Interplay Between Electricity Generation and It's Sales in New York: Correlations and Key Insights

Introduction:

The energy sector has evolved very significantly over the past few decades due to changes in technologies, policy changes, and the associated shifts in consumer demand. It's been a period of change that has emphasized the need for a deep understanding of the relationship between electricity generation and consumption. Consideration in increasing the proliferation of renewable sources of energy, changes in fuel type use, and sector-specific consumption patterns have greatly impacted the dynamics of energy systems. As the world strides toward a sustainable energy future, one of the greatest challenges is to achieve the optimal balance between supply and demand.

This report aims to provide insights into these dynamics by analysing historical trends in electricity production and its utilization across residential, commercial, industrial, and transportation sectors. Specifically, the report seeks to address the central question:

1. **How are changes in electricity generation across fuels affecting consumption patterns in various sectors?**
2. **what can we learn to optimize energy planning and infrastructure during this transition?**

Pipeline output and Data:

The ETL pipeline generates one structured SQLite database: GenerationToSales.sqlite. This database contains one table called Electricity Data in which the datasets concerning "Electricity Sales by Sector" and "Electricity Generation by Fuel Type" are brought together into a single table allowing easy querying and analysis of the merged information. Individual rows or columns may be retrieved depending on the requirements of the analyses through SQL queries.

```
gen_data = pd.read_csv(r'/content/Electric_Generation_By_Fuel_Type__GWh__Beginning_1960.csv')
gen_data.head()
```

	Year	Coal	Natural Gas	Petroleum	Conv. Hydro	PS Hydro	Nuclear	Net Imports	Other	Waste	LFG	Wood	Wind	Solar	Total
0	1980	19643	10766	37834	26241	0	19276	5575.370	0.0	0.0	0.0	0.0	0.0	0.0	119335.370
1	1981	19632	11566	36760	25658	0	17444	7136.554	0.0	0.0	0.0	0.0	0.0	0.0	118196.554
2	1982	20807	13974	32932	25329	0	14438	10196.177	0.0	0.0	0.0	0.0	0.0	0.0	117676.177
3	1983	20753	12428	34380	26162	0	16376	11888.300	0.0	0.0	0.0	0.0	0.0	0.0	121987.300
4	1984	21902	15395	28891	26586	0	21187	10812.065	0.0	0.0	0.0	0.0	0.0	0.0	124773.065

Figure 1.1 Above table has the Electricity Generation by Fuel Type. Contains data of generation over past 40 years.

```
Sales_data = pd.read_csv(r'/content/sample_data/Electricity_Sales_By_Sector__GWh__Beginning_1980.csv')
Sales_data.head()
```

	Year	Residential	Commercial	Industrial	Transportation	Total
0	1980	30583	40471	32110	2146	105310
1	1981	30702	42685	32240	2059	107686
2	1982	30626	42800	30484	2069	105979
3	1983	31803	43850	31424	2422	109498
4	1984	32836	47600	28789	2685	111910

Figure 1.2 Above table has the Electricity sales by sector. Contains data of generation over past 40 years.

```
(base) C:\Users\rithv\OneDrive\Desktop\MADE\MADE-FAU-Project-WS-2425>C:/Users/rithv/anaconda3/python.exe c:/Users/rithv/OneDrive/Desktop/MADE/MADE-FAU-Project-WS-2425
/project/pipeline.py
Downloaded data from https://data.ny.gov/api/views/pv7j-5nz8/rows.csv?accessType=DOWNLOAD to ./data/raw_csv/Electricity_Sales_by_Sector.csv
Downloaded data from https://data.ny.gov/api/views/h4gs-8qnu/rows.csv?accessType=DOWNLOAD to ./data/raw_csv/Electric_Generation_By_Fuel_Type.csv
Data transformation complete and saved to transformed_data.csv
(1980, 30583, 40471, 32110, 2146, 105310, 19643, 10766, 37834, 26241, 0, 19276, 5575.37, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 119335.37)
(1981, 30702, 42685, 32240, 2059, 107686, 19632, 11566, 36760, 25658, 0, 17444, 7136.554, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 118196.554)
(1982, 30626, 42800, 30484, 2069, 105979, 20807, 13974, 32932, 25329, 0, 14438, 10196.177, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 117676.177)
(1983, 31803, 43850, 31424, 2422, 109498, 20753, 12428, 34380, 26162, 0, 16376, 11888.3, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 121987.3)
(1984, 32836, 47600, 28789, 2685, 111910, 21902, 15395, 28891, 26586, 0, 21187, 10812.065, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 124773.065)
(1985, 32757, 48816, 28659, 2442, 112674, 24255, 15995, 25927, 26956, 0, 24092, 10154.529, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 127379.529)
(1986, 33771, 50430, 28107, 2601, 114908, 20815, 12471, 31911, 29480, 0, 22084, 13203.525, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 129964.525)
(1987, 35294, 52256, 28726, 2693, 118970, 25304, 16385, 31389, 27546, 0, 22926, 11372.669, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 134922.669)
(1988, 37460, 55305, 30155, 2722, 125643, 28698, 13936, 39931, 23901, 0, 24175, 9993.721, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 140634.721)
(1989, 37878, 56051, 31448, 2825, 128203, 31228, 17141, 40963, 24697, 0, 22847, 5538.569, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 142414.569)
(1990, 38574, 56025, 31929, 2795, 129324, 25913, 22724, 33885, 28188, 0, 23623, 4519.213, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 138852.672)
```

