Bakary Coulibaly
ECE-6610
Final Project: EM-CURE

1. <u>Data Processing:</u>

1.1. Identify Corrupted Data:

File Name Containing Corrupted Data	Row Number Containing Corrupted Data (The header is not counted. Row counting starts at 0)	What is the corrupted data?
fullservicerestaurant.csv	33921	out.electricity.fans.energy_consumption: “corrupted”
largehotel.csv	184, 316, 6781, 17264, 22382, 25161	Column — Corrupted Data out.electricity.heating.energy_consumption — with out.natural_gas.heating.energy_consumption — great out.natural_gas.water_systems.energy_consumption — data out.other_fuel.heating.energy_consumption — come out.other_fuel.water_systems.energy_consumption — great out.other_fuel.total.energy_consumption — responsibility
largeoffice.csv	108, 424, 2752, 30125	Column — Corrupted Data out.electricity.exterior_lighting.energy_consu... — bad out.electricity.heat_rejection.energy_consumption — data out.electricity.pumps.energy_consumption — really out.natural_gas.heating.energy_consumption — sucks

1.2. Plan your Mitigation Strategy:

1.2.1. What would happen if you replaced it with a 0?

Replacing the corrupted data point with a 0 would create a significant and unrealistic drop in the energy consumption data. Since the data represents time-series energy consumption in kWh, a zero value can drastically skew downstream analysis. For example, it may lead to incorrect conclusions when calculating consumption trends, peaks, or efficiency metrics.

1.2.2. What would happen if you deleted that row completely?

Deleting the entire row would result in the loss of a 15-minute interval from the dataset. This would create a gap in the time series, which could lead to issues if the data is used for time-based analysis like forecasting or trend identification.

1.2.3. What is the correct way to deal with this corrupted data?

The correct approach would be to use linear interpolation to estimate the corrupted data point. This technique involves computing an estimated value based on the preceding and succeeding data points. This way, the time series remains continuous, and the imputed value aligns more realistically with surrounding consumption patterns.

1.3. Mitigate Errors through Linear Interpolation:

File Name Containing Corrupted Data	Row Number Containing Corrupted Data	Preceding Data Point	Superseding Data Point	Replaced Number of Corrupted Data
fullservicerestaurant.csv	33921	514.96	561.5	1: From "Corrupted" to 538.23
largehotel.csv	Row Number	Preceding Data Point	Superseding Data Point	Replaced Value
	184	1760.74	1743.71	1752.225
	316	11809.35	18471.42	15140.385
	6781	2142.31	2175.09	2158.700
	17264	0.00	0.00	0.000
	22382	0.00	0.00	0.000
	25161	0.00	0.00	0.000
largeoffice.csv	Row Number	Preceding Data Point	Superseding Data Point	Replaced Value
	108	1276.74	1276.74	1276.740
	424	15.50	15.65	15.575
	2752	1799.33	1810.48	1804.905
	30125	10532.02	12202.22	11367.120

