Estimating the Global Electricity Demand (Gwh) by Country-wise

Thesis submitted in partial fulfillment of the

requirements for the

Post Graduate Diploma in Data Science

By

[redacted]

[redacted] Tejaswini

Reg. No. 21125760030

Under the guidance of

Akshatha Lakshmi U S

Senior Associate Faculty

Jigsaw Academy

MANIPAL ACADEMY OF HIGHER EDUCATION, MANIPAL

Estimating the Global Electricity Demand (Gwh) by Country-wise

Thesis submitted in partial fulfillment of the

requirements for the

Post Graduate Diploma in Data Science

By

[redacted]

[redacted] Tejaswini

Reg. No. 21125760030

Under the guidance of

Akshatha Lakshmi U S

Senior Associate Faculty

Jigsaw Academy

MANIPAL ACADEMY OF HIGHER EDUCATION, MANIPAL

Estimating the Global Electricity Demand (Gwh) by Country-wise

Thesis submitted in partial fulfillment of the

requirements for the

Post Graduate Diploma in Data Science

By

[redacted]

[redacted] **Tejaswini**

Reg. No. 21125760030

Examiner 1

Signature:

Name:

Examiner 2

Signature:

Name:

MANIPAL ACADEMY OF HIGHER EDUCATION, MANIPAL

(Industry letter head and signature for an intern, and Bonafide letter signed by the guide for a non-intern)

CERTIFICATE

This is to certify that the project work titled

Estimating the Global Electricity Demand (Gwh) by Country-wise

is a Bonafide record of the work done by

[redacted]

[redacted] Tejaswini

Reg. No. 21125760030

In partial fulfillment of the requirements for the award of **Post Graduate Diploma in Data Science** under Manipal Academy of Higher Education, Manipal, Manipal and the same has not been submitted elsewhere for any kind of certification/recognition.

(Signature)

Akshatha Lakshmi U S,

Senior Associate Faculty,

Jigsaw Academy.

ACKNOWLEDGEMENT

First, I would like to thank my advisor Akshatha Lakshmi, for her valuable support and exemplary guidance throughout the course of my project.

I would also like to express my gratitude to all the professors who are a part of the PGDDS course.

I take this opportunity to thank my colleague Meghana Chikka for sharing her knowledge and giving some insightful input.

Finally, I express my profound gratitude to my family and friends for their continuous support and encouragement throughout my graduate studies. This would not have been possible without all of you.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

The energy management system (EMS) is becoming increasingly crucial as global energy consumption rises. In EMS, energy projection is an important part of the initial phase in creating a management strategy.

The fastest-growing source of final energy demand is electricity. Its growth is expected to outstrip energy consumption during the next 25 years. More money is invested in the power sector than in oil and gas combined. The rise of renewable energy sources such as wind and solar is also transforming the worldwide electricity supply. As a result, electricity is at the front of clean-energy transitions.

For an energy generating organization, particularly one that generates electricity, precise forecasting of consumption, as well as the ability to analyze and predict demand, is critical. It serves as the foundation for power system planning and operating choices.

In this project, we will create a technique for forecasting electricity demand. In this study, we primarily do exploratory data analysis before employing various forecasting models to forecast monthly consumption.

The information used ranges from 2016 to 2020. The root mean squared error is used to evaluate models (RMSE).

1. INTRODUCTION

Global efforts to combat climate change are causing fast electrification of a wide range of end-users, from transportation to industry, resulting in a tremendous rise in power demand and the need to generate as much of it as feasible from renewable sources. As a result, electricity systems around the world are undergoing significant changes.

The rapid deployment of renewable energy sources such as wind and solar PV places electricity at the forefront of clean energy, bringing electricity to the 800 million people who are now without it and assisting in significantly decreasing air pollution and meeting climate targets.

Global energy demand is estimated to rise by 48 percent in the next 20 years as a result of projected population growth and growing industrial growth.

A major transition like this necessitates new methods to power system design and operation. Fossil fuels currently meet 80% of the world's energy consumption. However, fossil fuel stocks are diminishing fast. Furthermore, because sunlight and wind are not always available, a variety of backup generation choices, as well as smarter and better-connected grids, are required.

In the below Fig: 1.1, we can see the linear progression of electricity demand globally from 1990-2018.

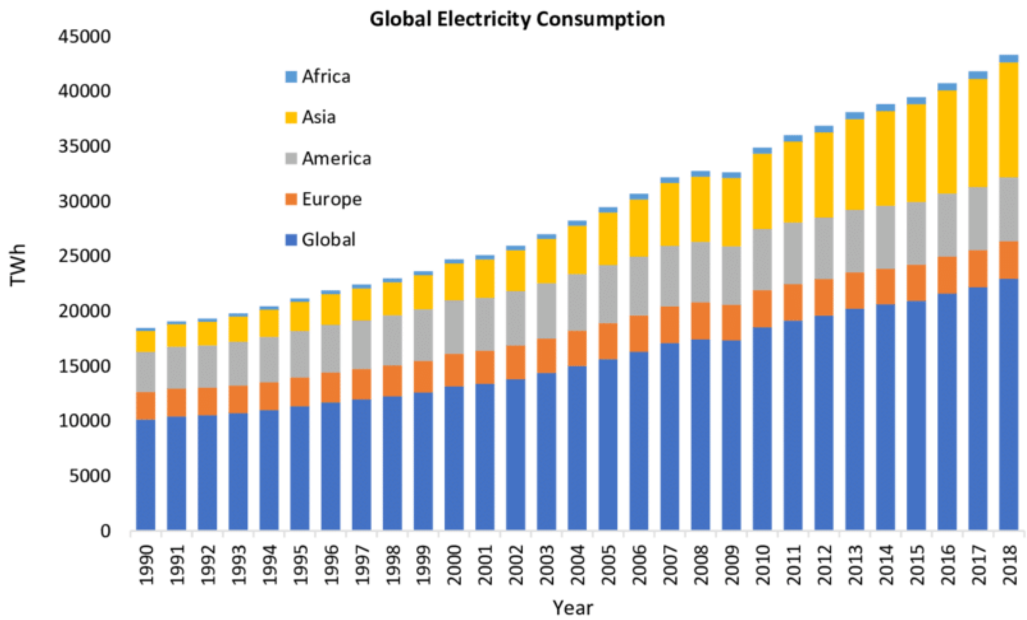


Fig: 1.1- Global Electricity Consumption from 1990-2018

Volatile energy costs and deregulation of the electricity generation business have made electricity load forecasting a major concern for power plant operations in recent years. Power plants employ a variety of time horizons and reliable projections to ensure plant functioning while preparing for possible facility expansions and to assess future demand optimally and securely. Building load forecasting has grown more significant around the world, as residential and commercial buildings consume almost 40% of total energy, according to the US Energy Information Administration. As a result, increasing the energy efficiency of buildings is crucial for reducing greenhouse gas emissions. Demand response (DR) measures have been offered as energy and cost-saving potential in the energy sector as a result of data power consumption analysis (forecasting, scheduling, and risk management in 2002). One of the fundamental jobs is to schedule the heating and cooling equipment in buildings because DR activities include turning off the electrical equipment. As a result, energy demand forecasting at the building level has become a hot topic in recent years. It is possible to design energy management to predict future load demand.

Long-term electricity demand predictions are also more economical for increasing electricity demand during an economic downturn. On a power system level, load forecasting can be divided into three categories based on time horizon: short-term (a few hours to a few weeks ahead), medium-term (a month to a year ahead), and long-term (a year or more ahead) (ranging from 1 to 20 years ahead). Long-term predictions will be made over the next 1 to 20 years, and they will be critical for strategic planning, new generation building, and transmission capacity. Mid-term forecasting will vary from a month to a year in advance, and will be used to schedule maintenance and power sharing agreements. The time frame for short-term load forecasting will range from a few hours to a few weeks. It is critical for real-time control, plant scheduling, and fuel purchasing schemes, short-term maintenance and short-term storage.

1.1 Project Goal

The purpose of this study is to see if a general, straightforward strategy based on Machine Learning models can outperform a difficult forecasting challenge. We also look into machine learning and construct data-driven models to anticipate monthly energy consumption.

Transmission Service Operators (TSOs) provide energy demand predictions once a month to ensure that energy demand is met for the following month. Every month, this is done all around the world to anticipate the expected maximum energy consumption. These projections assist the energy industry in maintaining grid balance and planning supply for month-ahead bidding processes.

1.2 Organization of the Report

- Overview of the data
- Data Cleaning
- Data Preparation
- Detailed description of the analysis
- Data Preprocessing
- Data Modeling
- Performance Evaluation of the models
- Conclusions of the study

2. PROJECT DESCRIPTION

2.1 Business Understanding

Transmission Service Operators (TSOs) provide energy demand predictions once a month to ensure that energy demand is met for the following month. Every month, this is done all around the world to anticipate the expected maximum energy consumption. These projections assist the energy industry in maintaining grid balance and planning supply for month-ahead bidding processes.

Predicting power demand with great precision could result in a significant set of values for a country, a city, or even a family. Stakeholders may change their energy production to lower costs, or they may purchase adequate energy if their power needs are not met by external sources. Stakeholders may produce additional benefit in various instances, such as during tendering processes in daily energy exchange.

2.2 Dataset Description

Data is taken from:

https://www.iea.org/data-and-statistics/data-product/monthlyelectricity-statistics

The IEA's Monthly Electricity Statistics provide member nations of the Organization for Economic Cooperation and Development with up-to-date data on electricity production, consumption, and trade (OECD). It also includes data on the production of energy in a number of other countries.

2.2.1 Basic Definitions:

Net Production: Net production does not include the utilization of power plants for their own purposes.

Combustible Fuels/ Conventional Thermal: Production of combustible renewables and wastes (primary coal, coal products, peat and peat products, oil shale, oil sands, crude oil, NGL, oil products, and natural gas) as well as fossil fuels (primary coal, coal products, peat and peat products, oil shale, oil sands, crude oil, NGL, oil products, and natural gas) (solid biofuels, biogases, liquid biofuels, industrial and municipal waste).

Coal: Primary coal, coal products, peat and peat products, oil shale and oil sands, and produced gas are all used in the manufacturing process (coke-oven gas, blast-furnace gas and gas works gas).

Oil: Crude oil, natural gas liquids, refinery feedstock's, and petroleum products are all used in production (refinery gases and fuel oil).

Natural Gas: Natural gas-based production (including gas distributed via the grid that may contain little amounts of blended other gases).

Combustible Renewables: Combustible renewable energy production (solid biofuels, biogases, liquid biofuels and municipal renewable waste).

Other Combustible Fuels (non-renewables): All other combustible fuels are used for production (industrial and non-renewable municipal solid waste).

Nuclear: Heat from nuclear fission is used to generate electricity.

Hydro: All hydro plants' net generation, including pumped storage production.

Wind: Wind kinetic energy is used to create electricity in both on-shore and off-shore wind turbines.

Solar: Solar energy, both thermal and photovoltaic, is used to generate electricity.

Geothermal: Heat radiated from within the earth's crust, usually in the form of hot water or steam, is used to generate electricity.

Other Renewables: Tide, wave, ocean, and other non-combustible sources of electricity

Non-Specified: Electricity generation was not reported anywhere else.

Renewables: Hydro, Wind, Solar, Geothermal, Other Renewables, and Combustible Renewables are all used to generate electricity.

Non-Renewables: Coal, oil, natural gas, other combustibles, nuclear, and unspecified sources of electricity are all included.

Total Net Production: The total quantity of net electricity produced in the country from all energy sources.

Imports/Exports: Amounts of electricity that have passed the country's political boundaries, also known as physical flows, whether or not customs clearance has occurred.

Net Exports: Exports minus Imports.

Electricity supplied: Imports minus indigenous production equals exports. Losses in both transmission and distribution are included.

Used for Pumped Storage: Pumping water into a reservoir consumes electricity in both mixed and pure pumped storage hydro facilities.

Transmission & Distribution Losses: All losses incurred as a result of electrical energy transportation and delivery. This includes losses in transformers that are not considered to be important elements of power plants.

Electricity consumed: Available electricity (calculated as: indigenous production + imports - exports - electricity used for pumped storage – both transmission and distribution losses)

Low-carbon Power Generation: Electricity generation from renewables and nuclear power.

Year: Unless otherwise specified, all dates are in calendar year (January to December).

2.2.2 Regional Groupings:

- Australia, Austria, Belgium, Canada, Chile, Colombia, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, the Slovak Republic, Slovenia, Switzerland, Turkey, the United Kingdom, and the United States are all members of the OECD Total.

- This report does not include Israel.
- Canada, Chile, Colombia, Mexico, and the United States are all members of the OECD Americas.
- Australia, Japan, Korea, and New Zealand are part of the OECD Asia Oceania region.
- Austria, Belgium, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, and the United Kingdom are among the countries that make up the OECD Europe.
- Except for Chile, Iceland, Latvia, Lithuania, and Slovenia, the IEA covers all nations that are included in the OECD Total.

2.2.3 Country Notes

2.2.3.1 OECD Member Countries

AUSTRALIA:

- Overseas territories are not included.

- Monthly data is modified based on data submitted on a fiscal year basis, which begins on July 1st and ends on June 30th.

- Large-scale solar generation is not included in the monthly reporting. As a result, the IEA Secretariat estimates it.

- Transmission and distribution losses, as well as hydro used for pumped storage, are not included in monthly reporting. As a result, the IEA Secretariat estimates it.

AUSTRIA

- Non-combustible fuels such as geothermal, solar, and wind are not included in monthly reporting. As a result, the IEA Secretariat has calculated it based on annual submissions from January 2003.

- Austria's final coal-fired power station closed in April 2020. Since then, all coal output has come from automakers' CHP units, and all coal products have been used.

BELGIUM

- The last major coal-fired power station was shut down in March of 2016. From then on, Belgium's coal production is solely based on produced gases.

CANADA

- From January 2017 to December 2018, the amount of wind and sun coverage has been steadily growing.

- Since April 2019, we've used a new methodology and expanded our coverage (starting with MES January- 2019 publication). Data was reviewed all the way back to January of 2016.

- The breakdown for conventional thermal production is not included in monthly reporting. As a result, the IEA Secretariat estimates it.

- Transmission and distribution losses, as well as hydro used for pumped storage, are not included in monthly reporting. As a result, the IEA Secretariat estimates it.

CHILE

- In February of 2011, Chile was added to the MES report.

COLOMBIA

- After becoming an OECD Member Country in April 2020, Colombia was included in the MES Report in November 2020.

- The gross values are used to present all monthly submissions. The MES report's values are net and were calculated by the IEA Secretariat.

- Auto output that is fed back into the grid is only included in monthly submissions. This covers auto manufacturing in the hydrocarbon sector.

CZECH REPUBLIC

- Other non-combustible fuels were not included in the monthly reporting. As a result, the IEA Secretariat assessed it based on annual submissions from January 2004 to December 2008.

