FORECASTING WIND ENERGY

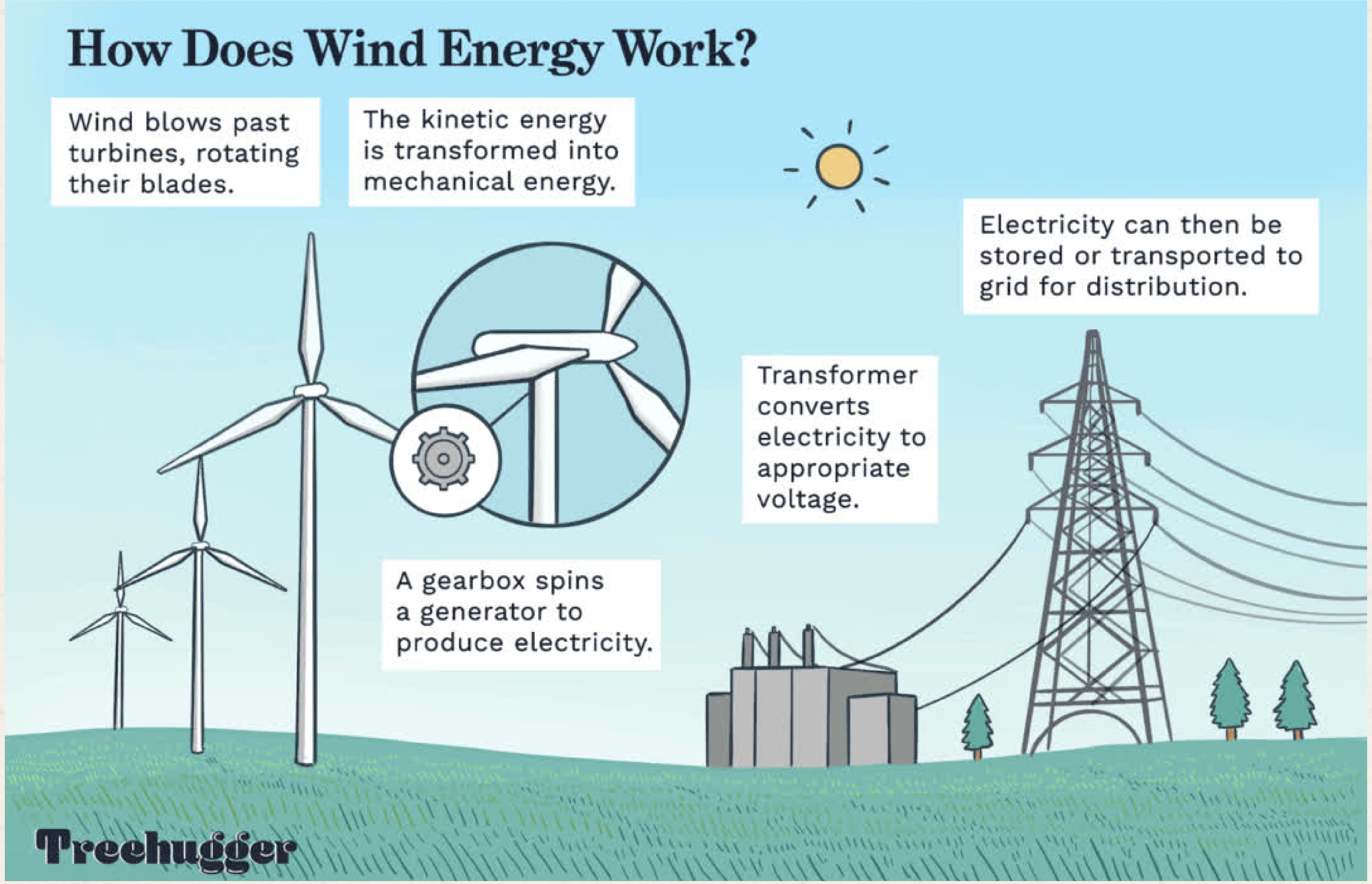
Wind energy is an important source of renewable energy production. Machine learning techniques can help us forecast, predict and plan the supply and demand of available wind energy.

