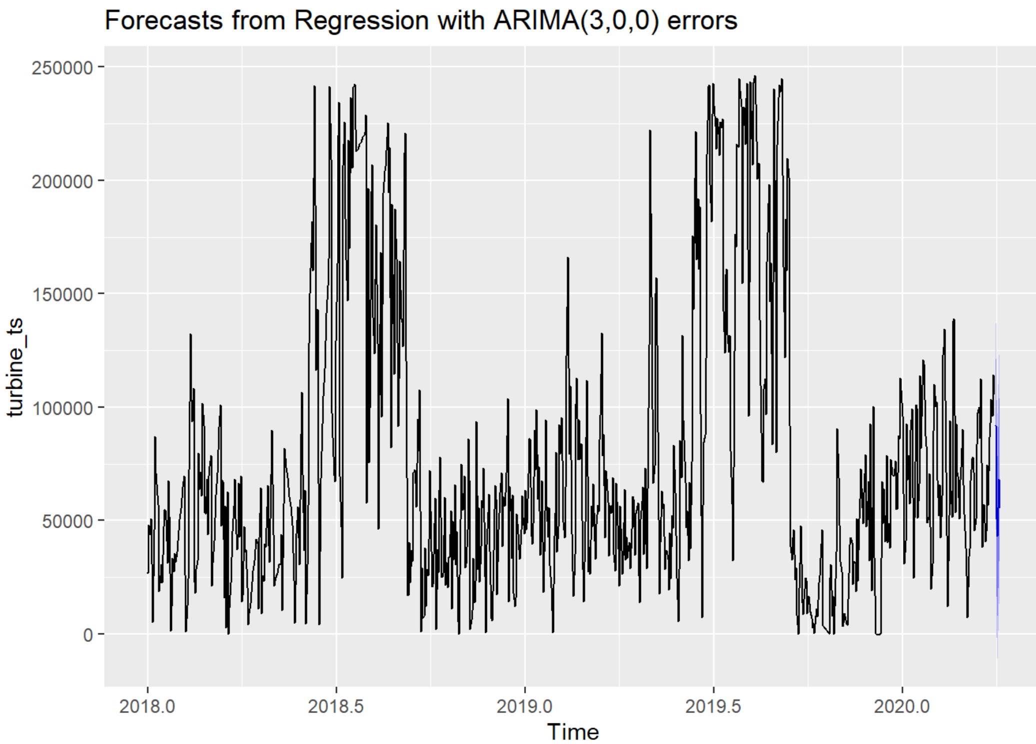
```
# Create matrix of covariates for next 5 time periods
xdat <- c(32.25, 5.73, 35.59, 4.03, 34.68, 3.88, 33.44, 5.01, 34.06, 4.51)
xdat1 <- matrix(xdat,nrow=5,ncol=2,byrow=TRUE)

xregmat = as.matrix(xreg_2[,2:3])

# re-run model with Arima and produce forecast for next 5 time periods

fit4_v2 <- Arima(turbine_ts, order=c(3, 0, 0), xreg=xregmat) #only works for matrix
#Arima could forecast variable

# Forecast value for the next 5 time period
forecast(fit4_v2, xreg = xdat1,
         h = 5)
```

```
## Warning in forecast.forecast_ARIMA(fit4_v2, xreg = xdat1, h = 5): xreg contains
## different column names from the xreg used in training. Please check that the
## regressors are in the same order.
```

```
##           Point Forecast     Lo 80     Hi 80        Lo 95     Hi 95
## 2020.2466       91774.85 62040.090 121509.60  46299.46112 137250.23
## 2020.2493       50918.45 16539.453  85297.45  -1659.68765 103496.59
## 2020.2521       43197.89  7928.455  78467.33 -10742.05700  97137.84
## 2020.2548       68307.84 32439.556 104176.13  13452.03038 123163.66
## 2020.2575       55455.97 19230.124  91681.81     53.32262 110858.61
```

```
autoplot(forecast(fit4_v2, xreg=xdat1, h=5))
```

```
## Warning in forecast.forecast_ARIMA(fit4_v2, xreg = xdat1, h = 5): xreg contains
## different column names from the xreg used in training. Please check that the
## regressors are in the same order.
```




```
summary(fit4_v2)
```

```
## Series: turbine_ts
## Regression with ARIMA(3,0,0) errors
##
## Coefficients:
##          ar1      ar2     ar3   intercept  daily_ambient_temperature
##       0.5803  -0.0719  0.1076  -150016.30                  2477.4677
## s.e.  0.0351   0.0401  0.0347    16525.82                   530.4599
##       daily_wind_speed
##             26279.5400
## s.e.          735.6432
##
## sigma^2 = 538339922:  log likelihood = -9414.85
## AIC=18843.69   AICc=18843.83   BIC=18876.67
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -5.660052 23117.22 14404.87 -164.9185 331.9818 0.3486352
##                      ACF1
## Training set -0.002239622
```

