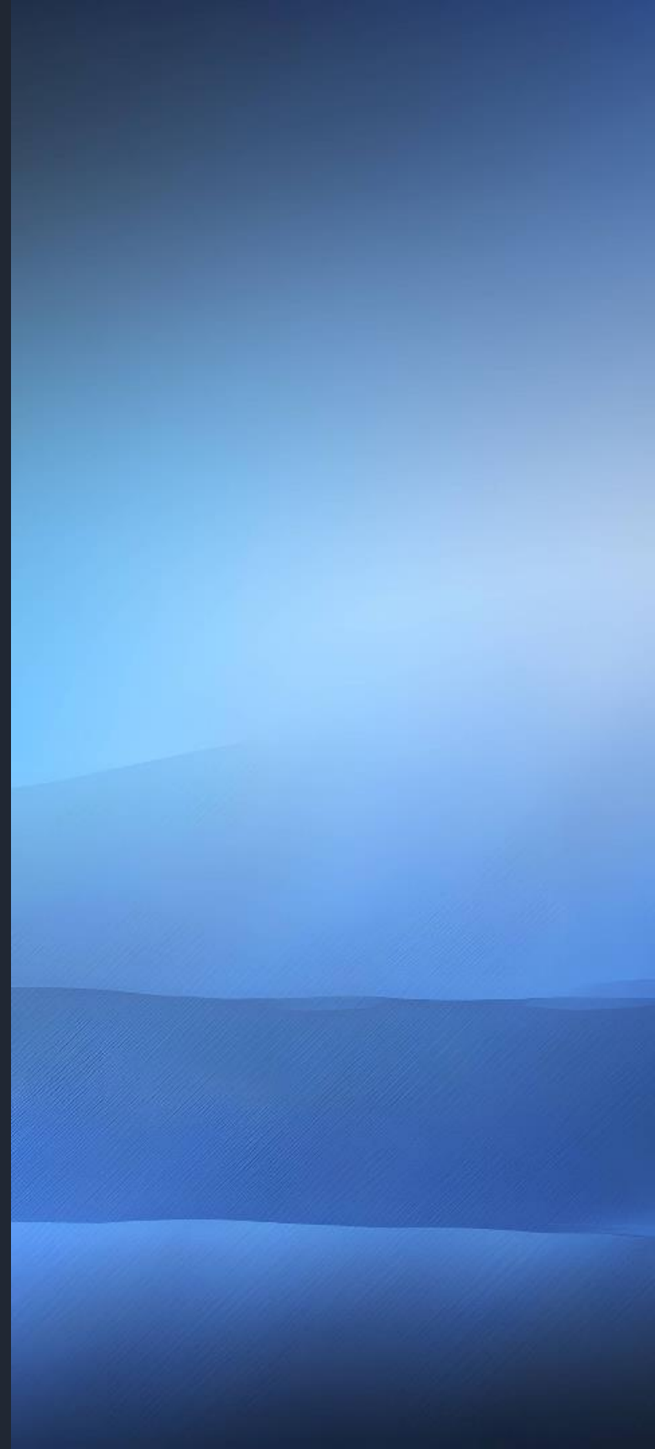
Preempting Potential Energy Shortages

Introduction

- The managing and predicting energy consumption has always been challenging and important.
- This Project presents an analysis where the primary objective is to understand and forecast the impact of rising temperatures on electricity demand, particularly during the month of July.
- The focus is on preempting potential energy shortages and blackouts that could arise from an overburdened electrical grid during an unusually hot summer.



