Analysis of residential energy consumption in London

Camila MalagónSuárez
Andrés Camilo Viloria García
Yesid Rivera
Oscar Nieto Garzón
Eduardo González Fierro
Didier Santander
Hollman Báez

July 2022

Contents

1 OVERVIEW

Energy is one of the main topics on the UN agenda for the following years, to assure global accessibility and reduce the related generation of pollution. According to the UN, energy currently provides 60% of the greenhouse gas emissions, although 13% of the global population has no access to electricity.

For these reasons, countries like the UK are making efforts to create public policies focused on converting their current energy source to clean alternatives.

To understand the dynamics of residential energy consumption in large cities, in 2014, the UK Government hired UK Power Networks for a project focused on collecting information about energy production and consumption through smart meters installed in a selected group of London households.

This information is useful to determine the current residential sector energy consumption characteristics. For UK Power Networks and the UK Government, it is important to know in detail the patterns of energy consumption in London's households, to create strategies to ease the transition to clean energy sources.

This project is focused on providing relevant information to the public and private entities, such as the government of the United Kingdom, London authorities, energy suppliers, network operators, researchers, and in general players of the energy market about energy consumption patterns and demand trends of London households to allow them to make better decisions in efficiently planning and operation of the electricity distribution networks, improving customer service and adopting of low carbon strategies. Last but not least, this study can be used as a guide for other countries that want to advance in the implementation of alternative energies.

1.1 Problem impact

An improved understanding of energy consumption patterns allows for studying the potential for smart approaches to assist distribution systems management. Considering the transition challenges to renewable energy sources imply to the supply requirements, this data analysis can result in insights needed to avoid energy deficits in the future. Also, it allows quantifying the consumer response of setting residential dynamic Time-of-Use tariffs, which could lead to peak reduction on the network and match electricity consumption to sustainable energy availability. Finally, this information will be useful to a Distribution Network Operator (DNO) to efficiently plan and operate electricity distribution networks, improving customer service and the adoption of low carbon technologies.

1.2 Potential audience

This project is focused on providing relevant information to the public and private entities, such as the government of the United Kingdom, London authorities, energy suppliers, network operators, researchers, and in general players of the energy market about energy consumption patterns and demand trends of London households to allow them to make better decisions in efficiently planning and operation of the electricity distribution networks, improving customer service and adopting of low carbon strategies. Last but not least, this study can be used as a guide for other countries that want to advance in the implementation of alternative energies.

2 OBJECTIVES

Specific problems are going to be defined in the following business questions:

- How are the variables correlated to energy consumption taking into account the classification presented by ACORN.

- What variables have an impact on energy consumption in London's households, considering the classification presented by ACORN, and how are they correlated.

- How does knowledge of energy consumption in the UK Power Networks-led Low Carbon London project contributes to a better understanding of general consumption patterns.
- How are energy consumption peaks affected by applying the Dynamic Time of Use (dToU) prices throughout the 2013-year calendar period.
- Once the energy transition is implemented, what will be the expectations of energy consumption for the next years considering that the data reaches until 2014.
- Relate results from London to Colombia (Use insights from the UK energy market that could be used in Colombia).

3 MATERIALS AND METHODS

3.1 Data

Database	Scheme	Table	Description
Rds-database	Public	Weather hourly darksky	Database with the behavior of the meteorological variables of the city of london every hour since november 11, 2011 and february 15, 2014
Rds-database	Public	Weather daily darksky	Database with the behavior of the meteorological variables of the city of london every day since november 11, 2011 and february 15, 2014
Rds-database	Public	Uk bank holidays	Database with holidays in london since november 11, 2011 and february 15, 2014
Rds-database	Public	Information households	Database with all the information on the households in the panel (their acorn group, their tariff) and in which block.csv file their data are stored
Rds-database	Public	Acorn details	Database with details on the acorn groups and their profile of the people in the Group
Rds-database	Public	Halfhourly dataset	Zip file that contains the block files with the half-hourly (0 to 111) smart meter measurement
Rds-database	Public	Hhblock dataset	Zip file that contains the block files with the half-hourly (0 to 111) smart meter measurement
Rds-database	Public	Daily dataset	Zip file that contains the block files (0 to 11) with daily information like the number of measures, minumum, maximum, mean, median, sum, and standard dev.

Table 2: Datasets Source: https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london

3.2 Visualizations and Models

To understand the different components of energy consumption, in the Exploratory Data Analysis (EDA) phase it is necessary to classify and visualize the patterns behind this variable, dis-aggregating the data for the next points of analysis according to the ACORN categories and groups:

- Histograms to analyze the distribution of the consumption per hour to explore the behavior throughout the day.
- Bar charts to see whether there are differences or not between the energy demand per day, especially in comparison to working days and holidays.
- Line plots for the months of the year with the average consumption per Group.
- Scatter plots to visualize patterns of consumption according to weather conditions like temperature.
- Heatmap for households consumption to display possible specific behaviors for ACORN categories and/or groups, and identify the most relevant of them, such as:
  - Technology use (devices and frequency of internet usage)
  - Housing size
  - Occupation (employment) of the population
  - Travel behavior (work & vacations)
  - Finance (income)
  - Family (structure and size)
  - Lifestyle (regular exercise)
  - Environmental care (member of a group)
- Scatter plots to see possible correlations between consumption and the kind of tariff applied.

