Table of contents

Part one: Introduction and data preparation

1. Introduction:

In the face of escalating global energy demand and growing concerns about the environmental impact of traditional fossil fuels, the exploration and utilization of renewable energy sources have emerged as pivotal components of sustainable energy strategies. Renewable energy, often interchangeably referred to as green or clean energy, encompasses a diverse array of naturally occurring resources that can be harnessed to generate power without depleting finite reserves or perpetuating environmental harm.

The primary allure of renewable energy lies in its inherent ability to tap into sources that are naturally replenished over time, such as sunlight, wind, water, and geothermal heat. Solar energy, derived from the radiant power of the sun, stands as an emblematic example, with photovoltaic cells converting sunlight into electricity with remarkable efficiency. Wind power capitalizes on the kinetic energy of moving air masses, propelling turbines to generate electricity, while hydropower exploits the gravitational force of flowing water to produce energy. Geothermal energy, drawn from the Earth's internal heat, and biomass, derived from organic materials, round out the renewable energy portfolio, each offering distinctive advantages in diverse contexts.

The adoption of renewable energy technologies not only mitigates the environmental impact associated with conventional energy sources but also contributes to energy security by diversifying the energy mix. As governments, industries, and communities worldwide intensify their efforts to combat climate change and transition towards sustainable energy solutions, an in-depth analysis of renewable energy datasets becomes imperative. This thesis endeavors to explore and elucidate patterns, trends, and implications within a specific dataset, aiming to contribute valuable insights to the ongoing discourse surrounding the pivotal role of renewable energy in shaping a greener and more sustainable future.

Apart from the apparent advantages of less ecological footprint and improved energy stability, the widespread adoption of renewable energy technology stimulates economic expansion and employment generation. The transition to renewable energy will create jobs in production, installation, maintenance, and research as well as need innovation and investment in new infrastructure. Furthermore, by encouraging localized energy production and a more resilient energy infrastructure, renewable energy projects frequently boost local economies and give people more power. This increases the societal impact of renewable energy and synchronizes economic growth with social welfare and environmental sustainability.

Objective 1: Research Objectives:

Basically the objective of the selected data set is to highlight the difference among the developed and undeveloped countries on the basis of renewable energy. The pursuit of renewable energy is a critical aspect of global efforts to address climate change, enhance energy security, and foster sustainable development. The approach to renewable energy adoption, however, diverges significantly between developed and non-developed countries, reflecting distinct challenges, opportunities, and priorities.

- **Developed Countries:**

In economically advanced nations, a robust infrastructure and financial capacity often facilitate substantial investments in renewable energy technologies. Developed countries typically have the resources to deploy solar, wind, and other renewable projects on a large scale. Government incentives, policy frameworks, and public awareness further propel the transition towards cleaner energy sources. Developed nations aim not only to reduce carbon emissions but also to establish themselves as leaders in the burgeoning green economy, fostering innovation and job creation in the renewable sector.

- **Under-Developed Countries:**

Conversely, many non-developed countries face hurdles such as limited financial resources, inadequate infrastructure, and competing socio-economic priorities. Despite possessing abundant renewable resources, these nations may struggle to harness them effectively. International assistance and partnerships play a pivotal role in supporting renewable energy initiatives in non-developed countries. Initiatives focus not only on improving energy access but also on leapfrogging traditional, carbon-intensive development pathways, thereby promoting sustainable growth.

- **Common Challenges:**

Both developed and non-developed countries grapple with challenges like intermittency, storage, and the integration of renewable energy into existing grids. Additionally, the need for technology transfer, capacity building, and equitable access underscores the global nature of the renewable energy transition. Collaboration between developed and non-developed countries is crucial to address these challenges and ensure a just and inclusive energy transition.

In essence, while developed countries showcase leadership in renewable energy deployment, non-developed countries hold immense potential for sustainable development through leapfrogging to cleaner technologies. Bridging the gap requires a concerted global effort, acknowledging the diverse contexts and needs of both developed and non-developed nations on the path to a more sustainable energy future.

Objective 2: Sample Of Countries And Features:

The data set contains following features:

- Country Name
- Time
- Renewable energy consumption (% of total final energy consumption)
- Renewable electricity output (% of total electricity output)
- Fossil fuel energy consumption (% of total)
- Electricity production from renewable sources, excluding hydroelectric (% of total)
- Electric power consumption (kWh per capita)

The detail explanation of each column describe below.

i. Country Name:

The selected sample of countries are:

- Australia
- Argentina
- Bahrain
- Nepal
- New Zealand
- Nepal
- Pakistan
- Qatar
- Sri-Lanka
- United Arab Emirates
- United states

ii. Time: The data is ten of years as per the demand starting from 2005 to 2014

iii. Renewable energy consumption:

Renewable Energy Consumption refers to the amount of energy generated from renewable sources such as solar, wind, hydropower, geothermal, and biomass. Renewable energy is derived from resources that are naturally replenished and have a minimal environmental impact compared to conventional, non-renewable sources like fossil fuels.

- **<u>Percentage (%):</u>** The percentage value indicates the proportion of electricity generated from renewable sources relative to the total electricity output. A higher percentage suggests a greater reliance on renewable energy in the electricity generation mix, signaling progress toward a more sustainable and environmentally friendly energy system.

iv. <u>Renewable electricity output:</u>

Renewable electricity output is a metric that expresses the percentage of electricity generated from renewable sources such as solar, wind, hydropower, geothermal, and biomass in comparison to the total electricity output within a specific region, country, or globally. This metric provides insights into the share of renewable energy in the overall electricity generation mix.

- **<u>Total Electricity Output:</u>** This represents the overall quantity of electricity generated within a specific area, including electricity from all sources, both renewable and non-renewable. It encompasses electricity produced for various purposes, including residential, commercial, industrial, and public sectors.

- **<u>Percentage (%):</u>** The percentage value indicates the proportion of electricity generated from renewable sources relative to the total electricity output. A higher percentage suggests a greater reliance on renewable energy in the electricity generation mix, signaling progress toward a more sustainable and environmentally friendly energy system.

v. <u>Fossil fuel energy consumption:</u>

Fossil fuel energy consumption (% of total) is a metric that quantifies the proportion of energy derived from fossil fuels in comparison to the total energy consumption within a specific region, country, or globally. This refers to the amount of energy obtained from fossil fuels, which include coal, oil, and natural gas. Fossil fuels are non-renewable resources that are formed from the remains of ancient plants and animals. They have been the dominant sources of energy for various applications, including electricity generation, transportation, and industrial processes.

- **<u>Percentage (%):</u>** The percentage value indicates the proportion of energy derived from fossil fuels relative to the total energy consumption. A higher percentage suggests a greater reliance on fossil fuels, while a lower percentage indicates a lower dependency, potentially reflecting a more diverse and sustainable energy mix.

vi. **ELECTRICITY PRODUCTION FROM RENEWABLE SOURCES, EXCLUDING HYDROELECTRIC:**

Electricity production from renewable sources, excluding hydroelectric (% of total) is a metric that measures the percentage of electricity generated from renewable sources, excluding hydroelectric power, relative to the total electricity production within a specific region, country, or globally. This metric provides insights into the diversification of the renewable energy mix, excluding the contribution of large-scale hydroelectric projects.

- **Total Electricity Production:** This represents the overall quantity of electricity generated within a specific area, encompassing electricity produced from all sources, both renewable and non-renewable.

- **Percentage (%):** The percentage value indicates the proportion of electricity generated from renewable sources other than hydroelectric relative to the total electricity production. A higher percentage suggests a greater reliance on non-hydro renewable energy technologies in the electricity generation mix, reflecting efforts to promote a more diversified and environmentally sustainable energy portfolio.

Monitoring this metric is essential for assessing progress in reducing reliance on conventional, non-renewable sources of electricity and promoting the adoption of a broader range of renewable technologies. It reflects a nuanced approach to renewable energy, recognizing the significance of various sources beyond hydroelectric power in the pursuit of a cleaner and more sustainable energy system.

vii. **Electric power consumption (kWh per capita):**

Electric power consumption is a metric that quantifies the average amount of electric power consumed per person in a given population over a specific period, usually measured in kilowatt-hours (kWh). This metric is a key indicator of the electricity consumption habits and efficiency within a region or country and is often used to assess energy

- **Power Consumption:** This refers to the total amount of electrical energy consumed within a specific geographic area, typically measured in kilowatt-hours (kWh). Electric power consumption includes electricity used for various purposes such as residential lighting, heating, cooling, industrial processes, and commercial activities.

- **Per Capita:** This Latin term translates to "per person." In the context of this metric, "per capita" means that the total electric power consumption is divided by the population of the region or country to determine the average consumption per individual.

- **kWh** (Kilowatt-Hours): Kilowatt-hours are a unit of electrical energy, representing the amount of energy consumed when a device with a power rating of one kilowatt operates for one hour.

The formula for calculating electric power consumption per capita is:

$$Power\ Consumption\ per\ capita = \frac{P}{N}$$

Where,

P = Total Power Consumption
N = Population

Objective 3: ANALYSIS:

i) EDA:

Exploratory Data Analysis (EDA) is a method used to summarize, visualize, and understand key characteristics and patterns within a dataset. It involves techniques like plotting, summarizing statistics, and examining relationships between variables to uncover trends, anomalies, and insights. EDA helps in identifying outliers, understanding the distribution of data, detecting patterns, and guiding further analysis or modeling. Its primary goal is to gain an initial understanding of the data's structure, allowing for informed decision-making and hypothesis formulation before more complex analyses are conducted.