Data Analysis

- In the data preparation phase, we focused on consolidating and preparing three key datasets for our analysis of household energy consumption.
- We pipelined the weather data, electricity data and housing data
- An observation from the exploratory analysis indicates a positive correlation between energy consumption and temperature. This trend is evident at both the individual house and county levels, suggesting a consistent pattern across different scales of data.
- The following visualizations illustrates electricity consumption at individual household and county level

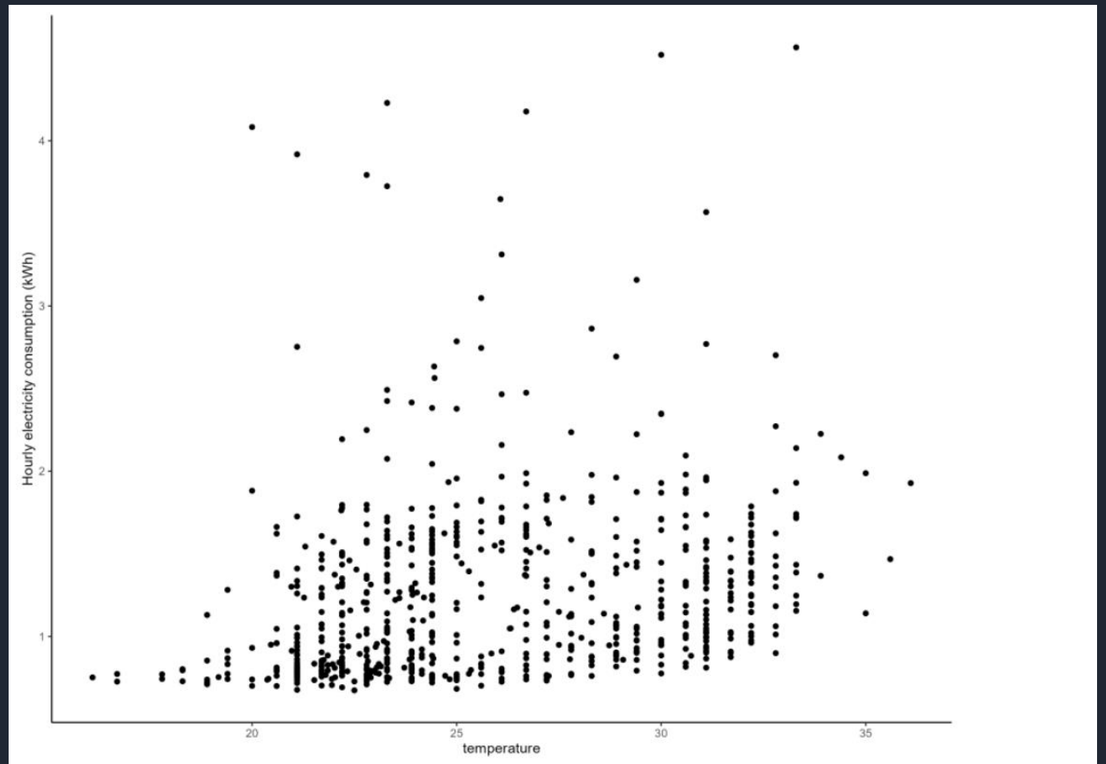


Figure 1: Electricity consumption at individual household level

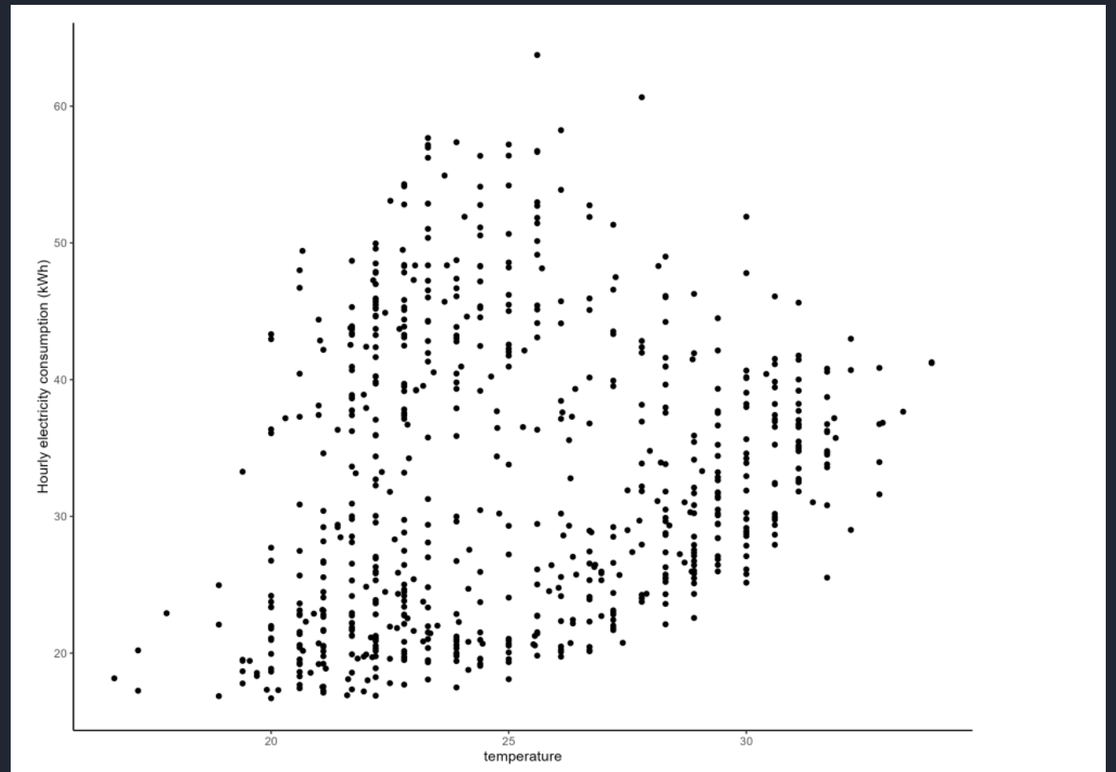


Figure 2: Electricity consumption at county level

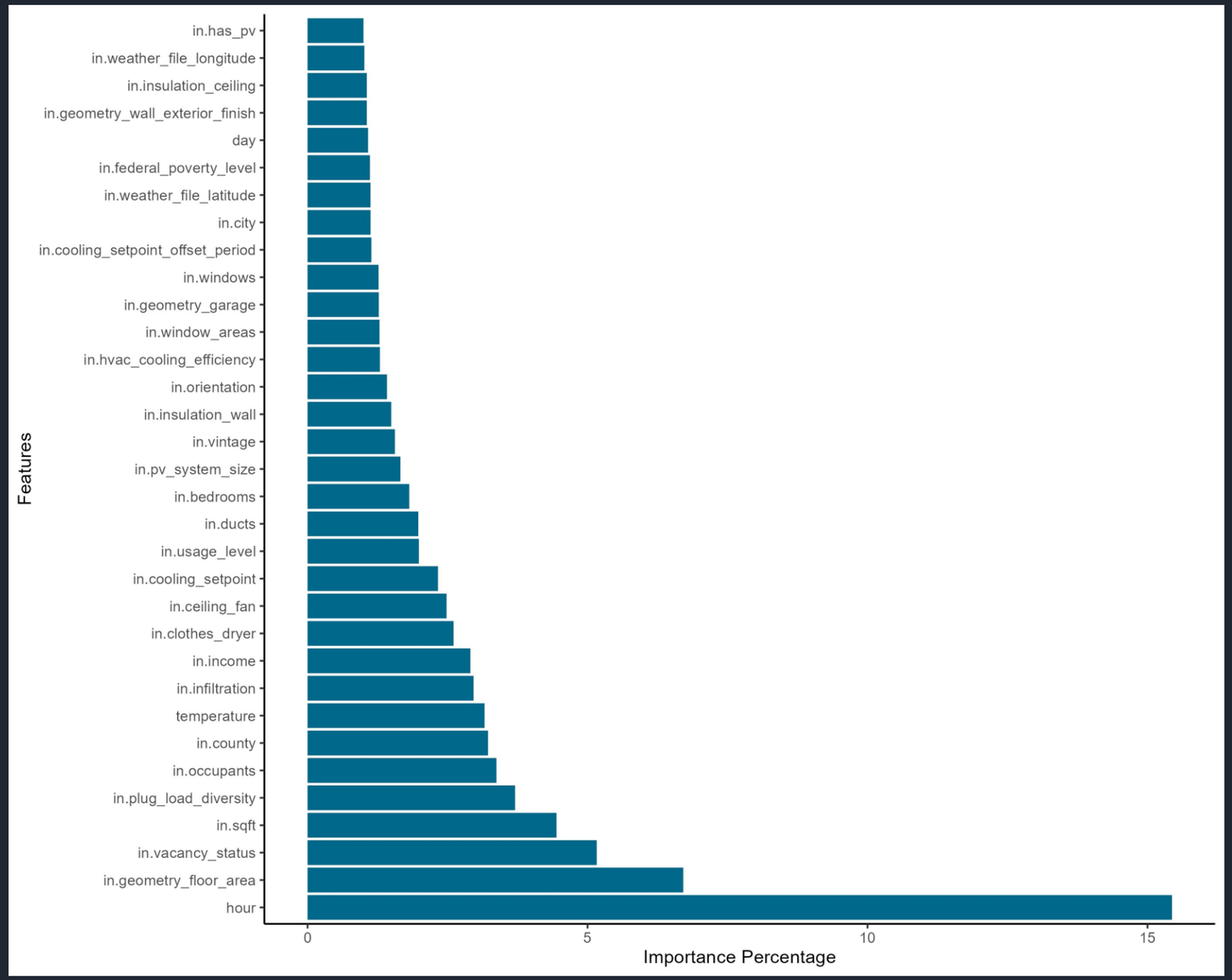
Feature Importance

1. To assess the importance of each house feature in energy consumption, we