2. <u>Data Visualization:</u>

2.1. Plot Selection:

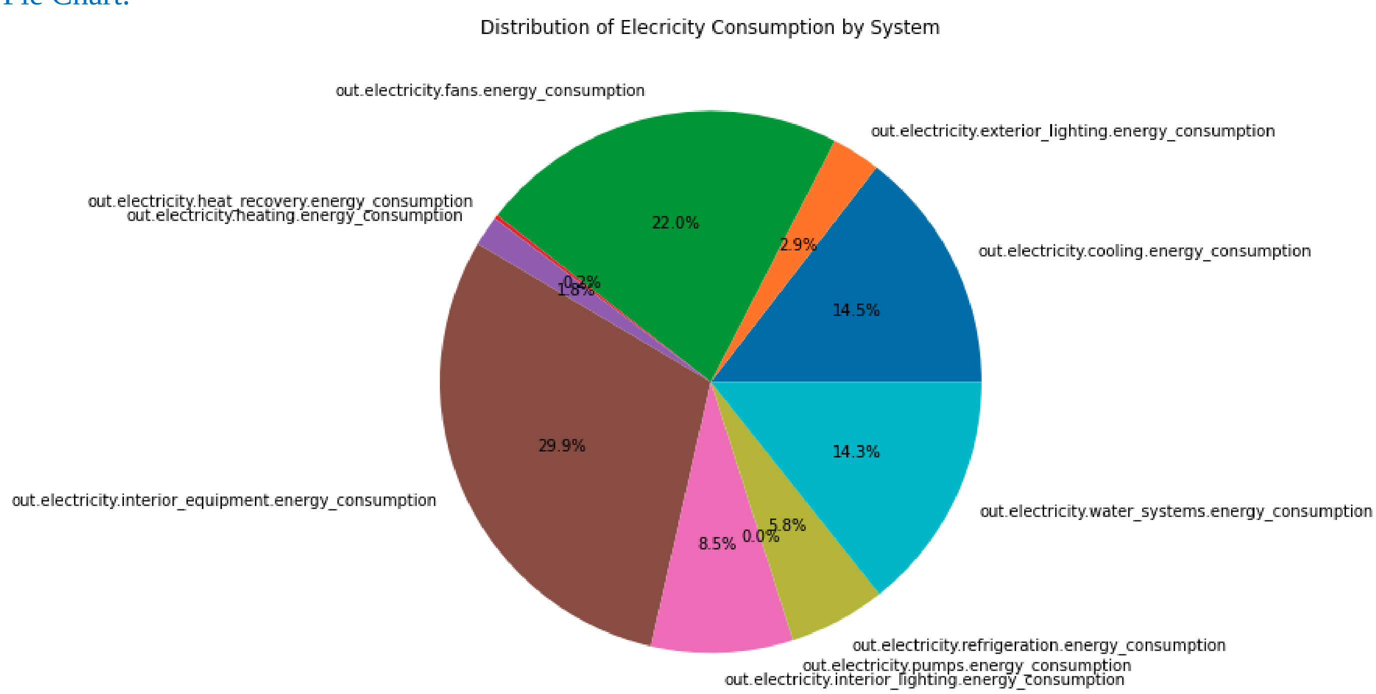
Plot Type	What aspect of the data does this plot type represent?	Why do you think this aspect is worth representing?
Pie Chart	Distribution of electricity consumption by different systems (e.g., fans, pumps, lighting)	This plot would visually depict the proportion of total electricity used by various systems. Understanding how energy is distributed can help in identifying which areas consume the most power and potentially offer the greatest opportunities for energy savings or efficiency improvements.
Time Series Line Chart (Interactive)	Hourly variations in energy consumption based on percentiles and averages for different energy sources.	The chart provides insights into the hourly patterns of energy usage, allowing for clear identification of peak periods. By using percentile bands and average lines, the plot highlights variability, and trends over time,

		facilitating more informed decision-making for energy management and efficiency planning. The interactive nature of the chart allows for easy comparison between different datasets, revealing consumption trends across different energy sources.
Heatmap (Interactive)	Daily distribution of energy consumption over the course of a year, segmented by hour.	The interactive heatmap allows for a quick visual assessment of energy usage patterns throughout the day across different times of the year. Users can select different energy consumption metrics via a dropdown menu, adapting the display to various aspects of the dataset. This functionality is essential for identifying trends, peak consumption times, and periods of low energy use across multiple variables, aiding in strategic energy management and operational optimization. The color coding enhances the readability of complex data, making it easier to spot anomalies and consistent patterns while the interactivity enriches user engagement and data exploration.

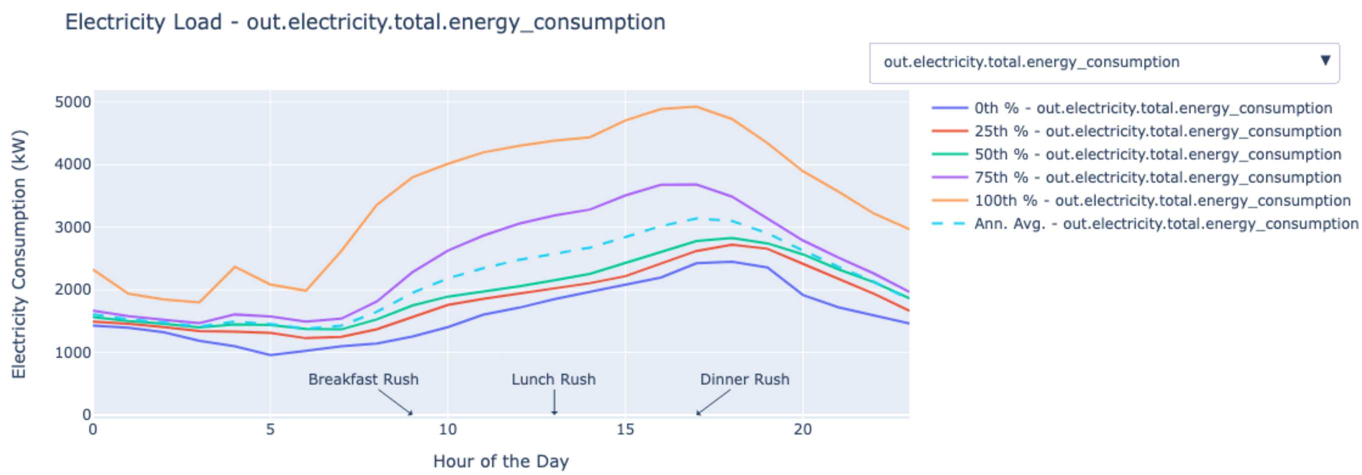
2.2. Generate Plots:

- For Full Service Restaurant:

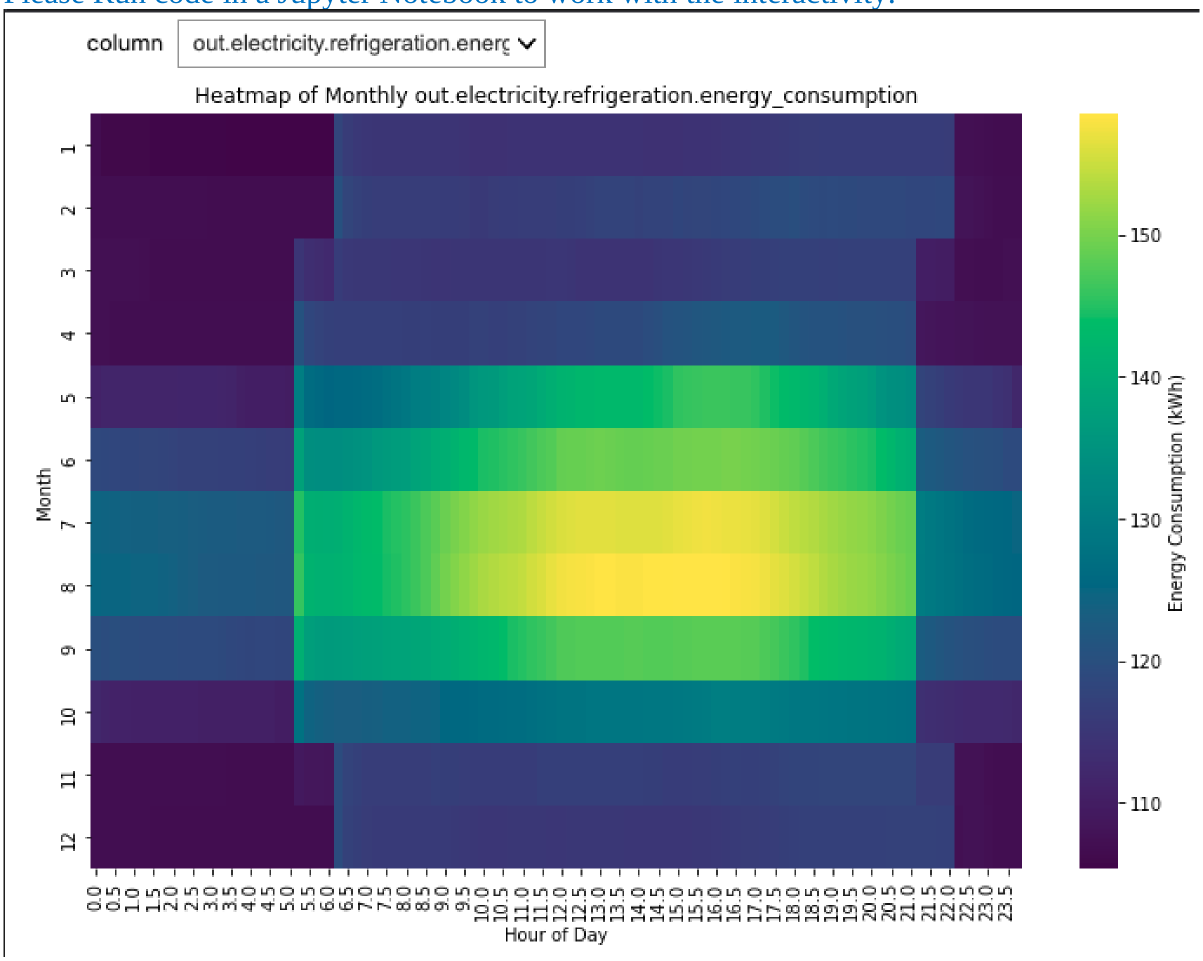
Pie Chart:



Time Series Line Chart:
Please Run code in a Jupyter Notebook to work with the interactivity.