DENMARK

- The countries of the Faroe Islands and Greenland are not included.

ESTONIA

- In April of 2011, Estonia was added to the MES report. Only data from January through 2006 is available in the archive.

- The breakdown for conventional thermal production is not included in monthly reporting. As a result, the IEA Secretariat estimates it.

- Transmission and distribution losses, as well as hydro used for pumped storage, are not included in monthly reporting. As a result, the IEA Secretariat estimates it.

FINLAND

- Other combustible fuels include peat power generation.

- Electricity generation is approximated based on annual submissions and is not included in other reports.

FRANCE

- Monaco is included. Since January 2018, it has additionally included French Polynesia, Guadeloupe, Guyann, Martinique, New Caledonia, Reunion, Saint Pierre, and Miquelon, as well as the French overseas territories of Guadeloupe, Guyann, Martinique, Martinique, New Caledonia, Reunion, Saint Pierre, and Miquelon.

- Other non-combustible fuels are not included in monthly reports. As a result, the IEA Secretariat estimates it based on annual submission.

- Based on annual submissions, production from alternative renewable sources is estimated.

- Electricity generation is approximated based on annual submissions and is not included in other reports.

GERMANY

- Electrical generation is estimated from monthly submissions of production from major electricity producers, which is not included in other reports.

GREECE

- Other combustible fuels are included in monthly reporting. As a result, the IEA Secretariat calculated it based on annual submissions.

HUNGARY

- Monthly gross production submissions were used to estimate net production from December to December 2015. Electricity generation is approximated based on annual submissions and is not included in other reports.

IRELAND

- Solar production is not included in monthly reports. As a result, the IEA Secretariat evaluated it based on the annual submission.

ITALY

- San Marino and the Vatican are among the countries represented.

- The breakdown for conventional thermal production is not included in monthly reporting. As a result, the IEA Secretariat estimates it.

- Transmission and distribution losses, as well as hydro used for pumped storage, are not included in monthly reporting. As a result, the IEA Secretariat estimates it.

- Electricity generation is estimated based on annual submissions in the Not Elsewhere reports.

JAPAN

- Japan's M-3 data isn't available. As a result, the IEA Secretariat estimates the most recent month.

- Monthly data is modified based on data submitted on a fiscal year basis, which begins on April 1st and ends on March 31st. The annual data displayed is the sum of months and may differ from publications referring to the fiscal year of Japan (April-March).

- Since April 2017, biomass energy generation has been reported in the coal category if co-fired with coal, and in combustible renewables if combustion is mono-fuel.

- Coal, bituminous mixture, natural gas (other than LNG), coke-oven gas, and city gas have all been included in the coal category since April 2017.

- Plants with a capacity of 1MW or more are classified as hydro generation.

- Trade data is not included in the monthly electricity generation data.

- The IEA Secretariat estimates transmission and distribution losses as well as hydro used for pumped storage.

KOREA

- Automakers are not included in monthly reports of electricity production. As a result, the IEA Secretariat estimates it based on the annual submission.

LATVIA

- In January 2017, Latvia was added to the MES report. Only data from January to December 2010 is available.

- The good hydro-performance in 2017-2018 is consistent with historical patterns and is related to environmental factors, particularly enhanced Daugava River input.

- Solar production is not included in monthly reports. As a result, the IEA Secretariat evaluated it based on annual submission.

- The IEA Secretariat estimates transmission and distribution losses and hydro used for pumped storage because monthly reports do not cover them.

LITHUANIA

- From January to December 2008, monthly data is available.

LUXEMBOURG

- Since April 2016, the Twinerg CCGT plant has been idle, and it was dismantled in October 2016. As a result, combustible fuel production has decreased, and as a result, exports have decreased.

MEXICO

- Automakers are not included in monthly reports of electricity production. As a result, the IEA Secretariat estimates it based on annual submission.

- Installations less than 500 kW are not included in the reporting data, hence the IEA Secretariat estimates it based on annual submission.

- Mexico's administration estimates electricity trading figures. It could be incompatible with data transfers in the United States. This information is currently being reviewed.

- In January of 2020, the data source was changed, and the data processing was upgraded to ensure that the data was consistent over the entire set.

- Because monthly reporting does not give a breakdown for conventional thermal production, the IEA Secretariat estimates it.

- The IEA Secretariat estimates trade data, transmission and distribution losses, and hydro used for pumped storage, which are not included in monthly electricity production figures.

NETHERLANDS

- Suriname and the Former Netherlands Antilles are not included.

- Because of plant maintenance, there was no nuclear power generation in May and June 2017. In July 2017, nuclear power was restarted.

- Electricity generation that hasn't been documented anywhere else is estimated based on annual reports.

NEW ZEALAND

- Only major electrical producers are included in monthly reports, implying 95 percent coverage. The IEA Secretariat estimates the rest of the power generation based on annual submissions.

NORWAY

- Because monthly reporting does not include solar production, the IEA Secretariat approximated it based on annual submission.

- Since November 2020, natural gas production numbers have been kept confidential, and natural gas output is listed as other combustible fuels.

POLAND

- Since January 2017, oil figures have been kept secret and are reported as other combustible fuels.

- From January 2017, mixed hydro numbers have been kept secret and reported as pure hydro.

- Electricity generation that hasn't been documented anywhere else is estimated based on annual reports.

PORTUGAL

- Since January 2015, the Açores and Madeira have been included.

SLOVAK REPUBLIC

- The IEA Secretariat estimates transmission and distribution losses and hydro used for pumped storage because monthly reporting does not cover them.

SLOVENIA

- In April 2011, Slovenia was added to the MES report. Only data from January to December 2009 is available.

SPAIN

- The Canary Islands, Balearic Islands, Ceuta, and Melilla are all included.

- Mixed hydro output is not included in monthly reporting. It's based on annual submissions, therefore it's a guess.

- Production from other renewable sources or production not stated elsewhere is not included in monthly reporting. It's based on annual submissions, therefore it's a guess.

SWEDEN

- Because monthly reporting does not include solar production, the IEA Secretariat estimates it based on annual submissions since January 2000.

- Because monthly reporting does not contain a breakdown for wind production, the IEA Secretariat estimates it based on annual submission.

- The IEA Secretariat estimates it based on annual submission because monthly reporting does not include the breakdown for hydro production and hydro used for pumped storage.

- Sweden's final coal-fired power plant shut down in April 2020.

SWITZERLAND

- Liechtenstein is not included.

- Because monthly reporting does not give a breakdown for hydro production, the IEA Secretariat calculated it based on annual submission.

TURKEY

- The IEA Secretariat estimates transmission and distribution losses because monthly reporting does not cover them.

UNITED KINGDOM

- The IEA Secretariat estimates production from other renewable sources because monthly reporting does not cover it.

UNITED STATES

- The 50 states, the District of Columbia, and Puerto Rico are all included.

- Since July 2017, the US administration has approximated electricity trade data. It could be in conflict with Mexican trade figures. This information is currently being reviewed.

2.2.3.2 Non–OECD Countries

ARGENTINA

- The Argentine Wholesale Electricity Market Clearing Company CAMMESA provides monthly data.

- Only the major electrical producers are included in the monthly production data. The IEA Secretariat estimates the rest of the power generation based on annual submissions.

- Paraguayan hydro production from the Yacyreta binational dam is included in monthly reporting. The IEA Secretariat estimates the Paraguayan contribution, which is subsequently subtracted from total Argentine hydro production.

BRAZIL

- The Argentine Wholesale Electricity Market Clearing Company CAMMESA provides monthly data.

- Only the major electrical producers are included in the monthly production data. The IEA Secretariat estimates the rest of the power generation based on annual submissions.

- Paraguayan hydro production from the Yacyreta binational dam is included in monthly reporting. The IEA Secretariat estimates the Paraguayan contribution, which is subsequently subtracted from total Argentine hydro production.

BULGARIA

- The Eurostat Database collects data on a monthly basis.

CHINA

- The National Bureau of Statistics provides monthly data.

- The National Bureau of Statistics provides raw monthly production data in gross numbers. Because the MES Report presents all statistics in net values, Chinese net production is computed using IEA Secretariat estimates.

- The raw production data for January and February is published together. Based on annual statistics and the FGE China Gas Monthly Report, the IEA Secretariat approximated output data for these two months.

- Because monthly reporting did not contain disaggregated production from combustible fuels, the IEA Secretariat calculated it using annual data.

- Because geothermal production was not reported on a monthly basis, the IEA Secretariat calculated it based on annual data.

- The IEA Secretariat estimates production from other renewables sources based on annual data because monthly reporting did not include it.

CYPRUS

Turkey's note:

- The material in this paper about "Cyprus" refers to the island's southernmost region. On the island, there is no single authority that represents both Turkish and Greek Cypriots. The Turkish Republic of Northern Cyprus is recognized by Turkey (TRNC). Turkey will maintain its stance on the "Cyprus issue" until a long-term and equitable solution is reached within the framework of the United Nations.

- All OECD and European Union member states should take note of the following:

- Except for Turkey, every member of the United Nations recognizes the Republic of Cyprus. The information in this paper pertains to the territory under the government of the Republic of Cyprus's effective control.

- The Eurostat Database collects data on a monthly basis.

INDIA

- The Central Electricity Authority and the National Power Portal provide monthly data. The Central Electricity Authority provides raw monthly production data in gross numbers. Because the MES Report presents all statistics in net numbers, Indian net production is computed using IEA Secretariat estimates.

- Monthly data is modified based on data submitted on a fiscal year basis, which begins on April 1st and ends on March 31st.

- Only the major electrical producers are included in the monthly production data. The IEA Secretariat estimates the rest of the power generation based on annual submissions.

- The IEA Secretariat estimates production from other combustible fuels based on annual data because monthly reporting did not include it.

MALTA

- The Eurostat Database collects data on a monthly basis.

3. OVERVIEW OF THE DATA

- The dataset downloaded consists of 120,898 records from Jan – 2010 to Oct – 2021 as on 23/01/2022 as shown in Fig: 3.1.

	Country	Time	Balance	Product	Value
0	Australia	October 2021	Net Electricity Production	Electricity	19914.331620
1	Australia	October 2021	Net Electricity Production	Nuclear	0.000000
2	Australia	October 2021	Net Electricity Production	Total Combustible Fuels	13378.485670
3	Australia	October 2021	Net Electricity Production	Coal, Peat and Manufactured Gases	9822.560196
4	Australia	October 2021	Net Electricity Production	Oil and Petroleum Products	893.288130
...	...	...	...	...	...
120893	IEA Total	January 2010	Total Imports	Electricity	34183.558000
120894	IEA Total	January 2010	Total Exports	Electricity	32045.283000
120895	IEA Total	January 2010	Used for pumped storage	Electricity	6755.476000
120896	IEA Total	January 2010	Distribution Losses	Electricity	61894.043000
120897	IEA Total	January 2010	Final Consumption (Calculated)	Electricity	865851.875000

120898 rows × 5 columns

Fig: 3.1- Dataset with records from Jan-2010 to Oct-2021

- In this project, we would only consider data from **Jan – 2016 to Dec – 2020** which are **58,286 records** in total as shown in Fig: 3.2.
- The dataset consists of **5 columns** - Country, Time, Balance, Product, and Value.

	Country	Time	Balance	Product	Value
0	Australia	December 2020	Net Electricity Production	Electricity	21231.533490
1	Australia	December 2020	Net Electricity Production	Nuclear	0.000000
2	Australia	December 2020	Net Electricity Production	Total Combustible Fuels	15210.694630
3	Australia	December 2020	Net Electricity Production	Coal, Peat and Manufactured Gases	11030.866570
4	Australia	December 2020	Net Electricity Production	Oil and Petroleum Products	858.100828
...	...	...	...	...	...
58281	Serbia	January 2016	Total Imports	Electricity	0.000000
58282	Serbia	January 2016	Total Exports	Electricity	0.000000
58283	Serbia	January 2016	Used for pumped storage	Electricity	0.000000
58284	Serbia	January 2016	Distribution Losses	Electricity	0.000000
58285	Serbia	January 2016	Final Consumption (Calculated)	Electricity	2987.162000

58286 rows × 5 columns

Fig: 3.2- Dataset with records from Jan-2016 to Dec-2020

- **'Country'** column is a **categorical** column & consists of 'Country' names.
- There are 52 countries represented in these records: Australia, Austria, Belgium, Canada, Chile, Colombia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States, Croatia, Romania, OECD Europe, OECD Americas, OECD Asia Oceania, OE.
- **'Time'** column is a **categorical** column & consists of values in format - 'Month-Year' which represents the corresponding month and year on which the electricity data is collected.
- **'Balance'** column is a **categorical** column & consists of categorical values like as shown in Fig.3.3.
- **'Product'** column is a **categorical** column & consists of categorical values of different means used to produce electricity as shown in Fig.3.3.

Balance	Product
• Net Electricity Production • Total Imports • Total Exports • Used for pumped storage • Distribution Losses • Final Consumption (Calculated)	• Electricity • Nuclear • Total Combustible Fuels • Coal • Peat and Manufactured Gases • Oil and Petroleum Products • Natural Gas • Combustible Renewables • Other Combustible Non-Renewables • Hydro • Wind • Solar • Geothermal • Other Renewables • Not Specified • Total Renewables (Geo, Solar, Wind, Other)

Fig: 3.3- Balance and Product values

♦ **'Value'** column consists of **numerical** values in gigawatt hours (**GWh**).

A GWh is a metric for one billion (1,000,000,000) watt hours of electricity.

♦ The detailed view of the dataset is shown in Fig. 3.4.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 58286 entries, 0 to 58285
Data columns (total 5 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   Country  58286 non-null  object 
 1   Time     58286 non-null  object 
 2   Balance  58286 non-null  object 
 3   Product  58286 non-null  object 
 4   Value    58286 non-null  float64
dtypes: float64(1), object(4)
memory usage: 2.2+ MB
```

Fig: 3.4- Dataset Overview

- The overview of the column 'Value' is shown in Fig. 3.5.

	count	mean	std	min	25%	50%	75%	max
Value	58286.0	17249.496764	76349.494706	0.0	15.069252	478.2405	3575.224555	983368.7927

Fig: 3.5- Value column summary

- The overview of the remaining columns is shown in Fig. 3.6.

	count	unique	top	freq
Country	58286	52	OECD Europe	1200
Time	58286	60	January 2018	1040
Balance	58286	6	Net Electricity Production	43500
Product	58286	15	Electricity	17906

Fig: 3.6- Country, Time, Balance, Product columns summary

4. DATA CLEANING

- From the Fig. 4.1, we see that there are no 'null' values in our data.

```
Country     0
Time        0
Balance     0
Product     0
Value       0
dtype: int64
```

Fig: 4.1- Null Value Check

- Also, there are no duplicates observed as well in our data.

