4/29/2020

# The United States Wind Turbine Analysis

Raghav Saxena

# 1. About:

The United States Wind Turbine Database (USWTDB) provides the locations of land-based and offshore wind turbines in the United States, corresponding wind project information, and turbine technical specifications.

This Report aims at pointing out the findings of a thorough analysis of the USWTDB for a better understanding of the Wind Turbines market in the US.

Since the information on individual turbines cannot be extracted the analysis is focused on the Turbine state and project year.

# 2. Cleaning Data:

- The data frame contains 27 columns of which 5 columns are unique identifiers of the data and one of the column have over 90% of missing values, so these columns will be dropped from our initial analysis.

```
% of missing values in each column:

case_id           0.000000
faa_ors           8.179867
faa_asn           7.233646
usgs_pr_id       45.962321
eia_id            8.181279
t_state           0.000000
t_county          0.000000
t_fips            0.000000
p_name            0.000000
p_year            0.865721
p_tnum            0.000000
p_cap             6.329793
t_manu            7.965202
t_model           8.158683
t_cap             7.739239
t_hh              8.727827
t_rd              8.380409
t_rsa             8.380409
t_ttlh            8.727827
retrofit          0.000000
retrofit_year    91.546153
t_conf_atr        0.000000
t_conf_loc        0.000000
t_img_date       11.744436
t_img_srce        0.000000
xlong             0.000000
ylat              0.000000
dtype: float64
```

*Figure 1: % Missing Values in Each column*

- The columns p_cap, t_cap, t_hh, t_rd, t_rsa, t_ttlh are highly correlated which other and may contribute a lot of information towards the turbine characteristics. The 8% of missing values are filled based on the columns mean.

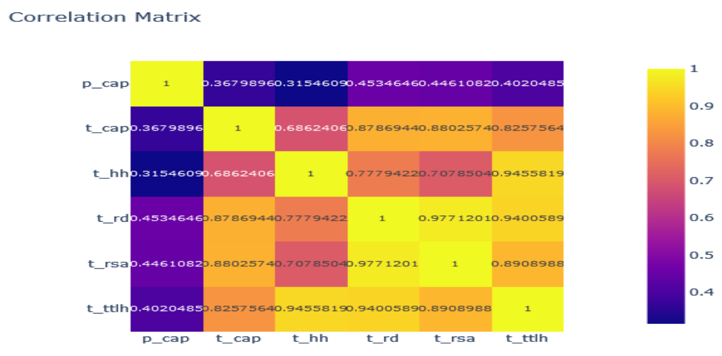



*Figure 2: Correlation Matrix*

Filling values based on the mean does not deteriorate the correlation between the columns. The correlation matrix reveals some important insights:

- The Turbine rated capacity(t_cap) is highly correlated with t_rd, t_rsa, t_ttlh.
- The Turbine hub heigh (t_hh) is highly correlated with Turbine total height from ground to tip of a blade (t_ttlh).
- Turbine rotor diameter(t_rd) is highly correlated to Turbine rotor swept area(t_rsa) and Turbine total height from ground to tip of a blade (t_ttlh).
- Turbine rotor swept area(t_rsa) is correlated with Turbine rotor diameter (t_rd) and turbine total height from ground to the tip of a blade (t_ttlh).
- Another Interesting insight is between the project capacity and the turbine rated capacity. Even though the turbine rated capacity depends more on the rotor swept area than the rotor diameter the reverse is true for a real-world situation.

- Number of Turbines Built In A Given Year:

Based on the visualization below we can see that there were a very limited number of turbines before the year 2005 and hence the total project capacity and the total rated capacity were very less.

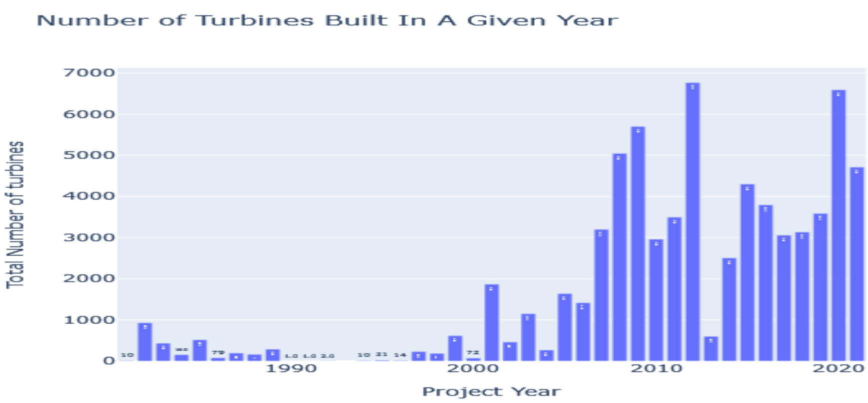



*Figure 3: Number of Turbines Built In A Given Year.*

A detailed analysis of the turbines and capacities between the years before and after 2005 showed that the data before the year 2005 can be ignored for analysing a prospect of the US market:

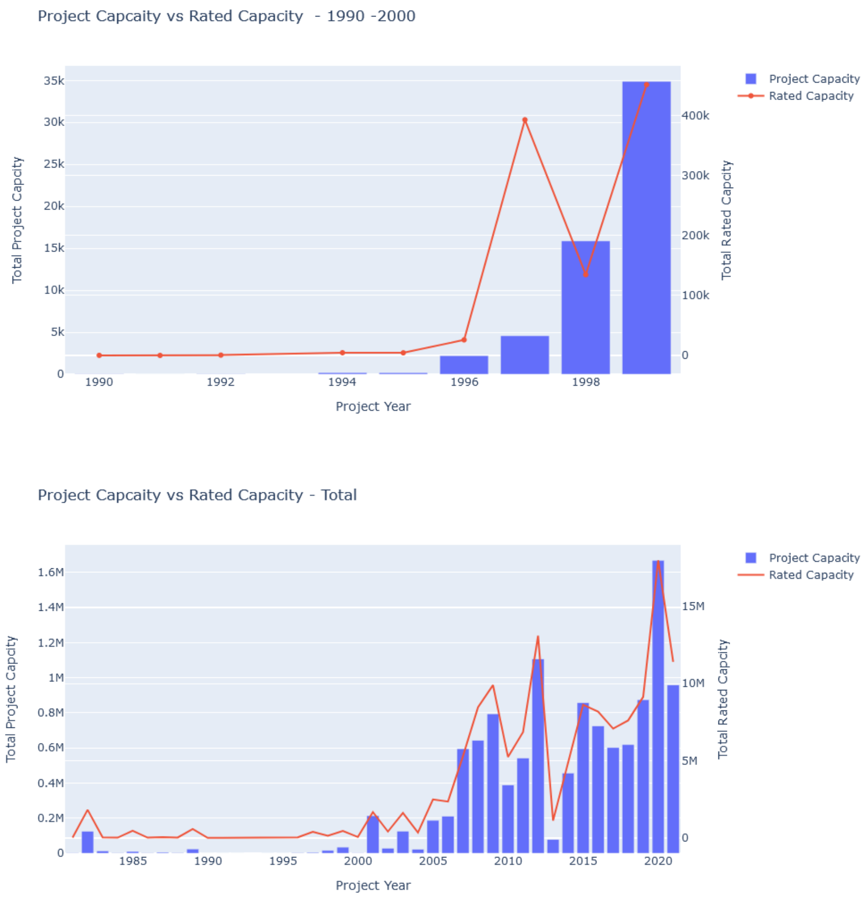



*Figure 4: Analysis of Projects before and After 2005*

All the analysis of the turbines is focused on the years after 2004.