3.3 Models

In this phase, statistical models of time series that could fit our data are explored, investigating their concepts, formulas, as well as the requirements and limitations of their use. Furthermore, the first attempts of data modeling are made, looking for identifying the model that better fits and has a great performance in the available data. Below, there is a short explanation of the considered time series forecasting models explored in this first phase of modeling.

3.3.1 ARIMA

ARIMA is a class of time series prediction models, and the name is an abbreviation for AutoRegressive Integrated Moving Average. The backbone of ARIMA is a mathematical model that represents the time series values using its past values. This model is based on two main features: Past Values and Past Errors. An important aspect here is that the time series needs to be standardized such that the model becomes independent from seasonal or temporary trends. The formal term for this is that we want the model to be trained on a stationary time series. AutoRegressive (AR): The parameter p tells us how many past values to consider for the expression of the current value. Essentially, we learn a model that predicts the value at time t as:

$$y_t = (\alpha_{t-p})(y_{t-p}) + (\alpha_{t-p+1})(y_{t-p+1}) + \ldots + (\alpha_{t-1})(y_{t-1}) \tag{3.1}$$

Moving Average (MA): How many of the forecast errors in the past should be considered. A new value is computed as:

$$y_t = (\theta_{t-q})(\varepsilon_{t-q}) + (\theta_{t-q+1})(\varepsilon_{t-q+1}) + \cdots + (\theta_{t-1})(\varepsilon_{t-1}) \tag{3.2}$$

The past prediction errors:

$$\varepsilon_i = y_i - \hat{y_i} \tag{3.3}$$

The combination of the three components gives the ARIMA(p, d, q) model. More precisely, we first integrate the time series, and then we add the AR and MA models and learn the corresponding coefficients. This could be the first option for us.

3.3.2 Prophet Forecasting model

Prophet is a time series forecasting model that is based on an additive model approach, where non-linear trends are fit with three main model components: growth (or trend) g(t), seasonality s(t), holidays h(t), and an error term is included to represent any changes which are not accommodated by the model (Taylor & Letham, 2017). One can tune the trend and seasonality hyperparameters to fit the model as well as possible, changing its value using cross-validation. The forecasting is phrased as a curve-fitting task, with time as the only regressor, so the model is univariate.

These components are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{3.4}$$

This formulation is similar to a generalized additive model (GAM), a class of regression models with potentially non-linear smoothers applied to the regressors, that has the advantage of being flexible, accurate, fast to implement, and interpretable parameters (Menculini et al., 2021). In this case, Prophet has some advantages compared to other time series models, such as its capacity to handle seasonal variations, missing data, and outliers.

This model is an open-source tool provided by Facebook Inc. through the prophet package, available in Python and R.

As mentioned before, the Prophet model could be implemented in Python, following the sklearn model API. So, after the installation, it is necessary to use an instance of Prophet class, which will be fitted and used for the forecasting. The input of Prophet is always a data frame with two columns, ds (datestamp column, with a format expected by pandas) and y (the numeric dependent variable), representing the measurement to forecast. By default, the package has integrated a cross-validation and forecasting tool, key features to the whole implementation process.

This process can be divided into three main steps: data preparation and fitting of the model, cross-validation, and hyperparameter tuning and forecasting.

Fitting the model is a very straightforward process but many parameters can be adjusted to optimize the model performance. We can outline that the type of trend, the trend changepoints (trend flexibility), the flexibility of the seasonality (Fourier order), and the holiday effects can be changed, however, it is recommended that only the flexibility of the trend and the seasonality must be tuned.

3.3.3 Exponential Smoothing - Holt- Winter model

The Exponential Smoothing Holt- Winter model is used for time series data that present trend and seasonal components, which means data should reflect an increasing or decreasing trend and shows peaks or falls with a certain frequency. In the exponential smoothing method, a greater weight is assigned to the last observations, and the weight will decrease as the observation gets older. This method generates a smoothing for each component in the time series: variation, trend, and seasonality (Mejía & Gonzales, 2019).

Below, there is the basic equation of the method:

Overall smoothing:

$$S_t = \alpha \frac{X_t}{L_{t-1}} + (1-\alpha)(S_{t-1} + b_{t-1}), \quad \alpha \ \epsilon \ (0,1) \tag{3.5}$$

Trend smoothing:

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}, \quad \gamma \ \epsilon \ (0,1) \tag{3.6}$$

Seasonal smoothing:

$$L_t = \delta \frac{X_t}{S_t} + (1-\delta)L_{t-k}, \quad \delta \ \epsilon \ (0,1) \tag{3.7}$$

Forecast:

$$X_{t+m} = (S_t + m * b_t)S_{t+m-k} \tag{3.8}$$

Where:
S_t: value of exponential smoothing for the time period t
α: parameter for exponential smoothing
X_t: real value of the time serie in the time period t
b_t: trend component for the time serie in the period t
γ:parameter of trend component in the exponential smoothing
L_t: seasonality component for the time serie in the period t
δ: parameter of seasonality component in the exponential smoothing

To improve the performance of the prediction, the parameters of the model should be adjusted to obtain the minimum mean square error (Banda & Garza, 2014). The exponential smoothing Holts-Winter model could be applied in python, through the statsmodel library and the Exponential Smoothing function.

4 EDA

Cleaning strategies:

As a first step before starting the exploration and analysis of data, it is needed to review searching null or duplicated values, and in general, issues that could cause a miss understanding of the reality, therefore a table with this overview is shown at figure 4.1 (cleaning process).