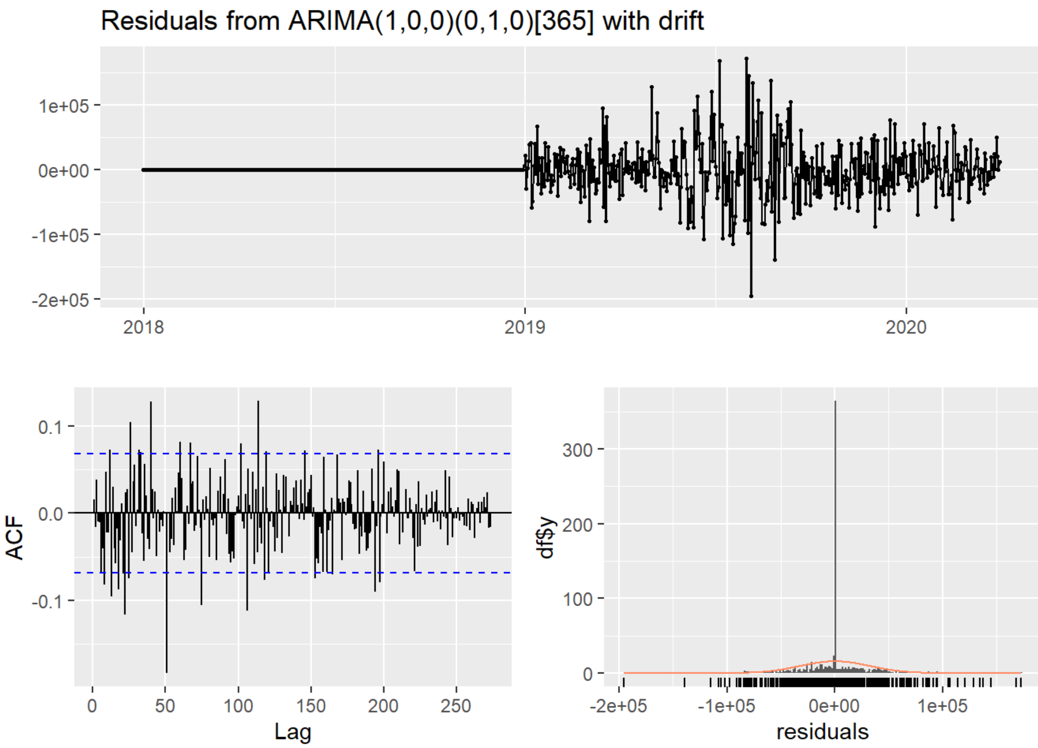
```
#almost all AR terms are significant (Ho: Term is not needed in the model; Ha: Term is needed in the model. As p-
value of ar1 & ar3 is smaller than 0.05, we reject null hypothesis, and state ar1 & ar3 are significant; ----only
one ar term (ar2) is not significant (Ho: Term is not needed in the model; Ha: Term is needed in the model. As p-
value of ar2 is bigger than 0.05, we fail to reject null hypothesis, and state all AR (2) term is insignificant.)
#all explanatory variables are significant (Ho: Term is not needed in the model; Ha: Term is needed in the model.
As p-value of all explanatory variables are smaller than 0.05, we reject null hypothesis, and state they are sign
ificant)
#has the lowest forecast error (RMSE)
#residual may not quite be white noise, but white noise at lower lag (According to p-values for Ljung-Box statist
ic and ACF of Residuals, the lower lag residuals are white noise. H0: white noise; Ha: not white noise. i. All lo
wer lags within the confidence bands, supporting that all lower lags of residuals are white noise. ii. All p-valu
e points of lower lags are above the alpha = 0.05 line, the lower lag residuals are white noises.)
#pattern on forecast plot (forecast values) for all five forecast days is reasonable
#this is the best model for for Active Power generation for the turbine for the next 5 day period
```