## 3. Analysis:

### a) Total Number of turbines in each State:

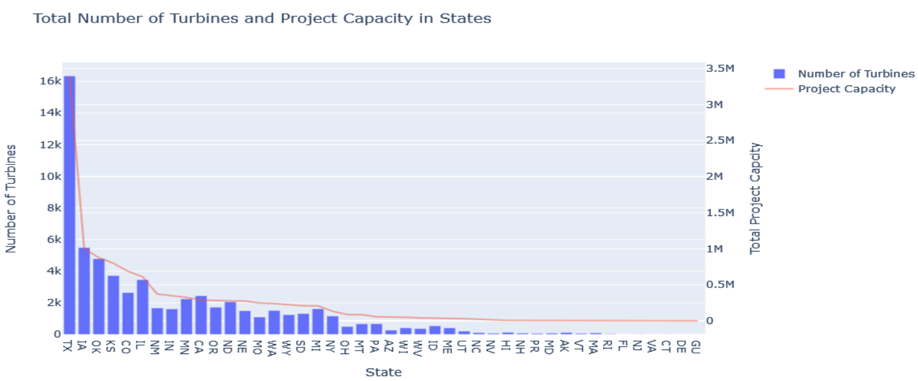



*Figure 5: Total Number of Turbines in Each State*

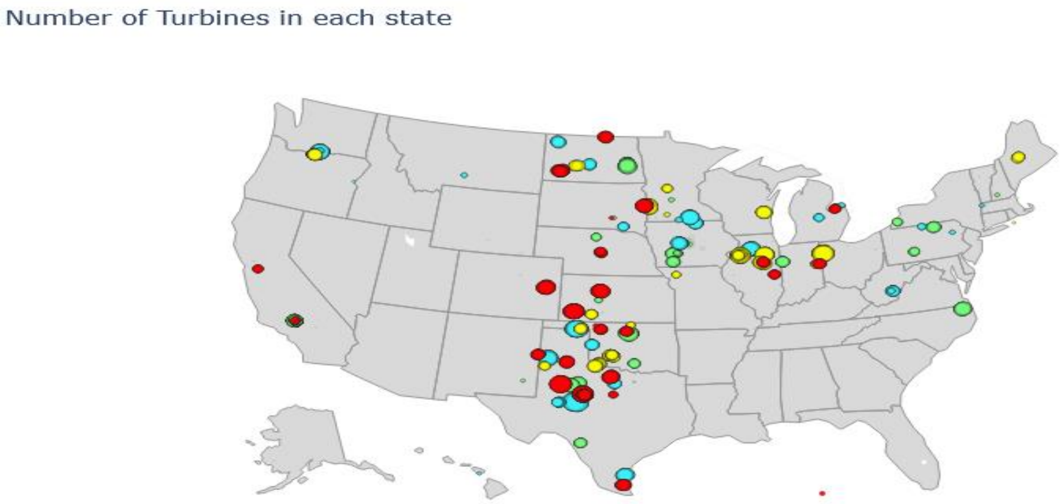



*Figure 6: Map of Turbines in each County*

The  State of Texas has the most number of turbines (around 16k), which is more than any other state. Texas is followed by Iowa which has 5.5K turbines in its state. The higher number of wind turbines can be attributed to the greater land availability and wind resources.

Also as supposed to the contrary assumption, more turbines do not necessarily mean greater project capacity. E.g.: The county of Colorado has a lesser number of turbines in its projects but produces greater capacity as compared to Illinois, which has more turbines but produces lesser capacity.

**b) Wind Turbines Project Capacity:**

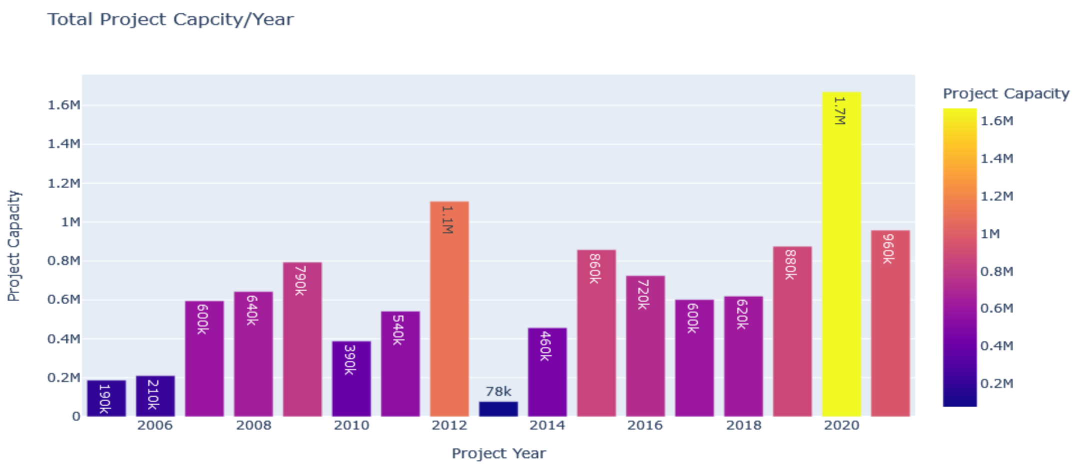



*Figure 7: Total Project Capacity/Year*

In the year 2020, the wind capacity reached a total of 1.7M. 42% of all electricity generation capacity in the US in the year 2020 came from land-based wind energy, which is more than any other source. Texas produces about four times more wind power than 3rd place California and three times more than 2nd place Iowa. (S&P Global, n.d.).

Further analysis showed that a lot of turbines were built in Texas and Oklahoma in the year 2019.