- Husain Miyala

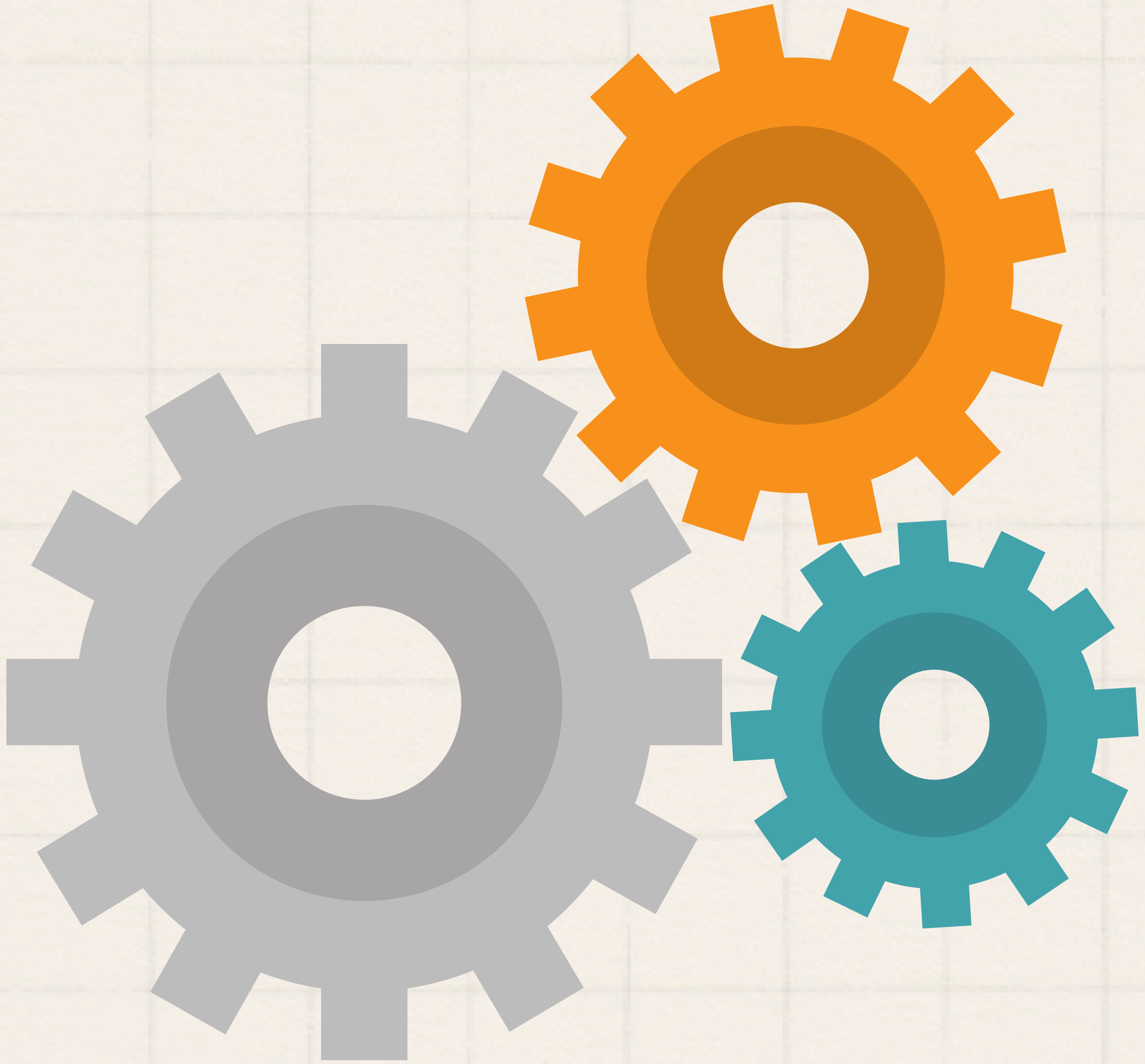
INDEX

01. INTRODUCTION

1. Wind energy is produced by the movement of air (wind) and converted into electricity.
2. Wind is a clean source of fuel. Turbines have no emissions and do not pollute the air.
3. Wind Energy constituted 7.32% of the total share of electricity production globally in 2022.
4. China, USA, Germany, India & Spain were the top 5 countries with installed wind turbine capacity in 2022.
5. Wind power capacity is forecasted to increase by 13% YoY between 2023 - 2030.



Machine Learning Applications in Wind Energy



Weather Forecasting & Wind Analysis

Predict Turbine Maintenance

Optimizing Wind System Production

Forecasting Wind Power Usage & Schedule

02. DATASET & HYPOTHESIS

Dataset: Wind Power Generation Data - Forecasting (Kaggle)

KEY ASSUMPTIONS

#1 Maximum power output of wind turbine is not controlled, i.e. there is no cap on power production.

#2 Technical factors such as blade angles, generator RPM, gearbox temperature, etc do not influence power production.

#3 Since data is collected hourly, it is assumed that variable values are measured at the end of each hour, i.e. there is no fluctuation by the minute.

Hypothesis: Winter months will produce the highest amount of wind power compared to other seasons.

03. REFINING DATA

A Understanding the Data

	power	temperature_2m	windspeed_100m	winddirection_100m
count	43800.000000	43800.000000	43800.000000	43800.000000
mean	0.405385	47.862911	6.284431	203.343676
std	0.288322	19.453691	2.685216	97.959852
min	0.000000	-14.400000	0.100000	0.000000
25%	0.148900	32.100000	4.380000	130.000000
50%	0.347650	47.300000	6.080000	226.000000
75%	0.659600	64.500000	7.990000	278.000000
max	0.991300	94.100000	20.650000	360.000000

```
#Check row count - 5 years of hourly data
total_hours = 24*365*5
total_hours == len(df1)
```

```
True
```

1. Power column has been normalized between 0 - 1, with 0 being no power production and 1 being max power production.
2. Based on the min/max of the temperature column, the units should be in Fahrenheit.
3. Similarly, wind direction units are in degrees.
4. Finally, wind speed has been decribed in the dataset as m/s.
5. Data contains 5 years of hourly data - verified row count.
6. Data contains no nulls or duplicates.

03. REFINING DATA

B Transforming Data

```
df1 = df1.astype({'time':'datetime64[ns]'})
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 43800 entries, 0 to 43799
Data columns (total 10 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   time                  43800 non-null  datetime64[ns]
 1   temperature_2m        43800 non-null  float64
 2   relativehumidity_2m   43800 non-null  int64
 3   dewpoint_2m           43800 non-null  float64
 4   windspeed_10m         43800 non-null  float64
 5   windspeed_100m        43800 non-null  float64
 6   winddirection_10m     43800 non-null  int64
 7   winddirection_100m    43800 non-null  int64
 8   windgusts_10m         43800 non-null  float64
 9   power                 43800 non-null  float64
dtypes: datetime64[ns](1), float64(6), int64(3)
```

```
df1 = df1.rename(columns={'Time':'time','Power':'power'})
df1
```

1. Time column data type was changed to datetime64[ns] format for further transformations.
2. Columns names converted to lower case for ease of coding.

