**Data Analysis of Solar Energy**

12201449 Choi Chan Woo

1. **Introduction**

In terms of making Renewable Energy Plants, assessing its profiles on land or sea will be important to see whether the area is suitable for certain renewable energy types or not based on thoroughly curated measurement practices on every factor of the area. For example, Solar Power does rely on weather, and in specific, solar irradiation does rely on other climate factors like Precipitation, Azimuth, Humidity and such. Whilst on wind power, Wind Profile (Speed and Direction) is the main factor for it and relies mostly on Geological Factors like topology of the area.

In this report, we`ll be seeing on a dataset of various factors and power on Solar energy to see which variable could be relatable for solar power generation.

1. **Before Starting (Analysis on calculation methods of Statistical Values)**

When starting to do statistical analysis on subject, we stumble upon finding statistical values like mean, standard deviation, kurtosis, skewness and much more.

In Python, several libraries do provide tools for finding statistical values. But before these libraries did came across mainstream, we had to make the function by ourselves to derive the values, which is known as “Manual Calculation”. So, in this part, there will be a comparison between manual calculation and calculation using standard libraries like NumPy, SciPy, Pandas.

텍스트, 스크린샷, 소프트웨어, 멀티미디어 소프트웨어이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

figure 1. Code With Manual Calculation)

텍스트, 소프트웨어, 멀티미디어 소프트웨어, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

figure 2. Code with calculation using standard libraries (NumPy, SciPy)

텍스트, 소프트웨어, 멀티미디어 소프트웨어, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

figure 3. Code with calculation using standard libraries (Pandas)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Manual | SciPy, NumPy | Pandas |
| Mean | 6.90117 | 6.90117 | 6.90117 |
| Variance | 35.68649 | 35.68649 | 35.68649 |
| Std.Deviation | 5.97381 | 5.97381 | 5.97381 |
| Skewness | 0.34314 | 0.34314 | 0.34314 |
| Kurtosis | -0.58603 | -0.58603 | -0.58603 |

figure 4. Chart of values derived from figure 1 to 3’s calculation method

In figure 4, the values of the statistical values are almost the same. But in terms of code length, figure 2 and figure 3 are already way ahead of having less lines than on figure 1, meaning that using Standard libraries is better than on Manual calculation, even if the output values are the same among those three methods. If there is a difference between figure 2 and figure 3 though, there could be a difference of the input. On figure 1 and 2, there is a process of converting input values into another new numerical dataframe. While in figure 3, you can use the dataset directly without having a new dataframe.

This is much the same result as on Correlation Coefficient and Covariance calculations as you would see in figure 5., figure 6., figure 7.

텍스트, 스크린샷, 소프트웨어, 멀티미디어 소프트웨어이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

figure 5. Code With Manual Calculation

텍스트, 소프트웨어, 멀티미디어 소프트웨어, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

figure 6. Code with calculation using standard libraries (SciPy. Stats)

|  |  |  |
| --- | --- | --- |
|  | Manual | SciPy. Stats |
| Covariance | 1779.34101 | 1779.34101 |
| Pearson | 0.29852 | 0.29852 |
| Spearman | 0.25454 | 0.25468 |

figure 7. Chart of values derived from figure 5 to 6’s calculation method)