**c) Project Capacity in Each Year:**

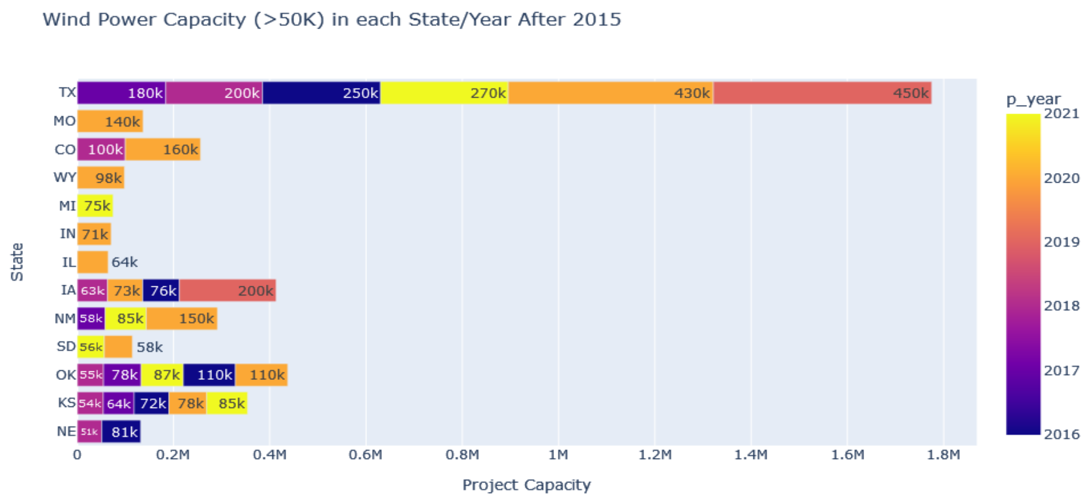



*Figure 8: Wind Power Capacity in each State/Year (capacity >50000)*

For the states that produced more than 50K capacity from the year 2016 to 2021, we can see a gradual increase in capacity for most of the states, with Texas having the highest capacity every year.

**d) Looking at the Manufacturers:**

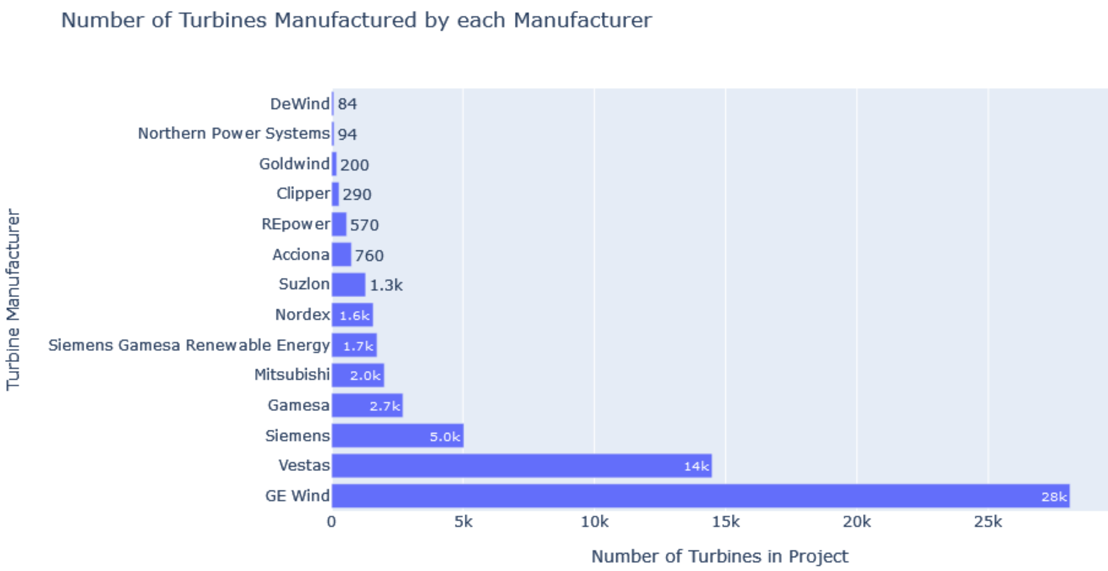



*Figure 9: 'Number of Turbines Manufactured by each Manufacturer.*

GE Wind manufactured the most number of turbines in the states, followed by Vestas and Siemens. Let's look at how many turbines each manufacturer made in different states:

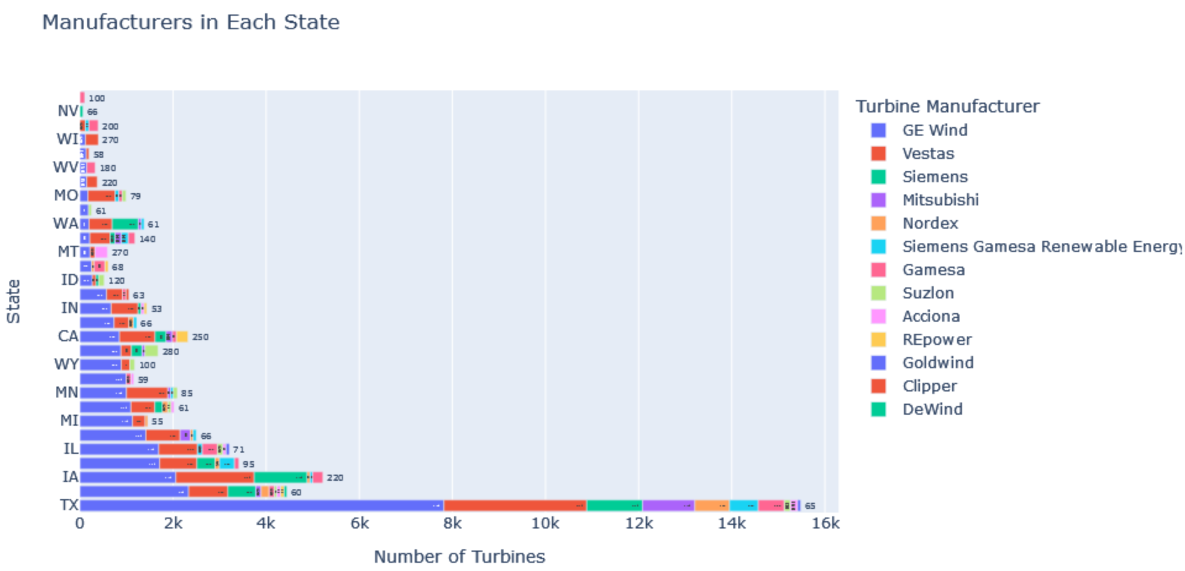



*Figure 10:Turbines Made by Manufacturers in Each State.*

GE Wind has its presence in almost every state and is also the major manufacturer of turbines. A further analysis into retrofitting revealed that GE wind is also the company who's turbines have been retrofitted the most. Although this number can be ignored as the ratio of turbines manufactured to the one's that need retrofitting is very less. (600:75).

## e) Projects

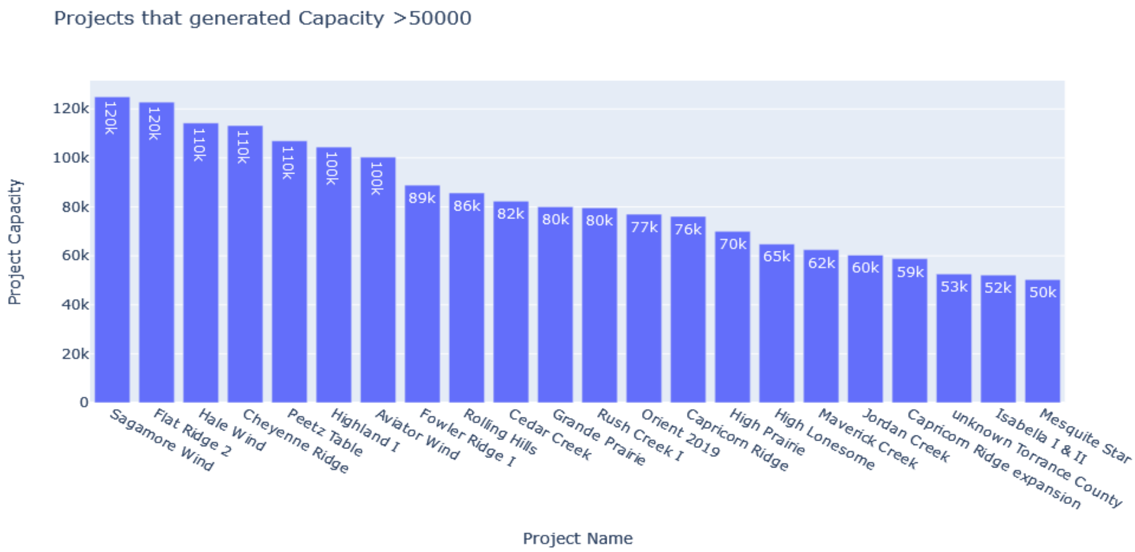



*Figure 11: Projects Generating More than 50K Capacity*

All the projects that generate more than 50K capacity can be further inspected on a state level.