One of the biggest findings is relative to the daily dataset with more than 11 thousand null values. Going deep into this information we can see there are some days with troubles for the smart-meter recording as 2014-02-28, where there is just one record for the 48 half-hours for every household for all of the households in the study, that is the explanation for 4,987 records with null standard deviation, as well as other days have less than 48 counts (half-hours) and this is a problem because we can have for one day one household with 24 records, but it is not possible to know what of the 48 half-hours we are looking at (ie. morning hours vs night hours). The total number of days by households with less than 48 records is 41,081 (1.18% of the whole data set), then those data will be dismissed.

For the 'hhblock' table the issues are focused on half-hour-30 for 5,460 records (weekday) of the total of 3.5 million, then, the null values will be completed using the average of the same half-hour-30 for that household using the consumption of the week of the missing day (avoiding a change of conditions like the season).

For households, there are 2 issues, one of them is households without a group (acorn-) where there is no way to find the right group, then, those records will be dismissed too. And there are 49 rows with group U, those will be recategorized as 'acorn-R' for the new category 'Not Private Households'.

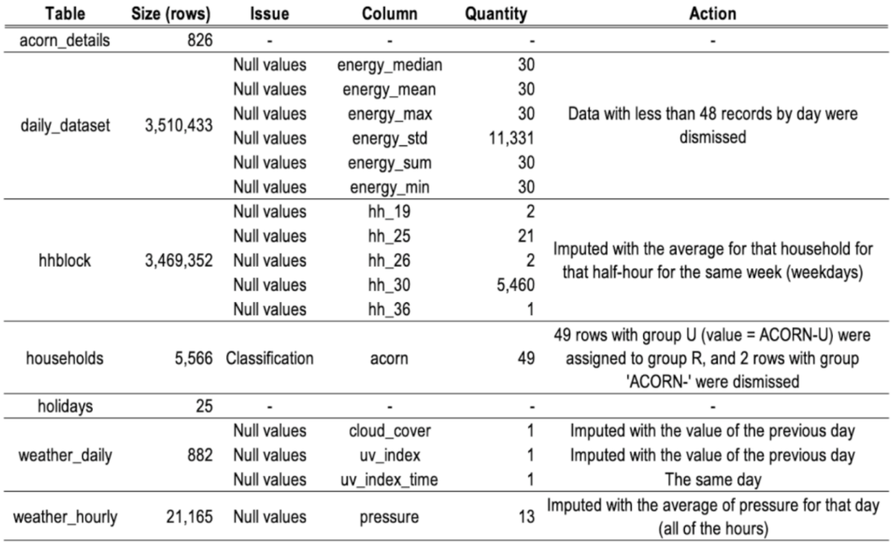


Table	Size (rows)	Issue	Column	Quantity	Action
acorn_details	826	-	-	-	-
daily_dataset	3,510,433	Null values	energy_median	30	Data with less than 48 records by day were dismissed
		Null values	energy_mean	30	
		Null values	energy_max	30	
		Null values	energy_std	11,331	
		Null values	energy_sum	30	
		Null values	energy_min	30	
hhblock	3,469,352	Null values	hh_19	2	Imputed with the average for that household for that half-hour for the same week (weekdays)
		Null values	hh_25	21	
		Null values	hh_26	2	
		Null values	hh_30	5,460	
		Null values	hh_36	1	
households	5,566	Classification	acorn	49	49 rows with group U (value = ACORN-U) were assigned to group R, and 2 rows with group 'ACORN-' were dismissed
holidays	25	-	-	-	-
weather_daily	882	Null values	cloud_cover	1	Imputed with the value of the previous day
		Null values	uv_index	1	Imputed with the value of the previous day
		Null values	uv_index_time	1	The same day
weather_hourly	21,165	Null values	pressure	13	Imputed with the average of pressure for that day (all of the hours)

Figure 4.1: cleaning process

For the weather information, there is a little group of null values that will be filled with data from the same day as the missing value or the day before (avoiding again a significant change in the climate conditions).

The review gives a 0 number of duplicate data for all the tables, then, this is not an issue to deal with.

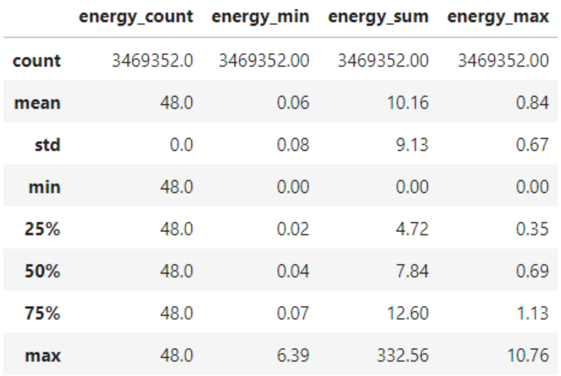
4.1 Analysis per dataset

4.1.1 Daily dataset

This dataset contains the records that were collected for the Smart meters. We have the daily consumption, which was obtained by adding up the 48 records each half-hourly while the 24 hours of the day. Also, the dataset has the date, and ID as categorical variables and the median, mean, maximum, minimum, and standard deviation of the records.

We are going to work with 3,469,352 rows after cleaning data process. One row is data for each column, and we have 17 columns.