```
a.duplicated().sum()
```

```
0
```

5. DATA PREPARATION

- 'Time' column is split into 'Month' & 'Year' columns as shown in Fig. 5.1.

	Country	Balance	Product	Value	Month	Year
0	Australia	Net Electricity Production	Electricity	21231.533490	12	2020
1	Australia	Net Electricity Production	Nuclear	0.000000	12	2020
2	Australia	Net Electricity Production	Total Combustible Fuels	15210.694630	12	2020
3	Australia	Net Electricity Production	Coal, Peat and Manufactured Gases	11030.866570	12	2020
4	Australia	Net Electricity Production	Oil and Petroleum Products	858.100828	12	2020
...	...	...	...	...	...	...
58281	Serbia	Total Imports	Electricity	0.000000	1	2016
58282	Serbia	Total Exports	Electricity	0.000000	1	2016
58283	Serbia	Used for pumped storage	Electricity	0.000000	1	2016
58284	Serbia	Distribution Losses	Electricity	0.000000	1	2016
58285	Serbia	Final Consumption (Calculated)	Electricity	2987.162000	1	2016

58286 rows × 6 columns

Fig: 5.1- Dataset after Time column split into Month, Year columns

- Product = 'Total Renewables (Geo, Solar, Wind, Other)' records are removed as they are redundant.
- Our purpose is to predict the net electricity consumption in a given month – year.
- Hence, only Balance = ‘Net Electricity Production', 'Final Consumption (Calculated)' records are considered for our project, other records are removed.
- For model building we will only consider 'Final Consumption (Calculated)' records because we are more interested in predicting the electricity consumption, but for analyzing the trend we would use both 'Net Electricity Production' and 'Final Consumption (Calculated)' records.

- We also remove the records where Product='Electricity' as it is nothing but the net electricity for the respective month in a year which is redundant.
- As last step, we round the values of 'Value' column for better interpretability.
- After performing all the above, the dataset consists of 43,500 records as shown in Fig. 5.2.

	Country	Balance	Product	Value	Month	Year
0	Australia	Net Electricity Production	Electricity	21232.0	12	2020
1	Australia	Net Electricity Production	Nuclear	0.0	12	2020
2	Australia	Net Electricity Production	Total Combustible Fuels	15211.0	12	2020
3	Australia	Net Electricity Production	Coal, Peat and Manufactured Gases	11031.0	12	2020
4	Australia	Net Electricity Production	Oil and Petroleum Products	858.0	12	2020
...	...	...	...	...	...	...
58276	Serbia	Net Electricity Production	Combustible Renewables	3.0	1	2016
58277	Serbia	Net Electricity Production	Other Combustible Non-Renewables	0.0	1	2016
58278	Serbia	Net Electricity Production	Hydro	853.0	1	2016
58279	Serbia	Net Electricity Production	Solar	0.0	1	2016
58285	Serbia	Final Consumption (Calculated)	Electricity	2987.0	1	2016

43500 rows × 6 columns

Fig: 5.2- Dataset with only 'Net Electricity Production' & 'Final Consumption (Calculated) records

6. EXPLORATORY DATA ANALYSIS

6.1 Outliers Check

♦ Outliers are observed in great extent for the numerical column - ‘Value’ as shown in Fig. 6.1.

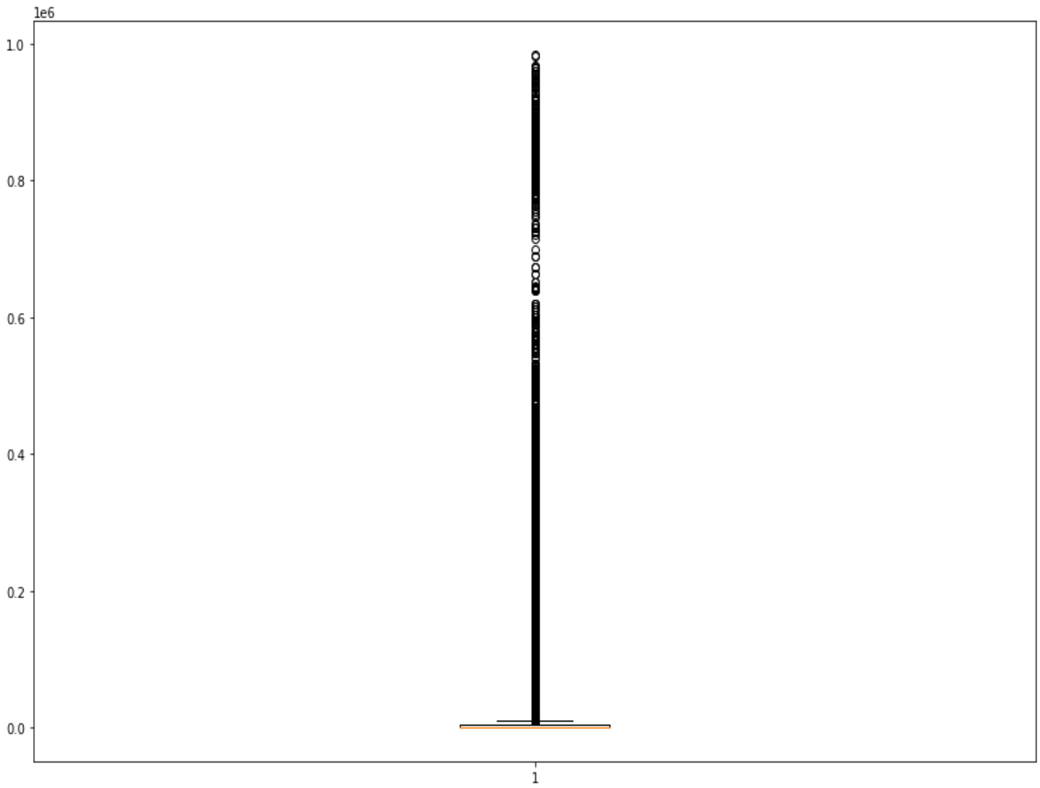


Fig: 6.1- Outlier Check

♦ These are not treated as they are essential for training the models.
♦ In case we thought to remove them, we will lose most of the data. Hence, we prefer not to.

6.2 Correlation

- As shown in Fig. 6.2, no strong correlation is observed either strong positive or strong negative.

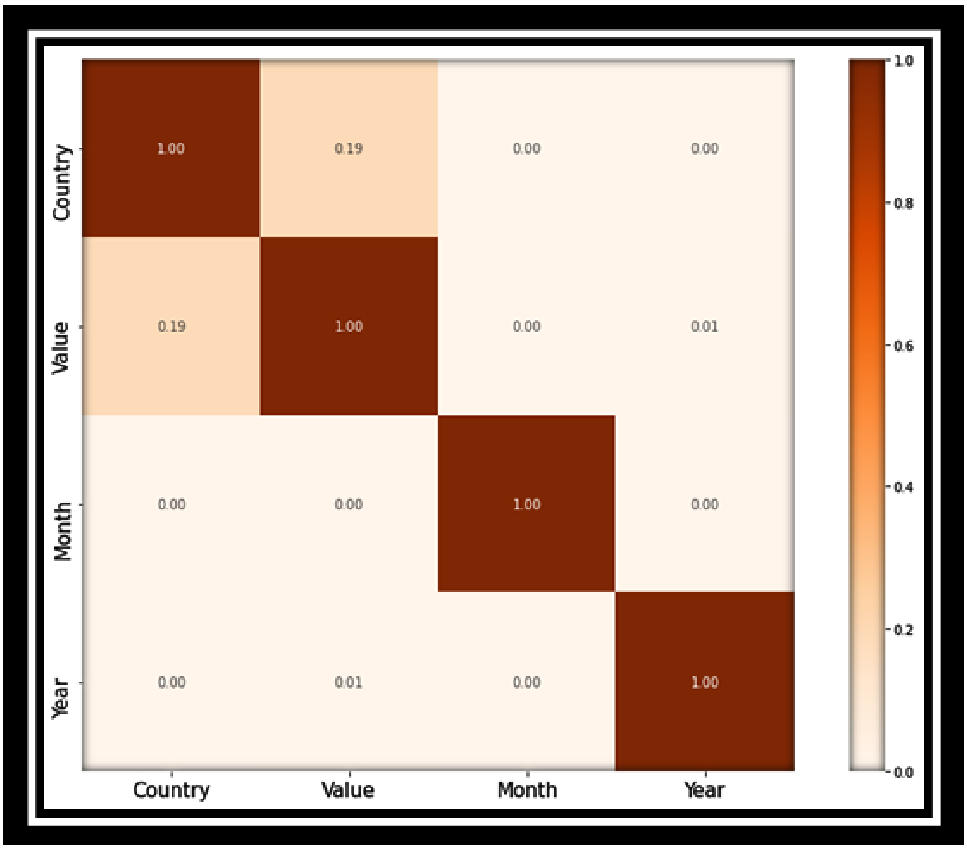