03. REFINING DATA

C Pre-Processing

speed_100m	winddirection_10m	winddirection_100m	windgusts_10m	power	hour	dayofweek	month	year	is_summer	is_fall	is_winter	is_spring	timeofday
1.26	146	162	1.4	0.1635	0	0	1	2017	0	0	1	0	00:00-04:00

```
conditions = [
    (df1['hour']>=0) & (df1['hour']<5),(df1['hour']>4) & (df1['hour']<9),(df1['hour']>8) & (df1['hour']<13),
    (df1['hour']>12) & (df1['hour']<17),(df1['hour']>16) & (df1['hour']<21),(df1['hour']>20) & (df1['hour']<=23)]

values = ['00:00-04:00','04:00-08:00','08:00-12:00','12:00-16:00','16:00-20:00','20:00-23:00']
df1['timeofday'] = np.select(conditions, values)
```

1. Time elements such as ‘hour’, ‘day of week’, ‘month’ and ‘year’ were added to compare wind power output with respect to time.
2. One-hot encoding was used to create seasonal indicators such as Summer, Winter, etc.
3. Time of day column indicates what window of time the particular row falls into, such as 12am - 4am, 4am - 8am, etc.

03. REFINING DATA

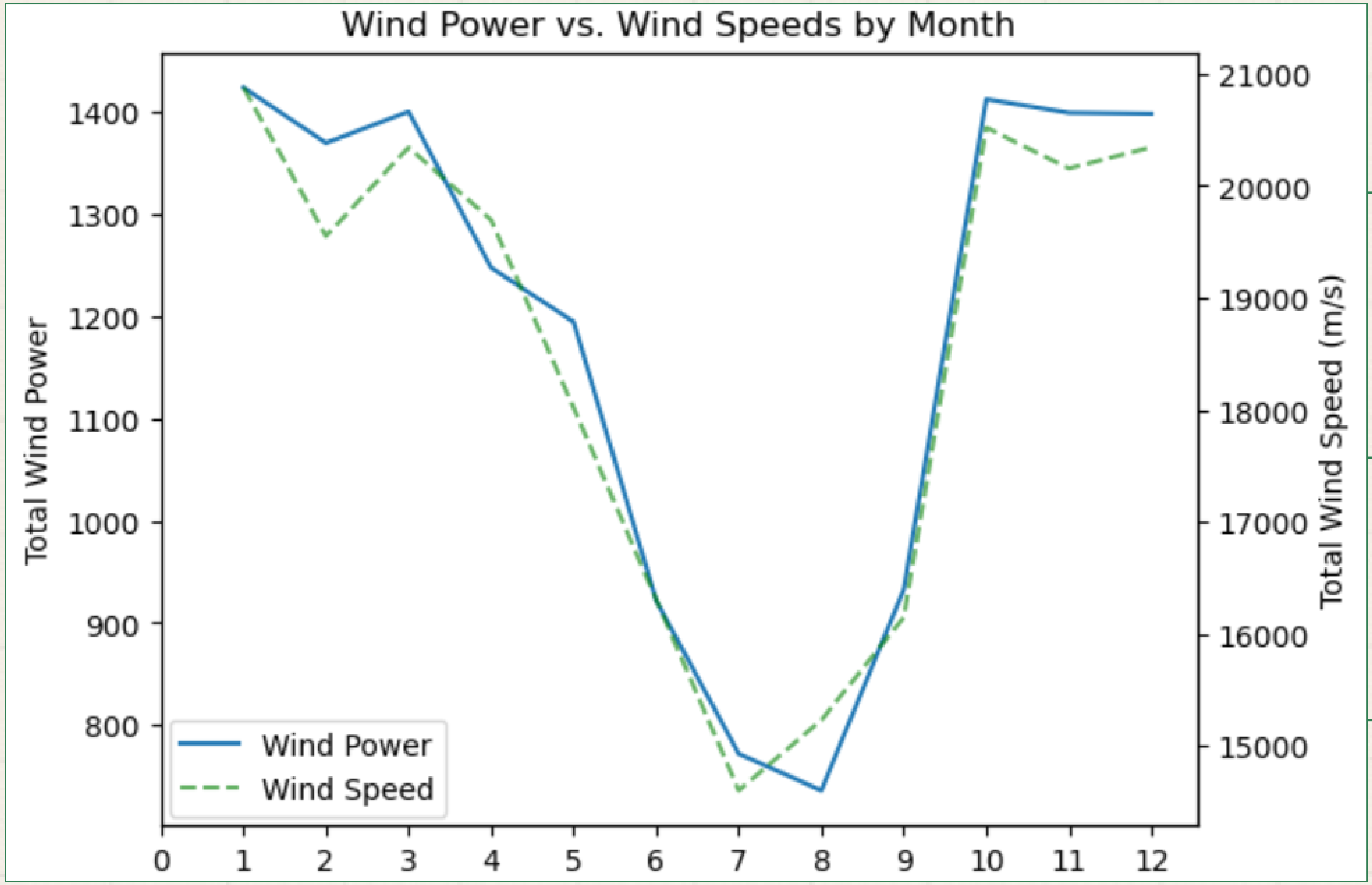
D Grouping Data

```
weekday_grouped = loc1_train.groupby('dayofweek')['power'].agg('sum')
```

```
dayofweek
0     2016.1529
1     2073.9297
2     2139.3028
3     2065.7054
4     1991.8886
5     1936.5821
6     1981.9246
Name: power, dtype: float64
```

1. Data was grouped by each time element, and aggregated on power output.
2. The resulting panda series were used in plotting and understanding the relationship between time and power output.

