IE 6200 Probability and Statistics

PROJECT REPORT

Presented by – Parth Malpure

Contents

OBJECTIVE

The main purpose of this project is to perform an analysis on wind energy consumption of various countries for the time-period of 1995 -2014, various plots were used to display the findings. In this report three statistical tests were performed, and a linear regression model was built to predict the wind energy consumption for Germany for the next 3 decades.

Data Collection

Data Description: The main dataset is collection of key metrics maintained in 'Our World in Data' (https://ourworldindata.org/energy) by Hannah Ritchie, Max Roser and Pablo Rosado. The dataset has information on energy consumption (primary energy, per capita, and growth rates), energy mix and other relevant metrics.

Dataset link: https://www.kaggle.com/datasets/pralabhpoudel/world-energy-consumption

The columns which were used for analysis in this report were:

- country: geographic location
- year: year of observation
- gdp_per_capita: per capita GDP of a country
- wind_share_energy: share of primary energy consumption that comes from wind

PART 1: ANALYSIS FOR % WIND ENERGY SHARE FOR VARIOUS COUNTRIES

1.1 Comparison of Total Energy Consumption

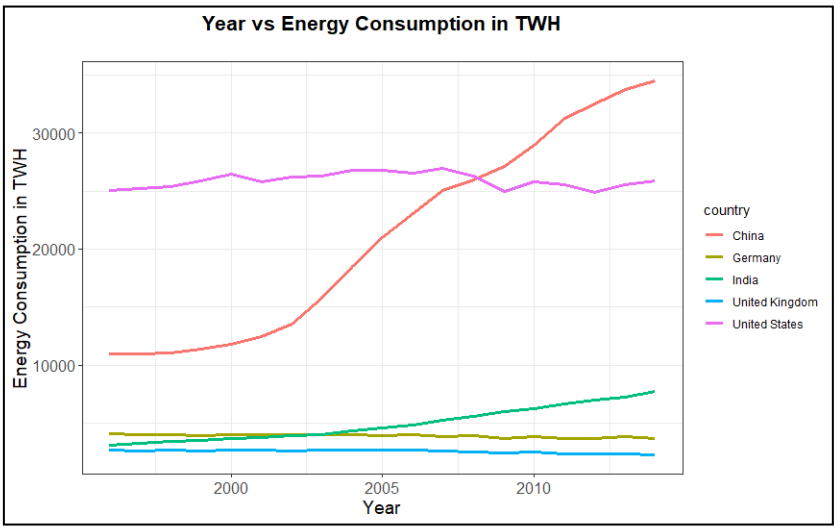

Figure 1: Year v/s Energy Consumption in TWH

It is evident from figure 1 that the total energy consumption is increasing for developing countries like China and India, whereas for developed countries like Germany, UK, and USA the total energy consumption is almost constant.

1.2 Comparison of Renewable Energy

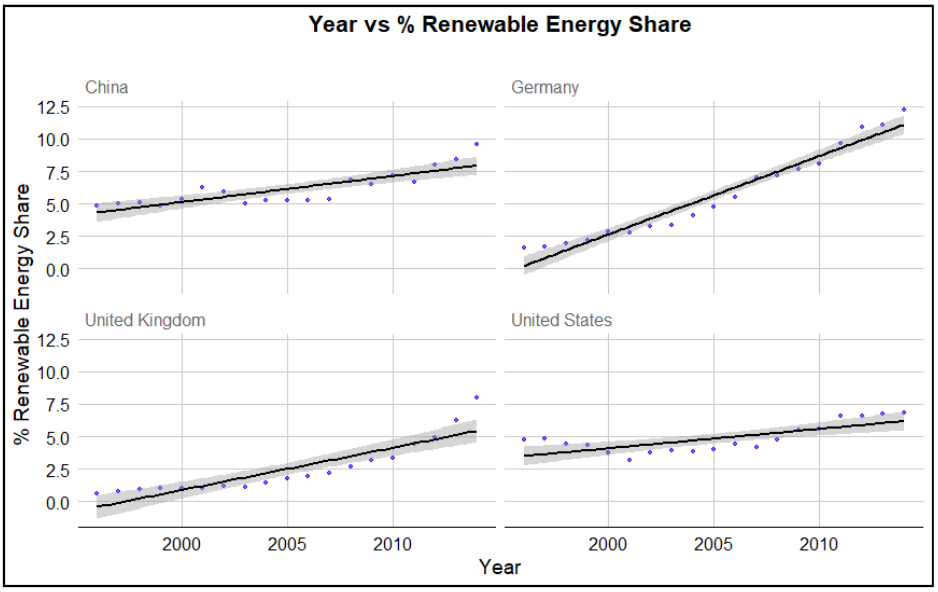

Figure 2: Year v/s renewable Energy Share

As many countries are moving towards renewable energy, it is observed that Germany has the highest growth rate in renewable energy share over the years, when compared to countries like China, UK, and USA.