1. **Interpretation and Analysis of Variables**

도표, 스크린샷, 그래프, 라인이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

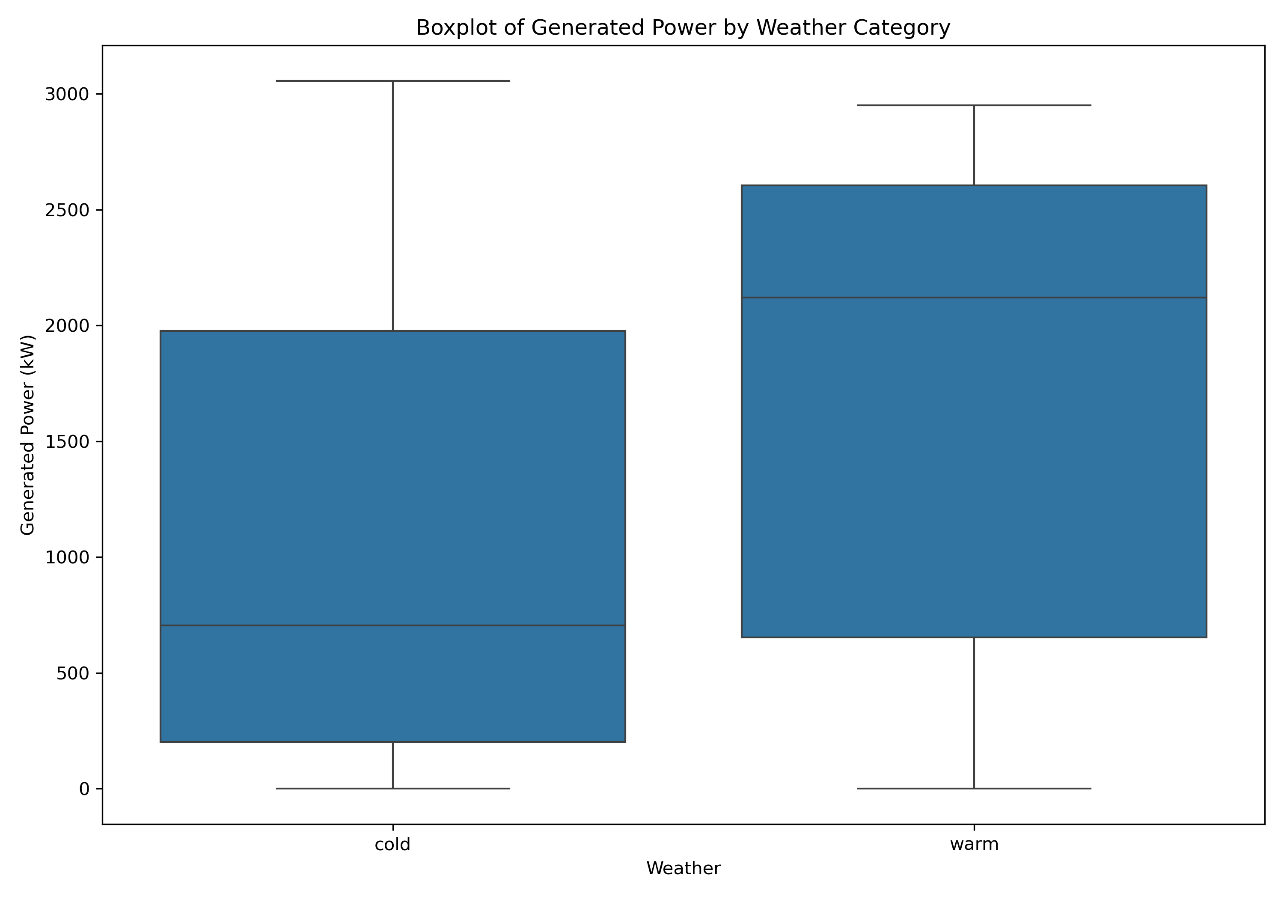
figure 8. Histogram of Generated Power by Weather Category

figure 9. Boxplot of Generated Power by Weather Category

In figure 8, you could interpret that in cold weathers, there is a massive amount of low energy generation and some observations of having a few High Energy Generation within special conditions. On the Warm side, there is a steady spread of frequency on Mid~High Energy Generation and a few moments where there is low energy generation. So could know that in cold weather, there is a high possibility of having low energy generation while in warm weather, even if the energy generation could vary, can be more than mid relatively.

figure 9 shows the box plot of warm weather and cold weather respectively. Comparing the Median of each plot could know that warm weather`s median is higher than on cold weather. Also, the IQR range is higher in Warm Weather.

So, we could say that for the most part, warm weather could make more energy generation than cold weather. But we must not forget that there is a small but possible probability that even cold weather could make more energy or warm weather with low energy generation since there is an outlier on each of them.