04. DATA ANALYSIS

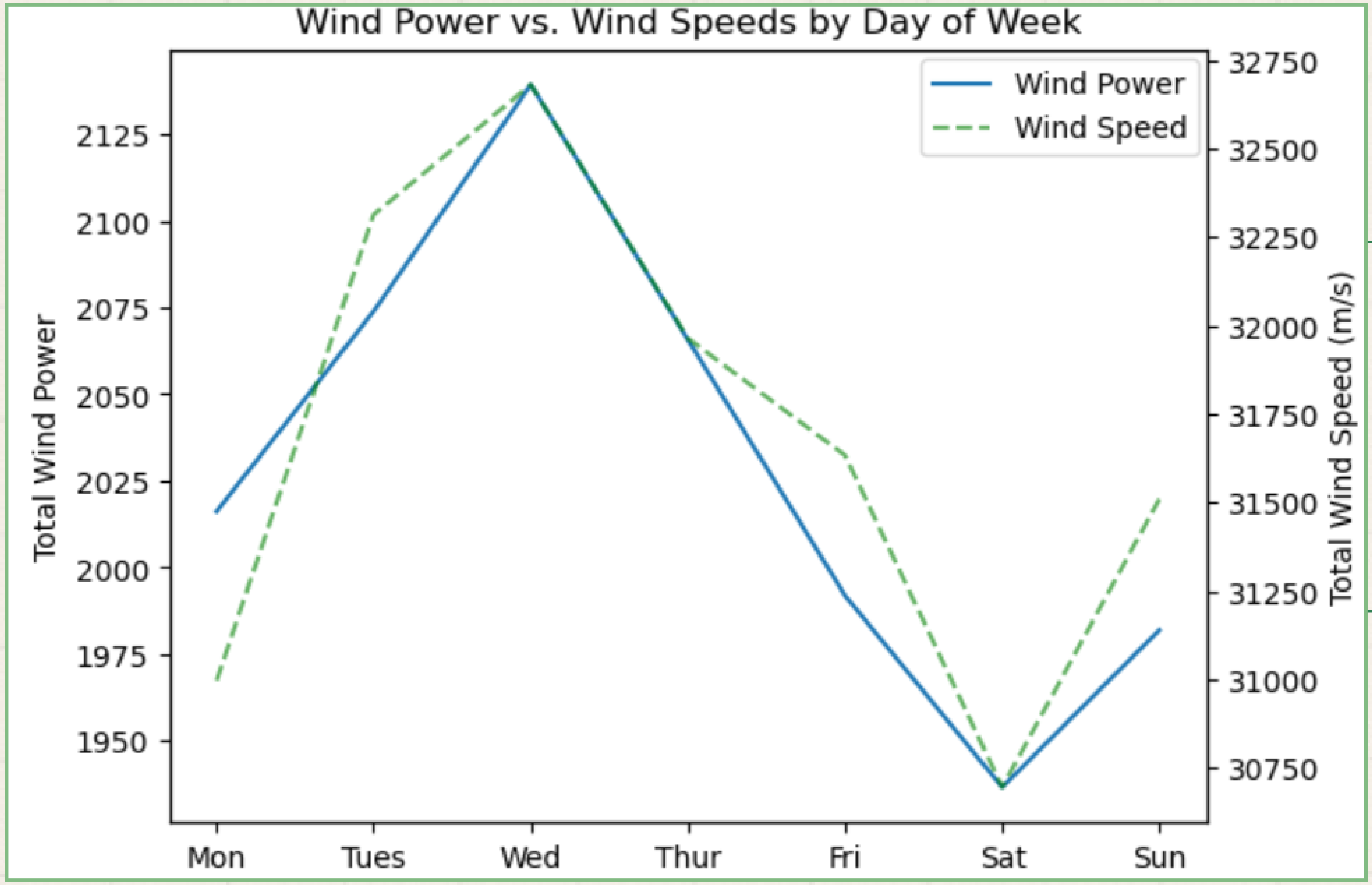


- Wind power is closely correlated to the wind speed.

- Power produced is highest in winter/spring months, which is consistent with average wind turbine performance.

- Cold air has a higher air density and produces more energy at the same wind speed as warm air.

