EDA + STATS PROJECT

Prepared by: Soha Mohamed Hosny

Table Of Content

Introduction

- This project focuses on energy usage analysis to identify patterns and optimize consumption.
- The analysis leverages data to track energy consumption trends and behavior, providing valuable insights into how energy is used over time.
- The project involves processing and analyzing large datasets related to energy usage, identifying key patterns

Objective

- The aim is to develop strategies for better energy management, enabling users or organizations to reduce costs, improve energy efficiency, and make informed decisions based on data-driven insights.

Methodology

- Data aggregated by hourly, daily, monthly usage.
- Python libraries used for analysis and visualization.
- Key steps include cleaning, aggregation, and trend analysis.
- Steps followed to achieve the objectives.

Detailed Methodology

- **Step-by-step process: Data collection and preprocessing**

- Why Use Both ffill and bfill?

I. Forward fill fills gaps after a valid value.

II. Backward fill fills gaps before a valid value.

- transformation of variables(DateTime).
- check missing values.
- check inconsistencies.

Detailed Methodology

- **Feature Engineering:**
- From consumption file & demand file extract new columns from column (DateTime) like: - Hour
  - Day
  - Month
  - Year
  - WeekDay
- And merge files depend on this columns.

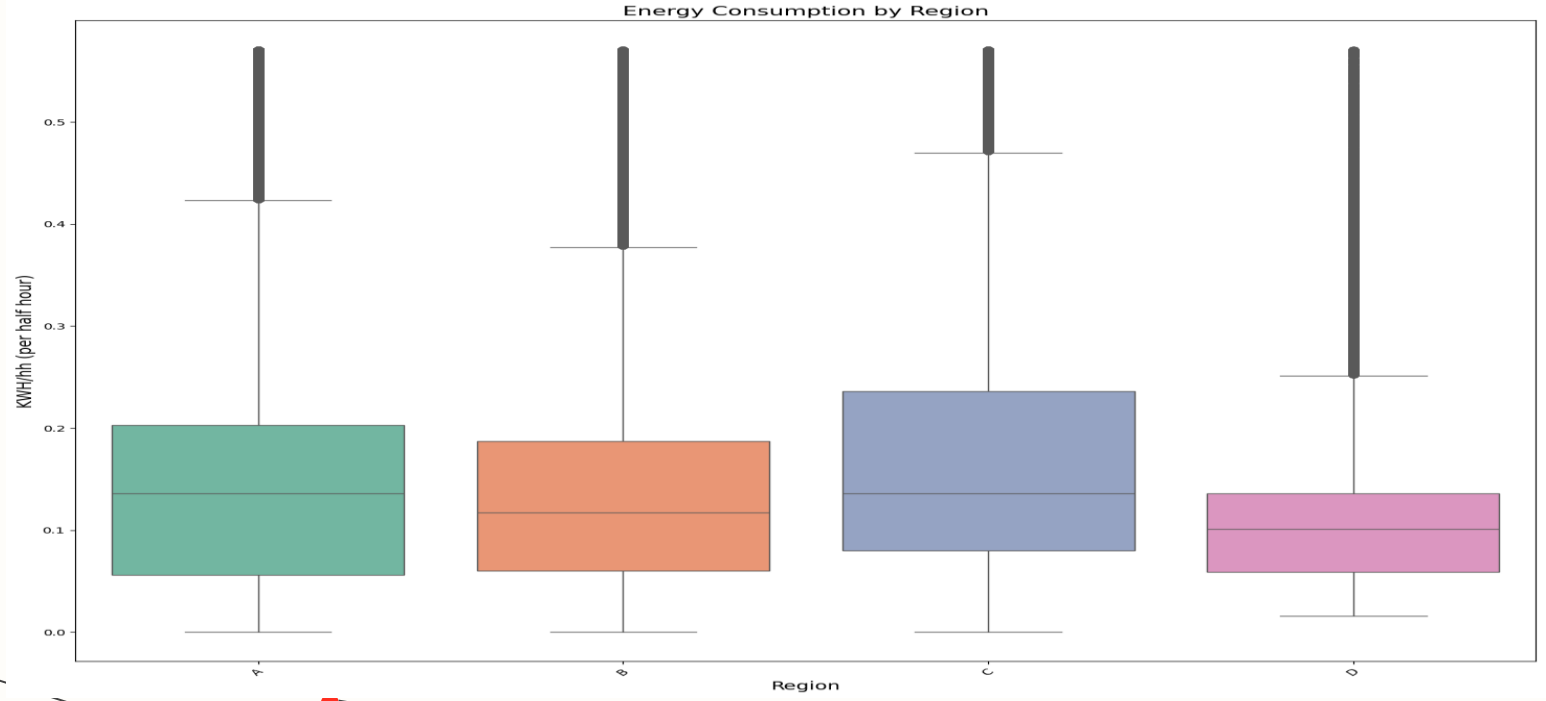
Detailed Methodology

- **And merge files depend on this extracted columns:**

	cell_id	DateTime	KWH/hh (per half hour)	site_id	region	Hour	Weekday	Month	Year	Day	KWH/hh (per half hour) _hourly_mean	Demand	Demand_Cost	Season
0	MAC000002	2013-01-01 00:00:00	0.219	A	A	0	Tuesday	1	2013	Tuesday	0.292001	Normal	19.46	Winter
1	MAC000002	2013-01-01 00:30:00	0.241	A	A	0	Tuesday	1	2013	Tuesday	0.292001	Normal	19.46	Winter
2	MAC000002	2013-01-01 01:00:00	0.191	A	A	1	Tuesday	1	2013	Tuesday	0.265175	Normal	19.46	Winter
3	MAC000002	2013-01-01 01:30:00	0.235	A	A	1	Tuesday	1	2013	Tuesday	0.265175	Normal	19.46	Winter
4	MAC000002	2013-01-01 02:00:00	0.182	A	A	2	Tuesday	1	2013	Tuesday	0.251485	Normal	19.46	Winter
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1379305	MAC000100	2013-12-20 18:30:00	0.260	N	B	18	Friday	12	2013	Friday	0.358768	Normal	19.46	Winter
1379306	MAC000100	2013-12-20 19:00:00	0.267	N	B	19	Friday	12	2013	Friday	0.337095	Normal	19.46	Winter