Fig: 6.2- Correlation

6.3 Total Electricity Production for 5 years

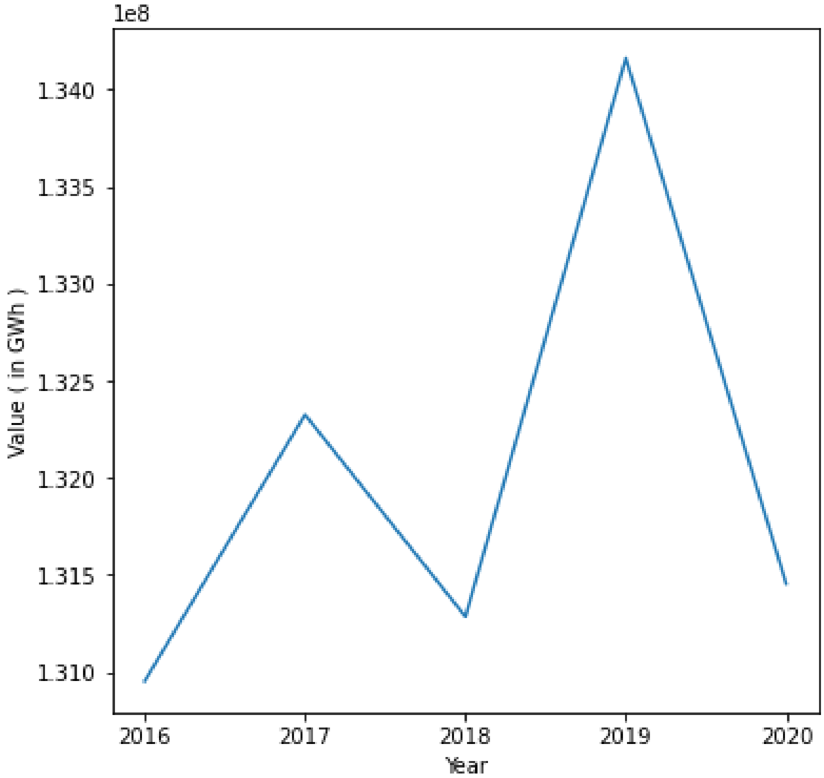


Fig: 6.3- Total Electricity Production for 5 years

- The production of electricity in 2016 is the lowest in the last 5 years.
- In 2019, production is at a peak.
- There is no regular pattern observed in such. But we can interpret that in one year it will be at its lowest, the next year it increases drastically and the next year it drops.
- This pattern is observed from 2016 – 2018 and 2018 – 2020.
- The reasons behind this pattern are not covered in this study. It involves studying countries' economic statistics and the related factors.

6.4 Total Electricity Consumption for 5 years

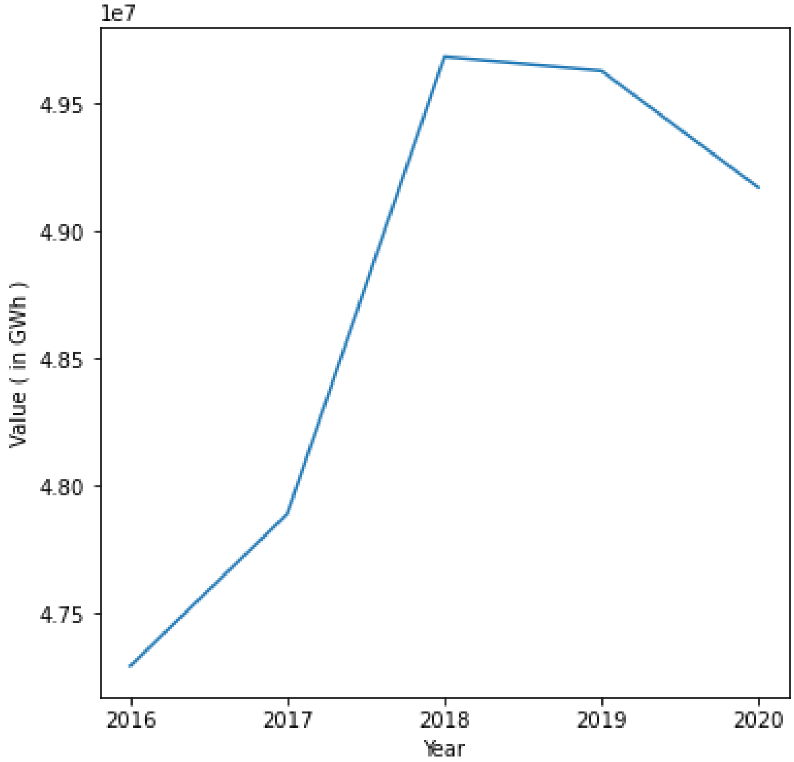
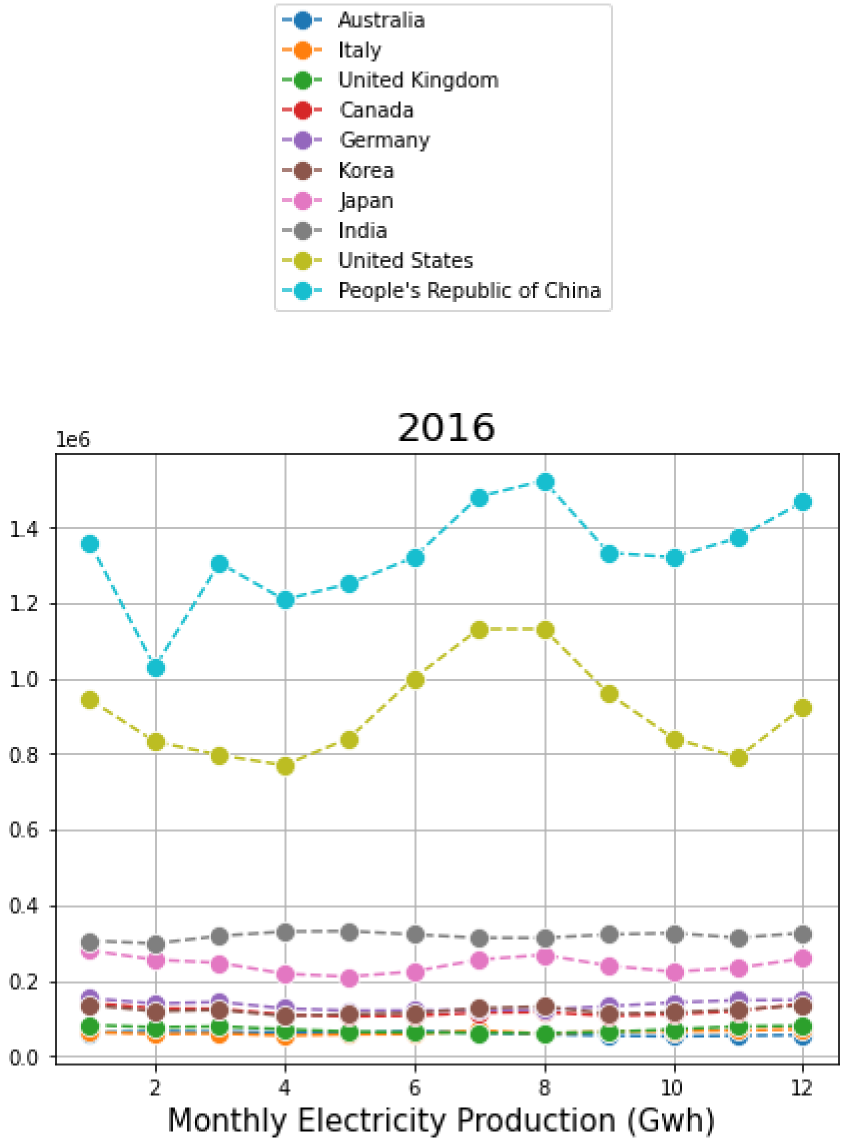


Fig: 6.4- Total Electricity Consumption for 5 years

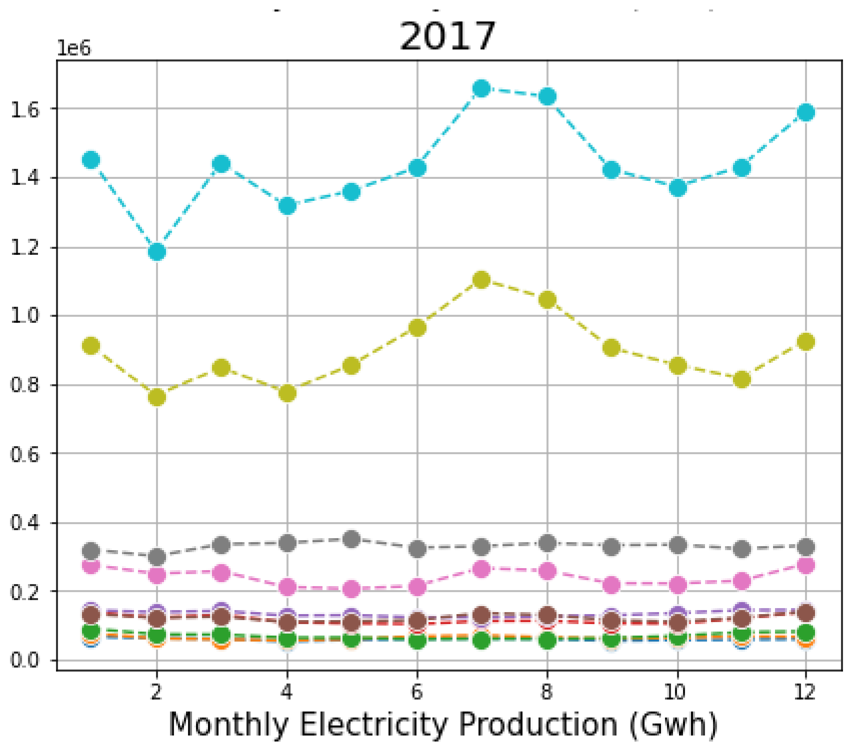
- In 2016, the consumption of electricity globally was the lowest.
- It drastically increased till 2018 and started down facing.
- The reasons behind this pattern are not covered in this study. It involves studying countries' economic statistics and the related factors.

6.5 Monthly overall production and monthly overall consumption of selective countries for last 5 years

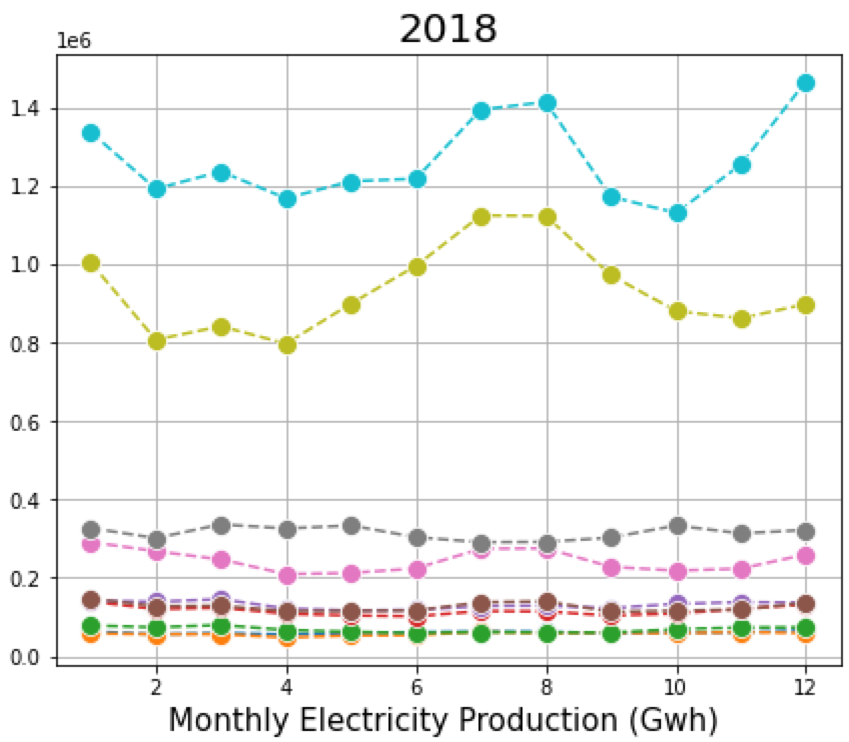
- Considered only few countries which are mentioned in the Fig. 6.5.
- These are the developed and rapidly developing countries.
- ‘People’s Republic of China’ holds the spot of highest producer and consumer of electricity for the past 5 years.
- It is followed by ‘United States’, ‘India’, ‘Japan’ and so on.



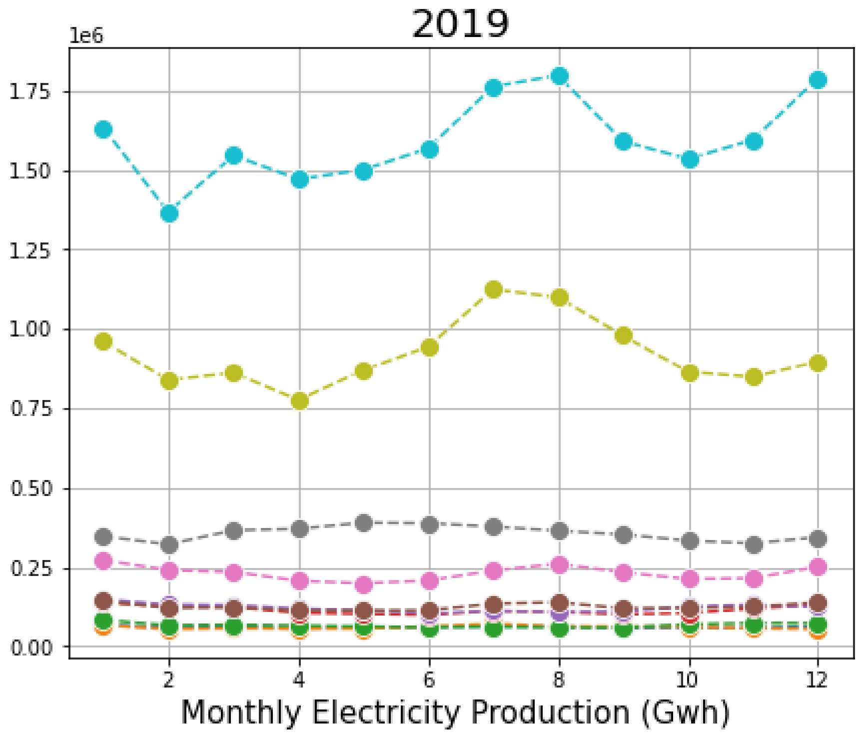
- In 2016, People's Republic of China produced large amount of electricity in July, August and December. It produced least in the month of February.
- It produced 15, 40,000Gwh in August which is the highest and the lowest was- 10, 20,000Gwh.
- The second largest producer of electricity- United States produced highest electricity in July, August and produced least in April.
- For these two countries, we can observe the pattern- Steady increase in production from April to August and then decrease in production later on and then again drastic jump in December.



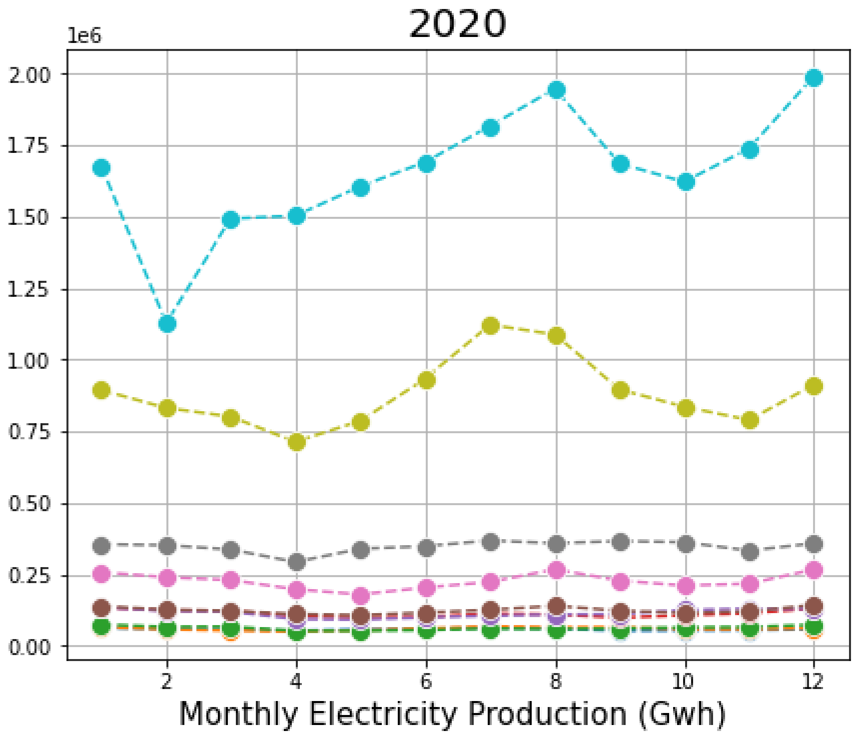
- In 2017, People's Republic of China produced high electricity in July, August and December. It produced least in the month of February. It followed almost same pattern as in 2016.
- The Highest electricity it produced is- 16, 00,000Gwh.
- United States produced more electricity in July, August and produced least in April.
- For these two countries, we can observe the pattern- Steady increase in production from April to July and then decrease in production later on and then again drastic jump in December.



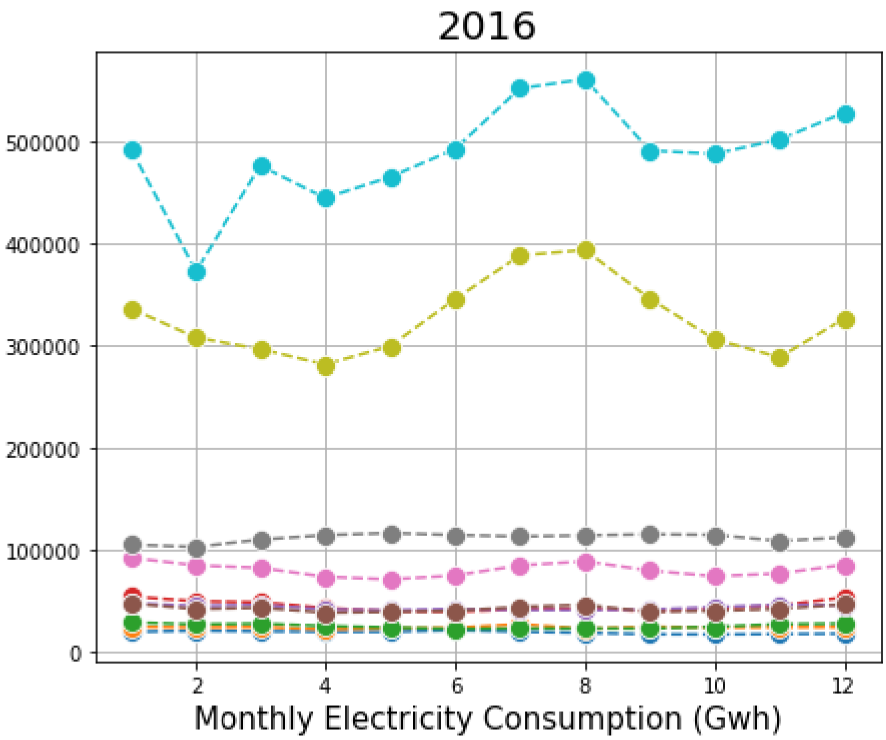
- In 2018, People's Republic of China produced high electricity in December. It produced least in the month of October.
- United States produced more electricity in July, August and produced least in April.
- United States produced same amount of electricity in July, August months.
- In these two countries, we can observe the pattern- drastic increase in production from April to July and then saw a decrease in production later on till November and then again jumped in December.



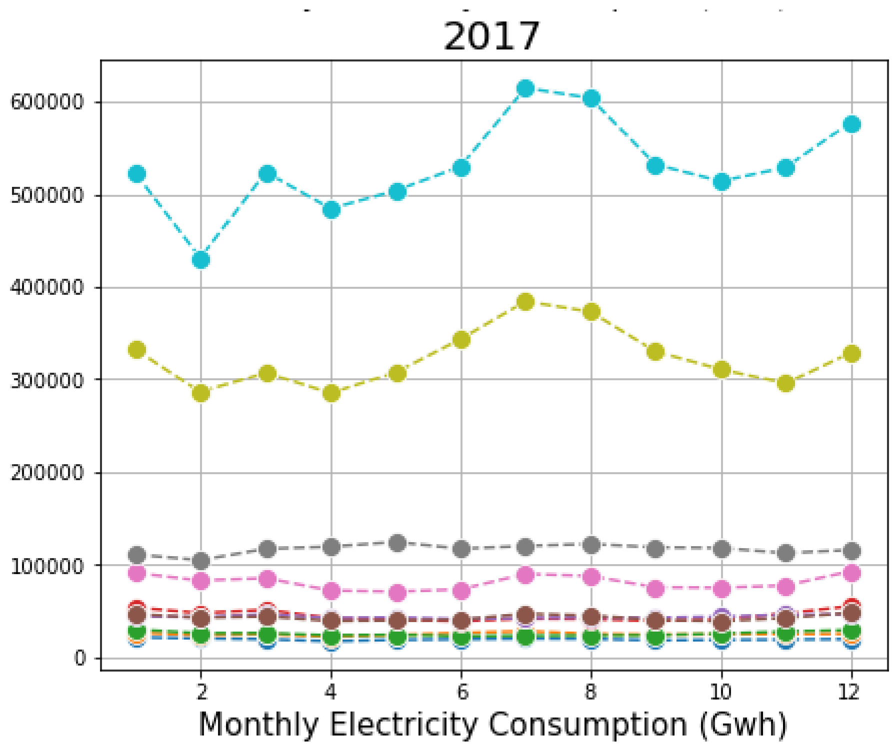
- In 2019, People's Republic of China produced high electricity in August i.e. 17, 50,000Gwh. It produced least in the month of February.