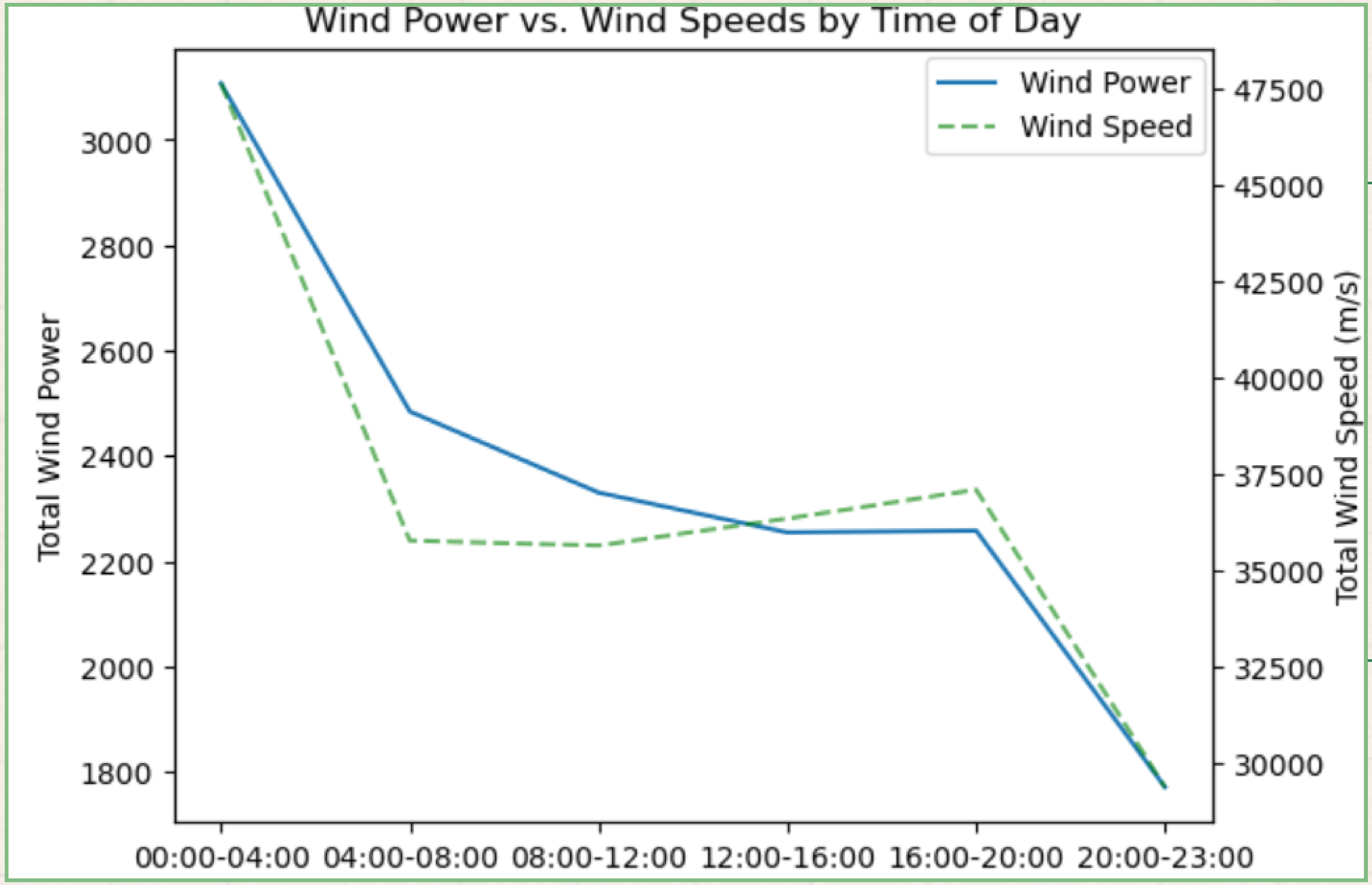
04. DATA ANALYSIS



- Turbines peak performance is during mid-week, and poorest performance is during weekends. This beahviour matches the wind speed patterns of the week.

- Alternatively, poor performance on weekends could be attributed to maintenance work, and peak performance mid-week can be attributed to optimal working conditions post-maintenance.

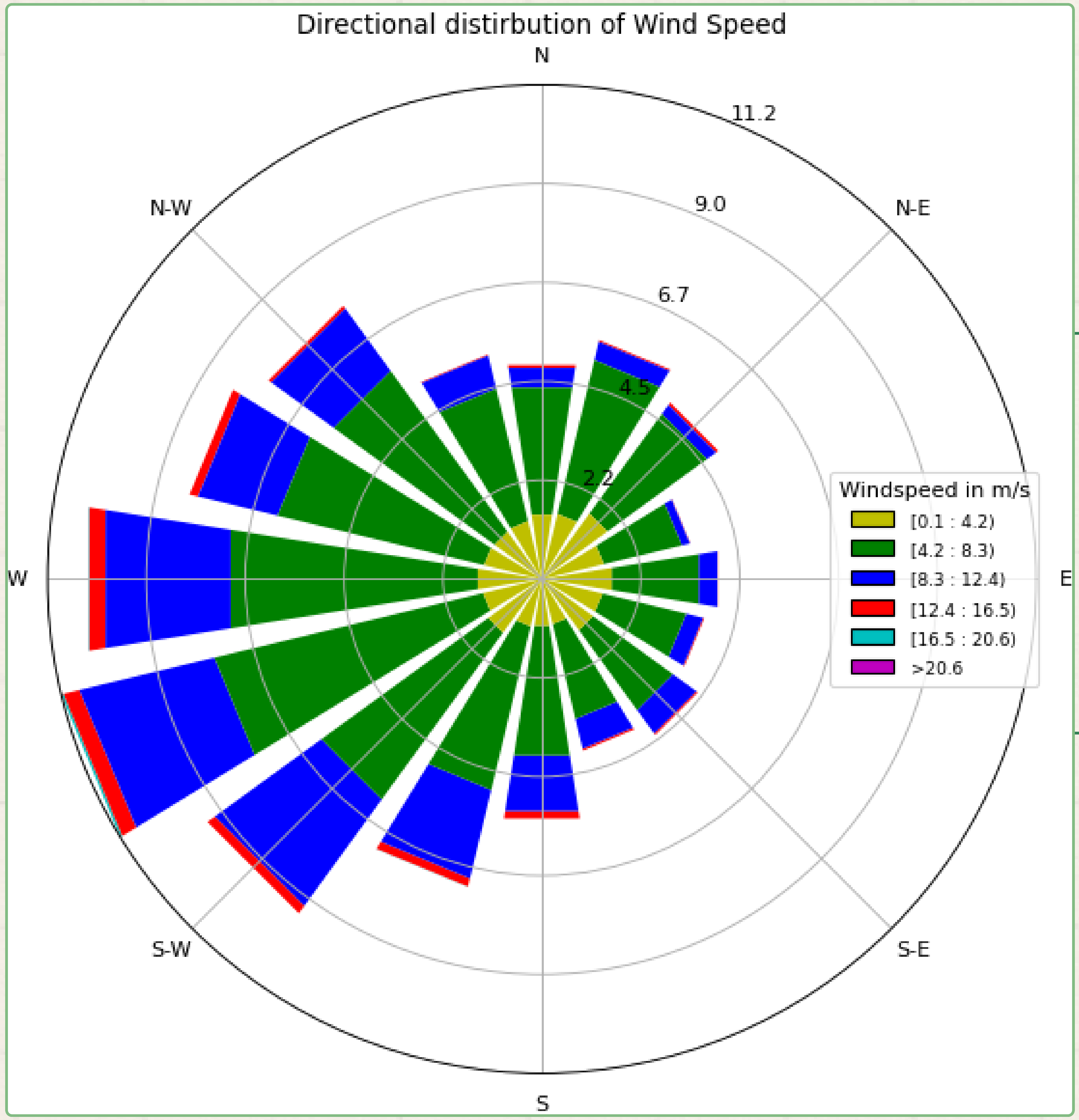
04. DATA ANALYSIS



- Wind power is highest during midnight hours (00:00-04:00), and lower during the daytime. Generally, daytime winds are stronger vs nighttime at ground level. However an interesting phenomenon occurs at higher altitudes where the turbine sits.