merged the static and dynamic data, selecting a sample of 5% of the rows to run a random forest model.

2. The most significant features identified are hour, the floor area and the vacancy status.

3. For further analysis, we only selected features with an importance higher than 0.95%.

Feature Importance



Feature engineering

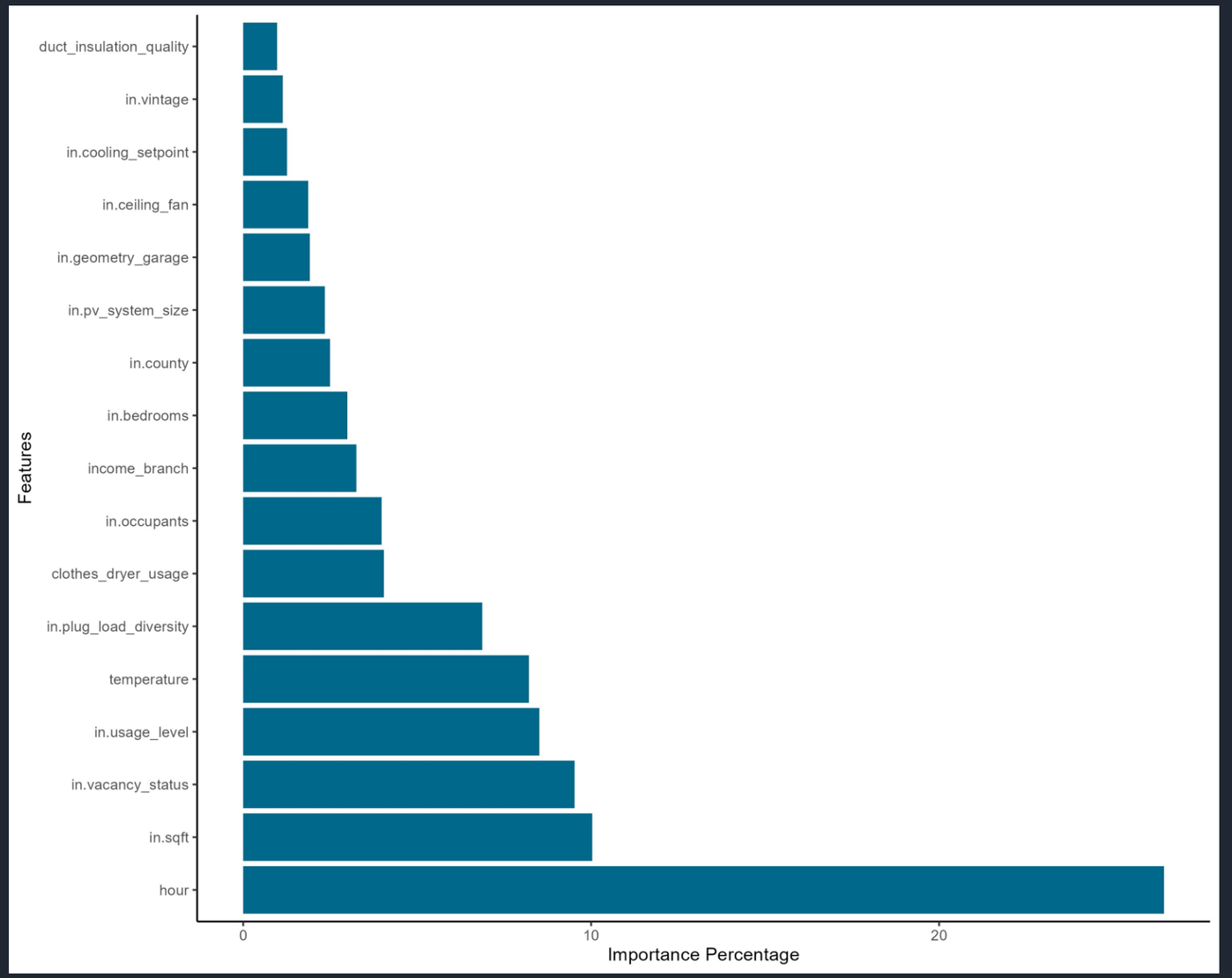
Once the most important features were identified, we developed a pipeline to refine the features.