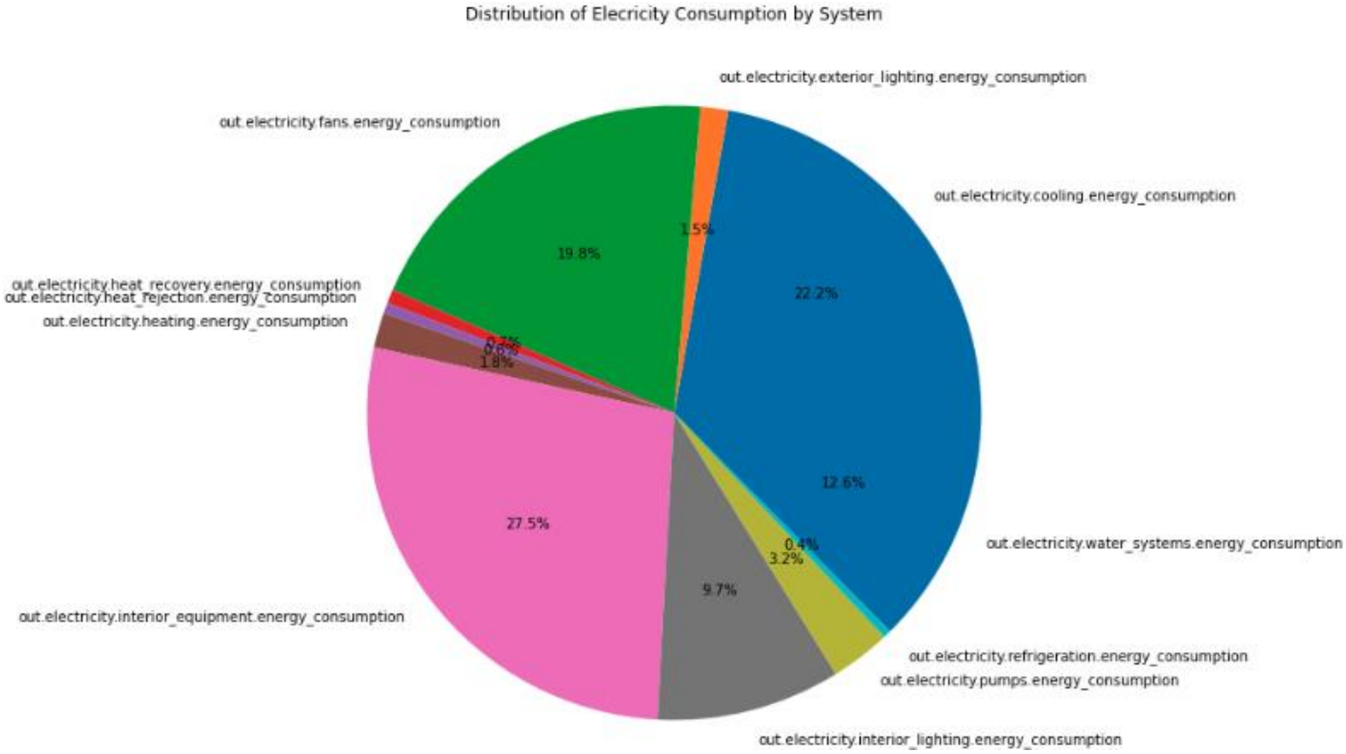


Heatmap:
Please Run code in a Jupyter Notebook to work with the interactivity.