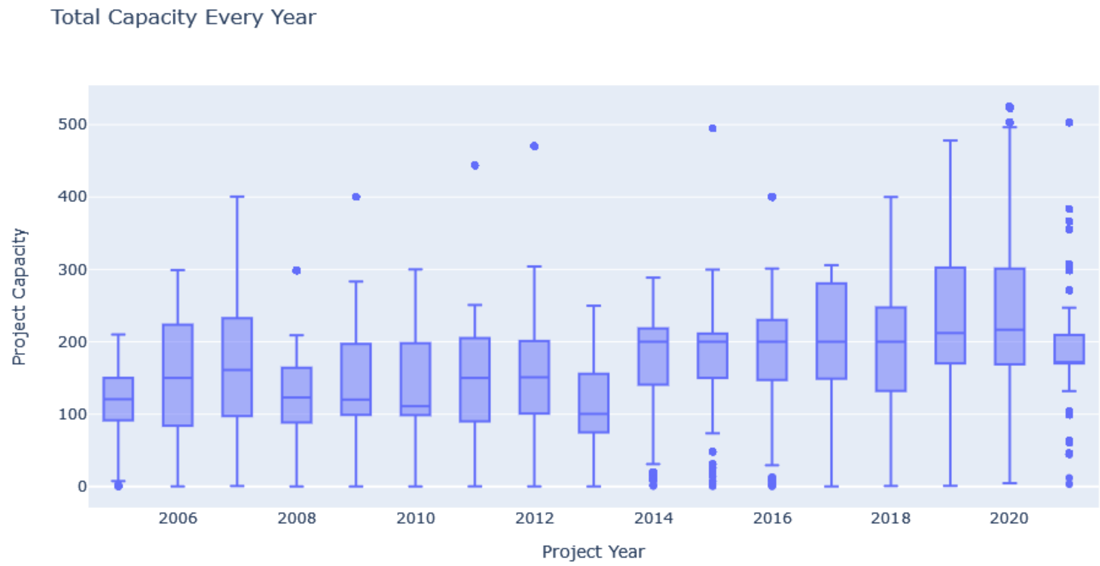
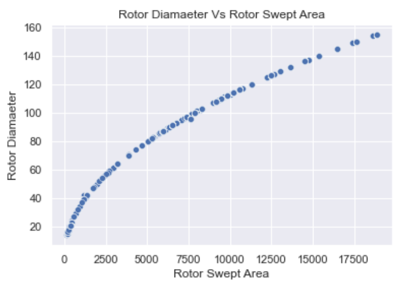
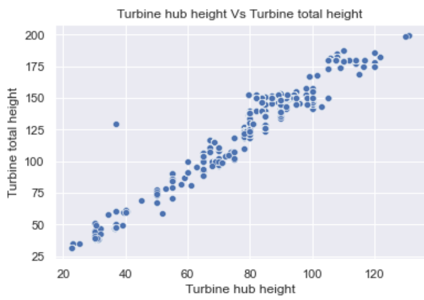


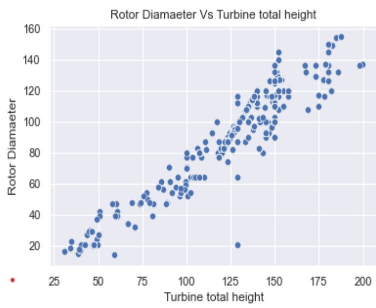

*Figure 12: Total Capacities in Each Year*

A high capacity project may need constant inspection to make sure that the turbines are functioning properly. Since the capacity is increasing on an annual basis we can use this to localize big projects in different states and target them accordingly.

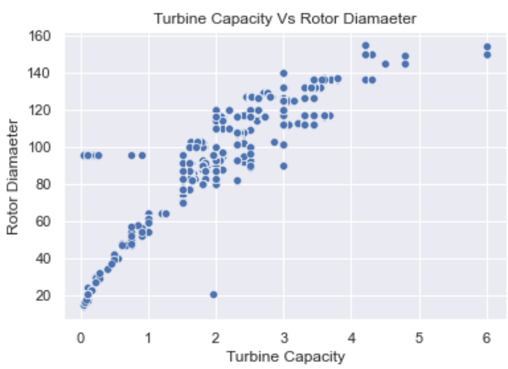
**f) Inspecting Turbines:**

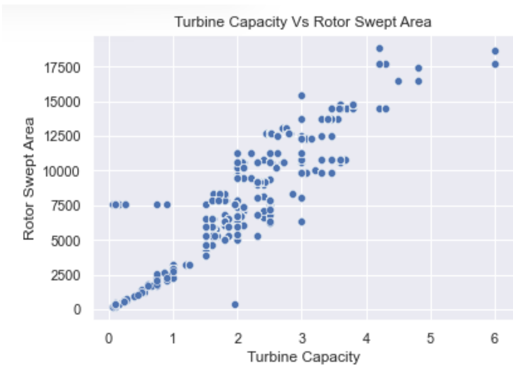



As seen from the plots the turbine characteristics have a linear relationship with each other and hence together affect the capacity of a turbine:



A plot between the Turbine Capacity and the Rotor Diameter and the Rotor Swept Area validates the previous assumption.