- The highest is the highest value ever observed in the last 3 years.
- United States produced more electricity in July and produced least in April.
- In these two countries, we can observe the pattern- drastic increase in production from April to August and then saw a decrease in production later on and then again increased in December.



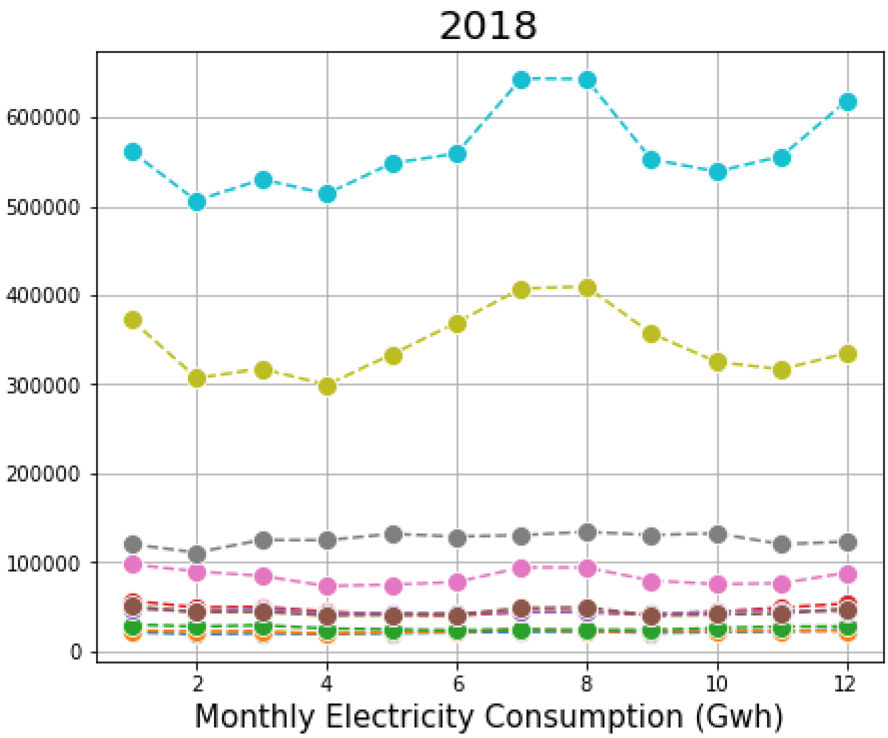
- In 2020, People's Republic of China produced high electricity in December i.e. 20, 00,000Gwh. It produced least in the month of February.
- The highest is the highest value ever observed in the last 4 years.
- With People's Republic of China, we can observe the pattern- drastic increase in production from April to August and then saw a decrease and then increased in December.
- United States produced more electricity in July and produced least in April.
- With United States, we can observe the pattern- drastic increase in production from April to July and then saw a decrease in production later on and then again increased in December.



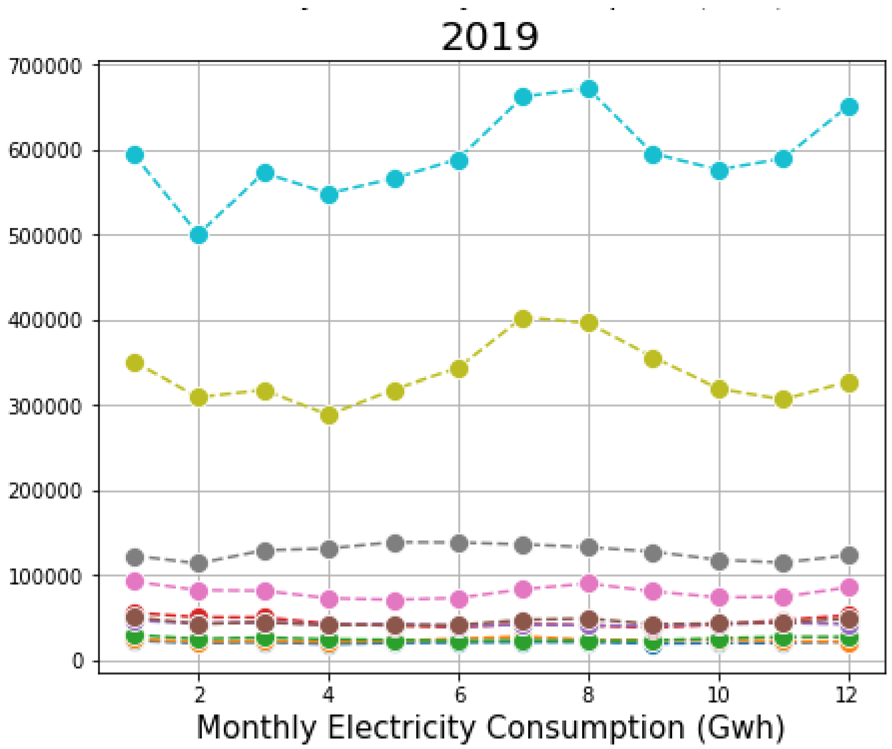
- In 2016, People's Republic of China consumed large amount of electricity in August. It consumed least in the month of February.
- People's Republic of China highest consumption is 5, 80,000Gwh.
- The second largest consumer of electricity- United States consumed high electricity in August and consumed least in April.
- For these two countries, we can observe the pattern- Steady increase in consumption from April to August and then decrease in consumption later on and then again drastic jump in December.



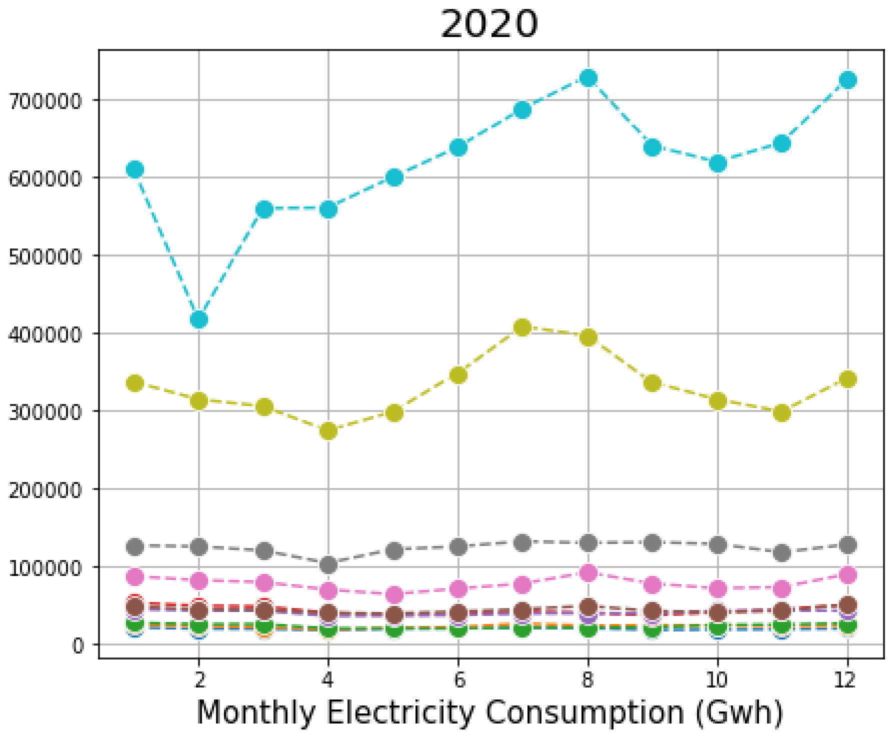
- In 2017, People's Republic of China consumed large amount of electricity in July. It consumed least in the month of February.
- The second largest consumer of electricity- United States consumed high electricity in July and consumed least in February, April.
- For these two countries, we can observe the pattern- Steady increase in consumption from April to August and then decrease in consumption later on and then again drastic jump in December.



- In 2018, People's Republic of China consumed large amount of electricity in July, August. It consumed least in the month of February.
- In July, August months we can observe the People's Republic of China consumption is equal.
- The second largest consumer of electricity- United States consumed high electricity in August and consumed least in April.
- For these two countries, we can observe the pattern- Steady increase in consumption from April to August and then decrease in consumption later on and then again drastic jump in December.



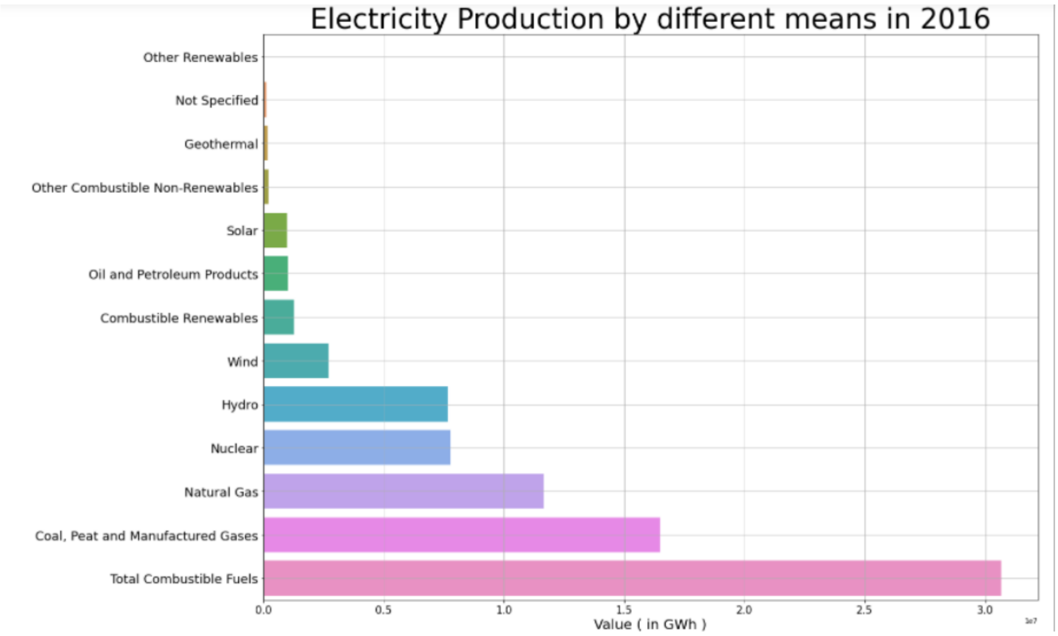
- In 2019, People's Republic of China consumed large amount of electricity in August. It consumed least in the month of February.
- People's Republic of China highest consumption is 6, 80,000Gwh.
- With People's Republic of China, we can observe the pattern- drastic increase in production from April to August and then saw a decrease and then increased in December.
- The second largest consumer of electricity- United States consumed high electricity in July and consumed least in April.
- With United States, we can observe the pattern- drastic increase in production from April to July and then saw a decrease and then increased in December.



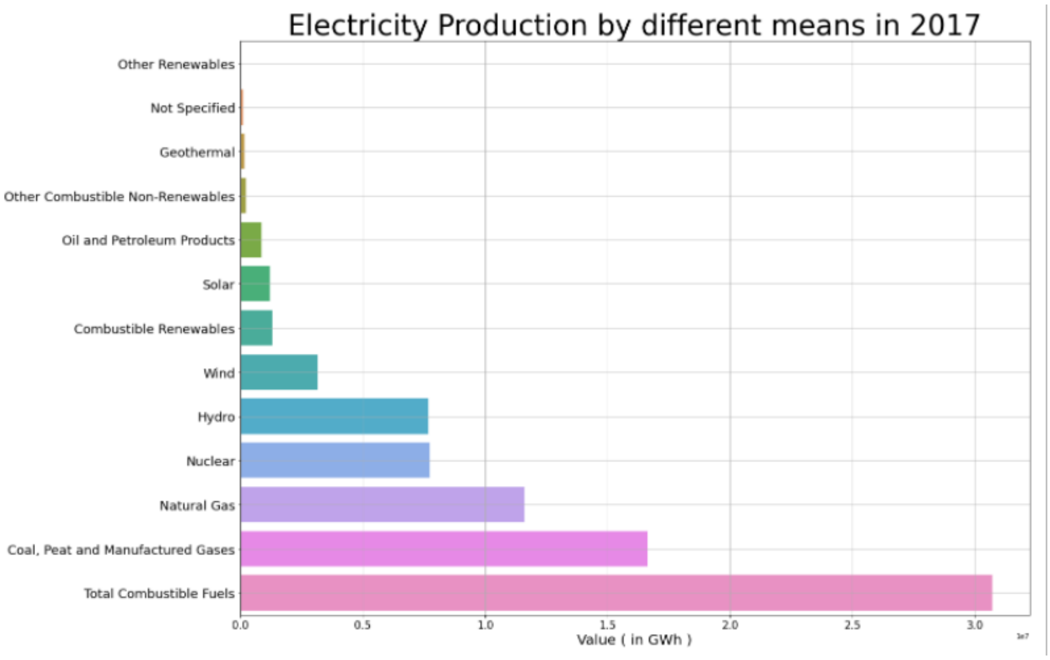