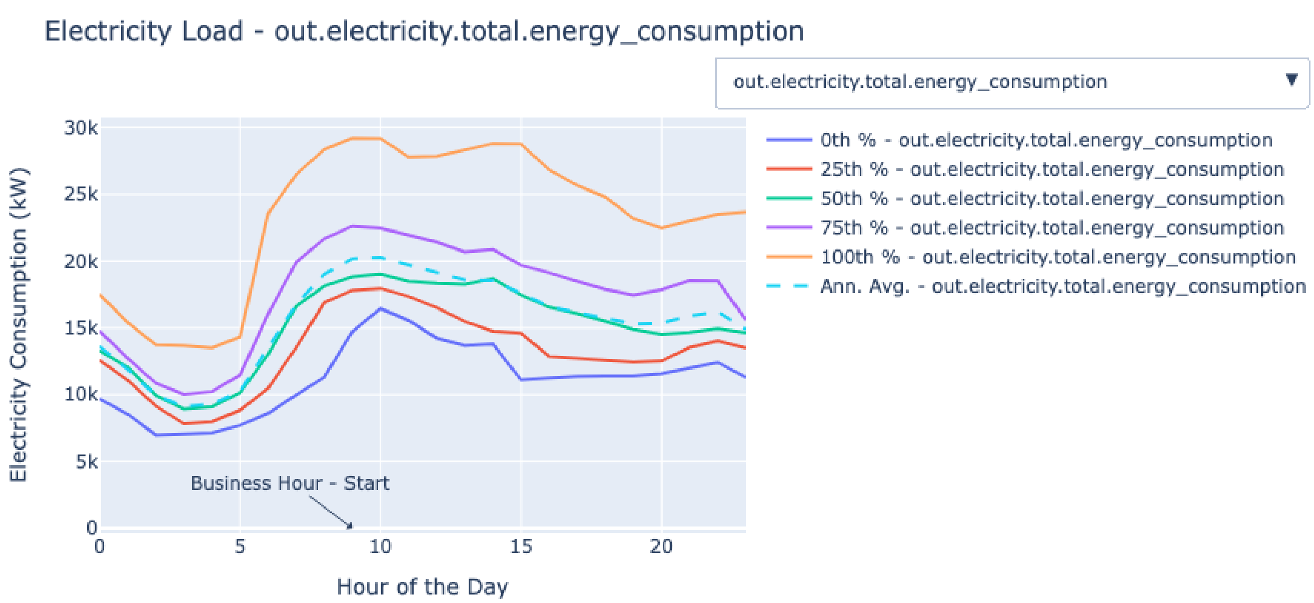
- For Large Hotel:

Pie Chart:

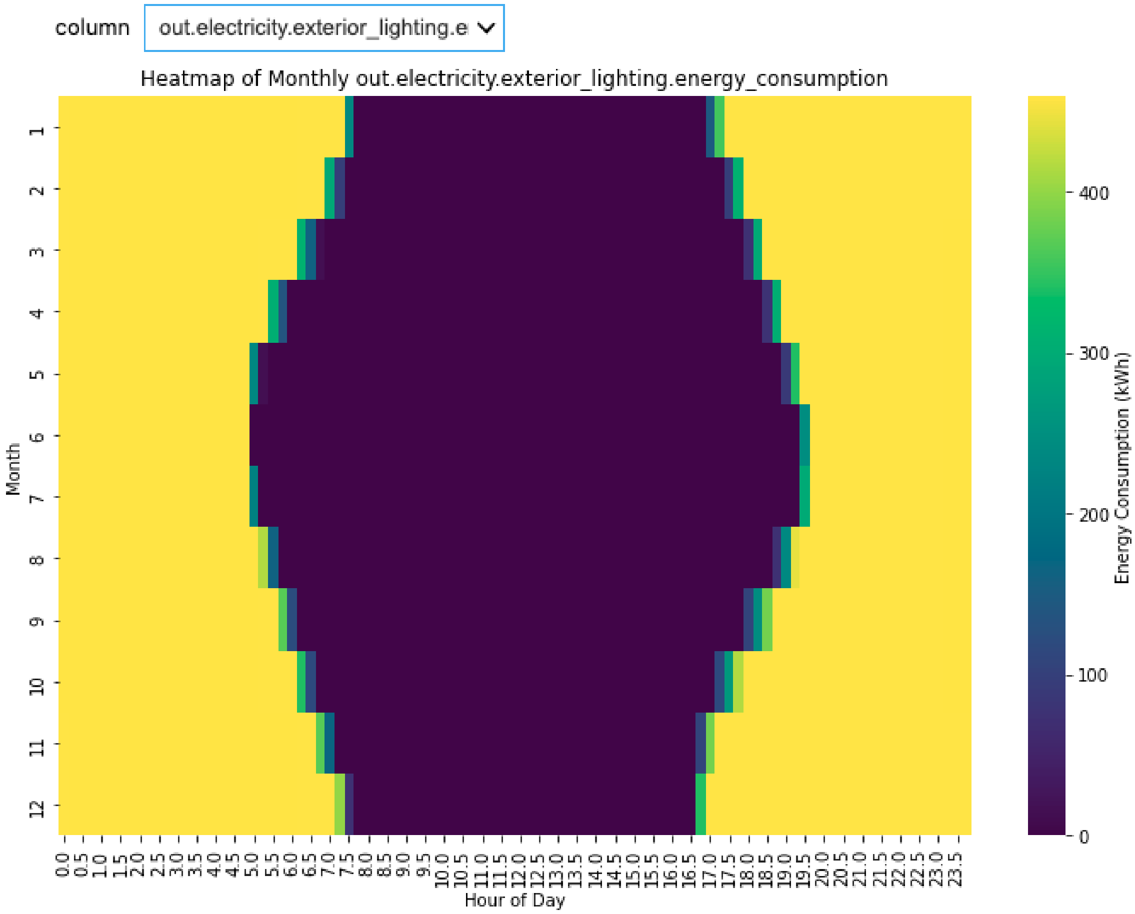


Time Series Line Chart:

Please Run code in a Jupyter Notebook to work with the interactivity.

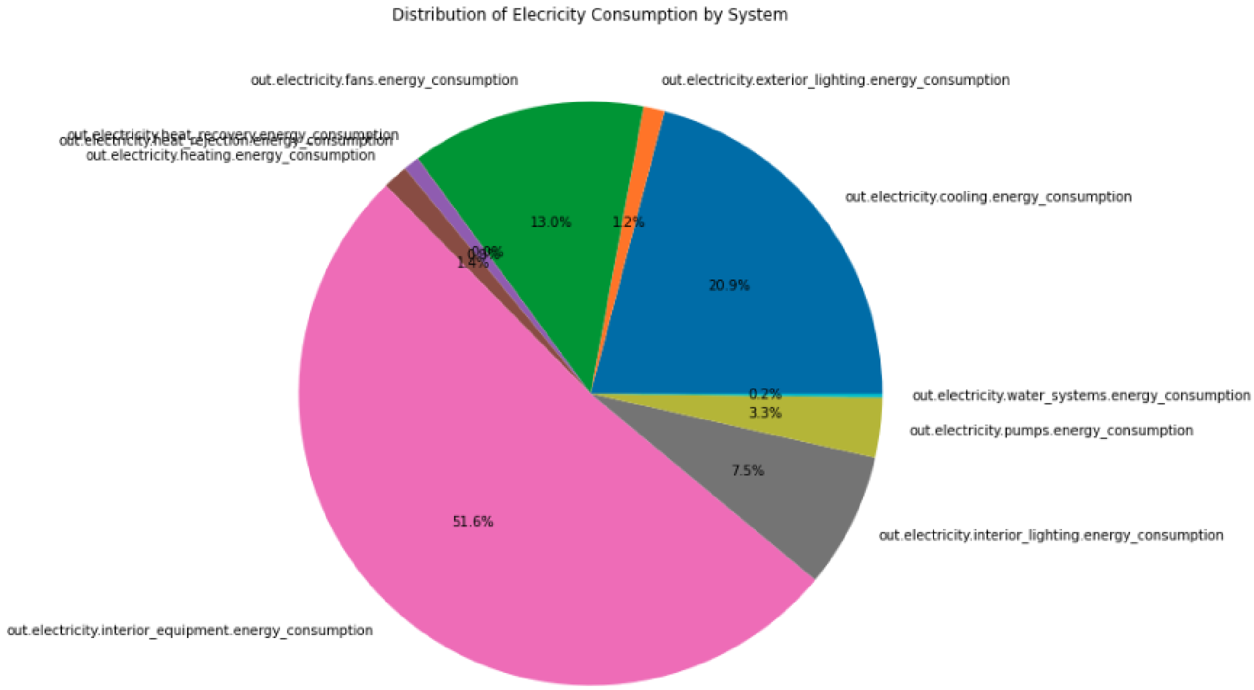


Heatmap:
Please Run code in a Jupyter Notebook to work with the interactivity.