## 4) Lightning Data:

### a) Combining and Cleaning:

- The first task in cleaning the dataset is to convert the States named into IDs to be able to combine them with the turbines dataset.
- The data in the lighting file contains information on lighting aggregated on the basis of the year from 2015 – to 2019 whereas the turbines dataset contains information on each state every year. This creates a discrepancy in combining the data and hence has to be addressed first. To tackle this issue the following steps were performed:
  - Aggregating the turbines data from the year 2015 to 2019 on the basis of states
  - Now we have the aggregated data according to the state, we can combine it with the lighting data.
  - Next process is to normalize the combined data.

### b) Insights:

- Average Total Lightning:

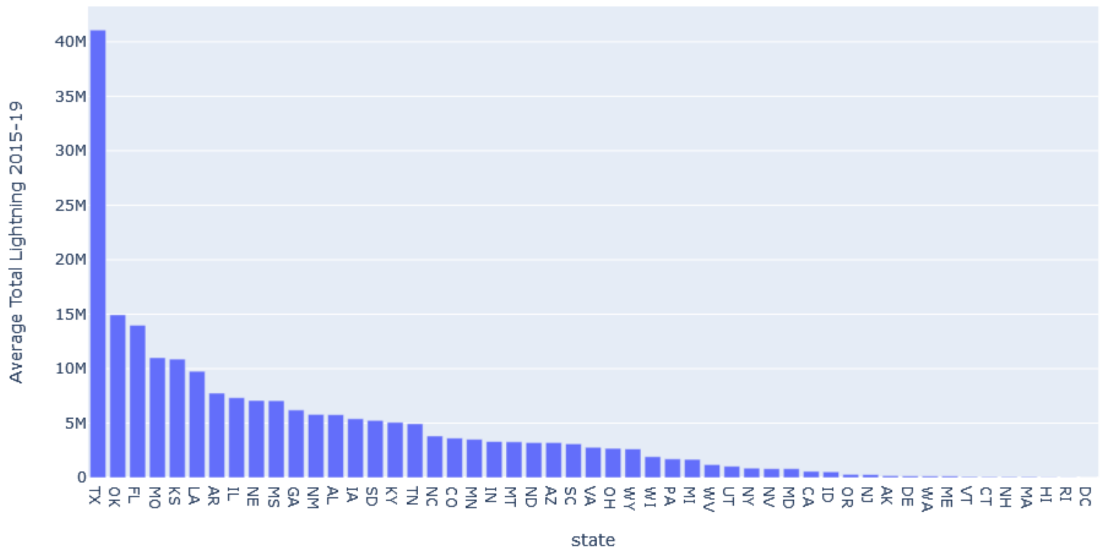



*Figure 13 Average Total Lightning in Each State:*

Texas received the most lighting in the year 2015-19, This may be attributed to the large area of Texas. Higher lighting may have an effect on the height of turbines.