1.3 Comparison of Germany's Renewable Energy

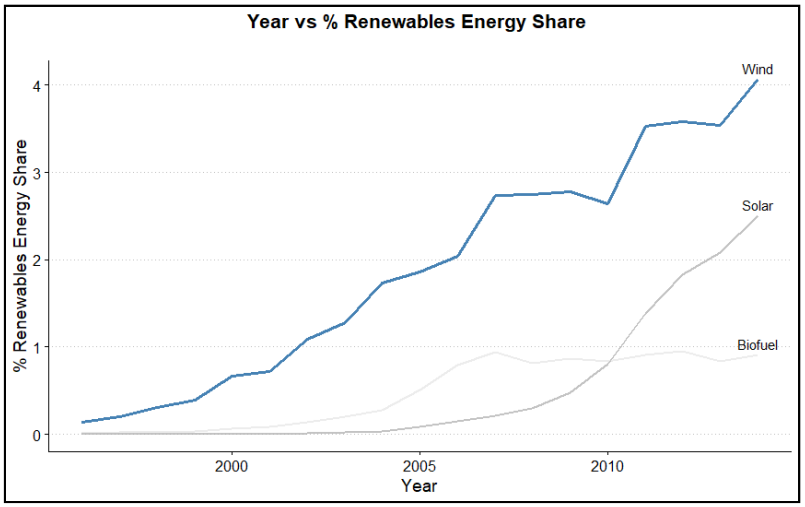

Figure 3: Year v/s Renewable Energy Share

In further analysis, the percentage of wind energy share for Germany is rapidly growing as compared to other sources of energy like solar and biofuel i.e., at the end of 2014, the %energy share of wind energy was 4.07, solar energy was 2.51, and biofuel energy was 0.91.

1.4 Comparison of GDP

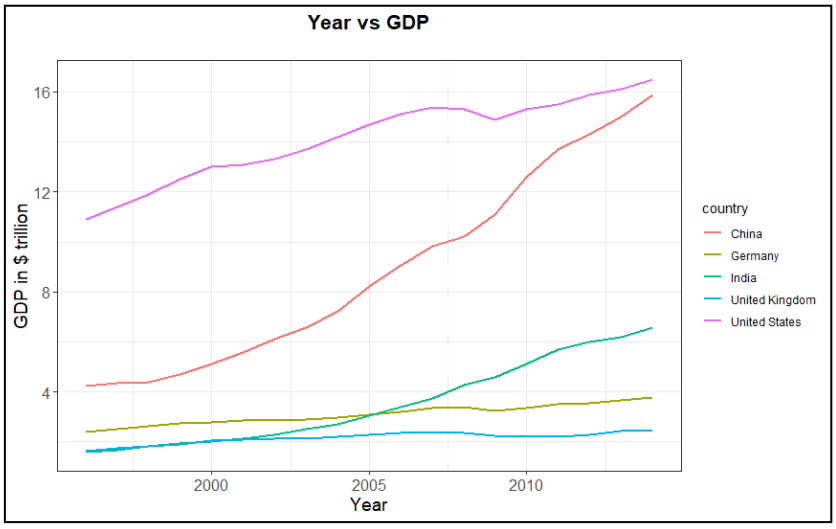

Figure 4: Year v/s GDP

It is observed that all the countries have shown an increase in GDP for the two decades. But the growth rate GDP for developing countries like China and India is very high, whereas for developed countries like Germany, and UK the growth rate of GDP is very small.

In addition to this, another method of performance comparison needs to be evaluated for comparing renewable energy resources over the time period.