CODE:

```
file_path <- "E:/tab/RE_data_upd.csv"
df <- read.csv(file_path)
head(df)
```

OUTPUT:

Country.Name	Time	Renewable.energy.consumption....of.total.final.energy.consumption.	Renewable.electricity.output....of.total.electricity.output.	Fossil.fuel.energy.cons
Australia	2005	8.71	8.803707	
Australia	2006	8.85	9.298308	
Australia	2007	8.95	8.848072	
Australia	2008	8.79	8.115258	
Australia	2009	7.15	7.497934	
Australia	2010	8.16	9.611980	

- **To find no of rows and columns:**

CODE:

```
num_rows <- nrow(df)
num_columns <- ncol(df)


cat("Number of rows:", num_rows, "\n")

cat("Number of columns:", num_columns, "\n")
```

OUTPUT:

```
Number of rows: 100
Number of columns: 7
```

- **Data Info:**

Str(df)

OUTPUT:

```
'data.frame':   100 obs. of  7 variables:
 $ Country.Name                                                           : Factor w/ 10 levels "Argentina","A
ustralia",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Time                                                                   : int  2005 2006 2007 2008 2009 2010
2011 2012 2013 2014 ...
 $ Renewable.energy.consumption....of.total.final.energy.consumption.     : num  6.71 6.85 6.95 6.79 7.15 8.16
8.32 8.3 9.22 9.33 ...
 $ Renewable.electricity.output....of.total.electricity.output.           : num  8.8 9.3 8.65 8.12 7.5 ...
 $ Fossil.fuel.energy.consumption....of.total.                            : num  94.2 94.3 94.3 94.4 95.5 ...
 $ Electricity.production.from.renewable.sources..excluding.hydroelectric....of.total.: num  2.1 2.46 2.75 3.22 2.74 ...
 $ Electric.power.consumption..kWh.per.capita.                            : num  10571 10617 10973 10749 10792
```

ii) Missing values from the set:

CODE:

missing_values <- colSums(is.na(df))
missing_values

OUTPUT:

```
              Country.Name    0
                      Time    0
Renewable.energy.c...         0
Renewable.electricit...       0
Fossil.fuel.energy.co...      12
Electricity.productio...      0
Electric.power.cons...        0
```

In the above observation fossil fuels column contains 12 null values

CODE:

```
column_name <- "Fossil.fuel.energy.consumption....of.total."
if (column_name %in% names(df)) {
  # Check if there are any non-missing values in the column
  if (any(!is.na(df[[column_name]]))) {
        if (!is.numeric(df[[column_name]])) {
      df[[column_name]] <- as.numeric(as.character(df[[column_name]]))
    }

    # Check if there are any missing values in the column
    if (any(is.na(df[[column_name]]))) {
      column_mean <- mean(df[[column_name]], na.rm = TRUE)
      df[[column_name]] <- ifelse(is.na(df[[column_name]]), column_mean, df[[column_name]])
    }
  } else {
   print("All non-missing values in the column are missing.")
  }
} else {
print("Column not found in the data frame.") }
print(df)
```

OUTPUT:

```
   Fossil.fuel.energy.consumption....of.total.
1                                    94.215852
2                                    94.283165
3                                    94.293104
4                                    94.351488
5                                    95.510059
6                                    94.424742
7                                    94.421656
8                                    94.377006
9                                    93.708049
10                                   93.386836
11                                   89.053284
12                                   88.653228
13                                   89.225378
14                                   90.651789
15                                   89.611993
16                                   89.503243
17                                   88.853811
18                                   89.024907
19                                   88.966064
20                                   87.722407
21                                   71.098654
22                                   71.098654
23                                   97.831772
24                                   71.098654
25                                   99.782448
26                                   71.098654
27                                   98.785946
28                                   71.098654
29                                   71.098654
30                                   99.365847
31                                   10.646797
32                                    8.831054
33                                    8.615627
34                                    9.886058
35                                   11.039139
```

Objective 4.1: Comprehensive Descriptive Statistical Analysis:

Comprehensive descriptive statistical analysis has been done in objective one. This analysis helped in finding the mean, mode, and median of the data set on renewable energy. Along with skewness graphs and graphs of descriptive statistical analysis.

CODE:

```
summary(df[[column_name]])
```

OUTPUT:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  8.616  60.022  71.099  71.099  92.103 100.000 
```

CODE:

```
mean_values <- colMeans(df[, 2:ncol(df)], na.rm = TRUE)
median_values <- sapply(df[, 2:ncol(df)], median, na.rm = TRUE)
mode_values <- sapply(df[, 2:ncol(df)], function(x) {
  ifelse(length(unique(x[!is.na(x)])) > 1, as.numeric(names(sort(table(x), decreasing =
TRUE)[1])), NA)
})
std_dev_values <- apply(df[, 2:ncol(df)], 2, sd, na.rm = TRUE)
skewness_values <- sapply(df[, 2:ncol(df)], skewness, na.rm = TRUE)
kurtosis_values <- sapply(df[, 2:ncol(df)], kurtosis, na.rm = TRUE)

summary_stats <- data.frame(
  Mean = mean_values,
  Median = median_values,
  Mode = mode_values,
  Std_Dev = std_dev_values,
  Skewness = skewness_values,
  Kurtosis = kurtosis_values
)
print(summary_stats)
```

OUTPUT:

Mean, Median and Mode

```
                                                                                   Mean
Time                                                                        2009.500000
Renewable.energy.consumption....of.total.final.energy.consumption.            24.640600
Renewable.electricity.output....of.total.electricity.output.                  29.581318
Fossil.fuel.energy.consumption....of.total.                                   71.098654
Electricity.production.from.renewable.sources..excluding.hydroelectric....of.total.    2.720021
Electric.power.consumption..kWh.per.capita.                                 8351.417231
                                                                                 Median
Time                                                                       2009.5000000
Renewable.energy.consumption....of.total.final.energy.consumption.            8.7600000
Renewable.electricity.output....of.total.electricity.output.                 20.6252090
Fossil.fuel.energy.consumption....of.total.                                  71.0986537
Electricity.production.from.renewable.sources..excluding.hydroelectric....of.total.    0.2396798
Electric.power.consumption..kWh.per.capita.                                9885.7448945
                                                                                   Mode
Time                                                                         2005.00000
Renewable.energy.consumption....of.total.final.energy.consumption.              0.00000
Renewable.electricity.output....of.total.electricity.output.                    0.00000
Fossil.fuel.energy.consumption....of.total.                                    71.09865
Electricity.production.from.renewable.sources..excluding.hydroelectric....of.total.    0.00000
Electric.power.consumption..kWh.per.capita.                                    75.48000
                                                                                Std_Dev
Time                                                                           2.886751
Renewable.energy.consumption....of.total.final.energy.consumption.            29.083677
Renewable.electricity.output....of.total.electricity.output.                  32.180204
Fossil.fuel.energy.consumption....of.total.                                   25.948640
Electricity.production.from.renewable.sources..excluding.hydroelectric....of.total.    5.081624
Electric.power.consumption..kWh.per.capita.                                 6689.827390
```

Skewness And Kurtosis:

```
                                                                               Skewness
Time                                                                         0.00000000
Renewable.energy.consumption....of.total.final.energy.consumption.           1.01889884
Renewable.electricity.output....of.total.electricity.output.                 0.98840884
Fossil.fuel.energy.consumption....of.total.                                 -0.97465613
Electricity.production.from.renewable.sources..excluding.hydroelectric....of.total.  2.50028044
Electric.power.consumption..kWh.per.capita.                                  0.08053619
                                                                               Kurtosis
Time                                                                         -1.2595800
Renewable.energy.consumption....of.total.final.energy.consumption.           -0.2924763
Renewable.electricity.output....of.total.electricity.output.                 -0.2073380
Fossil.fuel.energy.consumption....of.total.                                   0.1485916
Electricity.production.from.renewable.sources..excluding.hydroelectric....of.total.  5.6930984
Electric.power.consumption..kWh.per.capita.                                  -1.3345297
```

- **SKEWNESS:**

  Skewness is a measurement of the distortion of symmetrical distribution or asymmetry in a data set. Ther are three conditions for skewness.

  - **Positive skewness:** The tail of the distribution extends more to the right, indicating more extreme values on the positive side.

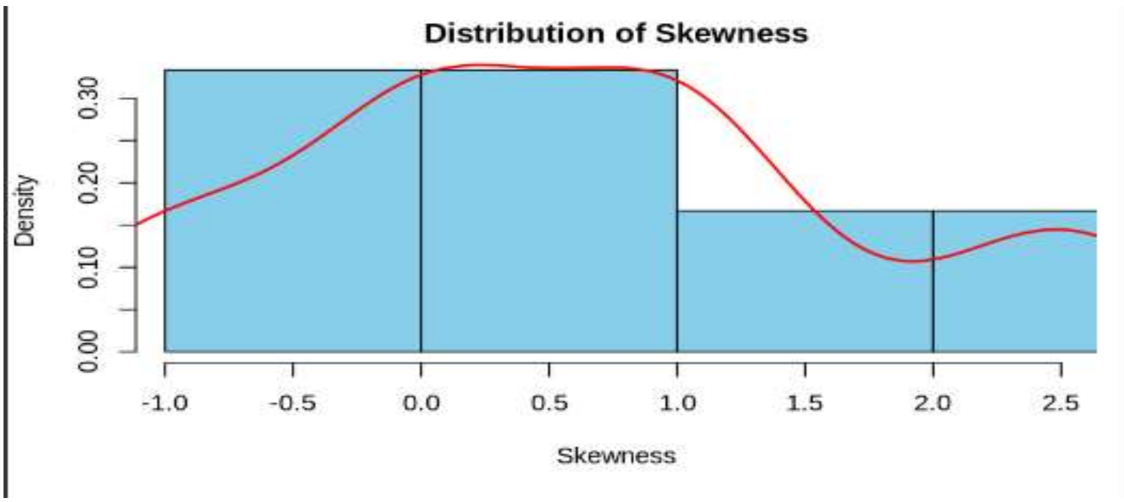
  - **Negative skewness:** The tail of the distribution extends more to the left, suggesting more extreme values on the negative side.

  - **Zero skewness:** The distribution is symmetric, with equal tails on both sides. Positive and negative skewness indicate the direction and extent of asymmetry, offering insights into the distribution's shape and tendencies toward higher or lower values.

CODE:

```
par(mfrow = c(2, 1), mar = c(4, 4, 2, 1))
hist(summary_stats$Skewness, main = 'Distribution of Skewness', col = 'skyblue',
xlab = 'Skewness', xlim = range(summary_stats$Skewness), prob = TRUE)
lines(density(summary_stats$Skewness), col = 'red', lwd = 2)
```

OUTPUTS:

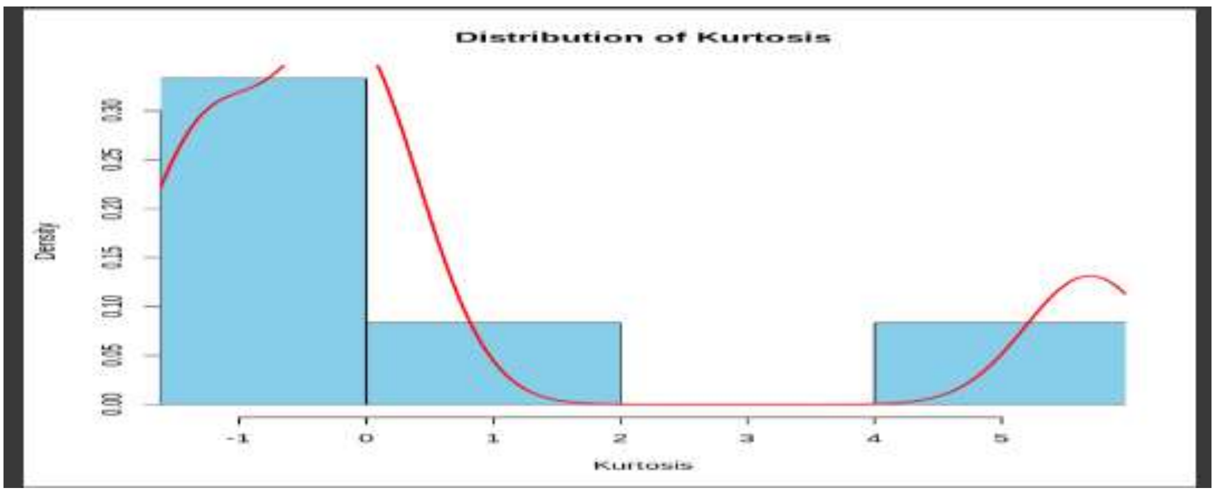


- **KURTOSIS:**

kurtosis is the weighting of the tails and the concentration of the data around the mean of a distribution. Kurtosis evaluates whether the distribution has lighter or heavier tails than a normal distribution.

CODE:

```
hist(summary_stats$Kurtosis, main = 'Distribution of Kurtosis', col = 'skyblue', xlab = 'Kurtosis', xlim = range(summary_stats$Kurtosis), prob = TRUE)
lines(density(summary_stats$Kurtosis), col = 'red', lwd = 2)
par(mfrow = c(1, 1))
```



Objective 4.2: <u>Correlation Analysis:</u>

Correlation is a statistical metric used to assess the link or relationship between two variables. It measures the degree to which changes in one variable are correlated with those in another.

There are 3 types of correlation:

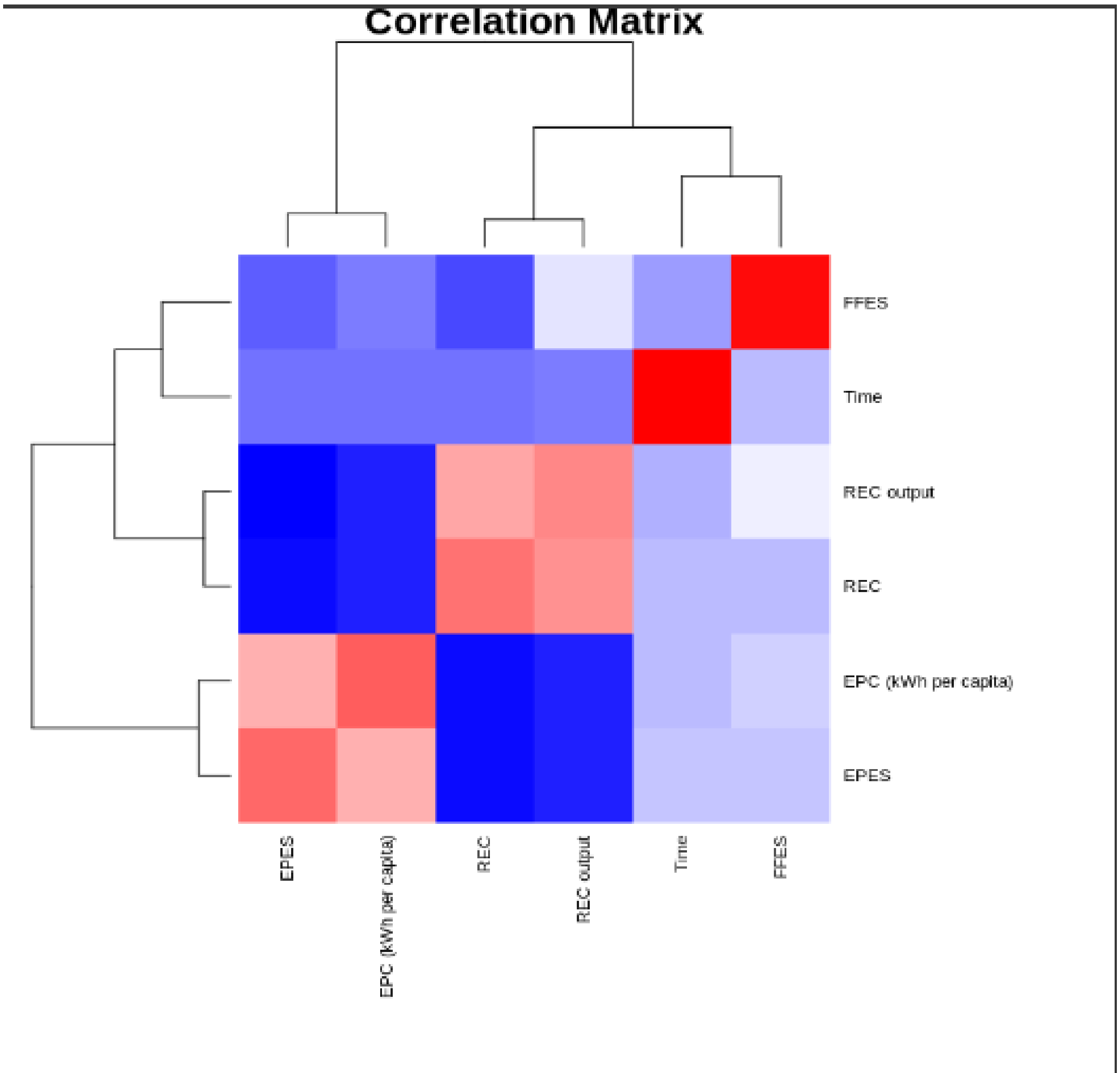
- **Positive correlation:** Data points form a pattern sloping upwards.
- **Negative correlation:** Data points form a pattern sloping downwards.
- **Zero correlation:** Data points appear randomly scattered

Let's examine how our dataset's columns correlate with each other.

CODE:

```
column_name_mapping <- c('Time' = 'Time',
'Renewable energy consumption (% of total final energy consumption)' = 'REC',
'Renewable electricity output (% of total electricity output)' = 'REC output',
'Electricity production from renewable sources, excluding hydroelectric (% of
total)' = 'EPES',
'Fossil fuel energy consumption (% of total)' = 'FFES',
'Electric power consumption (kWh per capita)' = 'EPC (kWh per capita)')
numeric_columns <- sapply(df, is.numeric)
numeric_df <- df[, numeric_columns]
correlation_matrix <- cor(numeric_df)
colnames(correlation_matrix) <- rownames(correlation_matrix) <- column_name_mapping
heatmap(correlation_matrix,
    col = colorRampPalette(c("blue", "white", "red"))(50),
    main = 'Correlation Matrix',
    cexRow = 0.8, cexCol = 0.8, margins = c(10, 10),
    labRow = names(correlation_matrix),
    labCol = names(correlation_matrix))
```

OUTPUTS:


Correlation Matrix
FFES
Time
REC output
REC
EPC (kWh per capita)
EPES
EPES
EPC (kWh per capita)
REC
REC output
Time
FFES

Objective 4.3: **Hypothesis Testing:**

Hypothesis testing is a statistical method for examining if there's enough evidence in a sample to support a claim about a population. It involves setting up two opposing hypotheses, collecting data, and using statistical tests to decide if the evidence supports rejecting the null hypothesis in favor of an alternative. This process helps in making informed decisions and validating assumptions in research or analysis

Hypothesis 1: Developing nations have a higher percentage of renewable energy consumption compared to underdeveloped nations.

•Null Hypothesis (H0): The percentage of renewable energy consumption is the same between developing and underdeveloped nations.

•Alternative Hypothesis (H1): Developing nations have a higher percentage of renewable energy consumption compared to underdeveloped nations.

CODE:

```
developing_countries <- c('Nepal', 'New Zealand', 'Sri Lanka')
underdeveloped_countries <- c('Australia', 'Argentina', 'Bahrain', 'Pakistan', 'Qatar','United
States','United Arab Emirates')
developing_nations <- df[df$`Country.Name` %in% developing_countries, ]
underdeveloped_nations <- df[df$`Country.Name` %in% underdeveloped_countries, ]
if (nrow(developing_nations) == 0 || nrow(underdeveloped_nations) == 0) {
  print("No matching countries found in the original dataset.")
} else {
  t_test_result <- t.test(
developing_nations$`Renewable.energy.consumption....of.total.final.energy.consumption.`,
underdeveloped_nations$`Renewable.energy.consumption....of.total.final.energy.consumption.`)
cat("Hypothesis 1: T-statistic =", t_test_result$statistic, ", p-value =", t_test_result$p.value,
"\n")}
```

OUTPUT:

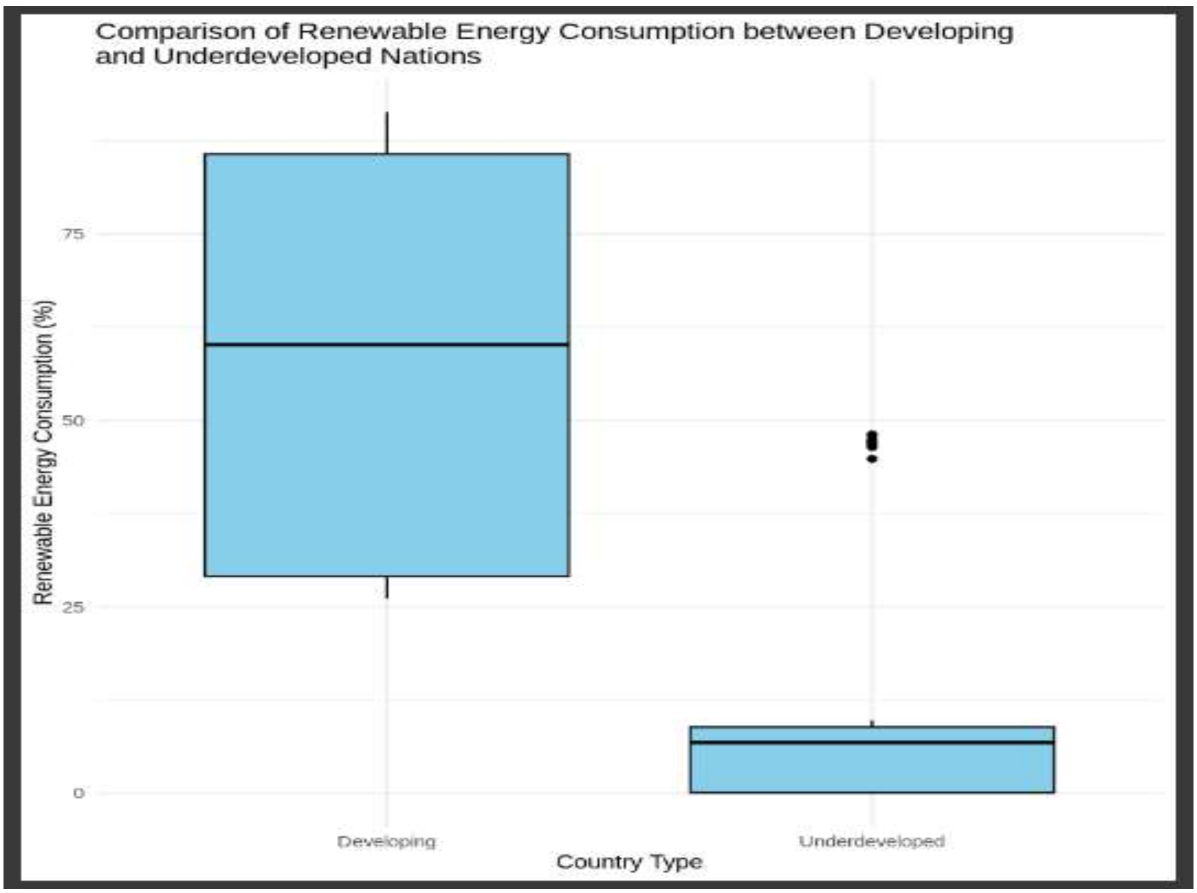
```
Hypothesis 1: T-statistic = 9.767608 , p-value = 4.959074e-12
```

- **CAT-PLOT:**
  Cat-plot is a categorical plot used to show the relationship between numerical and categorical variables using different types of plots.

CODE:

```
combined_data <- rbind(
  transform(developing_nations, Country_Type = "Developing"),
  transform(underdeveloped_nations, Country_Type = "Underdeveloped")
)
ggplot(combined_data, aes(x = Country_Type, y =
`Renewable.energy.consumption....of.total.final.energy.consumption.`)) +
geom_boxplot(fill = 'skyblue', color = 'black') +
labs(title = 'Comparison of Renewable Energy Consumption between Developing
and Underdeveloped Nations',
 x = 'Country Type',
 y = 'Renewable Energy Consumption (%)') +
theme_minimal()
```

OUTPUT:



Hypothesis 2: The percentage of renewable electricity output is positively correlated with the percentage of renewable energy consumption in both developing and underdeveloped nations.

•Null Hypothesis (H0): There is no correlation between the percentage of renewable electricity output and the percentage of renewable energy consumption.

•Alternative Hypothesis (H1): There is a positive correlation between the percentage of renewable electricity output and the percentage of renewable energy consumption

CODE:

```
correlation_result <- cor.test(
 df$`Renewable.electricity.output....of.total.electricity.output.`,
 df$`Renewable.energy.consumption....of.total.final.energy.consumption.`,
method = "spearman")
cat("Hypothesis 2: Spearman correlation coefficient =",
correlation_result$estimate, ", p-value =", correlation_result$p.value, "\n")
```

OUTPUT:

```
Hypothesis 2: Spearman correlation coefficient = 0.9308278 , p-value = 1.240995e-44
```

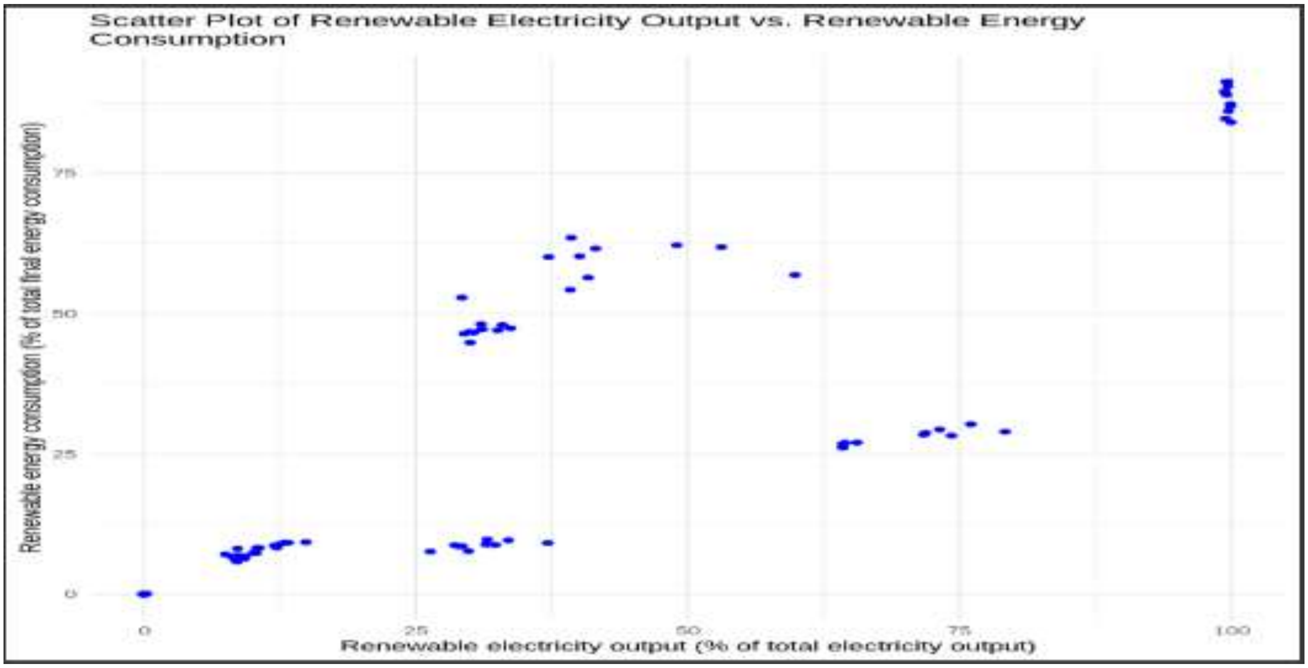
- **<u>SCATTER PLOT:</u>**

  A scatter plot displays the relationship between two numerical variables, depicting points on a graph to reveal patterns or correlations between them

CODE

```
ggplot(df, aes(x = `Renewable.electricity.output....of.total.electricity.output.`, y =
`Renewable.energy.consumption....of.total.final.energy.consumption.`)) +
 geom_point(color = 'blue') +
 labs(title = 'Scatter Plot of Renewable Electricity Output vs. Renewable Energy
Consumption',
 x = 'Renewable electricity output (% of total electricity output)',
 y = 'Renewable energy consumption (% of total final energy consumption)') +
 theme_minimal()
```

OUTPUT:



Objective 4.4: <u>Regression Analysis:</u>

Regression is a statistical process that can be used in different fields, such as machine learning, statistics, and data science, to model the relationship between a dependent variable and one or more independent variables. The goal of regression analysis is to understand the nature of this relationship and make predictions or infer insights based on the available data. If we say in simple words regression helps us in examine how changes in independent variables are interlinked with dependent variables.

There are different types of regression analysis:

- Linear regression
- Multiple regression
- Logistic regression
- Polynomial regression

- **MULTIPLE LINEAR REGRESSION:**

Multiple linear regression is a statistical method that extends simple linear regression to include multiple independent variables. It helps in understanding the relationship between a dependent variable and multiple independent variables, and is used to predict the outcome of the dependent variable. The method uses a linear equation to model the relationship between the variables, and the equation includes an intercept and coefficients for each independent variable

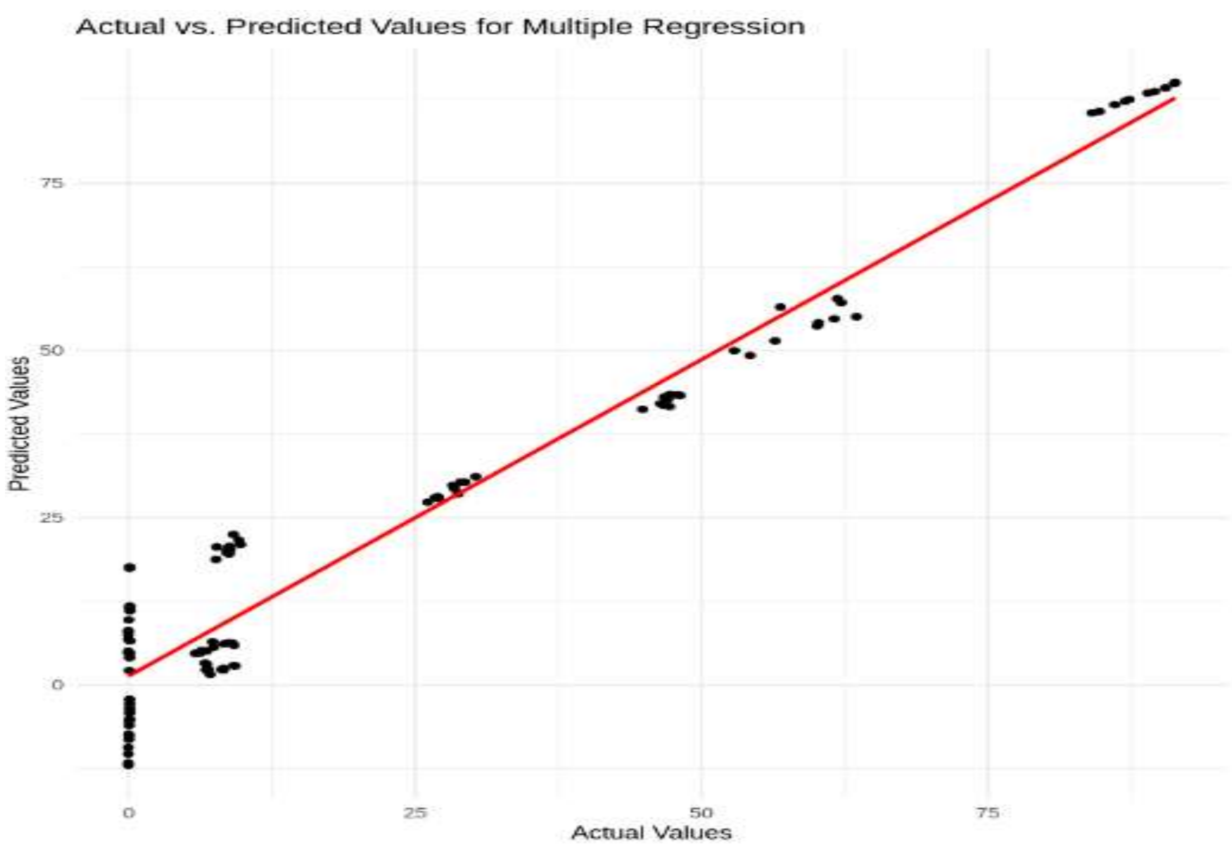
Let's understand how the code of Multi Linear Regression analysis works.

CODE

- X <- df[, c('Renewable.electricity.output....of.total.electricity.output.',