Descriptive statistics for consumption of energy:



	energy_count	energy_min	energy_sum	energy_max
count	3469352.0	3469352.00	3469352.00	3469352.00
mean	48.0	0.06	10.16	0.84
std	0.0	0.08	9.13	0.67
min	48.0	0.00	0.00	0.00
25%	48.0	0.02	4.72	0.35
50%	48.0	0.04	7.84	0.69
75%	48.0	0.07	12.60	1.13
max	48.0	6.39	332.56	10.76

Figure 4.2: Consumption summary

Consumption of energy by date

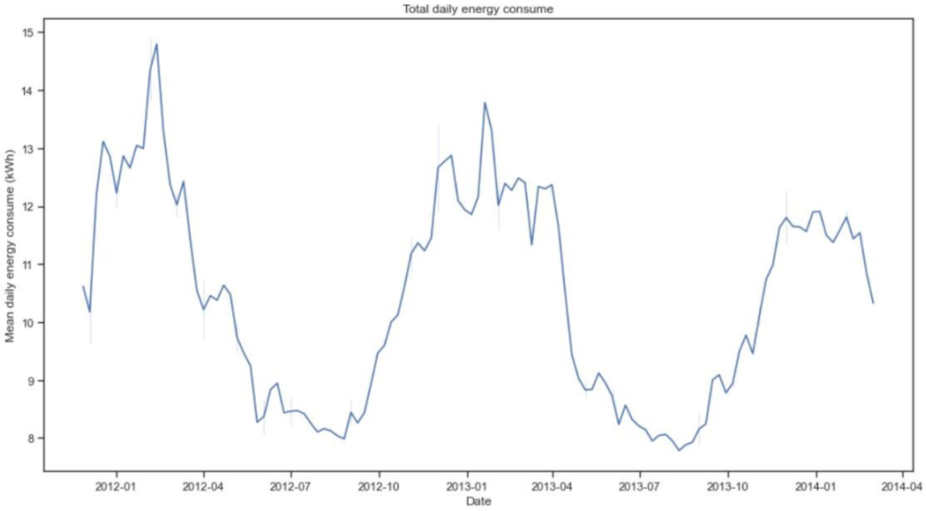


Figure 4.3: Daily energy consumption

In this plot, we can see that the energy demand has a seasonal behavior, but at this moment we can't affirm anything.

Maximum, minimum, and mean consumption of energy by month

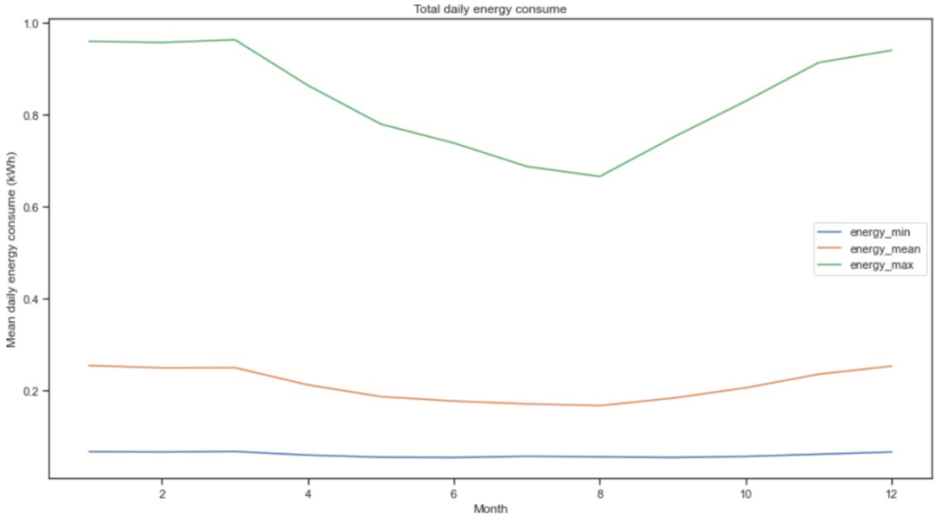


Figure 4.4: Energy consumption trends

This plot shows us the min, max, and mean consumption of energy. We see that the minimum is 0, it is said every hour there are places where they don't use energy. The maximum consumption was close to 1 kw/h. and the mean is 0.3 kw/h approximately.

Mean consumption of energy by month during 2011, 2012, and 2013.

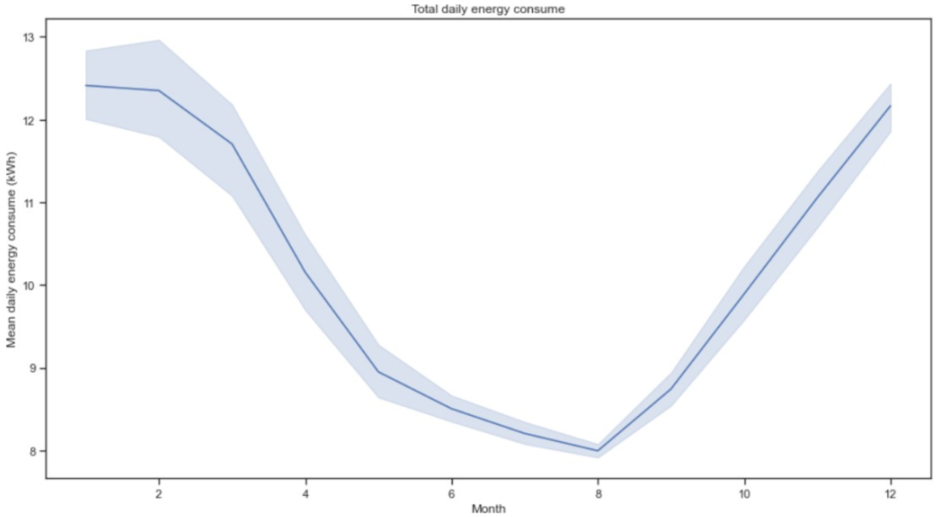
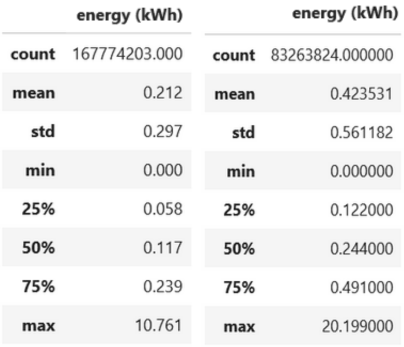