1. Converting some categorical variables into numeric ones, such as insulation quality, HVAC efficiency, and window area.
2. Dividing into multiple categories, especially those with composite information in their string values, like the "windows" feature.
3. Reducing the possible values of factors.
4. Numeric encoding some ordinal values, such as “usage_level”.

Overall, extensive use of regular expressions to extract information from strings.

Feature Importance

After feature engineering



Regression Analysis

Ultimate Goal: Predict hourly energy Consumption for each country due to a permanent 5° C increase in July

- Tree based models are discarded due to not being useful for extrapolation problems
- First approach: Panel data regression with fixed effects

Table 1: Regression Summary of Panel Model with Fixed Effects

	Dependent variable:
	total_electricity
temperature	0.032***
	(0.0004)
Observations	32,982
R^2	0.139
Adjusted R^2	0.138
F Statistic	5,328.392*** (df = 1; 32935)
Note:	*p<0.1; **p<0.05; ***p<0.01

- Not a good fit.
  - Interpretation: we need to breakdown dates and hours

Regression Analysis

Ultimate Goal: Predict hourly energy Consumption for each country due to a permanent 5° C increase in July

- Second approach: linear regression
  - $Y = \alpha + temperature\ \beta_1 + hour\ \beta_2 + county\ \beta_3 + weekday\beta_4$

Table 2: Summary Statistics of Linear Model

	Residual.Standard.Error	R.Squared	Adjusted.R.Squared	F.Statistic	DF1	DF2	P.Value
value	0.1009612	0.9920827	0.9920646	54977.1	75	32906	0

  - High R^2 means that the best prediction is the actual value, which makes sense for a time series approach with individual entities*

- After splitting data to produce a prediction model:

Table 3: Performance Metrics on Test Set

	Value
RMSE	0.1012035
Rsquared	0.9920848
MAE	0.0771170

Table 4: Performance Metrics on Test Set for non linearities in Temperature

	Value
RMSE	0.1008487
Rsquared	0.9921393
MAE	0.0768656

- Low RMSE corresponds to average error of 2.3%
  - Shows model has high level of accuracy

*If the county is excluded, R^2 drops to 0.07, so county explains 92% of variance

Forecasting Future Energy Demand

Outcome

- A 5° C increase in temperature would lead to a **12.5% increase in energy demand**

