

# Interplay Between Electricity Generation and Sales in New York: Correlations and Key Insights

## Project Description

The report explores the connection between electricity generation and consumption over the past several decades. It examines trends in electricity production across different fuel types and its utilization in residential, commercial, and transportation sectors, identifying key patterns and relationships. The insights aim to support strategic decision-making in energy planning and infrastructure development, striving to achieve an optimal balance between energy supply and demand during a period of significant energy transition.

## Project Question

How are changes in electricity generation across fuels affecting consumption patterns in various sectors, and what can we learn to optimize energy planning and infrastructure in the course of the transition?

## Data Sources

### Datasource1: Electricity Sales by Sector

- Metadata URL: <https://catalog.data.gov/dataset/electricity-sales-by-sector-gwh-beginning-1980-1f109/resource/65764672-4abb-4bcf-bd76-3eba55f77315>.
- Data URL: <https://catalog.data.gov/dataset/electricity-sales-by-sector-gwh-beginning-1980-1f109/resource/65764672-4abb-4bcf-bd76-3eba55f77315>.
- Data Type: CSV

The dataset offers information on the total electricity demand for New York State. It includes electricity sales to end consumers, categorized by sector (Residential, Commercial, Industrial, and Transportation) in gigawatt-hours (GWh).

### Datasource2: Electricity Generation by Fuel Type

- Metadata URL: <https://catalog.data.gov/dataset/electric-generation-by-fuel-type-gwh-beginning-1960/resource/df1528d1-bdd0-4594-8ef8-793b298cfd5>.
- Data URL: <https://catalog.data.gov/dataset/electric-generation-by-fuel-type-gwh-beginning-1960/resource/df1528d1-bdd0-4594-8ef8-793b298cfd5>.
- Data Type: CSV

The dataset contains information on New York State's total electricity demand and in-state generation measured in gigawatt-hours. It details electricity production from various sources,

including coal, natural gas, petroleum products, hydro, nuclear, waste, landfill gas, wood, wind, solar, and net electricity imports.

## Extract, Transform, Load (ETL) Pipeline

The **Extract Phase** focuses on getting some raw data from online sources. The URLs of the datasets "Electricity Sales by Sector" and "Electricity Generation by Fuel Type" are identified, and the data is retrieved through HTTP requests. The datasets are saved locally in a specific directory (./data/raw\_csv/) as CSV files to enable access for further process.

The **Transform Phase** focuses on cleaning, preparing, and merging of data for analysis. The raw CSV data files are imported using pandas, and info is merged using the common Year column in outer join format to retain all data entries.

The **Load Phase** ensures that all processed data is kept safe for future analysis. The formed dataset is carried to a SQLite database (ClimateData\_revised.sqlite), under the table ElectricityData. Data insertion is checked by querying the database, inspecting for records if the storage and integration have worked properly.

## Results and Limitations

A correlation of 0.973 signifies a very strong positive relationship. To explain:

---

Correlation between Total Generation and Total Sales: 0.9733213686616367

Figure-1.2

A high correlation is an indication of a strong linear relationship: In other words, a correlation coefficient equal to 0.973 implies that there is a very strong linear relationship between total electricity generation and total electricity sales, meaning electricity generation grows with increasing sales almost proportionally and vice versa.

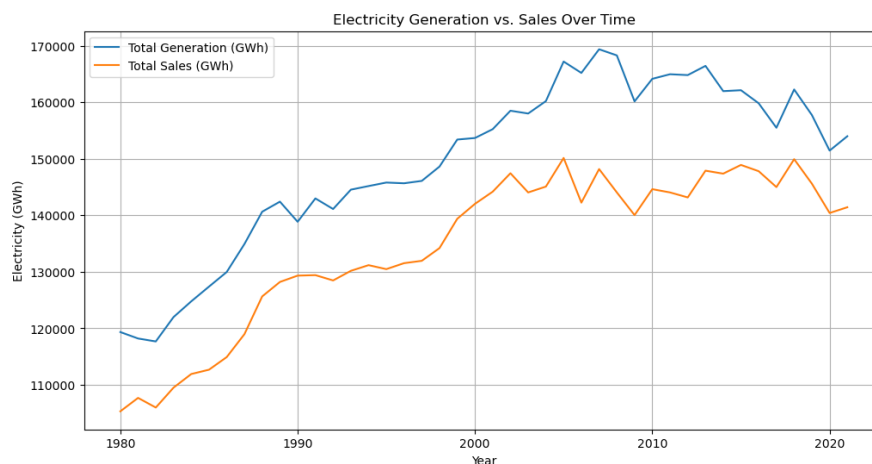


Figure-1.2

**Supply and Demand:** This strong correlation is a reflection of the direct relationship between a supply of generation and demand for sales in the electricity market. Utilities adjust generation to meet the demand of residential, commercial, industrial, and transportation sectors.

**Operational Efficiency-**The close match between generation and sales indicates operational management is efficient in that its electricity generation closely follows consumption requirements, thus reducing waste.

**Economic and Policy Implications:** This relationship could be influenced by economic growth, changes in energy policy, technological advances in energy efficiency, and shifts toward renewable energy sources that could affect both generation patterns and sales dynamics.

Below Heat Map suggests

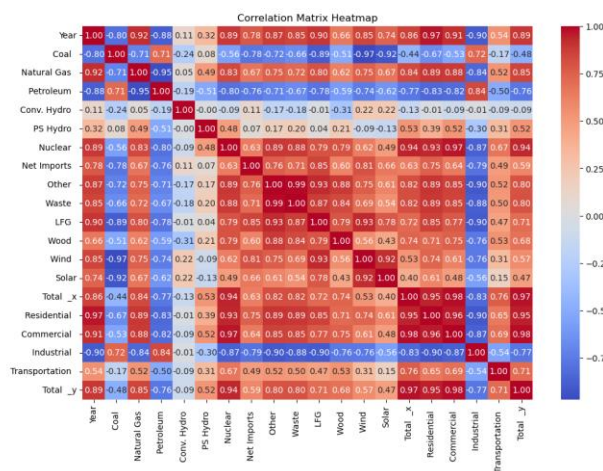


Figure-1.3

**The Energy Transition:** The inverse correlations of coal and petroleum with year indicate that these two sources were being displaced by others, probably natural gas, nuclear, and renewables.

**Changing Energy Consumption:** The strong correlations among residential, commercial, and industrial sectors suggest that some kind of economic or policy changes affecting one sector would quite likely be transmitted to the other.

**Increasing Renewable Adoption:** The fact that some high correlations exist among themselves indicative of the fact that something was in place favouring wind and solar energy, in general at the cost of traditional fossil fuels.

This Report thus shares views into trends of the production and consumption of energy across sectors such that they show greater shifts in either the establishment toward sustainability or efficiency.