Type 3. Compare to auto.arima

```
fit_auto <- auto.arima(turbine_ts)
summary(fit_auto)
```

```
## Series: turbine_ts
## ARIMA(1,0,0)(0,1,0)[365] with drift
##
## Coefficients:
##          ar1    drift
##       0.6539  24.1714
## s.e.  0.0353  15.9714
##
## sigma^2 = 1.88e+09:  log likelihood = -5515.09
## AIC=11036.17   AICc=11036.22   BIC=11048.54
##
## Training set error measures:
##                    ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
## Training set 48.42566 32238.71 17559.27 165.3137 1116.252 0.4249799 0.01603581
```

```
checkresiduals(fit_auto)
```



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0)(0,1,0)[365] with drift
## Q* = 321.4, df = 163, p-value = 1.913e-12
##
## Model df: 1.   Total lags used: 164
```

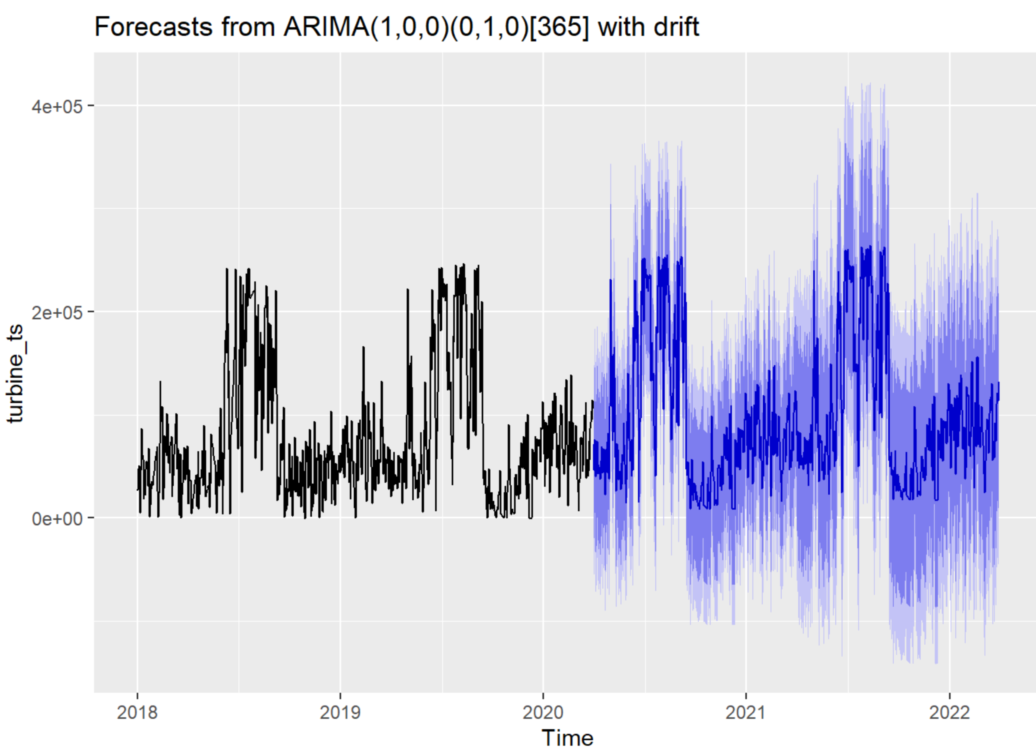
```
accuracy(fit_auto)
```

```
##                    ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
## Training set 48.42566 32238.71 17559.27 165.3137 1116.252 0.4249799 0.01603581
```

```
forecast(fit_auto, h=5)
```

```
##           Point Forecast       Lo 80    Hi 80     Lo 95    Hi 95
## 2020.2466       71478.27  15918.895 127037.6 -13492.46 156449.0
## 2020.2493       46402.82 -19981.499 112787.1 -55123.24 147928.9
## 2020.2521       75894.74   5386.817 146402.7 -31937.82 183727.3
## 2020.2548       52838.05 -19361.407 125037.5 -57581.49 163257.6
## 2020.2575       39816.47 -33094.365 112727.3 -71691.03 151324.0
```

```
fit_auto %>% forecast() %>% autoplot()
```




```
#relatively high forecast errors (RMSE)
#residuals are not white noises (According to p-values for Ljung-Box statistic of Residuals, residuals are white noise. H0: white noise; Ha: not white noise. As p-value is smaller than 0.05, we reject null hypothesis, and stat e residuals are not white noise.)
#seasonality pattern on forecast plot (forecast values) for all five forecast days
#this is not a good model for for Active Power generation for the turbine for the next 5 day period
```