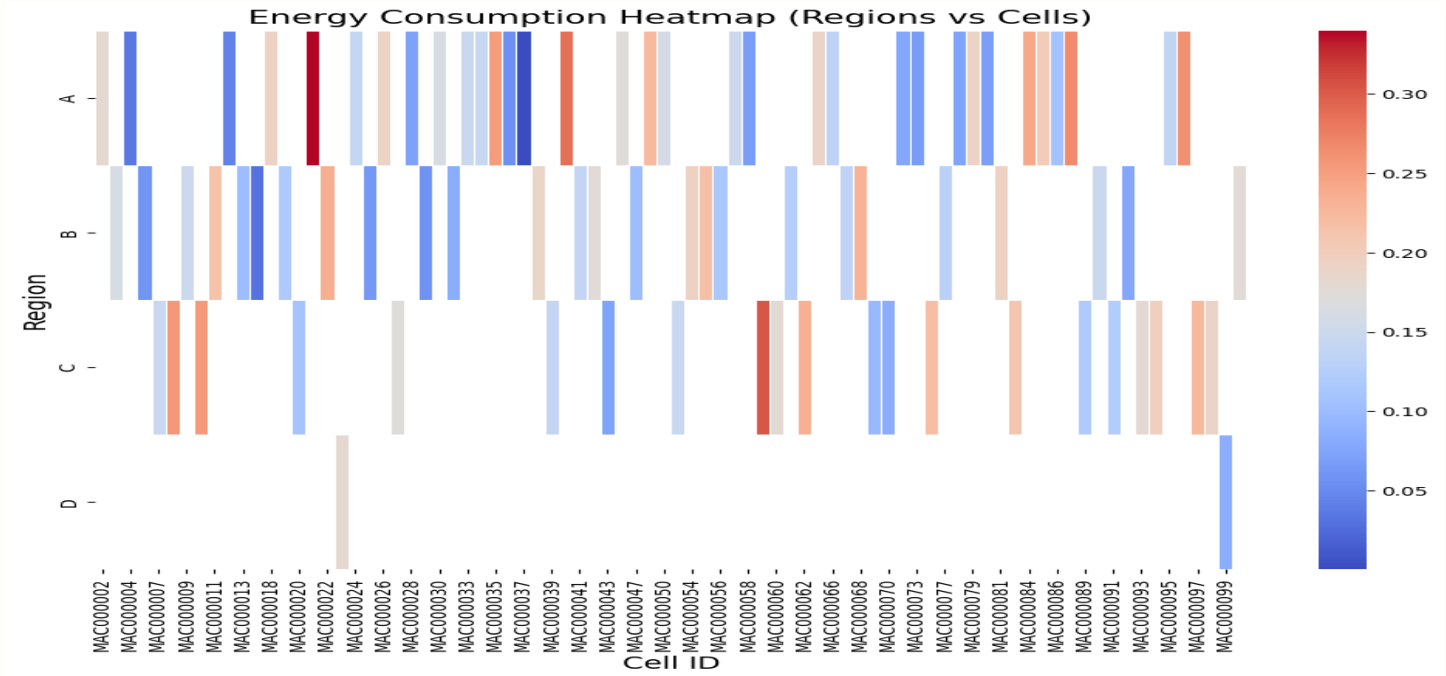
Visualization

- **Box plot of energy consumption by region:**



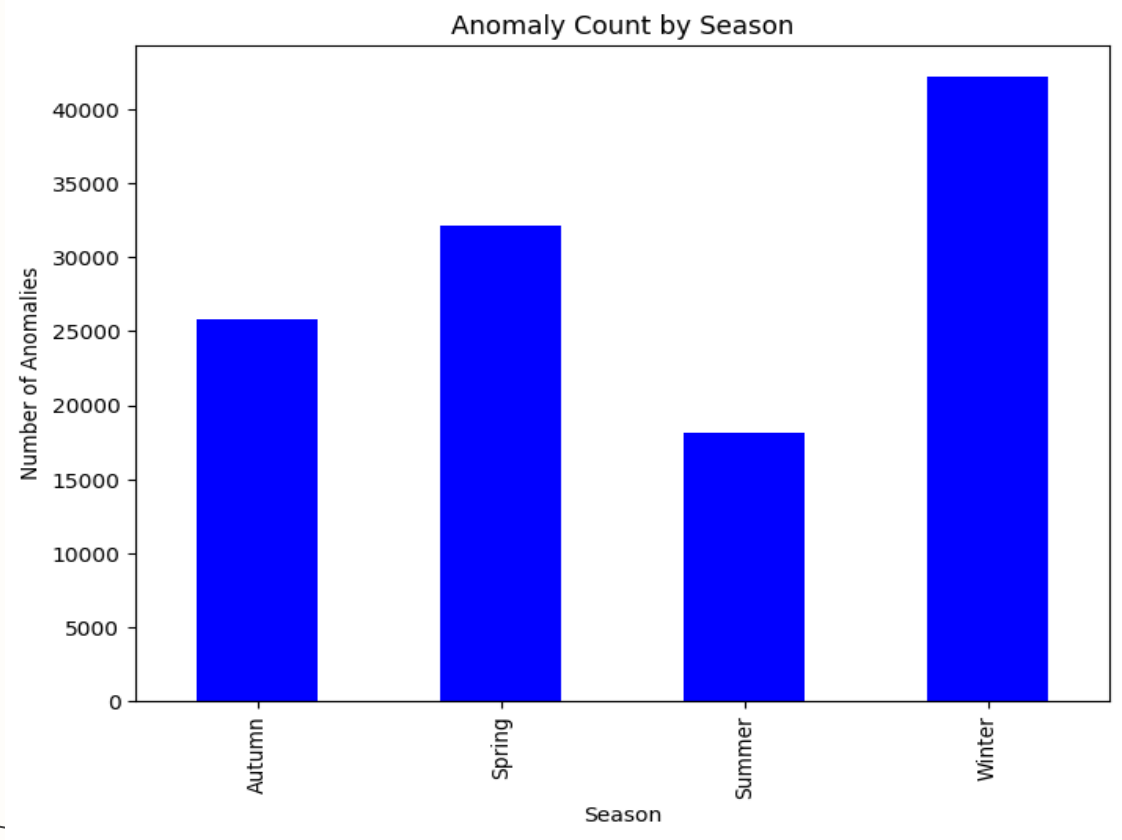
Visualization

- **Heatmap of energy consumption across sites and cells:**



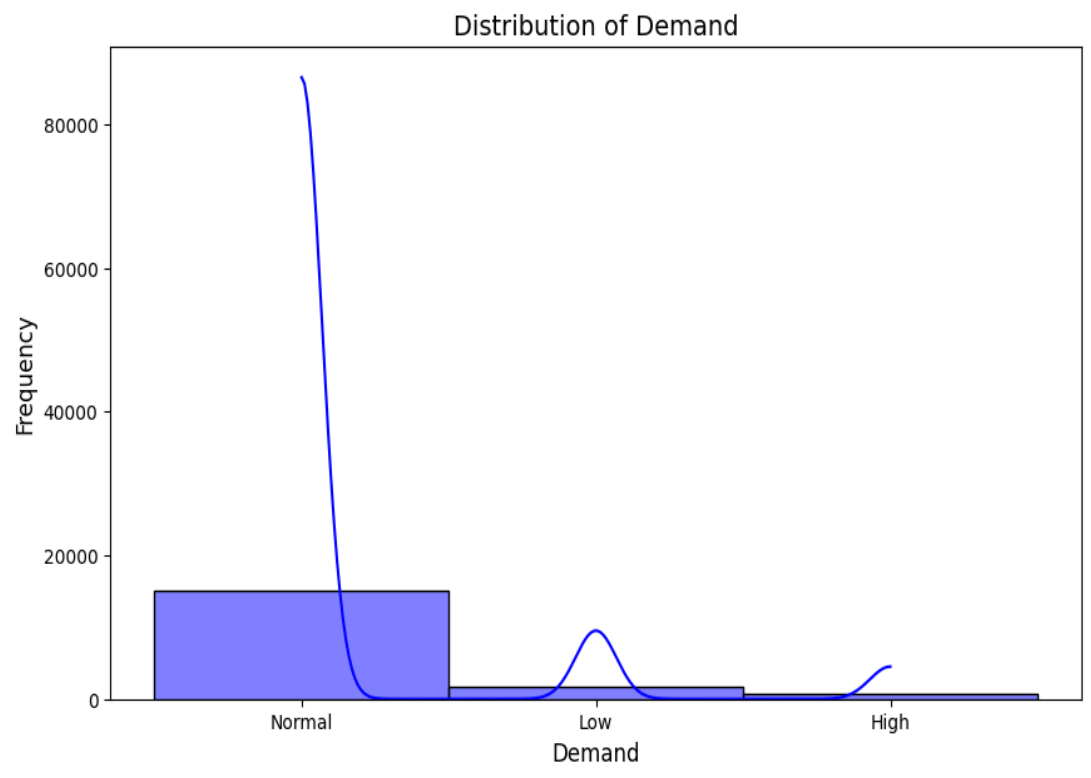
Visualization

- **Season analysis:**



Visualization

- **Demand analysis:**



Challenges

- Addressing missing or inconsistent data.
- Visualizing multi-dimensional patterns.
- Balancing regional and temporal analysis.

Insights

- Region A has the highest energy consumption, likely due to high population.
- June is the peak consumption month.
- Fridays show the highest daily energy usage.
- Most demand prices fall within a normal range.

Conclusion and Future Work

- Key insights provide a foundation for energy-saving strategies.
- Recommendations include targeting high-consumption regions.
- Future focus: Implement dynamic pricing models for demand management.


Thanks