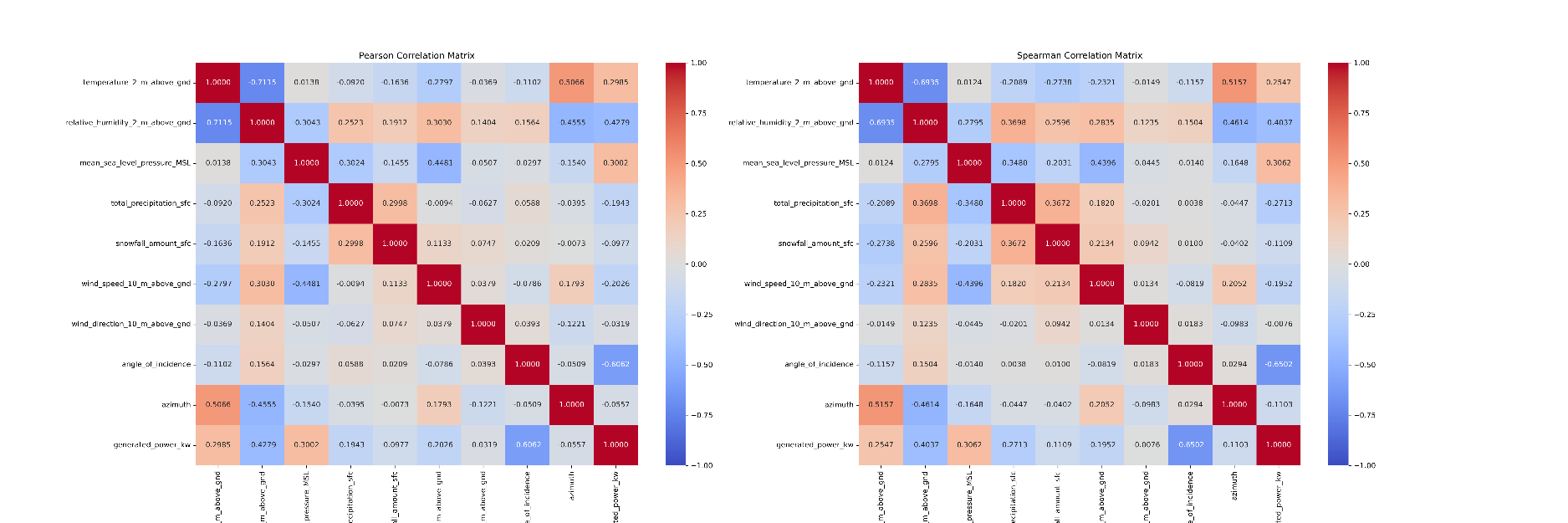


Figure 10. Heatmap of Correlation Matrix

Figure 10 shows the Correlation Matrix on a Heatmap. For some valuable findings in this plot (related to Energy Generation that we are focusing on), one of the kinds is the relation between ‘angle\_of\_incidence’ and ‘Generated Energy’. In both heatmap, those two variables do have the most negative relationship, meaning that when the angle could be higher, it could affect the energy generation for worse possible scenario. Also, ‘total\_precipitation,’ ‘relative\_humidity’ also could bring negative effect on energy generation when those values increase since they do have negative relationships on Energy Generation. On the positive side, temperature’ and ‘mean se level pressure’ do have positive relationship with Energy Generation, which means that if the values are higher, so does energy generation, but the increase is lower than the negative ones though.

텍스트, 스크린샷, 친필이(가) 표시된 사진

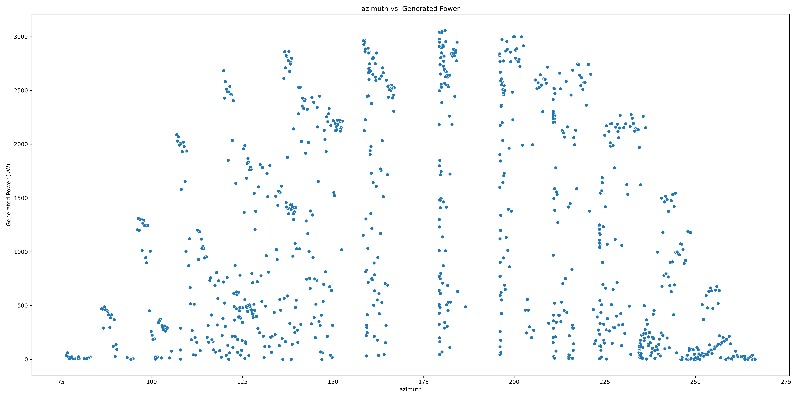
AI가 생성한 콘텐츠는 부정확할 수 있습니다.

figure 11. Scatterplot of G.Energy and Angle of incidence figure 12. Scatterplot of G. Energy and Azimuth