1.5 Correlation of GDP per capita and % Wind Energy Share for Germany

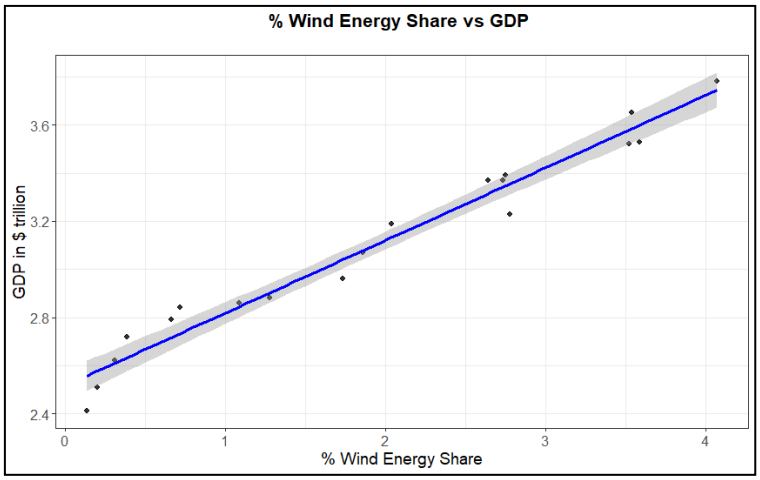

Figure 5: % Wind Energy Share v/s GDP

Pearson coefficient: 0.98

Kendall coefficient: 0.95

It can be inferred from the plot that there is a strong correlation between share of %wind energy share and GDP for Germany. Thus, it is concluded that an increase in GDP is directly related to the increase in wind energy consumption for Germany.

1.6 Outcomes

Energy Consumption for developing countries is increasing more rapidly as compared to some developed countries. Similarly, the GDP is also increasing more rapidly as compared to some developed countries. Therefore, we can say that there is correlation between energy consumption and GDP for developing countries.

Also, Germany is one of the leading countries in wind energy consumption, and there is a direct correlation between wind energy share and GDP.

PART 2: STATISTICAL ANALYSIS

2.1 Probability of % Wind Energy Share

We wanted to find out, if a random observation is picked from our dataset what is the probability that (a) more than 2% wind energy share, (b) less than 1% wind energy share, and (c) less than 0.5% wind energy share.

Assuming that the world's % wind energy share is normally distributed with mean = 0.5541424, sd = 1.24989, we determine the following:

X = R.V. of % Wind Energy Share

X ~ N (x; μ = 0.5541424, σ = 1.24989)

(a) more than 2% wind energy share

[P(X >= 2)] = 0.124

(b) less than 1% wind energy share

[P(X<=1)] = 0.64

(c) less than 0.5% wind energy share

[P(X<=0.5)] = 0.483

Therefore, if a random observation is picked from our dataset there is a 48.3% chance that the %wind energy share is less than 0.5%.

2.2 Hypothesis test for Equality of Proportions

We then performed a two-sample proportion test for the hypothesis that there is a significant difference between proportion of Germany's % wind energy share > 1.5% and rest of the World's % wind energy share between the year 1995 & 2015.

H_0: Germany_prop - World_prop = 0 (There is no significant difference)

H_1: Germany_prop - World_prop ≠ 0 (There is a significant difference)

```
        2-sample test for equality of proportions with continuity correction

data:  c(x1, x2) out of c(n1, n2)
X-squared = 35.889, df = 1, p-value = 2.089e-09
alternative hypothesis: two.sided
95 percent confidence interval:
 0.3122360 0.7930272
sample estimates:
   prop 1    prop 2
0.6842105 0.1315789
```

As p value is < 0.05, we reject the null hypothesis (H_0: Germany_prop - World_prop = 0) and conclude that there is a significant difference (H_1: Germany_prop - World_prop ≠ 0). Therefore the proportion of Germany's wind energy share > 1.5 is greater than proportion of rest of the World's wind energy share > 1.5.