```
 'Fossil.fuel.energy.consumption....of.total.',
 'Electricity.production.from.renewable.sources..excluding.hydroelectric....of.total.',
 'Electric.power.consumption..kWh.per.capita.')]
y <- df$`Renewable.energy.consumption....of.total.final.energy.consumption.`
X <- cbind(1, X)
model <- lm(y ~ ., data = as.data.frame(X))
y_pred <- predict(model, newdata = as.data.frame(X))
ggplot(data = as.data.frame(cbind(y, y_pred)), aes(x = y, y = y_pred)) +
 geom_point() +
 geom_smooth(method = 'lm', se = FALSE, color = 'red') +
 labs(title = 'Actual vs. Predicted Values for Multiple Regression',
 x = 'Actual Values',
 y = 'Predicted Values') +
 theme_minimal()
```

OUTPUT:



- **ERROR ANALYSIS:**

Error analysis metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) assess the performance of predictive models.

- ***MSE (Mean Square Error):*** Measures the average squared difference between predicted and actual values, emphasizing larger errors due to squaring.
- ***RMSE (Root Mean Square Error):*** Provides the square root of MSE, offering an easily interpretable metric in the same units as the dependent variable.
- ***R^2:*** Represents the proportion of variance in the dependent variable explained by the model, with values closer to 1 indicating better fit.

CODE:

```
mse <- mean((y - y_pred)^2)
rmse <- sqrt(mse)
rsquared <- summary(model)$r.squared
cat("Mean Squared Error (MSE):", mse, "\n")
cat("Root Mean Squared Error (RMSE):", rmse, "\n")
cat("R-squared:", rsquared, "\n")
```

OUTPUT:

```
Mean Squared Error (MSE): 44.66528
Root Mean Squared Error (RMSE): 6.683209
R-squared: 0.9466621
```

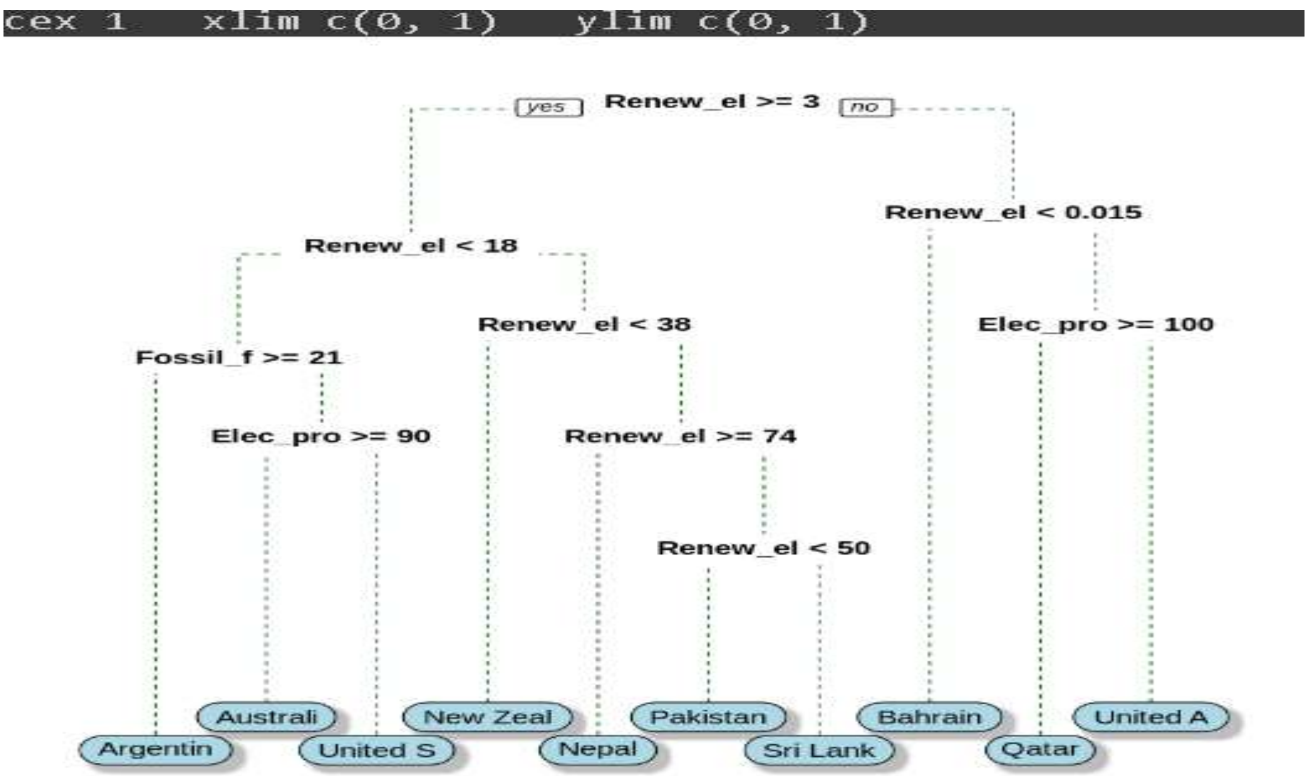
DECISION TREE:

A Decision Tree model is a predictive algorithm that uses a tree-like structure to make decisions by dividing data into smaller subsets based on different attributes. It operates by sequentially splitting the data into branches, identifying the most significant features, and creating a hierarchical tree to predict outcomes or classifications for new data based on these decisions

CODE:

```
df_tree <- df
new_names <- c("Target", "Renew_cons", "Renew_elec", "Fossil_fuel", "Elec_prod_renew",
"Elec_power_capita")
colnames(df_tree) <- new_names
tree_model <- rpart(Target ~ Renew_cons + Renew_elec + Fossil_fuel +
                  Elec_prod_renew + Elec_power_capita,
                data = df_tree, method = "class")
par(mar = c(0, 2, 0, 5) )
prp(tree_model, box.col = "lightblue", branch.lty = 2, branch.col = "darkgreen", shadow.col =
"gray", fallen.leaves = TRUE, trace = TRUE)
```

OUTPUT:



Objective 4.5: <u>**Time Series Analysis:**</u>

Data gathered over time at regular intervals is examined and modeled through time series analysis. It seeks to comprehend the behaviors, patterns, and trends present in the data in order to forecast or predict future values. Using statistical approaches to find seasonality, trends, and any underlying patterns within the time-based data, this research looks at how previous values affect future results. Following are the models of Time-series:

- ARIMA
- Exponential Smoothing
- STL (Seasonal Decomposition of Time Series)
- SARIMA VAR/VECM (Vector Autoregression/Vector Error Correction Models)
- Prophet
- LSTM Networks

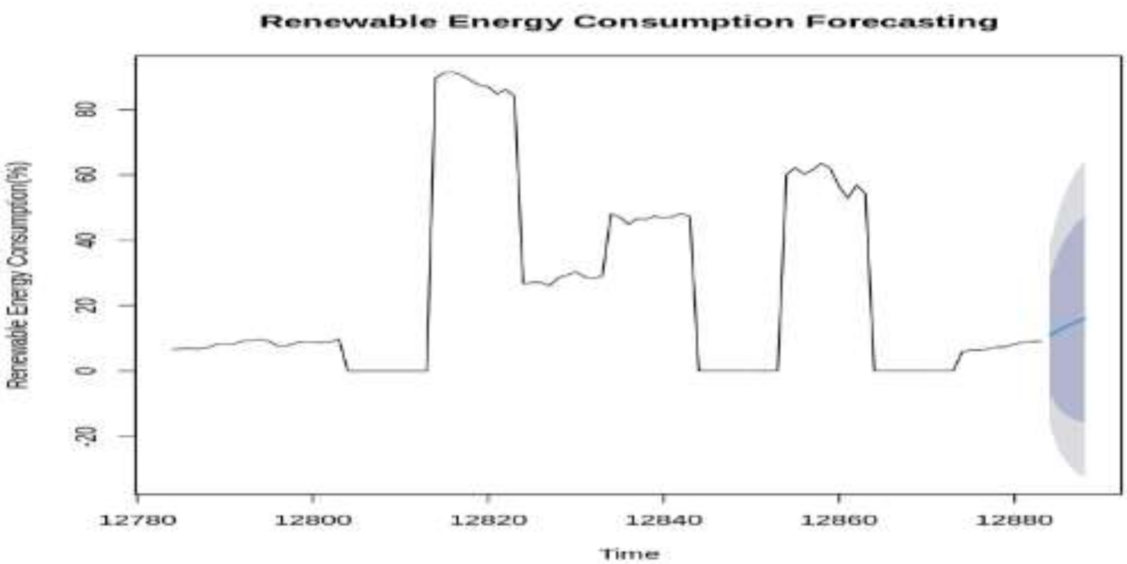
- **ARIMA:**

ARIMA (AutoRegressive Integrated Moving Average) is a time series forecasting model used to predict future values based on past observations. It combines autoregression (AR) of past values, differencing to achieve stationarity (I for Integrated), and moving average (MA) of past forecast errors. The model considers the data's trend, seasonality, and noise to forecast future values, making it effective for analyzing and predicting time-dependent data patterns.

CODE:

```
df$Time <- as.Date(paste0(df$Time, "-01-01"))
ts_data <- ts(df$`Renewable.energy.consumption....of.total.final.energy.consumption.`,
frequency = 1, start = min(df$Time))
model_consumption <- auto.arima(ts_data)
forecast_consumption <- forecast(model_consumption, h = 5)
plot(forecast_consumption, main = 'Renewable Energy Consumption Forecasting',
xlab = 'Time', ylab = 'Renewable Energy Consumption(%)')
plot_consumption <- ggplot(df, aes(x = Time,
y = `Renewable.energy.consumption....of.total.final.energy.consumption.`,
color = `Country.Name`)) + geom_line() + labs(title = 'Renewable Energy Consumption Over
Time',  x = 'Time',  y = 'Renewable Energy Consumption (%)') + theme_minimal()
```

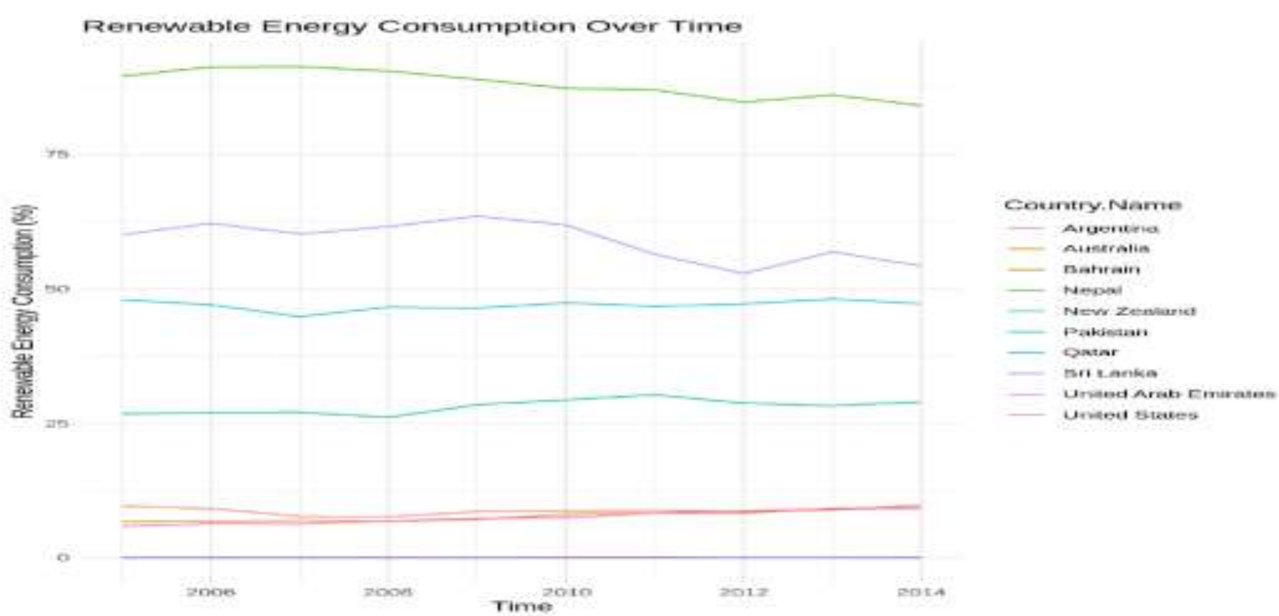
OUTPUT:



CODE:

```
plot_consumption <- ggplot(df, aes(x = Time, y =
`Renewable.energy.consumption....of.total.final.energy.consumption.`, color = `Country.Name`))
+ geom_line() + labs(title = 'Renewable Energy Consumption Over Time',
 x = 'Time',
 y = 'Renewable Energy Consumption (%)') + theme_minimal()
print(plot_consumption)
```

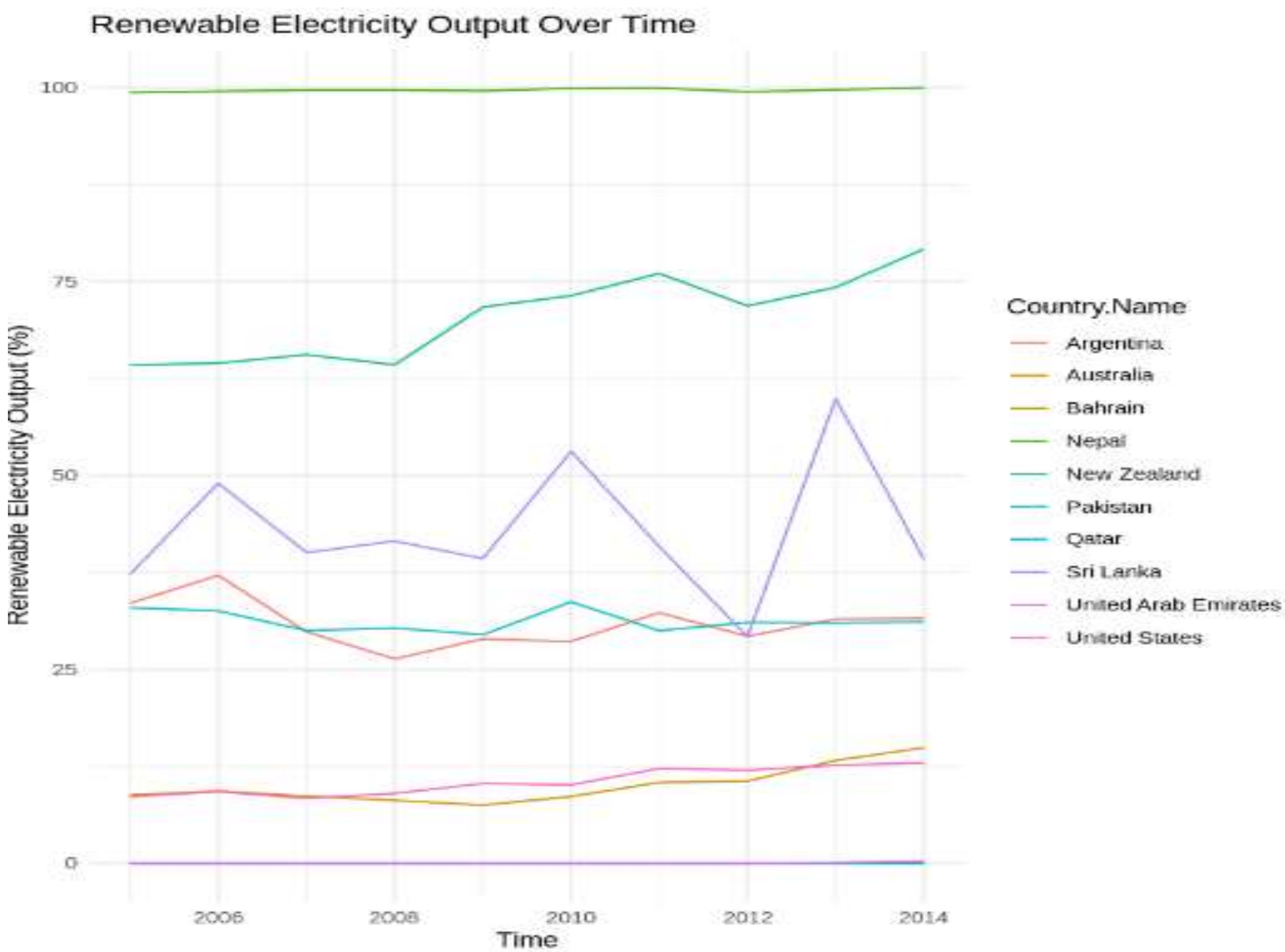
OUTPUT:



CODE:

```
plot_output <- ggplot(df, aes(x = Time, y =
`Renewable.electricity.output....of.total.electricity.output.`, color = `Country.Name`)) +
 geom_line() +
 labs(title = 'Renewable Electricity Output Over Time',
 x = 'Time',
 y = 'Renewable Electricity Output (%)') +
 theme_minimal()
print(plot_output)
```

OUTPUT:



Part 2: Interactive Dashboard Design:

Interactive dashboard design involves creating visual user interfaces that allow users to dynamically explore and interact with data. It includes features such as filters, drop-down menus, clickable elements and sliders that allow users to manipulate and analyze data in real time. The design focuses on user-friendly layouts, intuitive navigation and meaningful visualizations that effectively communicate knowledge. These dashboards often use graphs, charts and tables that update instantly based on user interactions, providing a more immersive and personalized experience for exploring data and gaining actionable insights.

Commonly used tools for dashboard design include

- Power BI
- Tableau
- Google Data Studio
- Sisense. Among these

Power BI is the primary tool utilized for our dashboard design needs.

POWER-BI:

Power BI is a powerful business analytics tool by Microsoft that allows users to visualize and share insights from their data. It enables easy transformation of raw data into visually compelling and interactive reports, dashboards, and visualizations. Commonly task we can done in Power-BI.

- **Data Cleaning:** Power BI facilitates data cleaning by allowing users to identify and handle missing or inconsistent data, remove duplicates, and apply transformations to prepare data for analysis.

- **Data Visualization**: It offers a wide range of visualization options like charts, graphs, maps, and tables to represent data visually, aiding in the interpretation of trends, patterns, and insights.

- **Data Transformation**: Power BI enables the transformation of raw data by shaping, combining, and modeling it to create meaningful datasets for analysis, using features like Power Query and DAX (Data Analysis Expressions).

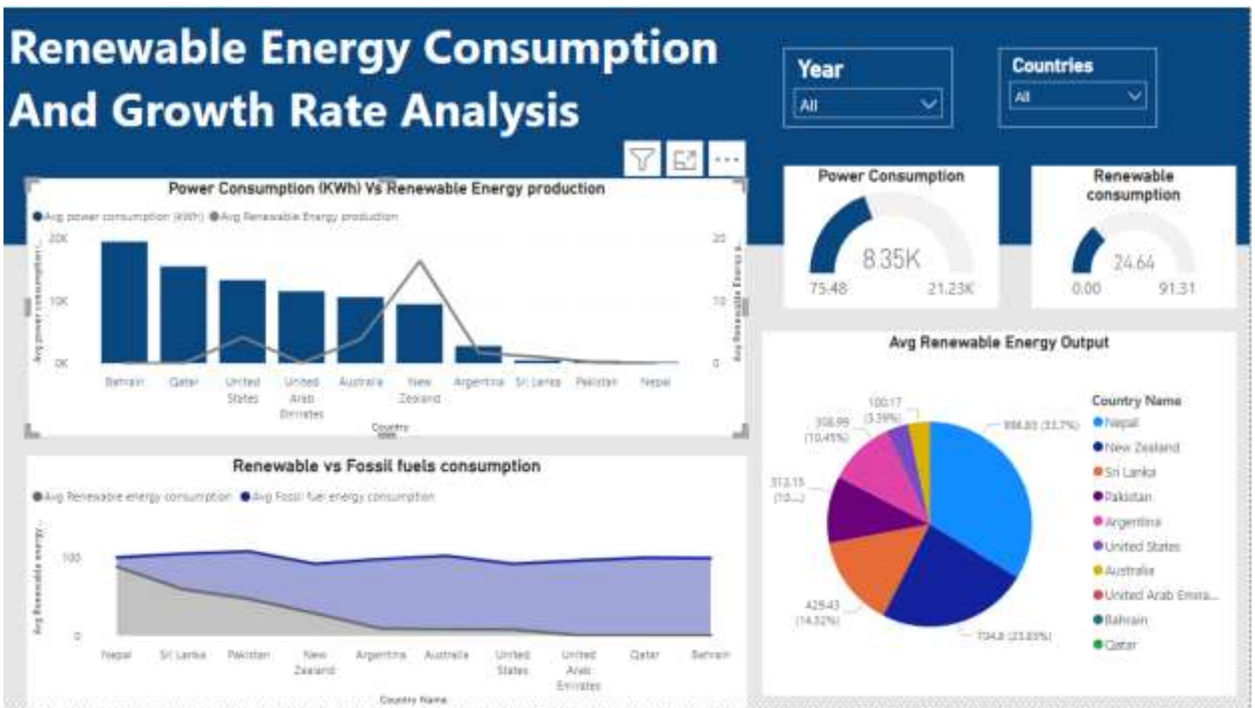
- **Data Modeling**: Users can create relationships between different data tables, build hierarchies, and define measures and calculated columns for in-depth analysis using Power BI's modeling capabilities.

- **Dashboard Creation**: Power BI allows the creation of interactive and personalized dashboards by combining multiple visualizations and reports into a single view, enhancing data exploration and decision-making.

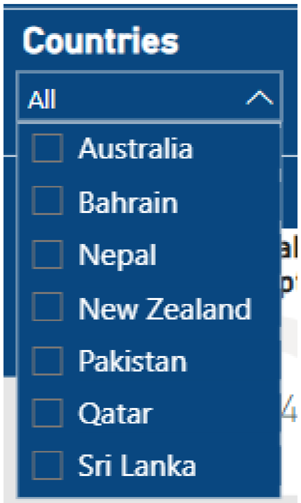
- **Report Generation:** It enables the creation of comprehensive reports with customized layouts, formatting, and storytelling features to communicate insights effectively.

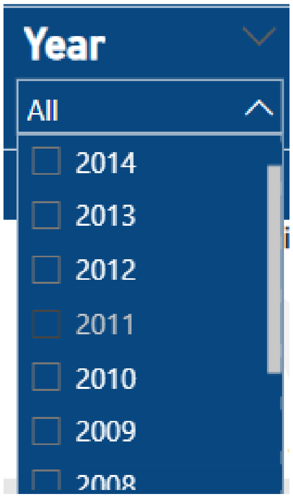
DASHBOARD USER INTERFACE (UI):

The Power BI dashboard showcasing our dataset appears as follows.



The analysis of various parameters varies based on both countries and the respective years under consideration.





Let's explore the individual visualizations within our dashboard

- <u>**RELATIONSHOP BETWEEN RENEWABLE ENERGY PRODUCTION VS POWER CONSUMPTION OF DIFFERENT COUNTRIES:**</u>

  The link between renewable energy production and power consumption in various countries represents the interaction between the creation of renewable energy sources and the general use of power in those countries. The purpose of this study is to determine how the output of renewable energy interacts with the quantity of power utilized. Understanding this link sheds light on how much countries rely on renewable energy sources to meet their energy needs. It provides insights into each country's energy environment and sustainability efforts by identifying trends, patterns, and potential linkages between renewable energy generation and electricity use. Furthermore, this investigation aids in determining the effectiveness of renewable energy techniques in satisfying the energy needs of various nations.

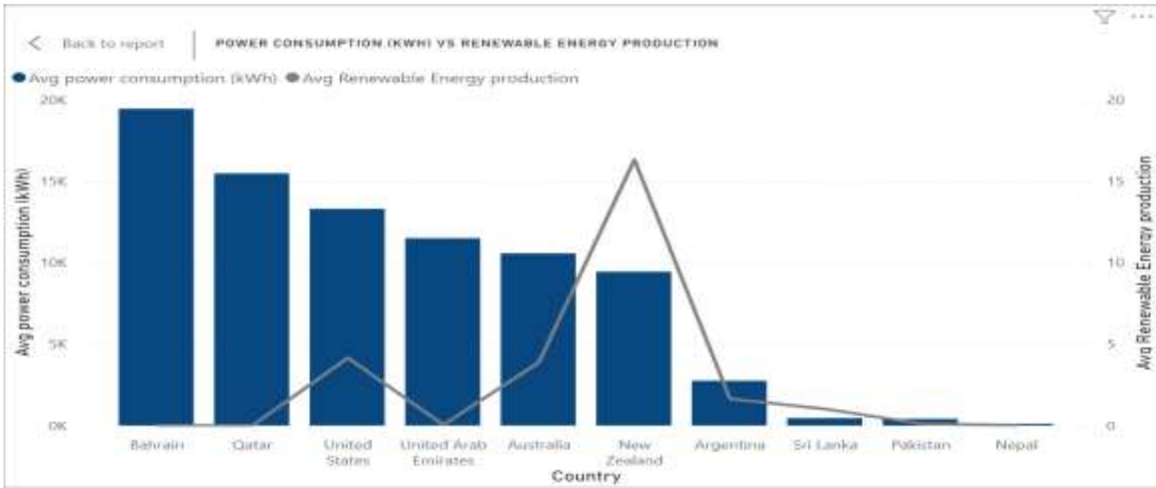


Fig: Comparison between power consumption and Renewable energy production

❖ ***ANALYSIS:***

The graph shows the average power consumption and average renewable energy production for each country in kilowatt-hours (kWh).

- The countries with the highest average power consumption are the United States, Australia and New Zealand.
- The countries with the highest average renewable energy production are the United States, Australia, and Qatar.
- The countries with the highest growth rate of renewable energy production are Qatar, Argentina, and the United Arab Emirates.
- The countries with the lowest growth rate of renewable energy production are New Zealand and Australia.

Overall, the graph shows that renewable energy is playing an increasingly important role in the global energy mix. However, there is still significant variation in renewable energy consumption and production between different countries.

- **<u>RELATIONSHIP BETWEEN RENEWABLE ENERGY CONSUMPTION VS FOSSIL FUEL ENERGY CONSUMPTION:</u>**

The relationship between renewable energy consumption and fossil fuel energy consumption is pivotal in understanding the transition towards sustainable energy sources. As renewable energy consumption increases, there's often an inverse impact on fossil fuel consumption. A rise in renewable energy usage typically correlates with reduced reliance on fossil fuels, reflecting efforts to mitigate environmental impact and address climate change. Analyzing this relationship helps gauge the effectiveness of renewable energy initiatives in displacing fossil fuel usage and promoting a greener energy mix. It showcases the pivotal shift from traditional, carbon-intensive energy sources towards cleaner, renewable alternatives, providing insights into the balance and dynamics between these contrasting energy types.

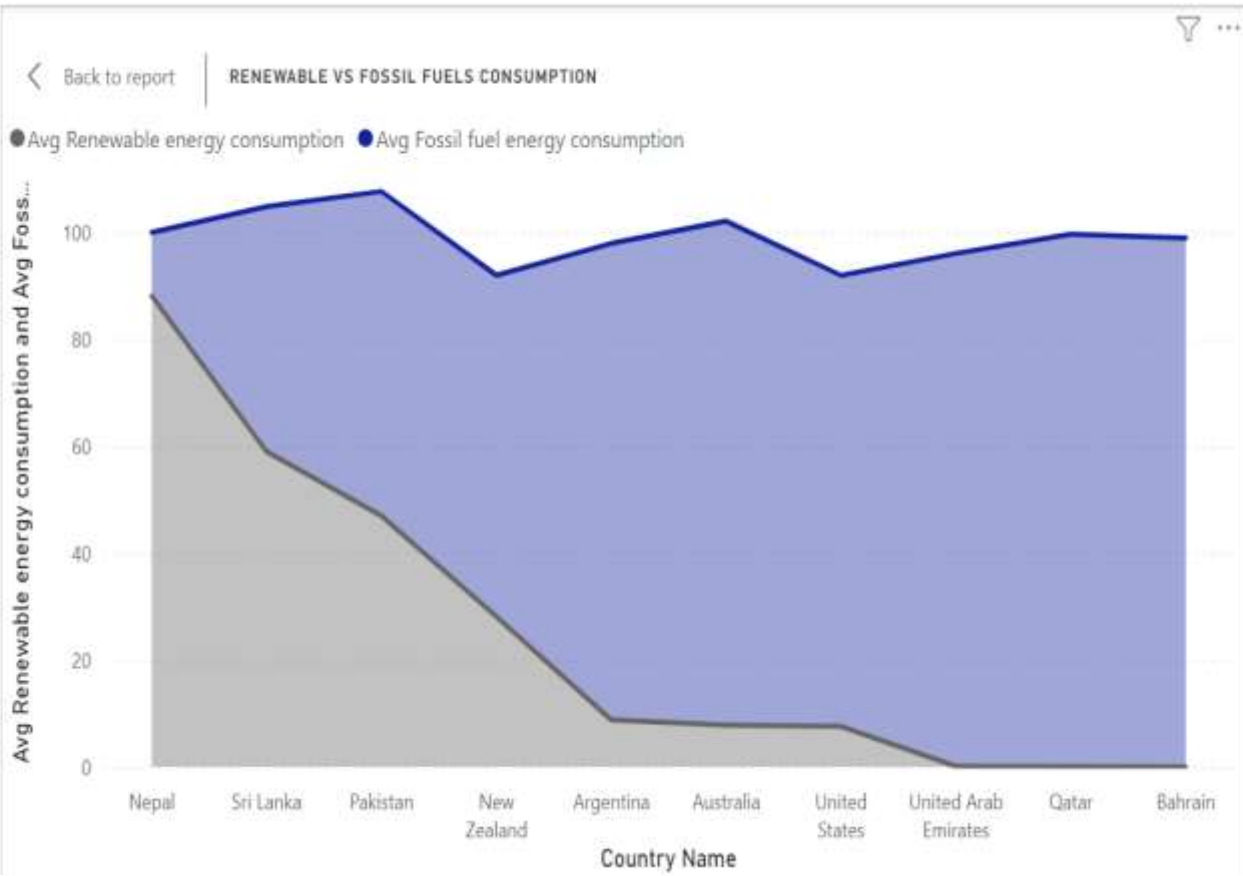


Fig: Comparison between Fossil Fuel production and Renewable energy production

❖ **ANALYSIS:**

The graph shows that fossil fuel consumption is still significantly higher than renewable energy consumption. However, renewable energy consumption is growing faster than fossil fuel consumption in countries. The following table shows the average renewable energy consumption and average fossil fuel consumption for each country:

- The countries with the highest renewable energy consumption are Nepal, Sri Lanka, and Pakistan, while the countries with the lowest renewable energy consumption are the United Arab Emirates, Qatar, and Bahrain.

- The countries with the fastest growth rate of renewable energy consumption are Nepal, Sri Lanka, and Pakistan, while the countries with the slowest growth rate of renewable energy consumption are the United Arab Emirates, Qatar, and Bahrain.

- **<u>GLOBAL COMPARISON OF RENEWABLE ENERGY OUTPUT:</u>**

Analyzing and comparing the production of renewable energy across different countries worldwide. It involves examining the output of renewable energy sources like solar, wind, hydroelectric, geothermal, and biomass, assessing the quantities generated by each country. By conducting this comparison, trends, disparities, and patterns in renewable energy production can be identified. This analysis provides insights into which countries are leading in renewable energy production, the distribution of various renewable sources among nations, and the global landscape of sustainable energy generation.