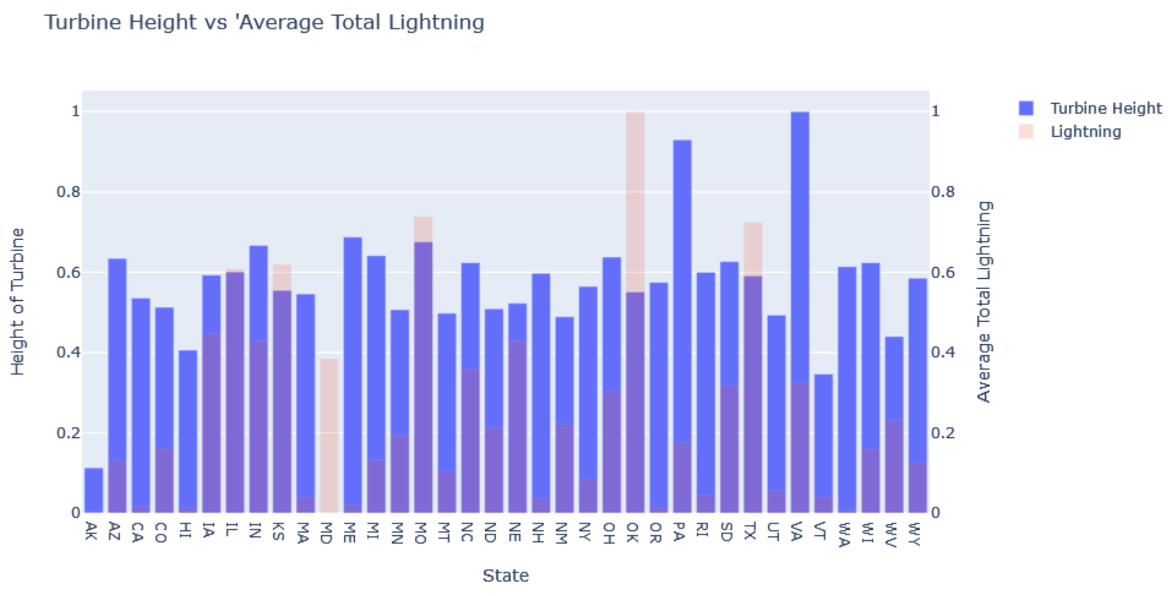



*Figure 14: Lighting vs Height*

Our assumption that the Height of turbines may be affected in states more prone to lightning can be validated with this graph. Texas and Oklahoma are two states with the majority of wind turbines and yet In these states, the height of turbines is smaller than in the states in which there is less lightning.

Higher lighting may also mean more stormy weather and hence more wind, which in turn makes the states with more lighting have more number of projects.

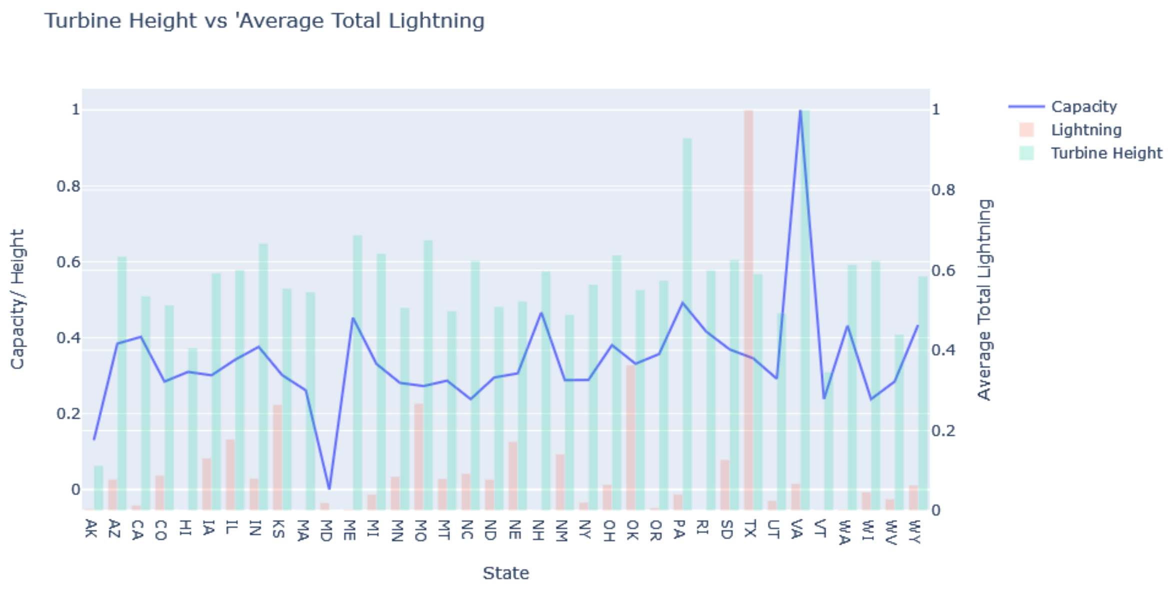



*Figure 15: Turbines Capacity with respect to Lighting  and Height*

## 5. Further Analysis:

Since the United States is a contrary that is prone to natural disasters, as an inspection company we can look at states that are most affected by natural disasters and aim at exploiting the market in these areas. The data on the Federally Declared Natural Disasters can be sourced from: https://worldpopulationreview.com/state-rankings/states-with-the-least-natural-disasters.

A second Dataset can be sourced from the US Energy Information Administration website: https://www.eia.gov/electricity/data/state/, website contains information on yearly production, revenue, and consumption of energy through different sources of energy (wind, hydroelectric, coal, etc.) in every state. We can use this to see how much wind energy was used in the states in comparison to other sources, which could give us an understanding of the working costs of turbines. Since the data is openly available, there will be no difficulty in sourcing it. However, some difficulties can arise when using this data with the turbines dataset because of a difference in the structure and key/value pairs of this data.

6. **Conclusion:**
   Based on the analysis we can see that a potential market exists in places where the states where there are more turbines and the places where the production capacity is high. Also, since lighting causes a huge amount of turbine blade damage, a potential market would be to exploit places where lighting storms are more frequent. Another avenue would be to look at places where the height of turbines is less and are more close to city areas, since these turbines are also more prone to damages from external factors.