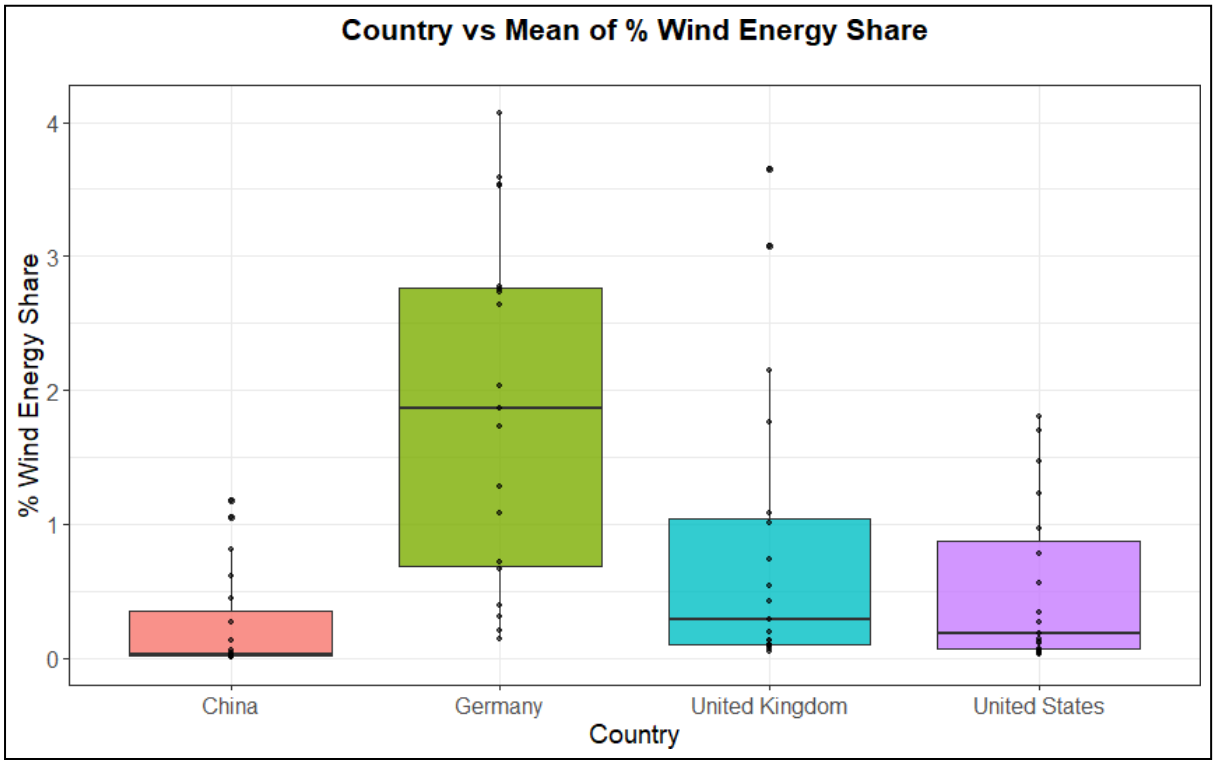

Figure 6: Country v/s Mean of % Wind Energy Share

Figure 6 describes the quartile range, minimum values and maximum values for the countries comparing their medians of % wind energy share. For Germany, it is observed that the median

of % wind energy share is 1.86, whereas for rest of the World it is observed that the median of % wind energy share is 0.0515.

2.3 Hypothesis test for Equality of Means

We then performed a two-sample t-test for the hypothesis that there is a significant difference between the mean of mean of %wind energy share of Germany and Rest of the World the best, when the population variances are unknown.

H_0: Mean_Germany – Mean_World = 0

H_1: Mean_Germany – Mean_World ≠ 0

```
        Welch Two Sample t-test

data:  sample1 and sample2
t = 3.9498, df = 12.952, p-value = 0.001673
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.6889332 2.3537935
sample estimates:
mean of x mean of y
 2.062083  0.540720
```

As p value is < 0.05, we reject the null hypothesis (H_0: Mean_Germany – Mean_World = 0) and conclude that there is a significant difference (H_1: Mean_Germany – Mean_World ≠ 0). We found that mean of %wind energy share of Germany is greater than Rest of the World.