- In 2020, People's Republic of China consumed large amount of electricity in August and December. It consumed least in the month of February.
- People's Republic of China highest consumption is 7, 30,000Gwh.
- Here, People's Republic of China consumed equal amounts in the months of August and December.
- With People's Republic of China, we can observe the pattern- drastic increase in production from April to August and then saw a decrease and then increased in December.
- The second largest consumer of electricity- United States consumed high electricity in July and consumed least in April.
- With United States, we can observe the pattern- drastic increase in production from April to July and then saw a decrease and then increased in December.

Fig: 6.5- Monthly overall production and monthly overall consumption of selective countries for last 5 years

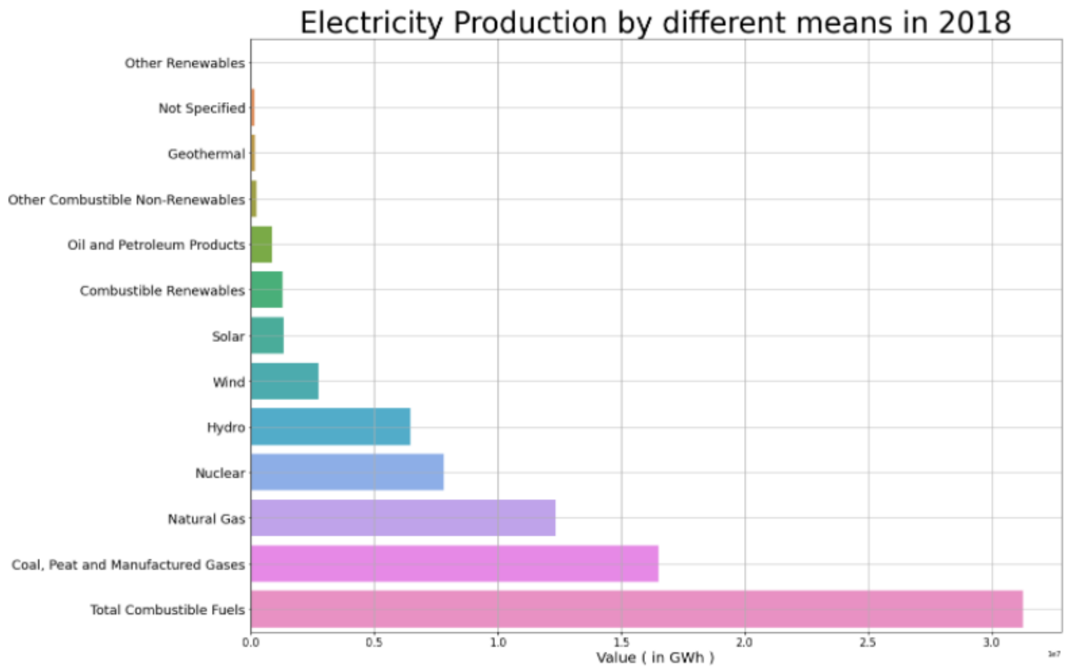
6.6 Electricity Production by different means in last 5 years



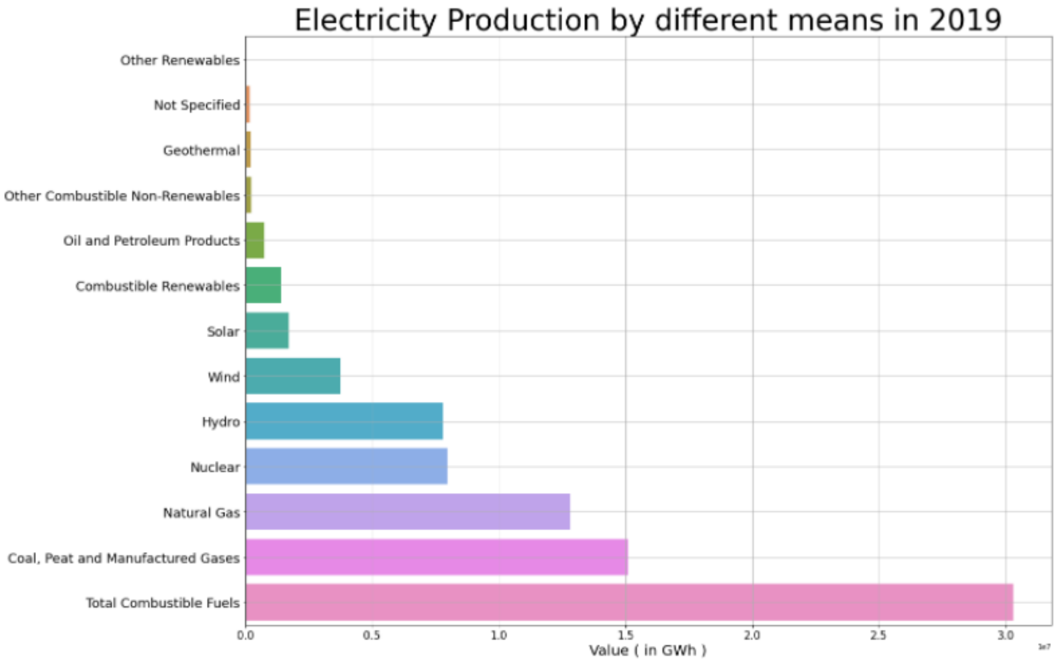
- In 2016, Total Combustible Fuels are the major resource for production of electricity.
- Coal, Peat and Manufactured Gases are the next major resources which are used to produce electricity in 2016. These are in half the value of Total Combustible Fuels used.
- Later followed by Natural Gas, Nuclear, Hydro, and Wind.
- Here, we can also observe that Hydro and Nuclear used are mostly in equal amounts.



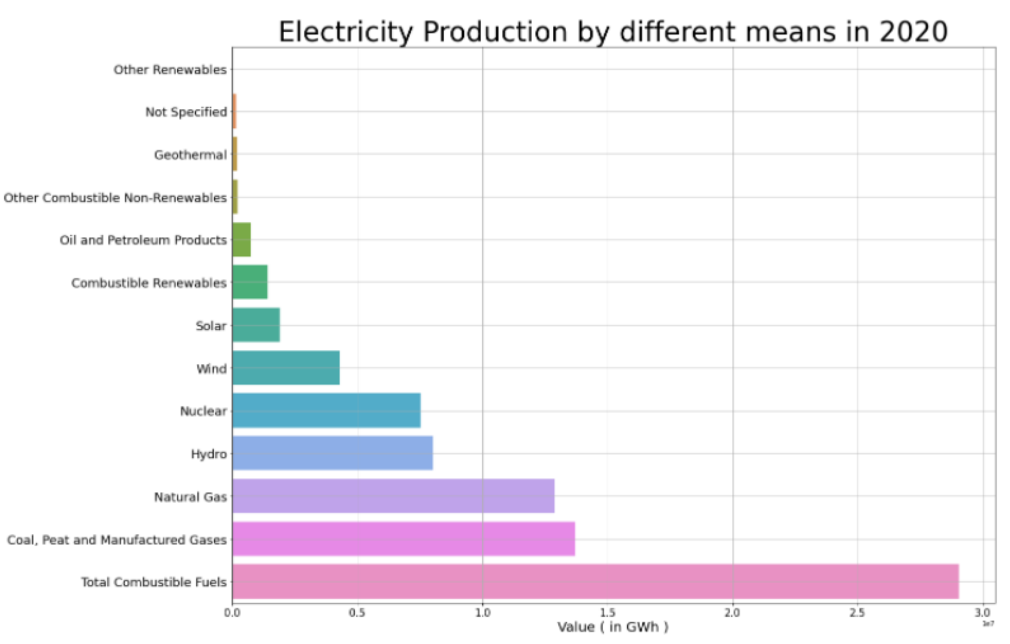
- In 2017, Total Combustible Fuels are the major resource for production of electricity.
- Coal, Peat and Manufactured Gases are the next major resources which are used to produce electricity in 2017. These are in half the value of Total Combustible Fuels used.
- Later followed by Natural Gas, Nuclear, Hydro, and Wind.
- Here, we can also observe that Hydro and Nuclear used are mostly in equal amounts.



- In 2018, Total Combustible Fuels are the major resource for production of electricity.
- Coal, Peat and Manufactured Gases are the next major resources which are used to produce electricity in 2018. These are in half the value of Total Combustible Fuels used.
- Later followed by Natural Gas, Nuclear, Hydro, and Wind.



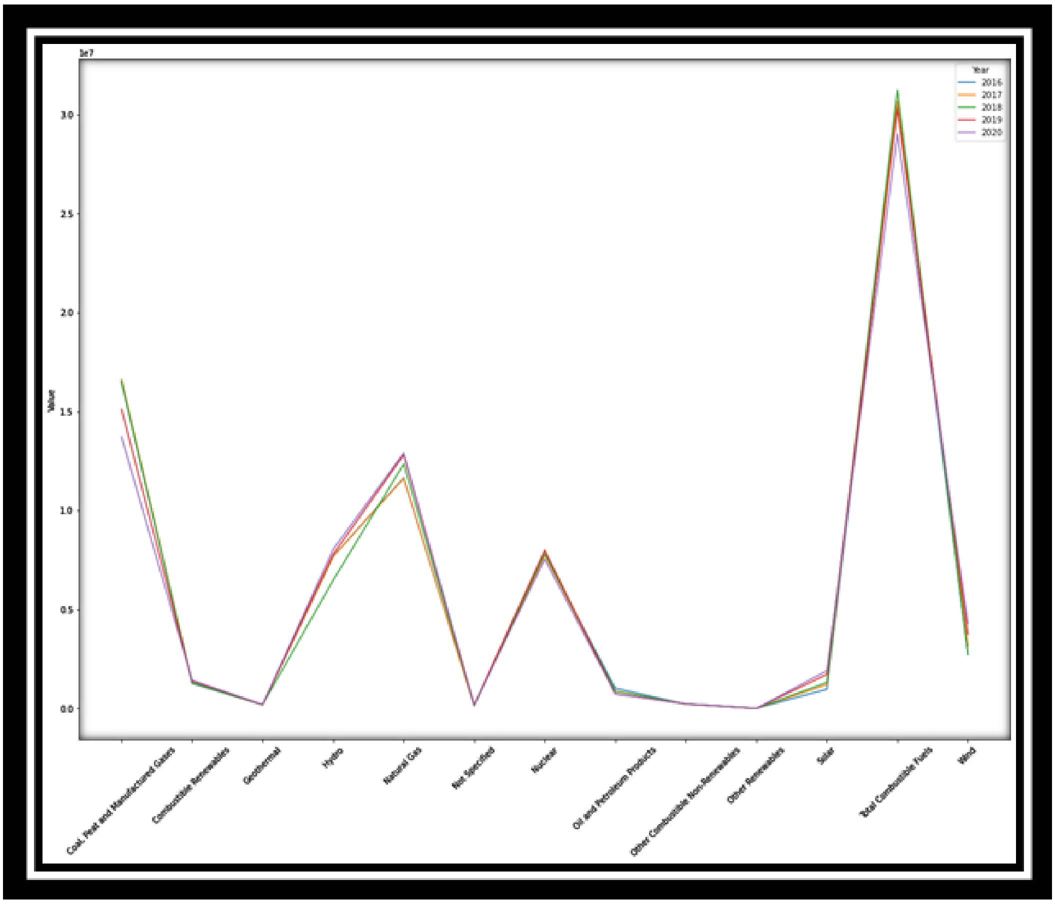
- In 2019, Total Combustible Fuels are the major resource for production of electricity.
- Coal, Peat and Manufactured Gases are the next major resources which are used to produce electricity in 2019. It is the exact half of the Total Combustible Fuels used.
- Here, we can also observe that Hydro and Nuclear used are mostly in equal amounts.
- The usage of Natural Gas is increased compared to last years.



- In 2020, Total Combustible Fuels are the major resource for production of electricity. This value got decreased than in previous year.
- Coal, Peat and Manufactured Gases are the next major resources which are used to produce electricity in 2020.
- Here, we can also observe that Hydro and Nuclear used are mostly in equal amounts.
- Also, Natural Gas value is almost meeting up with the Coal, Peat and Manufactured Gases.

Fig: 6.6- Electricity Production by different means in last 5 years

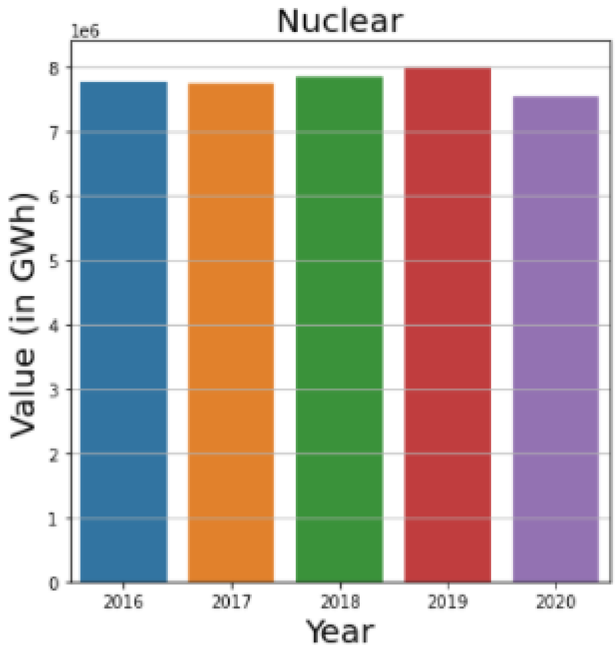
6.7 Comparison of Electricity Production by different means for the last 5 years



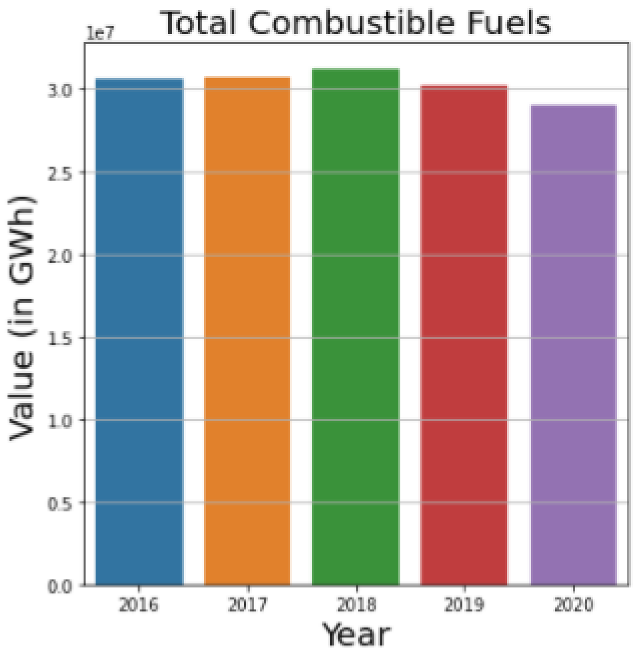
- For last 5 years, Total Combustible Fuels were the large contributors for the electricity production.