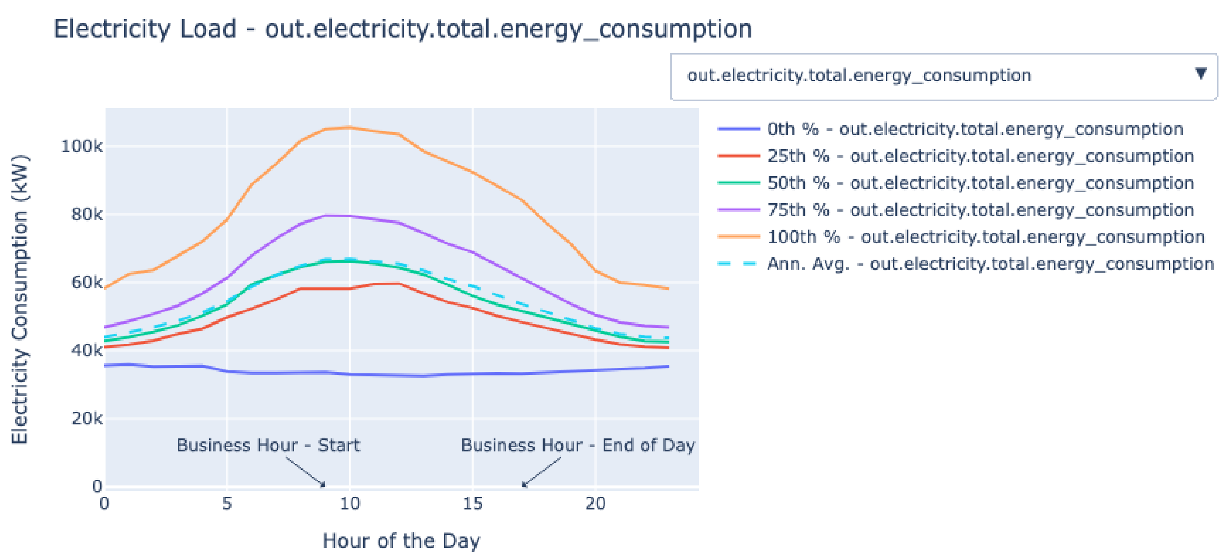
- For Large Office:

Pie Chart:

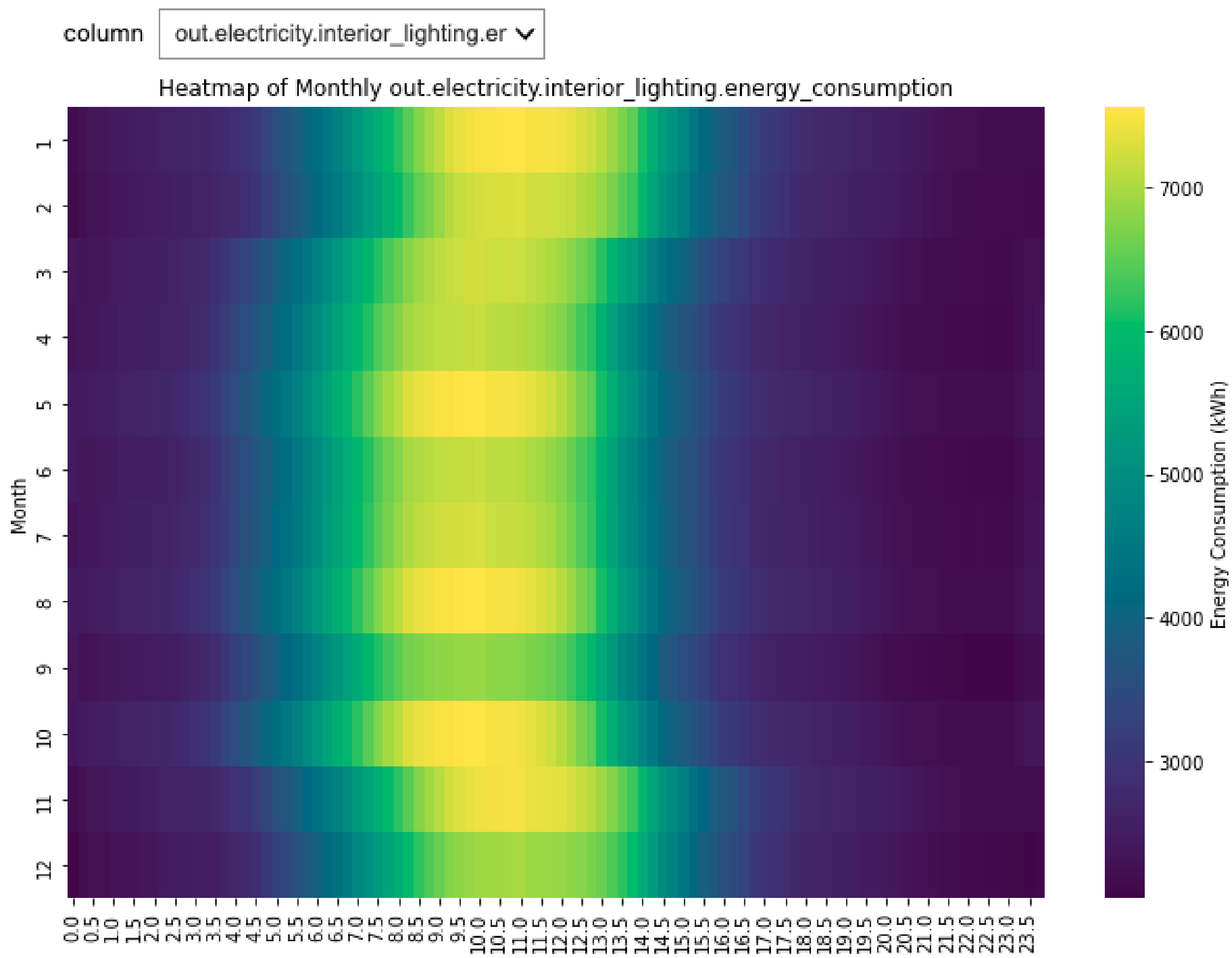


Time Series Line Chart:

Please Run code in a Jupyter Notebook to work with the interactivity.



Heatmap:
Please Run code in a Jupyter Notebook to work with the interactivity.



2.3. Plot Analysis:

2.3.1. For each plot type, what is a unique conclusion or insight it offers that other plots do not?

Pie Charts: Provide a clear breakdown of which systems consume the most electricity, offering insights into potential areas for energy efficiency improvements.
For example, in the pie chart of the hotel; the chart indicates significant energy consumption in interior equipment (27.5%) and fans (19.8%). This reflects high usage in rooms and ventilation needs.

Line Graphs: Illustrate the daily operational patterns and peak usage times, helpful for demand management and scheduling maintenance.
For example, in the line chart of the office; the consumption peaks are sharp during business hours, particularly at the start and end of the day, aligning with operational hours and employee presence.

Heatmaps: Show seasonal variations and time-of-day dependencies in energy usage, crucial for understanding impacts of external factors like weather on energy demands.
For example, in the heatmap of the restaurant; the map shows higher energy use in warmer months, which is likely due to increased demand on refrigeration systems to maintain temperatures.

2.3.2. Now that you have different plots for each of the three buildings, what can you notice about the differences in building energy use? What reasons could justify the differences in consumption?

The energy consumption patterns across different types of buildings—restaurants, hotels, and offices—reveal distinct differences due to their specific operational needs and business models.

Restaurants show high energy usage in kitchen equipment and cooling systems, peaking during mealtimes to accommodate cooking and customer comfort. This reflects their operational model which varies significantly throughout the day. Hotels, on the other hand, exhibit a more even distribution of energy use, with significant consumption by fans and lighting systems, necessary for the continuous comfort and safety of guests. Offices display pronounced energy consumption during business hours, particularly in electronic equipment and air conditioning, aligning with standard office operating times and the intensive use of technology.

These variations can be attributed to operational hours, with restaurants experiencing high variability, hotels requiring around-the-clock operation, and offices operating within fixed hours. The type of services provided also influences energy use; restaurants need extensive refrigeration and cooking facilities, hotels maintain constant climate control and lighting, and offices consume substantial energy in powering electronic devices.

Additionally, building size and occupancy patterns differ, with hotels and offices generally having larger and more consistent occupancy compared to the fluctuating customer presence in restaurants.