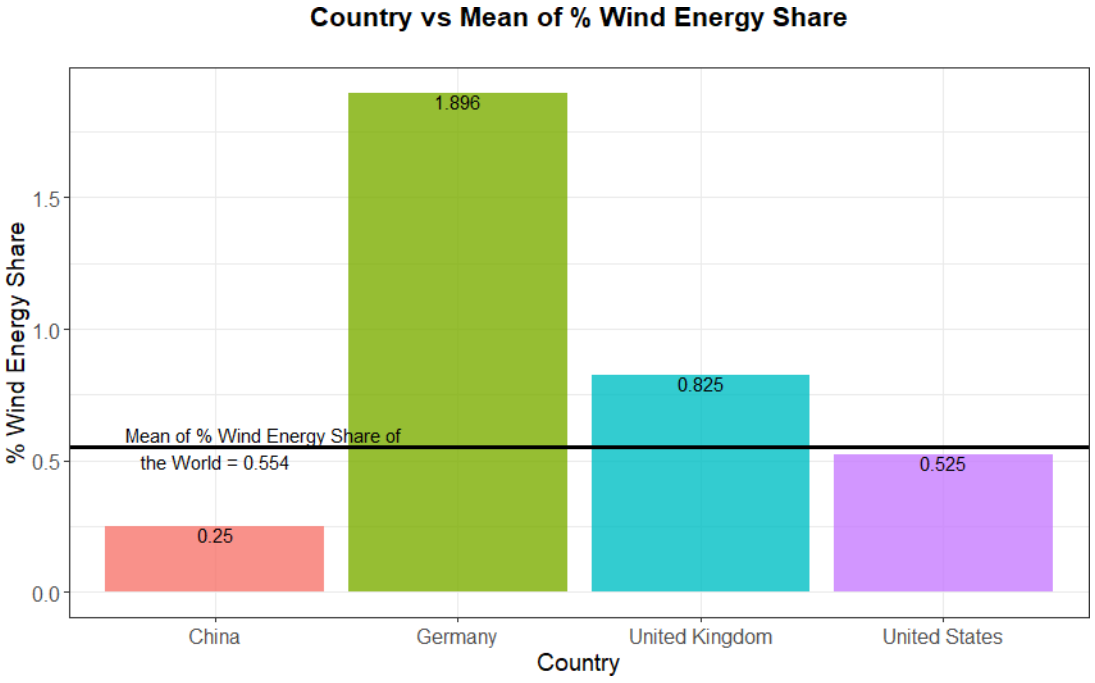

Figure 7: Country v/s Mean of % Wind Energy Share

2.4 Outcomes

It clear from the above statistical tests that Germany's wind energy usage is much higher than rest of the world's wind energy usage.

PART 3: PREDICTION ANALYSIS

Building a linear regression model for predicting wind energy share for next 30 years for Germany using the existing data with the help of Linear Regression Model.

3.1 Assumptions for Linear Regression

Linear regression is a useful statistical method we can use to understand the relationship between two variables, x and y. However, before we conduct linear regression, we must first make sure that four assumptions are met:

- **Linear relationship:** There exists a linear relationship between the independent variable, x, and the dependent variable, y.
- **Independence:** The residuals are independent. There is no correlation between consecutive residuals in time series data.
- **Homoscedasticity:** The residuals have constant variance at every level of x.
- **Normality:** The residuals of the model are normally distributed.

3.2 Linear Regression Model

Creating a linear regression model to predict % Wind energy Share for Germany for the years 2021-2050, using the existing data.

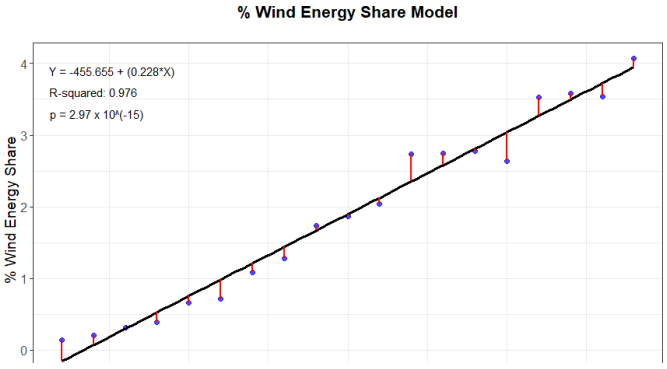

Figure 8: % Wind Energy Share Model

Model Summary

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.39739 -0.13244 -0.03419  0.12653  0.38122