- It is followed by Coal, Peat and Manufactured gases as the second highest contributor for the electricity production.
- Usage of Total Combustible Fuels slowly deterred from 2016-2020.
- Minuscule Contributors were Other Renewables, Geothermal, Not specified and Other Combustible Non-Renewables.

Fig: 6.7- Comparison of Electricity Production by different means for the last 5 years

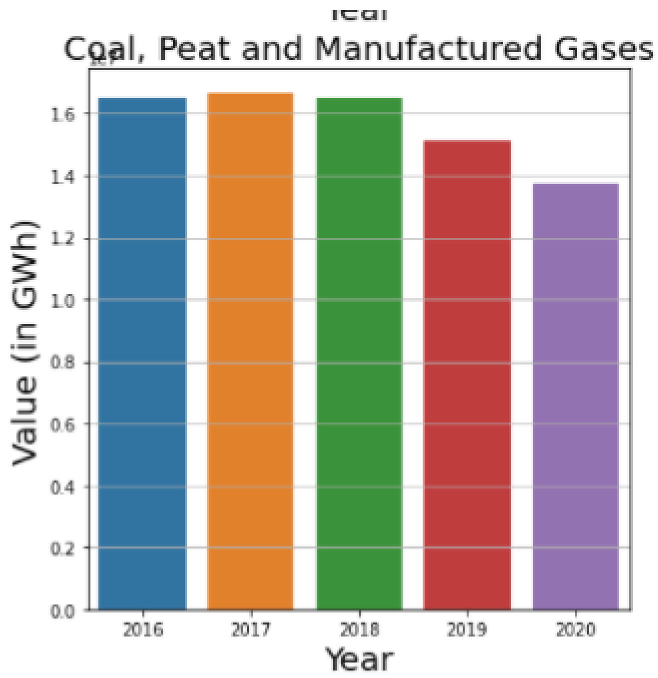
6.8 Overview of electricity production by each means for the last 5 years



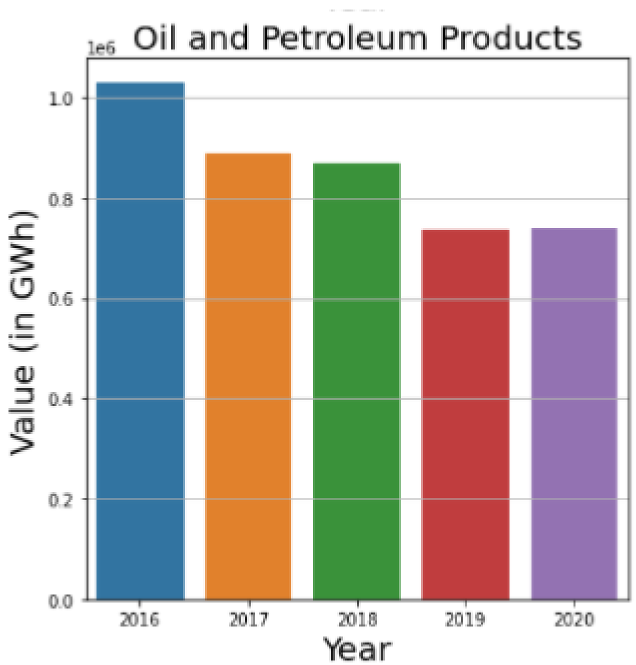
- Usage of Nuclear in the production of electricity increased from 2016-2019 and drastically fell in 2020.



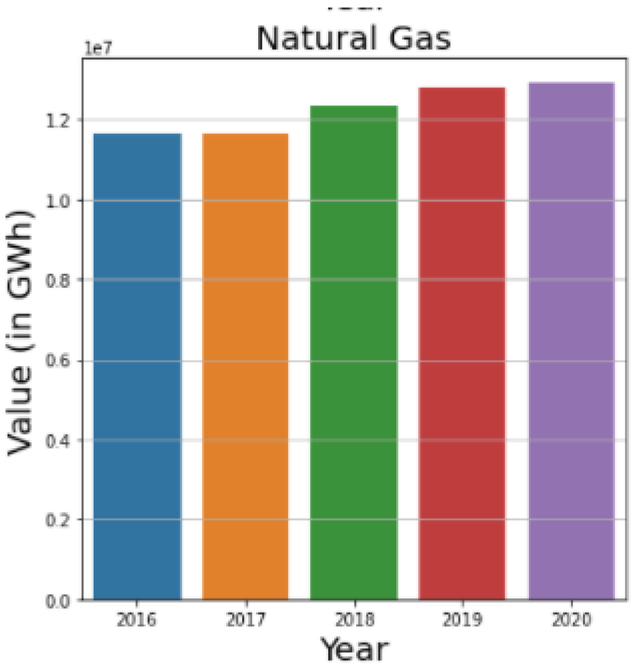
- Usage of Total Combustible Fuels in the production of electricity increased from 2016-2018, but then fell in the following years.



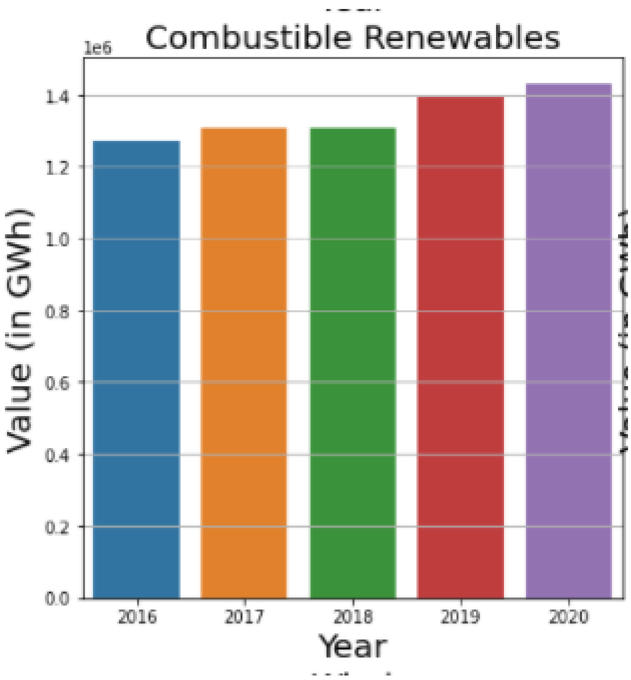
- Usage of Coal, Peat and Manufactured Gases in the production of electricity increased from 2016-2017, but then fell drastically in the following years.



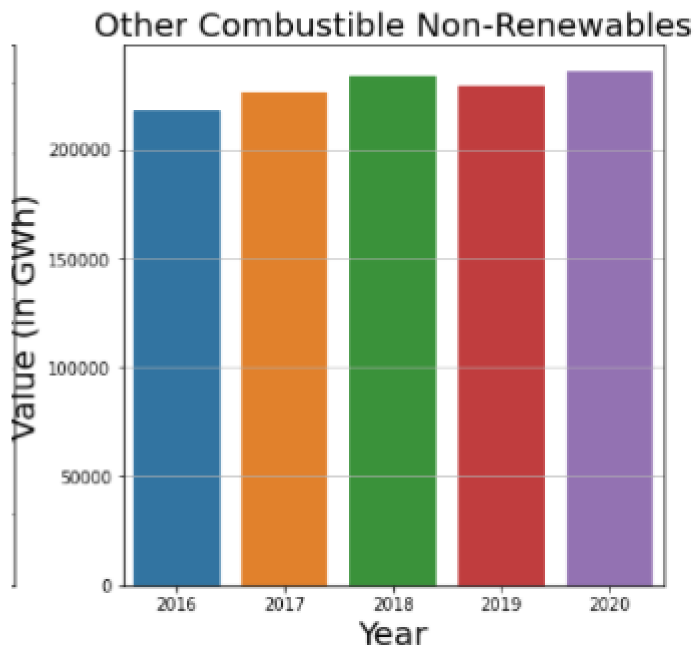
- Usage of Oil and Petroleum Products in the production of electricity decreased from 2016-2019 and then was in constant in 2020.



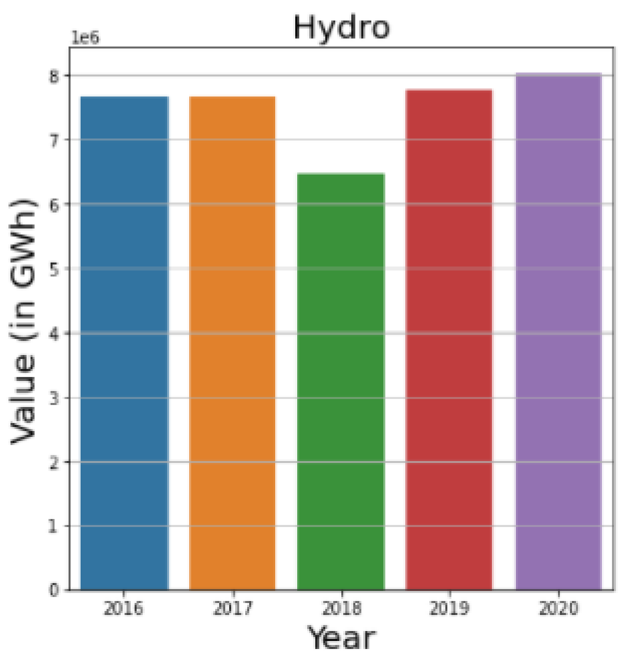
- Usage of Natural Gas in the production of electricity saw an increase.



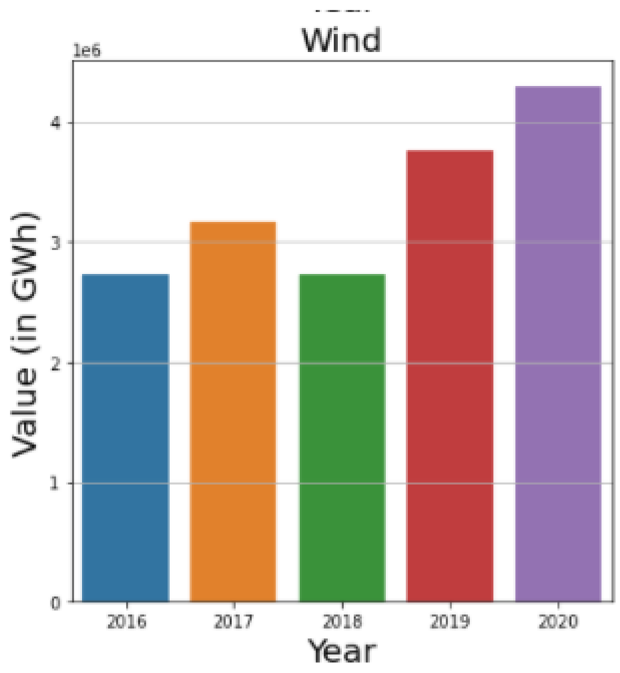
- Usage of Combustible Renewables in the production of electricity saw an increase too.



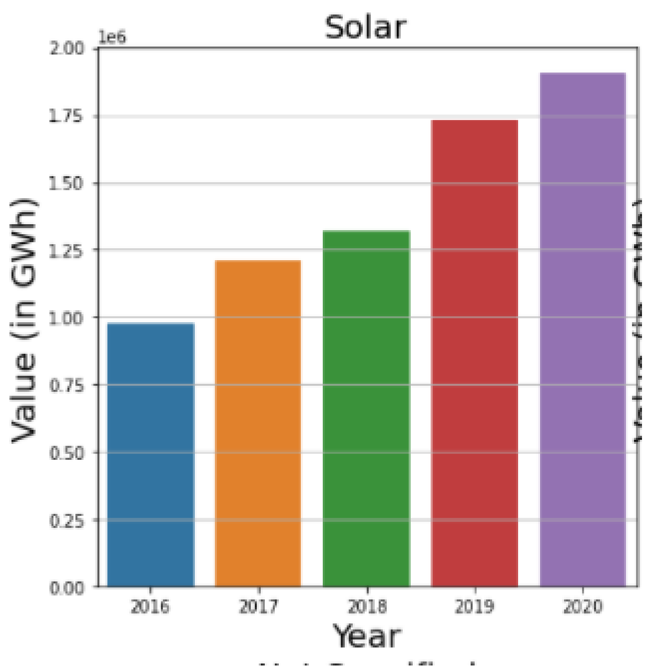
- Usage of Other Combustible Non-Renewables in the production of electricity saw an increase from 2016-2018, but then fell for a year and then increased in 2020.



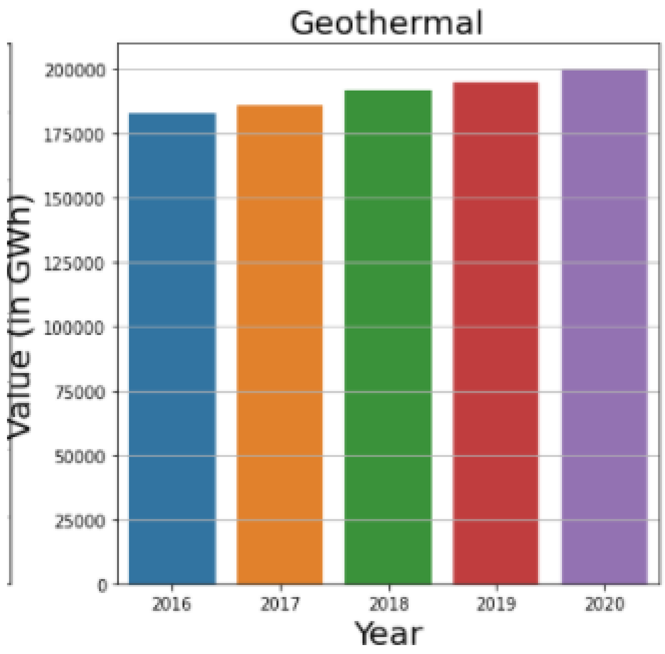
- Usage of Hydro in the production of electricity decreased from 2017-2018, but then drastically jumped in 2019 and saw a linear increase later on.



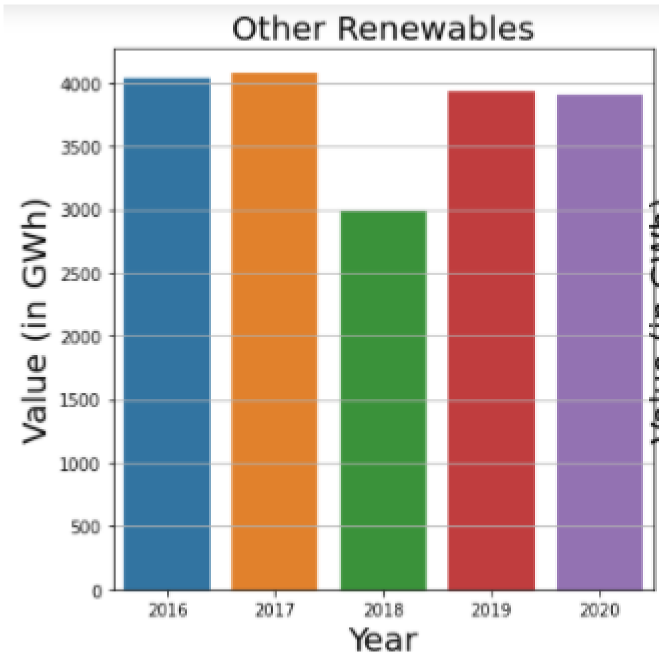
- Usage of Wind in the production of electricity saw an increase from 2016-2020, it took a dig in 2018.



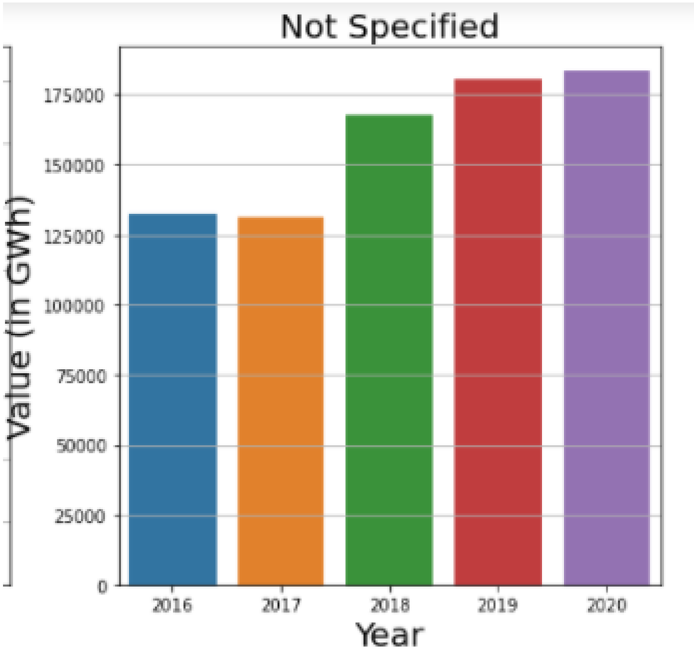
- Usage of Solar in the production of electricity saw a linear increase from 2016-2020.



- Usage of Geothermal in the production of electricity saw a slight increase from 2016-2020.



- Usage of Other Renewables in the production of electricity saw an increase from 2016-2017 and then fell drastically in 2018. Later it jumped up in 2019 and saw a stable stance



- Usage of Not Specified in the production of electricity saw a increase from 2017-2020.

Fig: 6.8- Overview of electricity production by each means for the last 5 years

7. DATA PRE-PROCESSING

7.1 Encoding

- Columns - 'Year', 'Country' are encoded using get-dummies and the resultant dataset is as shown in Fig. 7.1.

	Value	Month	Year_2016	Year_2017	Year_2018	Year_2019	Year_2020	Country_Argentina	Country_Australia	Country_Austria	...	Country_Romania
19	20497.0	12	0	0	0	0	1	0	1	0	...	0
39	6304.0	12	0	0	0	0	1	0	0	1	...	0
59	7558.0	12	0	0	0	0	1	0	0	0	...	0
79	51498.0	12	0	0	0	0	1	0	0	0	...	0
99	6884.0	12	0	0	0	0	1	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...
58238	3247.0	1	1	0	0	0	0	0	0	0	...	0
58247	364.0	1	1	0	0	0	0	0	0	0	...	0
58256	168.0	1	1	0	0	0	0	0	0	0	...	0
58270	685.0	1	1	0	0	0	0	0	0	0	...	0
58285	2987.0	1	1	0	0	0	0	0	0	0	...	0

3120 rows × 59 columns

Fig: 7.1- Data after encoding

7.2 Train – test split

- The data is split into train- test with test-size=0.3 respectively as shown in Fig. 7.2.