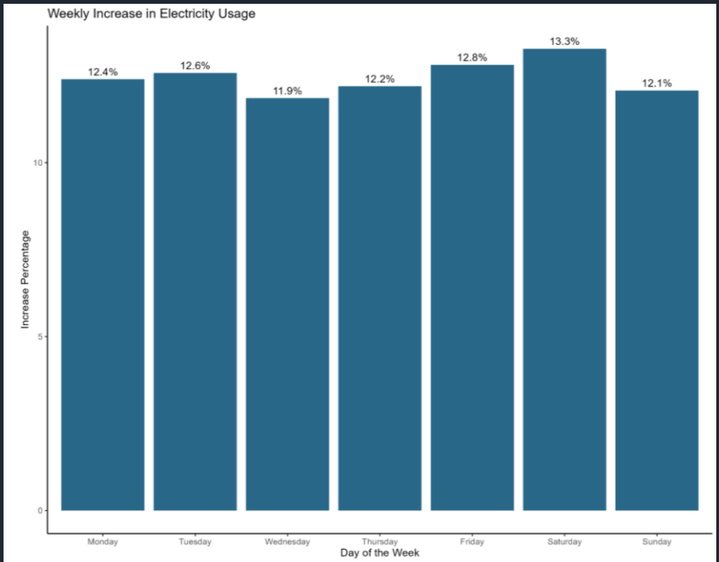


Figure 5: Electricity consumption increase by day of the week

Energy usage increases more on Friday/Saturday

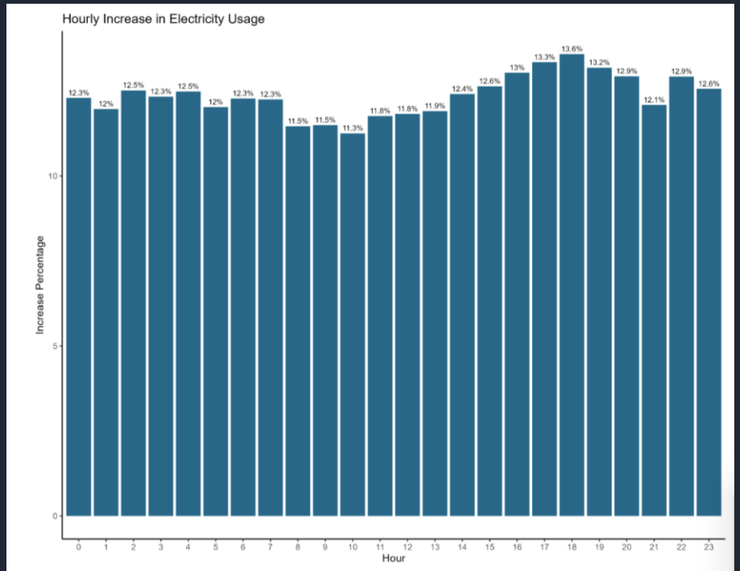


Figure 6: Electricity consumption increase by hour

Energy usage increases spike around 6 PM

Approaches to reduce Energy Demand

Efficient Energy Markets

Prices can act as signals to drive demand down.

Reduce Energy Consumption

The most impactful strategy for reducing electricity consumption may lie in altering household consumption patterns. Practical measures might include reducing air conditioning usage or adjusting the set temperature.

Renewable Energy Sources

Increase reliance on clean and sustainable energy options to reduce strain on the grid.

Conclusions

- The analysis focused on understanding the drivers of electricity demand, particularly in the face of potential temperature increases during the peak month of July.
- One of the primary findings is that the major determinants of energy usage are fixed features of residential properties, such as floor area and room count.
- the analysis highlighted the necessity of incorporating electricity pricing into the predictive models. The inclusion of price data would provide a more realistic and nuanced understanding of consumer behavior and energy consumption patterns.
- a comprehensive strategy that also considers dynamic pricing mechanisms would offer a more balanced and effective way to manage energy consumption patterns in the face of rising temperatures.