Figure 1.3 Above is a structured SQLite database named GenerationToSales.sqlite, it contains a single table, Electricity Data, which combines information from the "Electricity Sales by Sector" and "Electricity Generation by Fuel Type" datasets.

During the transformation phase, an outer join on the column Year was applied between the two datasets in order to create the Electricity Data table. This ensured that no data points were lost in the process. The final table has exhaustive entries for electricity generation by fuel type and respective consumption by sector. Although the dataset was cleaned to a great extent, some columns having either low data relevance or high missing values were dropped in order to preserve the accuracy and reliability.

The dataset complies with standard data-sharing practices, making it suitable for further analysis and reporting. All processed data is securely stored within the SQLite database, ensuring consistent access and seamless integration into energy-related studies or strategic planning processes. Data integrity was verified through rigorous checks, confirming the successful completion of the ETL process.

Analysis:

The analysis begins by loading two datasets:

The datasets are read into pandas Data Frames for further processing.

To ensure consistency and compatibility: Missing values in both datasets are replaced with zeros. The Year column in each dataset is converted to an integer data type, establishing a common field for merging. The datasets are merged using an outer join on the Year column, resulting in a combined dataset (merged_data) that includes all entries from both sources. This comprehensive dataset allows for a holistic analysis of electricity generation and consumption trend

```
correlation_matrix = merged_data.corr()
print("Correlation Matrix:\n", correlation_matrix)
```

```
Correlation Matrix:
                     Year      Coal    Natural Gas     Petroleum     \
Year             1.000000 -0.800339       0.918241     -0.880120
Coal            -0.800339  1.000000      -0.708329      0.711339
Natural Gas      0.918241 -0.708329       1.000000     -0.948347
Petroleum       -0.880120  0.711339      -0.948347      1.000000
Conv. Hydro      0.114461 -0.240387       0.050191     -0.192531
PS Hydro         0.324919  0.083122       0.486298     -0.508231
Nuclear          0.887526 -0.564733       0.831756     -0.797597
Net Imports      0.782168 -0.775211       0.673759     -0.759643
Other            0.871861 -0.721248       0.753219     -0.713921
Waste            0.851751 -0.661547       0.716550     -0.673375
LFG              0.895483 -0.886494       0.803136     -0.779319
Wood             0.660260 -0.508921       0.622183     -0.587590
Wind             0.853213 -0.968716       0.748360     -0.740742
Solar            0.742436 -0.919810       0.665324     -0.615701
Total  _x        0.858953 -0.437484       0.836859     -0.767346
Residential      0.967634 -0.668374       0.891697     -0.831345
Commercial       0.909646 -0.532843       0.878509     -0.815882
Industrial      -0.899312  0.724211      -0.841964      0.835413
Transportation   0.542982 -0.172799       0.524103     -0.502347
Total  _y        0.887831 -0.478488       0.847328     -0.763837
```