```
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=12,test_size=0.3)
```

```
Size of x_train is:  2184 records
Size of y_train is:  2184 records
Size of x_test is:  936 records
Size of y_test is:  936 records
```

Fig: 7.2- Data after Train-test split

7.3 Scaling

- As the last step, scaling is done on the y_train, y_test and after scaling they are shown in Fig. 7.3.

```
y_train
```

```
array([[-0.41897298],
       [-0.39527015],
       [-0.29268705],
       ...,
       [-0.25539453],
       [-0.40493575],
       [-0.40996776]])
```

```
y_test
```

```
array([[-0.30310582],
       [-0.15405578],
       [ 0.02386007],
       [-0.4247909 ],
       [-0.4135371 ],
       [-0.28428764],
       [-0.38112923],
       [-0.37299178],
       [ 3.86988581],
       [-0.41613497],
       [-0.42107966],
       [-0.18873407],
       [-0.36303146],
       [-0.26303532],
       [-0.41356439],
       [ 4.08466303],
       [-0.40451005],
       [ 4.57807268],
       [-0.39600694],
```

Fig: 7.3- Data after Scaling

8. MODELLING

- Linear Regressor, Decision Tree Regressor, Random Forest Regressor, Ada – Boost Regressor and XG – Boost models are built and then trained on the x_train and y_train.

Linear Regression

```
lr=LinearRegression()
```

```
lr.fit(x_train,y_train)
```

```
LinearRegression()
```

Decision Tree Regressor

```
dt=DecisionTreeRegressor()
```

```
dt.fit(x_train,y_train)
```

```
DecisionTreeRegressor()
```

Random Forest Regressor

```
rf=RandomForestRegressor()
```

```
rf.fit(x_train,y_train.ravel())
```

```
RandomForestRegressor()
```

Ada-Boost Regressor

```
ab=AdaBoostRegressor()
```

```
ab.fit(x_train,y_train.ravel())
```

```
AdaBoostRegressor()
```

Xgboost Regressor

```
xg=xgboost.XGBRegressor()
```

```
xg.fit(x_train,y_train)
```

```
XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
             colsample_bynode=1, colsample_bytree=1, enable_categorical=False,
             gamma=0, gpu_id=-1, importance_type=None,
             interaction_constraints='', learning_rate=0.300000012,
             max_delta_step=0, max_depth=6, min_child_weight=1, missing=nan,
             monotone_constraints='()', n_estimators=100, n_jobs=4,
             num_parallel_tree=1, predictor='auto', random_state=0, reg_alpha=0,
             reg_lambda=1, scale_pos_weight=1, subsample=1, tree_method='exact',
             validate_parameters=1, verbosity=None)
```

Fig: 8.1- Machine Learning models

- Snippets of these models are shown in Fig: 8.1.
- Neural network model is also built to compare whether machine learning models perform well or neural networks.
- Snapshot of the neural network build and compiled is as shown in Fig: 8.2.

```
model=keras.Sequential([
        layers.Dense(64,activation='relu',input_shape=[len(x_train.keys())]),
        layers.Dense(64,activation='relu'),
        layers.Dense(1)
    ])

optimizer=tf.keras.optimizers.RMSprop(learning_rate=0.001) # default learning rate is 0.001
model.compile(loss='mse',
             optimizer=optimizer,
             metrics=['mse'])
```

```
In [229]: model.fit(x_train,y_train,batch_size=100,epochs=100,validation_split=0.2,verbose=1)
          Epoch 92/100
          18/18 [==============================] - 0s 3ms/step - loss: 0.0085 - mse: 0.0085 - val_loss: 0.0093 - val_mse: 0.0093
          Epoch 93/100
          18/18 [==============================] - 0s 3ms/step - loss: 0.0078 - mse: 0.0078 - val_loss: 0.0178 - val_mse: 0.0178
          Epoch 94/100
          18/18 [==============================] - 0s 3ms/step - loss: 0.0080 - mse: 0.0080 - val_loss: 0.0286 - val_mse: 0.0286
          Epoch 95/100
          18/18 [==============================] - 0s 3ms/step - loss: 0.0090 - mse: 0.0090 - val_loss: 0.0089 - val_mse: 0.0089
          Epoch 96/100
          18/18 [==============================] - 0s 3ms/step - loss: 0.0091 - mse: 0.0091 - val_loss: 0.0078 - val_mse: 0.0078
          Epoch 97/100
          18/18 [==============================] - 0s 3ms/step - loss: 0.0078 - mse: 0.0078 - val_loss: 0.0318 - val_mse: 0.0318
          Epoch 98/100
          18/18 [==============================] - 0s 3ms/step - loss: 0.0104 - mse: 0.0104 - val_loss: 0.0105 - val_mse: 0.0105
          Epoch 99/100
          18/18 [==============================] - 0s 3ms/step - loss: 0.0076 - mse: 0.0076 - val_loss: 0.0104 - val_mse: 0.0104
          Epoch 100/100
          18/18 [==============================] - 0s 3ms/step - loss: 0.0085 - mse: 0.0085 - val_loss: 0.0172 - val_mse: 0.0172

Out[229]: <keras.callbacks.History at 0x2200356ebe0>
```

Fig: 8.2- Neural network model

9. Model Evaluation

- RMSE metric is used to evaluate a model in this study.
- The performance of the above created models on training data is shown below.

```
The rmse of Linear Regressor on train data is:  0.0868604924944168

The rmse of Decision Tree Regressor on train data is:  0.0

The rmse of Random Forest Regressor on train data is:  0.015513950075327623

The rmse of Ada-Boost Regressor on train data is:  0.42535857052246484

The rmse of Xgboost Regressor on train data is:  0.010108731985520448

The rmse of Neural Network model on train data is:  0.06672895809412563
```

Fig: 9.1- RMSE of built models on training data

- The performance of the above created models on testing data is shown below.

```
The rmse of Linear Regressor on test data is:  0.08693956912978669

The rmse of Decision Tree Regressor on test data is:  0.04703727903530587

The rmse of Random Forest Regressor on test data is:  0.04761529800768241

The rmse of Ada-Boost Regressor on test data is:  0.44551571589360767

The rmse of Xgboost Regressor on test data is:  0.03647528208620669

The rmse of Neural Network model on test data is:  0.12130447956494268
```

Fig: 9.2- RMSE of built models on testing data

- The less the RMSE value of the model on test data, the better the performance of the model is.
- Here, **Xgboost Regressor is having less RMSE value** compared with the other models, so we conclude that it is a better performer.

10. CONCLUSION

The hourly power load usage is employed in this research to anticipate future load electricity demands. We summarized the importance of demand forecasting and related research in the Introduction chapter. We began by performing exploratory data analysis and providing descriptive information to learn more about the dataset's properties. After that, there are chapters on data cleansing, data preparation, and data pre-processing. In the Modeling chapter, the Python programming language is used to apply five machine learning and a neural network technique to the dataset. In the chapter Model Evaluation, the results of each model are reported. This experiment demonstrates that energy consumption can be predicted using machine learning methods, and that the models can be used to forecast future demand.

11. APPLICATIONS

The models created in this study have a wide range of applications. Managers in the energy business can utilize them for the following purposes:

Measures to save energy and money should be quantified

The models can be used to make energy consumption forecasts under present conditions. The values under new conditions may be compared to those in the model, and the differences can be simply estimated once the energy measures take effect.

Taking Reasonable Decisions

The models can be used to forecast future values. If the consumption levels rise to a critical level in the forecast, then decisions can be made to recommission or renovate relevant parts of the power system. Improved forecasting is beneficial to the deployment of renewable energy, planning for

59 high/low load days, and reducing wastage from polluting on reserve standby generation (typically inefficient gas or coal-fired power plants).

12. APPENDIX

Python Coding:

```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder,StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor,AdaBoostRegressor,GradientBoostingRegressor
import xgboost
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from sklearn.metrics import mean_squared_error
#dataset loading
a=pd.read_csv('C:\\Users\\Teja\\Desktop\\MAHE-project\\Project\\Data\\Monthly_Electricity_Statistics.csv')
```

```python
def month(x):
    a=x[:3].lower()
    if a=='jan':
        return 1
    elif a=='feb':
        return 2
    elif a=='mar':
        return 3
    elif a=='apr':
        return 4
    elif a=='may':
        return 5
    elif a=='jun':
        return 6
    elif a=='jul':
        return 7
    elif a=='aug':
        return 8
    elif a=='sep':
        return 9
    elif a=='oct':
        return 10
    elif a=='nov':
        return 11
    else: return 12
def year(x):
    return '20'+x[-2:]
a['Month']=a['Time'].apply(lambda x : month(x))
a['Year']=a['Time'].apply(lambda x : year(x))
a.drop(columns='Time',inplace=True)
a=a[a['Product']!='Total Renewables (Geo, Solar, Wind, Other)'] # 'Total Renewables (Geo, Solar, Wind, Other)' is redundant
a=a[(a['Balance']=='Net Electricity Production') | (a['Balance']=='Final Consumption (Calculated)')]
#Only 'Net Electricity Production' and 'Final Consumption (Calculated)' is considered for our project,
#so removing the other records.
```

```python
#For model building we will only consider 'Final Consumption (Calculated)' records because we are more interested
#in predicting the electricity consumption, but for analysing the trend we would use both 'Net Electricity Production'
#and 'Final Consumption (Calculated)' records.
a=a[((a['Balance']=='Net Electricity Production')&(a['Product']!='Electricity'))|((a['Balance']=='Final Consumption (Calculated)')&(a['Product']=='Electricity'))]
#Removing the records where Product='Electricity' as it is nothing but the net electricity for the respective month in a year
a['Value']=a['Value'].apply(lambda x: np.round(x,0))
data=a[a['Balance']=='Final Consumption (Calculated)'].drop(['Balance','Product'],axis=1)
data=pd.get_dummies(data,columns=['Year','Country'])
x=data.drop('Value',axis=1)
y=data['Value']
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=12,test_size=0.3)
s=StandardScaler()
ss=s.fit(np.array(y_train).reshape(-1,1))
y_train=ss.transform(np.array(y_train).reshape(-1,1))
y_test=ss.transform(np.array(y_test).reshape(-1,1))
lr=LinearRegression()
lr.fit(x_train,y_train)
rmse_lr=mean_squared_error(y_test, lr.predict(x_test), squared=False)
dt=DecisionTreeRegressor()
dt.fit(x_train,y_train)
rmse_dt= mean_squared_error(y_test, dt.predict(x_test), squared=False)
rf=RandomForestRegressor()
rf.fit(x_train,y_train.ravel())
rmse_rf= mean_squared_error(y_test, rf.predict(x_test), squared=False)
ab=AdaBoostRegressor()
ab.fit(x_train,y_train.ravel())
rmse_ab= mean_squared_error(y_test, ab.predict(x_test), squared=False)
xg=xgboost.XGBRegressor()
xg.fit(x_train,y_train)
rmse_xg= mean_squared_error(y_test, xg.predict(x_test), squared=False)
```

```python
model=keras.Sequential([
        layers.Dense(64,activation='relu',input_shape=[len(x_train.keys())]),
        layers.Dense(64,activation='relu'),
        layers.Dense(1)
    ])

optimizer=tf.keras.optimizers.RMSprop(learning_rate=0.001) # default learning rate is 0.001
model.compile(loss='mse',
              optimizer=optimizer,
              metrics=['mse'])
model.fit(x_train,y_train,batch_size=100,epochs=100,validation_split=0.2,verbose=1)
rmse_nn=mean_squared_error(y_test, model.predict(x_test), squared=False)
print('The rmse of Linear Regressor on test data is: ',rmse_lr)
print('The rmse of Decision Tree Regressor on test data is: ',rmse_dt)
print('The rmse of Random Forest Regressor on test data is: ',rmse_rf)
print('The rmse of Ada-Boost Regressor on test data is: ',rmse_ab)
print('The rmse of Xgboost Regressor on test data is: ',rmse_xg)
print('The rmse of Neural Network model on test data is: ',rmse_nn)
```