- The lowest layer of air closest to the ground is susceptible to high turbulence due to solar heating. The turbulence causes slow-moving air at ground level to rise up during the day, hence reducing wind power generation. In the absence of solar heating at night, slow-moving air is not pushed up, and hence turbines experience stronger winds.

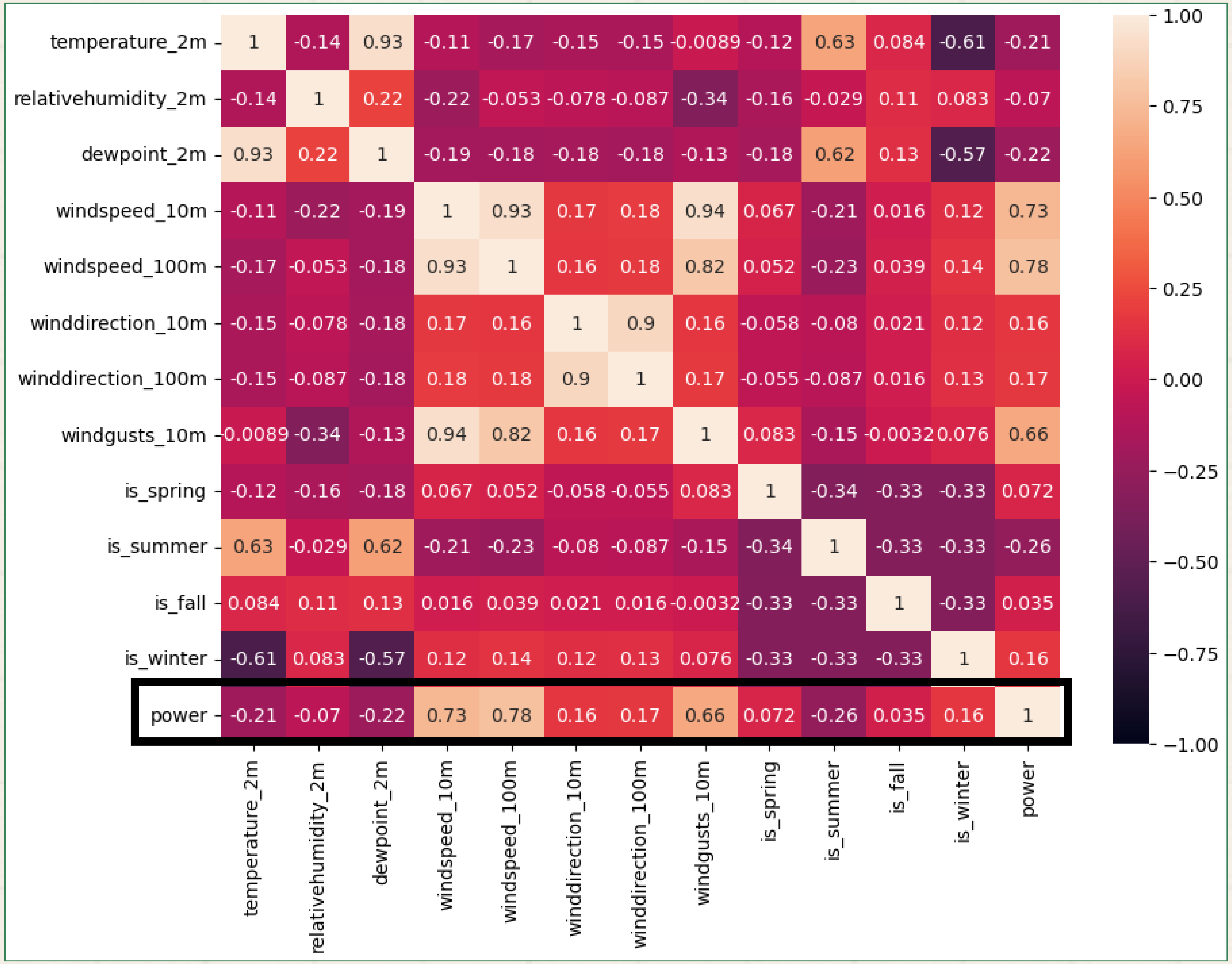
04. DATA ANALYSIS

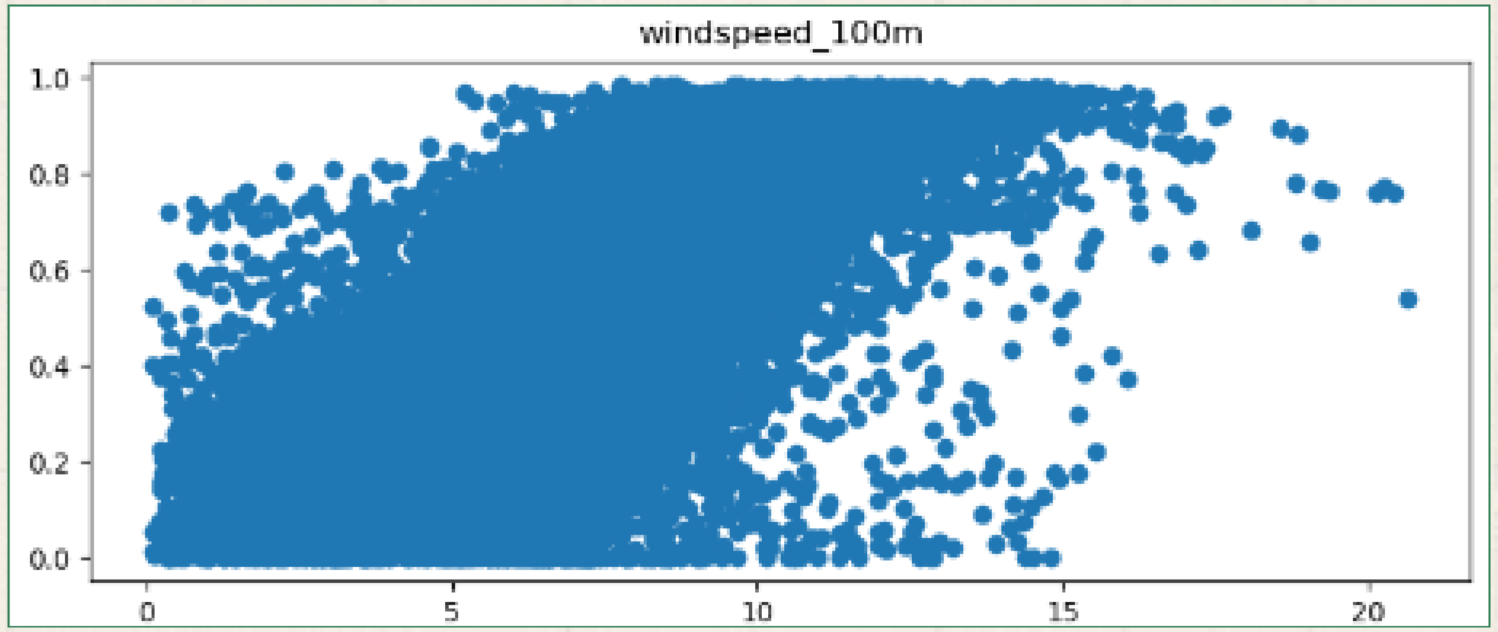


- West to South-West produces the highest distribution of winds as well as the strongest winds.

- For optimal production of wind power, the wind turbine should ideally face West to South-West directions.

05. MODEL & VALIDATION





05. MODEL & VALIDATION

Model = Statsmodel - Multi-Linear Regression

Features = Windspeed_100m, Windspeed_10m, Windgusts_10m, Temperature_2m, winddirection_100m, is_spring, is_summer, is_fall, is_winter

Target Variable = Wind Power

Training/Test Data= 2017-2020/2021

Dep. Variable:	power	**R-squared:**	0.617
Model:	OLS	**Adj. R-squared:**	0.617
Method:	Least Squares	**F-statistic:**	7051.