	Conv. Hydro	PS Hydro	Nuclear	Net Imports
Year	0.114461	0.324919	0.887526	0.782168
Coal	-0.240387	0.083122	-0.564733	-0.775211
Natural Gas	0.050191	0.486298	0.831756	0.673759
Petroleum	-0.192531	-0.508231	-0.797597	-0.759643
Conv. Hydro	1.000000	-0.002018	-0.087626	0.110523
PS Hydro	-0.002018	1.000000	0.478301	0.065066
Nuclear	-0.087626	0.478301	1.000000	0.633660
Net Imports	0.110523	0.065066	0.633660	1.000000
Other	-0.166473	0.169084	0.885440	0.755181
Waste	-0.183593	0.198258	0.882914	0.707852
LFG	-0.010864	0.035231	0.792756	0.847462
Wood	-0.314965	0.207467	0.793461	0.602238
Wind	0.215651	-0.088199	0.624351	0.805318
Solar	0.218663	-0.133843	0.489052	0.659142
Total _x	-0.128765	0.527313	0.944588	0.626223
Residential	-0.014572	0.385473	0.934986	0.748244
Commercial	-0.087732	0.524213	0.965674	0.641034
Industrial	-0.010415	-0.297562	-0.869704	-0.785993
Transportation	-0.088940	0.314202	0.668336	0.489947
Total _y	-0.090304	0.524369	0.935368	0.591183

Figure 1.4 Above correlation matrix shows how generation patterns influence consumption.

A correlation matrix is computed to identify relationships between variables in the combined

dataset. The correlation between total electricity generation and total electricity sales is specifically calculated and printed, providing insights into how generation patterns influence consumption.

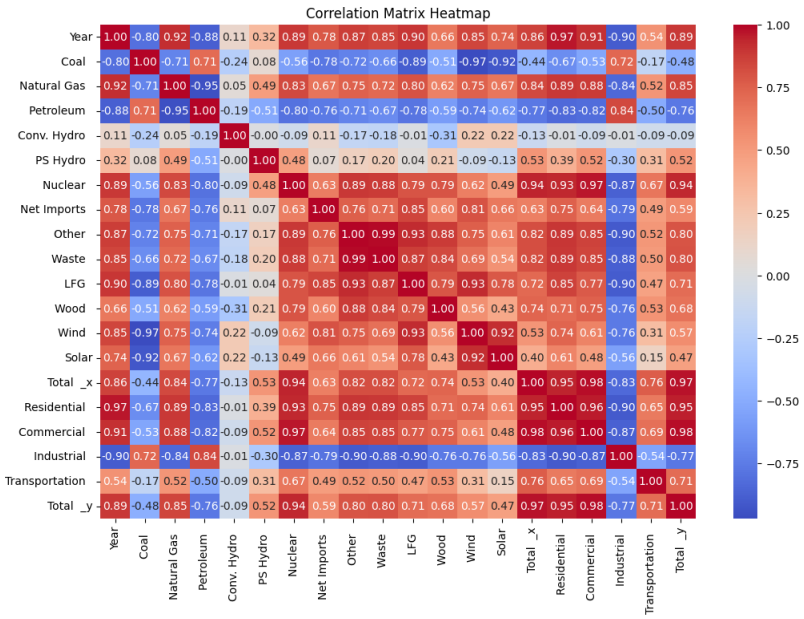
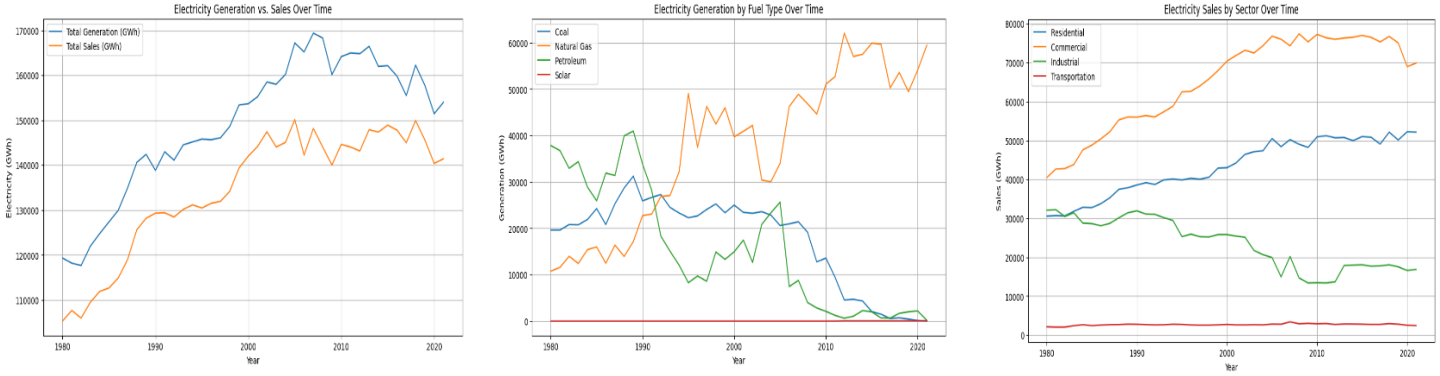


Figure 1.5 Above heatmap of the correlation matrix is generated using Seaborn. This visualization highlights the strength and direction of relationships between Electricity Generated by various types and Electricity sales by sector, making it easier to identify significant patterns.