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -455.65518   17.22390  -26.45 2.97e-15 ***
year           0.22820    0.00859   26.57 2.77e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2051 on 17 degrees of freedom
Multiple R-squared:  0.9765,    Adjusted R-squared:  0.9751
F-statistic: 705.7 on 1 and 17 DF,  p-value: 2.771e-15
```

3.3 Diagnostic Plots

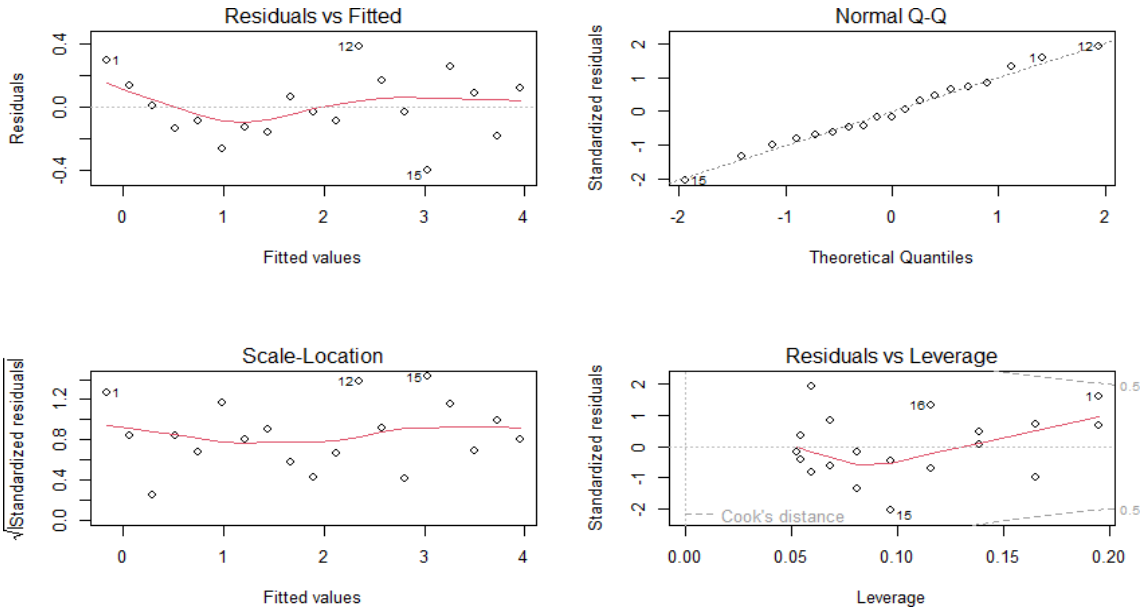

Figure 9: Diagnostic Plots

1. Residuals vs Fitted: Used to check the linear relationship assumptions. A horizontal line, without distinct patterns is an indication for a linear relationship, what is good.

2. Normal Q-Q: The QQ plot of residuals can be used to visually check the normality assumption. The normal probability plot of residuals should approximately follow a straight line. In our example, all the points fall approximately along this reference line, so we can assume normality.

3. Scale-Location (or Spread-Location): Used to check the homogeneity of variance of the residuals (homoscedasticity). Horizontal line with equally spread points is a good indication of homoscedasticity. This is not the case in our example, where we have a heteroscedasticity problem.

4. Residuals vs Leverage: Used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis. This plot will be described further in the next sections.

3.4 Prediction using the model

The following equation was obtained using Linear regression model:

$$Y = -455.655 + (0.228*X)$$

% Wind energy share = -455.655 + (0.228*year)

The Predicted values are:

Year	Wind Energy Share	Year	Wind Energy Share
2021	5.55	2036	8.97
2022	5.78	2037	9.2
2023	6	2038	9.43
2024	6.23	2039	9.66
2025	6.46	2040	9.88
2026	6.69	2041	10.11
2027	6.92	2042	10.34
2028	7.15	2043	10.57
2029	7.37	2044	10.8
2030	7.6	2045	11.02
2031	7.83	2046	11.25
2032	8.06	2047	11.48
2033	8.29	2048	11.71
2034	8.51	2049	11.94
2035	8.74	2050	12.17

Table 1: Prediction Values for 2021-2050

Conclusion

After the analysis of various parameters, we can say that the Energy Consumption for developing countries is increasing more rapidly as compared to some developed countries. Similarly, the GDP is also increasing more rapidly as compared to some developed countries. Therefore, we can say that there is correlation between energy consumption and GDP for developing countries. Germany is one of the leading countries in wind energy consumption, and there is a strong positive correlation between wind energy share and GDP.

A two-sample proportion hypothesis test was performed to find out whether there is a significant difference between proportion of Germany's % wind energy share > 1.5% and rest of the World's % wind energy share between the year 1995 & 2015. We found that proportion of Germany's wind energy share > 1.5 is greater than proportion of rest of the World's wind energy share > 1.5. Similarly, a two-sample t-test was performed for the hypothesis that there is a significant difference between the mean of mean of %wind energy share of Germany and Rest of the World the best, when the population variances are unknown. We found that mean of %wind energy share of Germany is greater than Rest of the World.

Lastly, a linear regression model was built to predict % Wind energy Share for Germany for the years 2021-2050, using the existing data.

References

1. https://ourworldindata.org/energy
2. https://www.kaggle.com/datasets/pralabhpoudel/world-energy-consumption
3. https://sustainabledevelopment.un.org/index.php?page=view&type=99&nr=24&men=1449
4. https://en.wikipedia.org/wiki/Renewable_energy_in_Germany
5. https://cran.r-project.org/web/packages/available_packages_by_date.html