Date:	Wed, 07 Feb 2024	**Prob (F-statistic):**	0.00
Time:	11:14:19	**Log-Likelihood:**	11001.
No. Observations:	35040	**AIC:**	-2.198e+04
Df Residuals:	35031	**BIC:**	-2.191e+04
Df Model:	8		
Covariance Type:	nonrobust		

	coef
Intercept	-0.0779
windspeed_100m	0.0805
windspeed_10m	-0.0323
windgusts_10m	0.0159
temperature_2m	-0.0007
winddirection_100m	8.715e-05
is_spring	-0.0133
is_summer	-0.0481
is_fall	-0.0076
is_winter	-0.0088

- R-squared value of ~0.62 indicates a decent measure of fit for the model, however model can be improved.
- Coefficients of interest here are:
  - Windspeed_100m: A one m/s increase in windspeed at 100m will increase wind power by 0.0805 units or **8%** (keep in mind power is in range 0-1 units).
  - is_summer: Increase in summer months lead to a decline in wind power by 0.0481 units or **-4.8%**.

05. MODEL & VALIDATION

Model: ScikitLearn - Multi-Linear Regression

Features = Windspeed_100m, Windspeed_10m, Windgusts_10m, Temperature_2m, winddirection_100m, is_spring, is_summer, is_fall, is_winter