Figure 4.5: Energy consumption per month

There, we can see the months and this consumption. The behavior shown in the plot is that between January and March the demand is high and between April and August the consumption decreases until the minimum. Then, between September and December the consumption increases.

4.1.2 Half-hourly dataset

The dataset that contains half-hourly data has more than 160 million entries, with 3 columns. The numeric column with the half-hourly energy consumption per household contains 5558 missing values



	energy (kWh)
count	167774203.000
mean	0.212
std	0.297
min	0.000
25%	0.058
50%	0.117
75%	0.239
max	10.761

	energy (kWh)
count	83263824.000000
mean	0.423531
std	0.561182
min	0.000000
25%	0.122000
50%	0.244000
75%	0.491000
max	20.199000

Figure 4.6: half-hourly and hourly summary

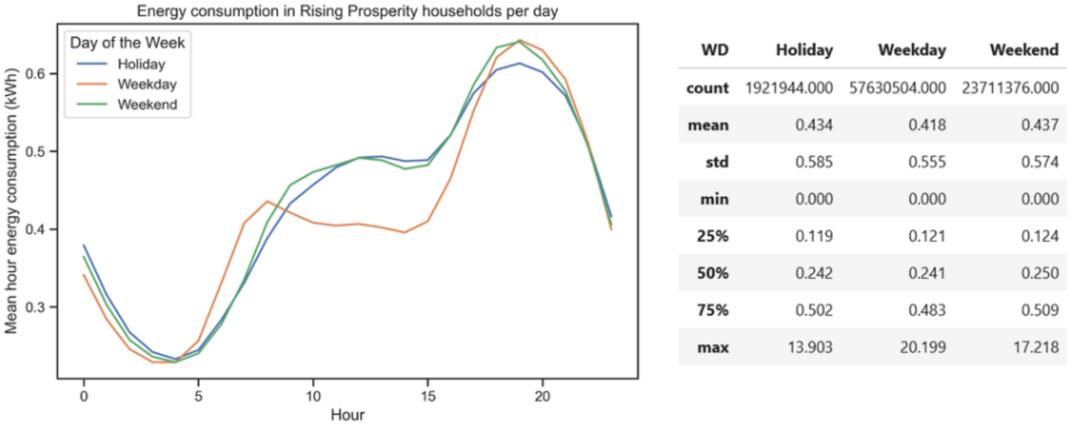
in total, of which a major part corresponds to observations registered on a non-standard frequency of observation, i.e. 15:13:37 instead of 15:30:00. Also, this table contains 5566 unique households, ids, matching the total number of households reported in other datasets.

The mean of the same half-hour of the corresponding week of the day with the missing value was computed to impute these values. This mean was used to fill in the missing values. Then, we aggregated the data into an hourly dataset, aggregating the recorded value of the same hour of the corresponding date and household. From this process, we obtain a new dataset with almost half the entries, around 80 million without missing values

We can see that the number of observations was reduced almost to half and that the mean is almost double the half-hourly data set mean. Also, there is a great difference between the 75th percentile value and the maximum value in the two-time scales.

Basic statistics of half-hourly and hourly datasets This new hourly dataset was used to perform the Exploratory Data Analysis on different time scales.

Analysis for day of the week First, the analysis per day of the week was done, considering weekdays, weekends, and holidays across the entire period of study. The dates of holidays were extracted from an additional dataset that listed the UK bank holidays, between 2011 and 2014.



WD	Holiday	Weekday	Weekend
count	1921944.000	57630504.000	23711376.000
mean	0.434	0.418	0.437
std	0.585	0.555	0.574
min	0.000	0.000	0.000
25%	0.119	0.121	0.124
50%	0.242	0.241	0.250
75%	0.502	0.483	0.509
max	13.903	20.199	17.218

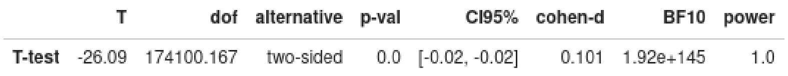
Figure 4.7: Energy consumption per type of day

Mean hourly energy consumption by day of the week The hourly consumption pattern shows that there is a peak of energy consumption, between 18:00 and 20:00 hours, in the evening. Also, it seems that holidays and weekends have a similar consumption profile, with an increase in the values in the morning and mid-day. In contrast, on weekdays household energy consumption tends to rise earlier in the morning but maintains a stable value until late afternoon.