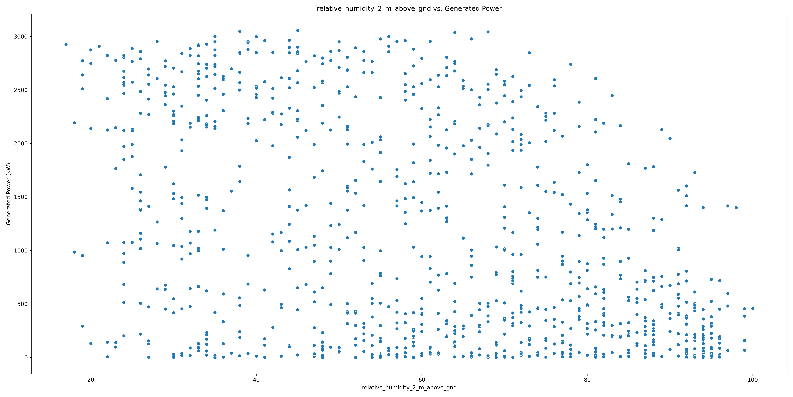
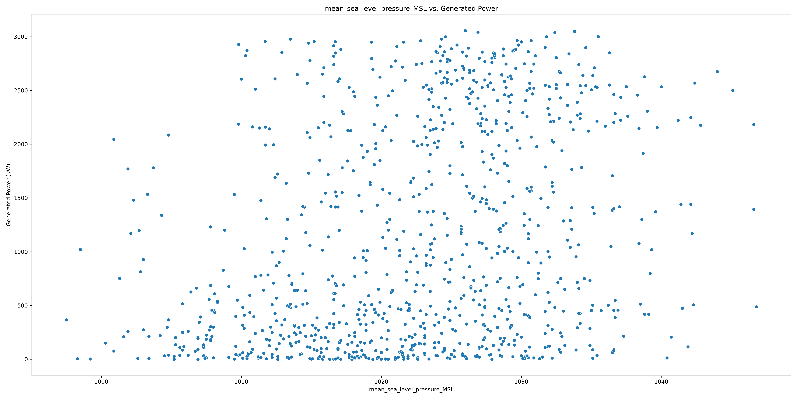


figure 13. Scatterplot of G.Energy and Mean Sea Level pressure figure 14. Scatterplot of G.Energy and humidity (2m above)

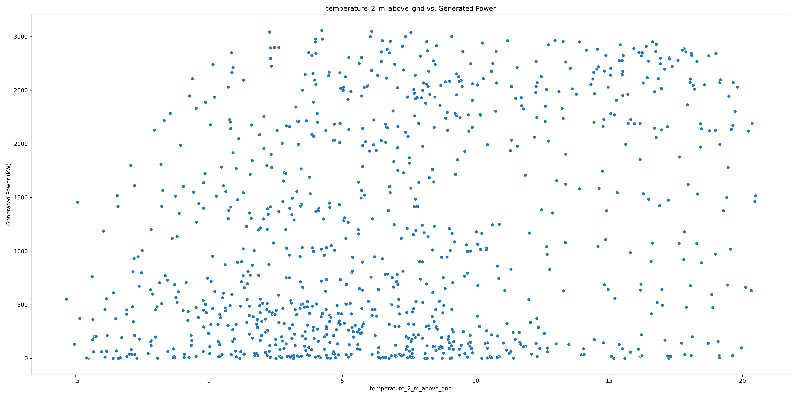
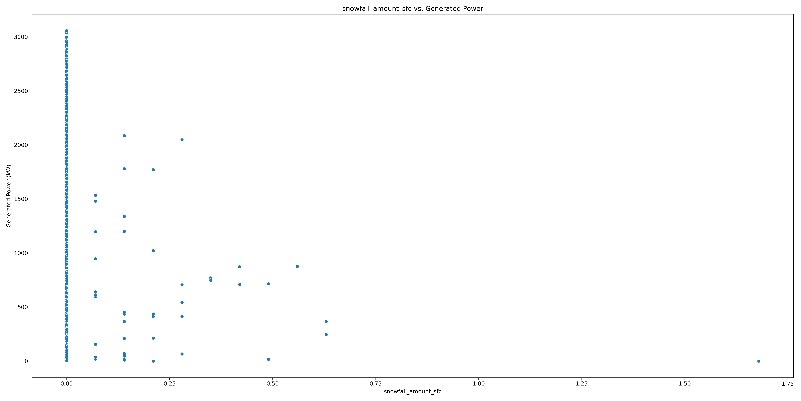


figure 15. Scatterplot of G.Energy and Snowfall amount figure 16. Scatterplot of G.Energy and temperature (2m above)