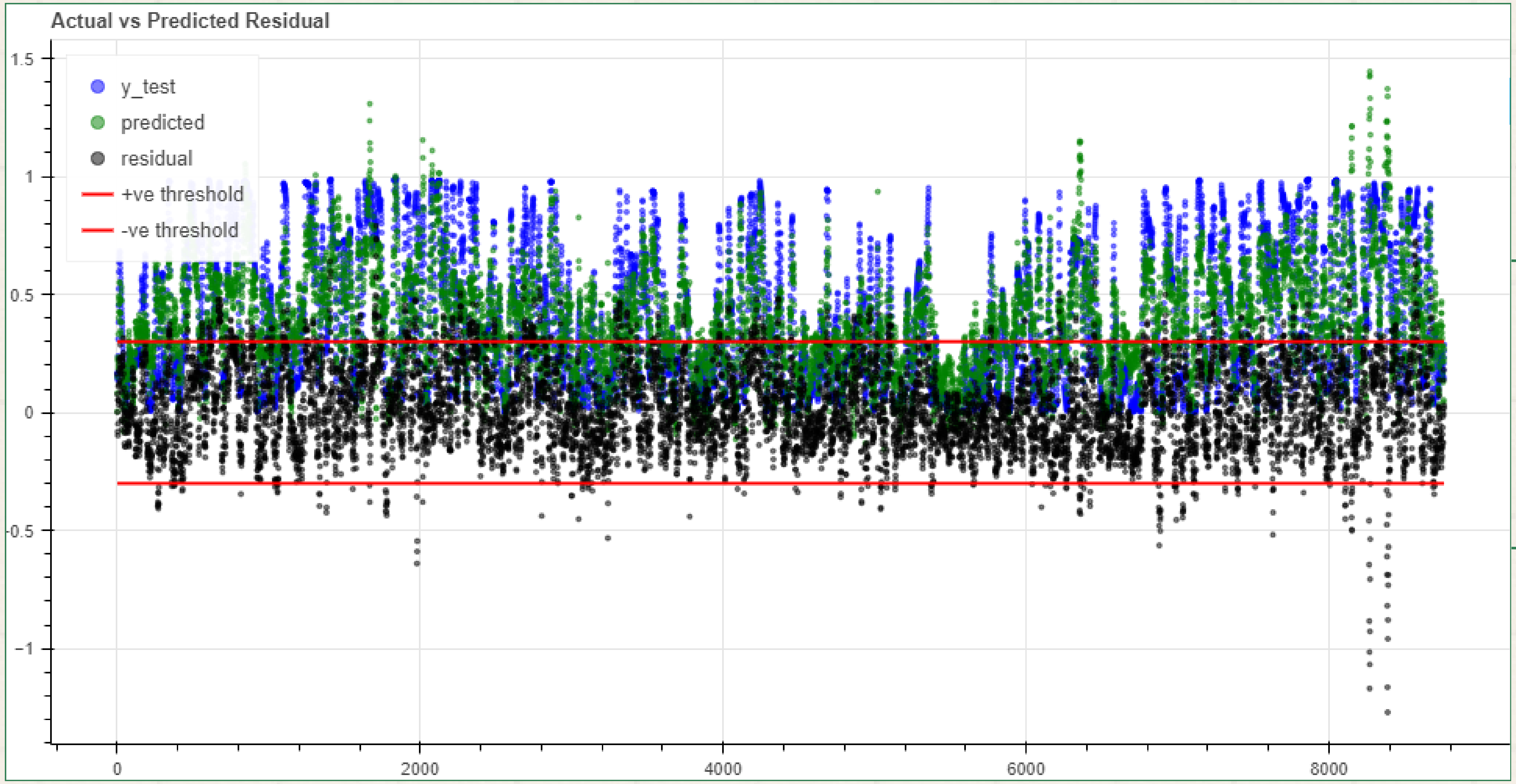
Target Variable = Wind Power

Training/Test Data = 2017-2020/2021

Model	MAE	RMSE
Statsmodel Multi-Linear Regression	0.148694	0.032288
Scikit Learn Multi-Linear Regression	0.144918	0.181866

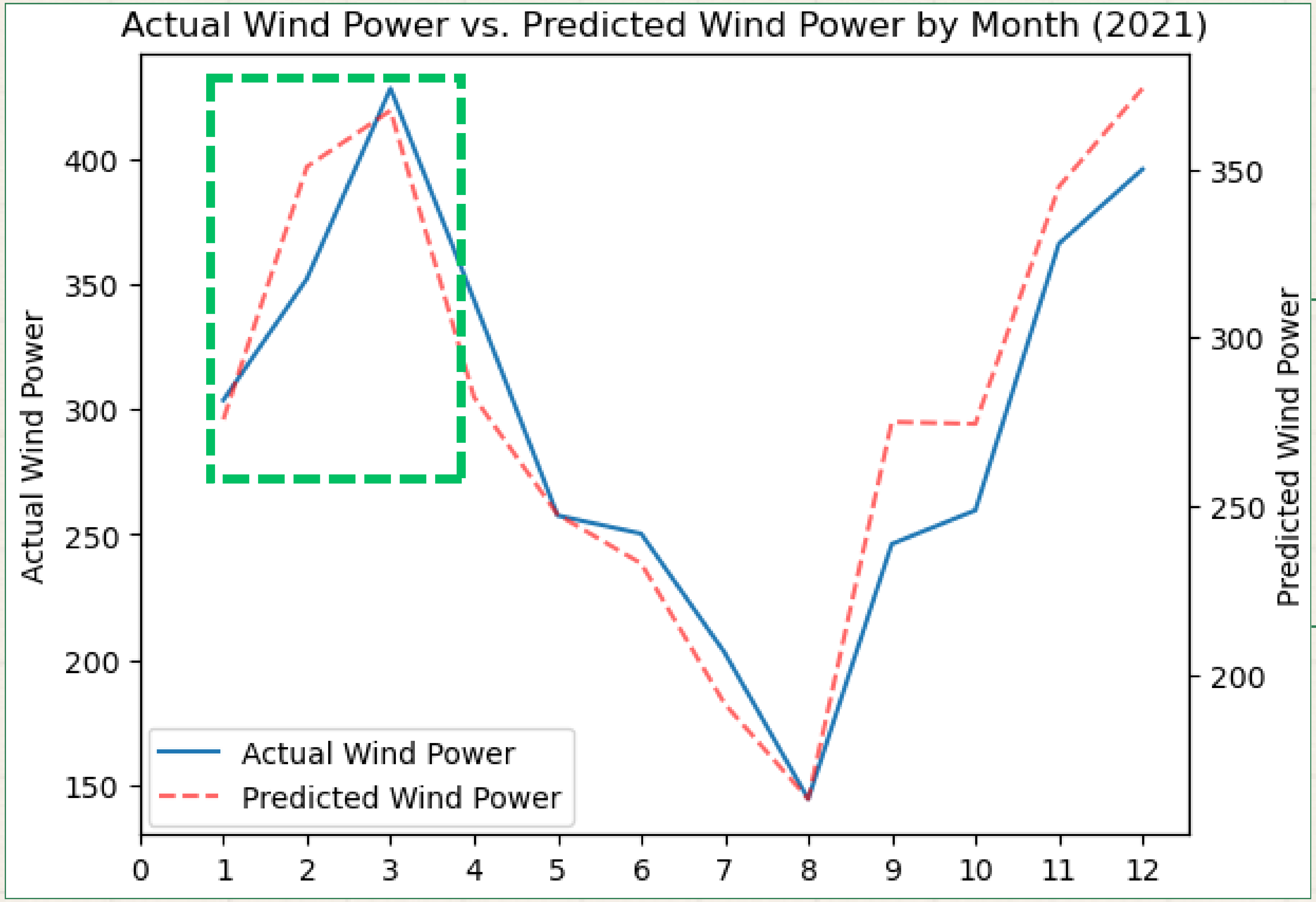
- Based on the selected evaluation metrics, both models seem to perform well with a low MAE and RMSE score.
- Overall, Statsmodel seems to perform better, however, with the addition of more features, both models only slightly improve in their evaluation scores.

06. INSIGHTS



- Predicted values of wind power overlap well over the actual values of wind power.
- Error rate (actual - predicted) falls between 30% to -30%, which is a large error rate.

06. INSIGHTS



- Hypothesis was to prove that winter months would produce the highest amount of wind power in 2021 (our test data).

- **We can confirm from the visualization that the predicted wind power peaks during the winter months of January, February and March.**

THANK YOU!