Also, the number of observations shows that a major part of the dataset corresponds to weekday values, with more than half of the total observations. The mean of each three groups of days is similar but on weekdays and weekends, the maximum value is larger than on holidays.

Additionally, to test that there is a meaningful difference in the hourly energy consumption between weekdays and weekends t-test is used, quantifying the difference between their means. In this case, the significance is defined as $\alpha = 0.05$.

Test 1: the output of the t-test between the hourly energy consumption per household between weekdays and weekends is presented with the corresponding statistics.



	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-26.09	174100.167	two-sided	0.0	[-0.02, -0.02]	0.101	1.92e+145	1.0

Figure 4.8: t-test weekdays vs weekends

We can see that the p-value (0.0) $<$ α (0.05), so the null hypothesis can be rejected, concluding that there is a significant difference in the energy consumption per day of the week and their means in the studied households.

Analysis for hour of the day In the line of describing the consumption per hour of the day, one relevant variable to include is the temperature, then, to see the possible pattern changes an analysis of two scenarios will be performed as follows:

Scenario 1: The coldest day for the period of the study has been selected on 2012-02-11, and no matter the category or group, the next chart shows the consumption for that day, broken down per hour:

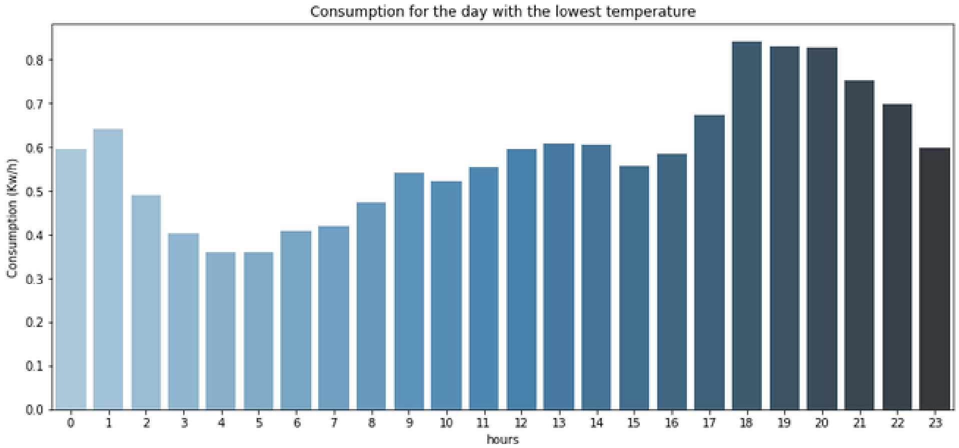


Figure 4.9: Energy consumption per hour (coldest day)

Scenario 2: Now the hottest day has been selected on 2013-07-31, and the same structure for information is shown below:

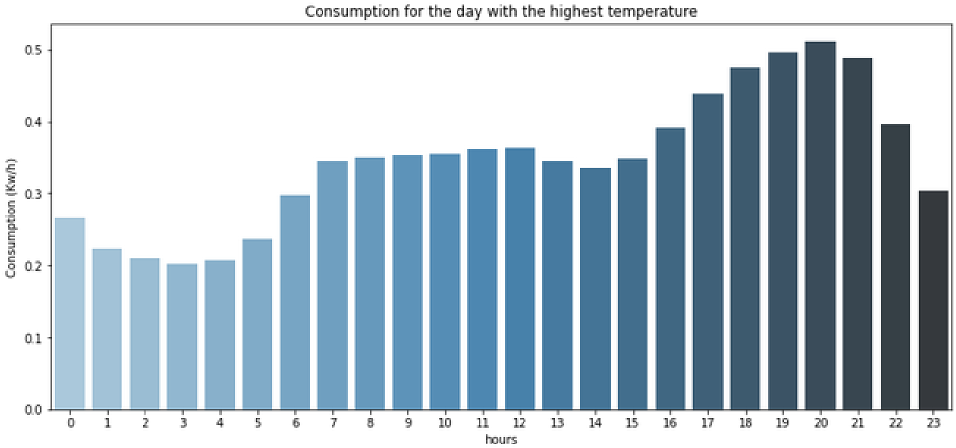


Figure 4.10: Energy consumption per hour (warmest days)

It is illustrative to see the general consumption with valley hours in the first hours of the morning (before 6 am, light bars) and then a constant rising to get the highest values at the beginning of the night (around 8 pm, dark bars); but for the coldest day the gap in lower and the minimum consumption is recorded around 4 am to 5 am with an average value of 0.4 Kw/h, as well as there is no huge reduction between 8 pm and 1 am; meanwhile the hottest day has a higher variation through the day, with the lowest point at 1 am to 5 am with a consumption a little greater than 0.2 Kw/h (2 times longer and 50% lower versus the coldest day). In addition, there is a flat range between 7 am and 3 pm (a significant part of the day) and the trend after 8 pm is for a fast decrease to get the minimum again.

The maximum value for the coldest day is greater than 0.8 Kw/h while for the hottest day it is lower than 0.5 Kw/h, which tells many things about the patterns of consumption in function of temperature, over the hour patterns even.