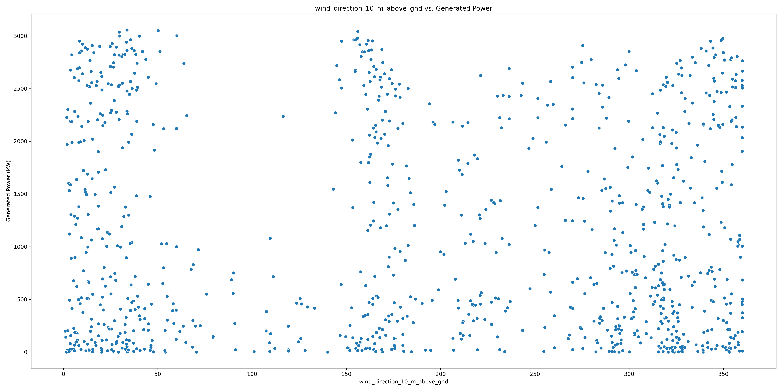
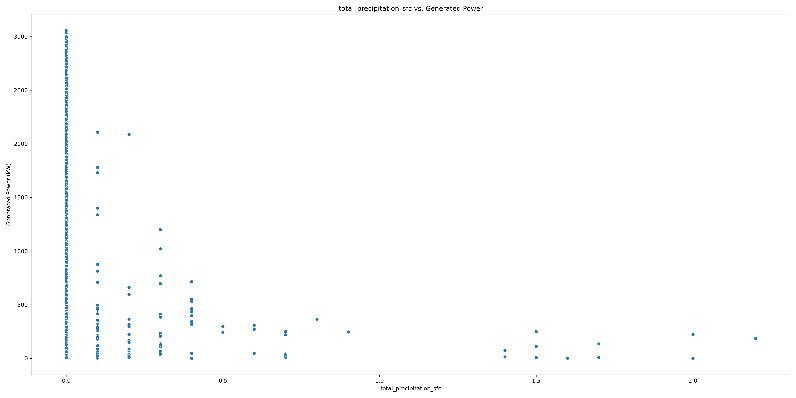


figure 17. Scatterplot of G.Energy and total precipitation figure 18. Scatterplot of G.Energy and wind direction (10m above)

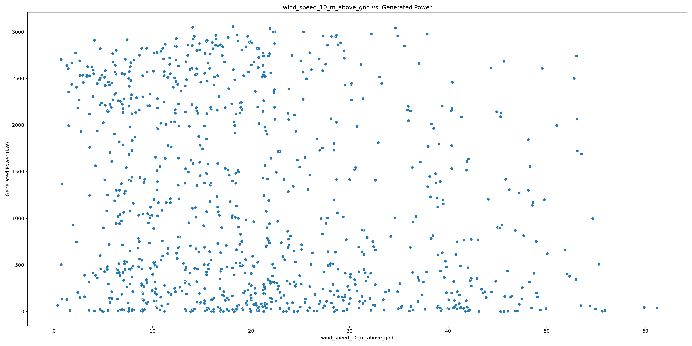


figure 19. Scatterplot of G.Energy and wind speed. (10m above)

Following figures from 11 to 19 is the scatter plot of other variables versus Generated Energy.

Figure 12, 15, 17, 18, 19 shows nothing much, but sparce among all areas.

Figure 11, 14 shows some parts where high on value means more scatter on lower energy generation

Figure 13, 16 shows parts where high on value means more scatter on higher energy generation, but not that distinctive than on figure 11.

And with figure 10 combined, we could see that having less relationships (near zero) do tend to have indescribable scatters, which are wind direction, azimuth, snowfall amount. On the other hand, having near negative 1 like ‘angle of incidence’ do tend to have some ‘group’ when the angle changes like when angle is lower, then there is a group of scatters located in high energy generation and vice versa. (figure 11). ‘humidity’ also shows a similar plot as that, but not that as distinctive as ‘angle of incidence’, meaning that on thinking of ‘negative’ parts, ‘Angle of Incidence’ will be the priority and then following with ‘humidity’. On the positive side, ‘temperature’, and ‘mean sea level pressure’ does show the plot when the value is high, then the generated energy scatter could locate on higher position. But the problem is that this doesn’t show that distinctive as figure 11.

So, in conclusion, when trying to assess the solar energy potential on certain areas and you get a full dataset of the area, first you will need to see will be the angle of incidence will be your first looking onto it. Then goes other factors like ‘humidity’, ‘mean sea level pressure’, and ‘temperature’ will be your second finding to effectively assess the potential of solar power.