A. **Electricity Generation vs. Sales Over Time:** This line plot illustrates the year-wise changes in total electricity generation and total sales, highlighting overall growth and potential disparities.
B. **Electricity Generation by Fuel Type:** This Line plot for each fuel type (e.g., coal, natural gas, wind, solar) show the evolution of generation contributions over the years.
C. **Electricity Sales by Sector:** This Line plots for residential, commercial, industrial, and transportation sectors display changes in electricity demand across different user groups.

Scatter plots are created to investigate how specific fuel types influence sectoral electricity consumption. As Shown below

- The relationship between coal generation and sectorial electricity sales is visualized, providing insights into dependency patterns.
- The relationship between nuclear generation and sectorial electricity sales is visualized, providing insights into dependency patterns.

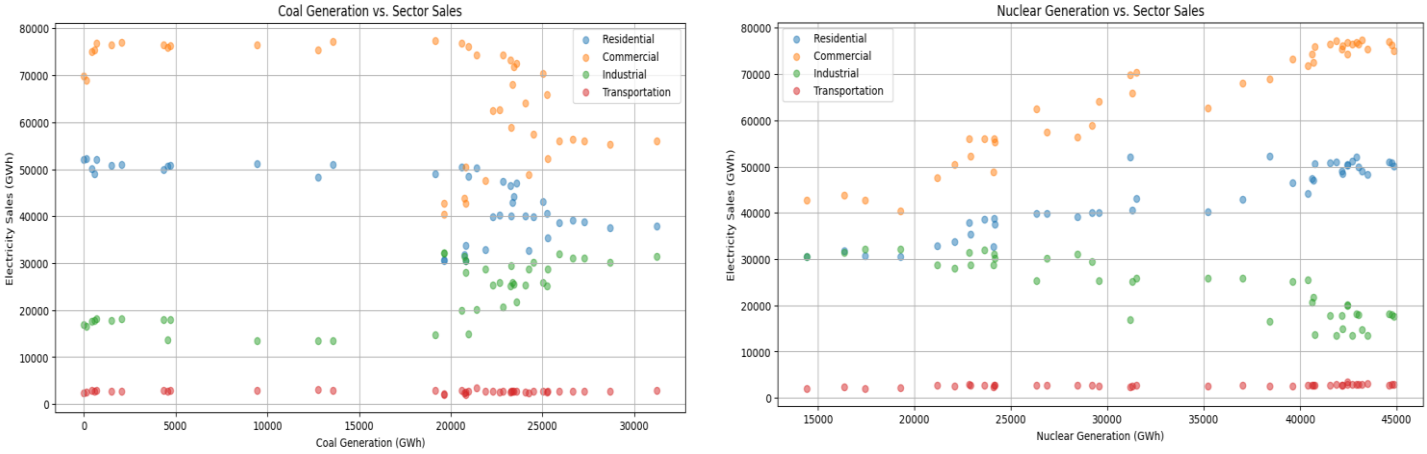


Figure 1.6: Scatter plots of a relationship between one of each renewable and non-renewable electricity generation and sectorial electricity sales.

Conclusion:

The analysis shows big shifts in energy generation and consumption patterns over the decades. Inverse correlations of coal and petroleum with time reflect their gradual replacement by substitute sources like natural gas, nuclear power, and renewable energies a slight transition to cleaner energy.

The strong interconnections among residential, commercial, and industrial sectors suggest that changes in economic conditions or energy policies affecting one sector are likely to influence the others, emphasizing the need for integrated approaches to energy planning. Furthermore, the high correlations observed among renewable energy sources, such as wind and solar, indicate a growing preference for sustainable energy solutions, often at the expense of traditional fossil fuels.

This report provides valuable insights into the evolving trends of energy production and consumption, showcasing a clear shift toward sustainability and efficiency. These findings can support informed decision-making to foster a more sustainable energy future.