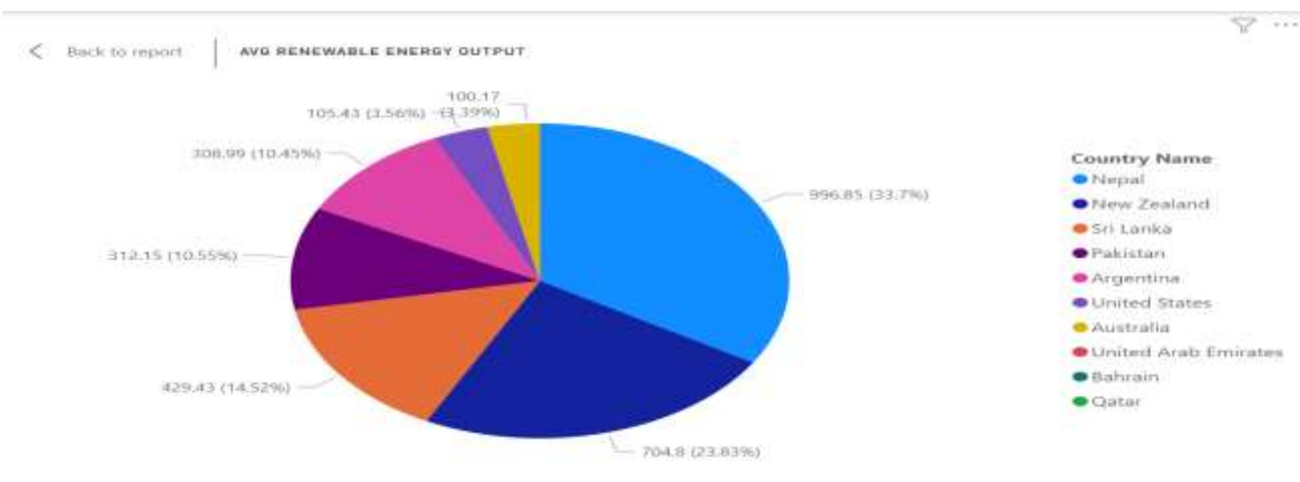


Fig: Analysis of the Pie Chart Showing the Percentage of Renewable Energy Output by Country

- **ANALYSIS:**

The pie chart shows that the percentage of renewable energy output varies significantly across countries.

- The countries with the highest percentage of renewable energy output are Nepal (99.7%), Sri Lanka (91.3%), and Pakistan (66.9%).
- The countries with the lowest percentage of renewable energy output are the United Arab Emirates (0.0%), Qatar (1.0%), and Bahrain (0.0%).
- The countries with the highest percentage of renewable energy output are typically developing countries with abundant renewable energy resources, such as hydropower and solar power.

These countries have made significant investments in renewable energy infrastructure in recent years. The countries with the lowest percentage of renewable energy output are typically oil-rich countries with limited renewable energy resources. These countries have less incentive to transition to renewable energy, as they are heavily reliant on oil exports for revenue. The pie chart also shows that the majority of countries in the dataset have a renewable energy output of less than 50%. This suggests that there is still significant room for growth in the global renewable energy sector.

In-Depth Insight:

Exploring In-Depth Insights within our Dashboard

NEPAL:

Nepal has the highest proportion of renewable energy consumption of any country in the dataset, with 92.6% of its energy coming from renewable sources. The majority of Nepal's renewable energy comes from hydropower, with 99% of its electricity generation coming from hydropower plants. Nepal has a significant amount of untapped hydropower potential, with an estimated 42,000 MW of potential capacity.

SRI LANKA:

Sri Lanka has the second-highest proportion of renewable energy consumption of any country in the dataset, with 91.3% of its energy coming from renewable sources. The majority of Sri Lanka's renewable energy comes from hydropower, with 85% of its electricity generation coming from hydropower plants. Sri Lanka also has a significant amount of untapped hydropower potential, with an estimated 5,000 MW of potential capacity.

- **PAKISTAN:** Pakistan has the third-highest proportion of renewable energy consumption of any country in the dataset, with 66.9% of its energy coming from renewable sources. The majority of Pakistan's renewable energy comes from hydropower, with 43% of its electricity generation coming from hydropower plants. Pakistan also has a significant amount of untapped wind and solar energy potential.

- **NEW ZEALAND:** New Zealand has the fourth-highest proportion of renewable energy consumption of any country in the dataset, with 83.3% of its energy coming from renewable sources. The majority of New Zealand's renewable energy comes from hydropower, with 60% of its electricity generation coming from hydropower plants. New Zealand also has a significant amount of untapped geothermal, wind, and solar energy potential.

- **ARGENTINA:** Argentina has the fifth-highest proportion of renewable energy consumption of any country in the dataset, with 83.4% of its energy coming from renewable sources. The majority of Argentina's renewable energy comes from hydropower, with 50% of its electricity generation coming from hydropower plants. Argentina also has a significant amount of untapped wind and solar energy potential.

- **AUSTRALIA:** Australia has the sixth-highest proportion of renewable energy consumption of any country in the dataset, with 24.6% of its energy coming from renewable sources. The majority of Australia's renewable energy comes from hydropower, with 6% of its electricity generation coming from hydropower plants. Australia also has a significant amount of untapped wind and solar energy potential.

- **UNITED STATES:** The United States has the seventh-highest proportion of renewable energy consumption of any country in the dataset, with 13.1% of its energy coming from renewable sources. The majority of the United States' renewable energy comes from hydropower, with 7% of its electricity generation coming from hydropower plants. The United States also has a significant amount of untapped wind and solar energy potential.

- **UNITED ARAB EMIRATES:** The United Arab Emirates has the lowest proportion of renewable energy consumption of any country in the dataset, with 0.0% of its energy coming from renewable sources. The United Arab Emirates relies heavily on fossil fuels, particularly oil and natural gas. The United Arab Emirates has been investing in renewable energy in recent years, but it still has a long way to go to transition to a clean energy future.

- **QATAR:** Qatar has the second-lowest proportion of renewable energy consumption of any country in the dataset, with 1.0% of its energy coming from renewable sources. Qatar relies heavily on fossil fuels, particularly natural gas. Qatar has been investing in renewable energy in recent years, but it still has a long way to go to transition to a clean energy future.

- **BAHRAIN:** Bahrain has the third-lowest proportion of renewable energy consumption of any country in the dataset, with 0.0% of its energy coming from renewable sources. Bahrain relies heavily on fossil fuels, particularly oil and natural gas. Bahrain has been investing in renewable energy in recent years, but it still has a long way to go to transition to a clean energy future.

Overall Insights Renewable energy consumption is growing faster than fossil fuel consumption in all 10 countries in the dataset. Developing countries have a higher proportion of renewable energy consumption than developed countries. There is a significant amount of untapped renewable energy potential in all 10 countries. Governments can support the growth of the renewable energy sector by investing in renewable energy infrastructure, providing subsidies and tax breaks for renewable energy companies, and creating regulations that promote the use of renewable energy. The world needs to accelerate the transition to renewable energy in order to achieve a sustainable energy future.

CONCLUSION:

In conclusion, developing nations generally exhibit higher growth rates in renewable energy consumption and production compared to underdeveloped nations. This is attributed to factors such as government policies, economic development, and access to technology. However, underdeveloped nations face significant challenges in adopting renewable energy, including poverty, energy infrastructure limitations, and political instability. Despite these challenges, there is a growing recognition of the importance of renewable energy for sustainable development in underdeveloped nations. International cooperation and support are crucial to accelerate the transition to renewable energy in these countries.

REFERENCES:

- Accelerating industry transition in developing countries: a comparative analysis" by Leadership Group for Industry Transition (2022)
- Relationship between Economic Growth and Energy Consumption from the Perspective of Sustainable Development" by PMC (2022)
- A comparative study of the relationship between circular economy, economic growth, and oil price across South Asian countries" by Frontiers (2022)