Checking the trends per hour and throughout the years, we realize a possible pattern for hours on the chart below:

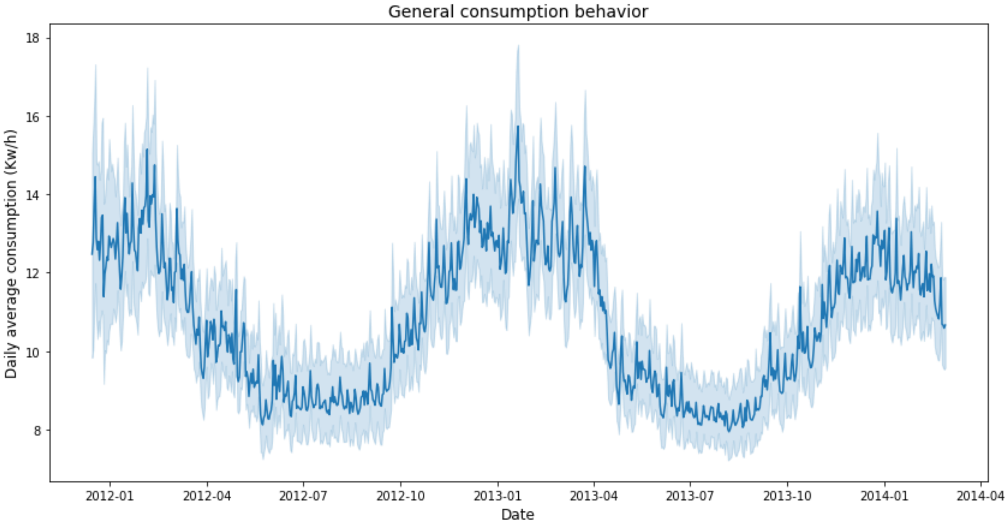


Figure 4.11: Energy consumption across the time

There could be an explanation since this could be correlated to the stations and days of the week.

Let's check the average consumption per hour:

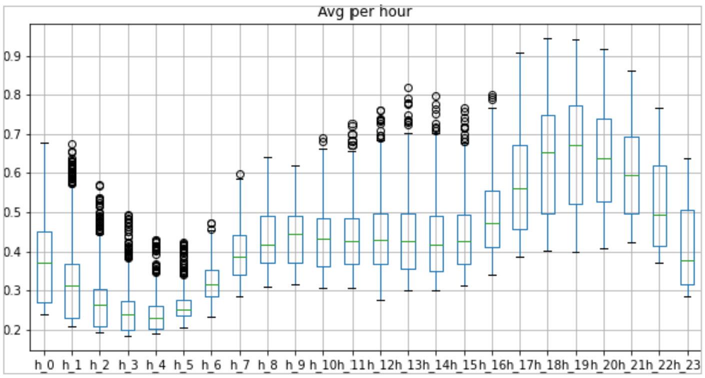


Figure 4.12: Box plot per hour

This graph shows the average consumption starts increasing from 17h to 22h, this will be the peak hour for us. When it comes to checking the day with the highest consumption, it is explained in this graph, Wednesday, Thursday, and Sunday. Now, there won't be a 0 demand for energy, so, checking the minimum consumption for the people, businesses, and the company, it will look like this, stable and almost perfectly aligned. Let's check those consumptions:

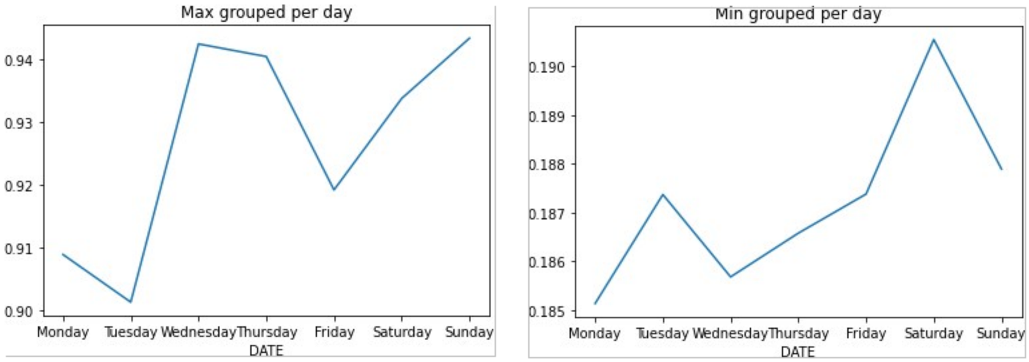


Figure 4.13: Energy consumption per week day (Max vs Min)

This will be interpreted together to get a hint on how the demand will be covered when having peaks and valleys, ups and downs. And to finish, we will plot the mean of the days to check if the weekends increase the consumption:

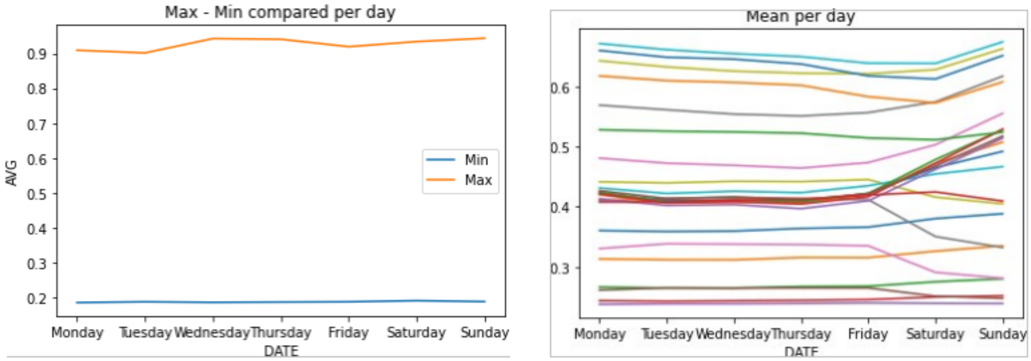
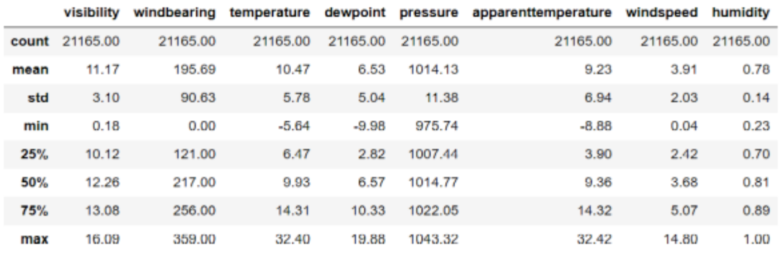


Figure 4.14: Energy consumption per week day (Max vs Min) and Average per weekday

Takeaways so far, the consumption increases on weekends, and weekdays starting 16 to 22 hours. This will give us a hint on how to forecast the consumption and the days when we need to back up and guarantee the service.

Hourly weather dataset In the hourly weather dataset, there is meteorological information about the daily climate in London, such as minimum and maximum temperature, cloud cover, visibility, time in which the highest and lowest temperature was registered, and wind. The dataset has 21.165 records, one for each hour in which the study was conducted, only 13 records have null data in the pressure column. To complete this information, an imputation of the mean values of total records from pressure was made. Below is a table with the summary of the most important descriptive statistics of the dataset:



	visibility	windbearing	temperature	dewpoint	pressure	apparenttemperature	windspeed	humidity
count	21165.00	21165.00	21165.00	21165.00	21165.00	21165.00	21165.00	21165.00
mean	11.17	195.69	10.47	6.53	1014.13	9.23	3.91	0.78
std	3.10	90.63	5.78	5.04	11.38	6.94	2.03	0.14
min	0.18	0.00	-5.64	-9.98	975.74	-8.88	0.04	0.23
25%	10.12	121.00	6.47	2.82	1007.44	3.90	2.42	0.70
50%	12.26	217.00	9.93	6.57	1014.77	9.36	3.68	0.81
75%	13.08	256.00	14.31	10.33	1022.05	14.32	5.07	0.89
max	16.09	359.00	32.40	19.88	1043.32	32.42	14.80	1.00

Figure 4.15: Basic statistics of hourly weather dataset

In this project, we are interested in discovering the relationship between energy consumption per household and meteorological data. The first variable to consider is temperature due to the different seasons where the London population lives in.

The following graph shows the total count of hours recorded for each season, indeed it was expected that the count would be higher in the winter because this season is longer than the others.

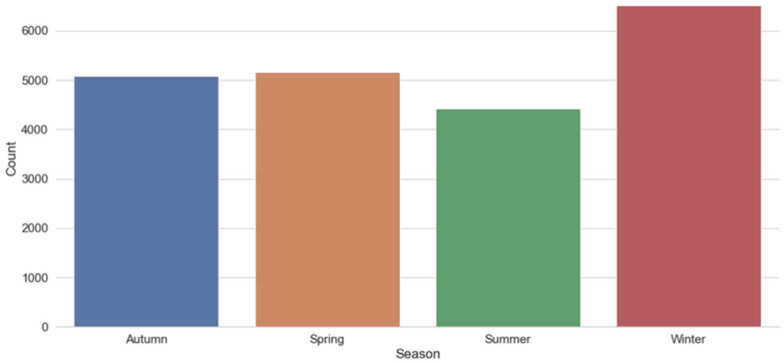


Figure 4.16: Count of hours per season

Regarding the most important variables that could have a potential in the incidence of energy consumption in households in London, this correlation is shown below.

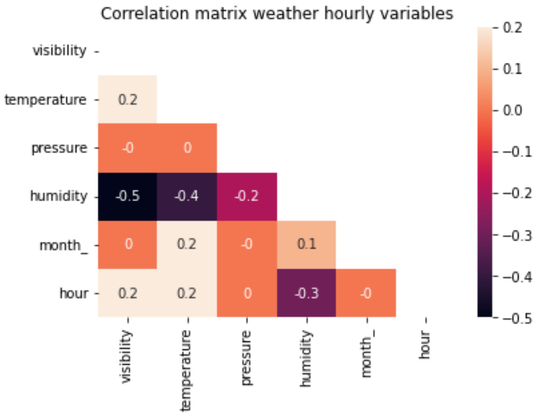


Figure 4.17: Correlation matrix for weather hourly variables

It is without a doubt that energy consumption is influenced by temperature, and in turn, this variable has a high dependence on other variables such as humidity, since the higher the temperature, the lower the humidity, therefore the correlation can be seen in red, on the other hand, visibility also shows a positive correlation with temperature. Therefore, as it is already known how the temperature influences the humidity, in subsequent analyzes the temperature is compared with the visibility.

The following graph shows the temperature differences in each of the hours of the day in each season. It is evident in chronological order that as you go through the temperature increases and the same behavior in each season is very similar, forming a behavior like in S where from the first hour of the day until around 8 am there is a drop in temperature, but then from this time until about 2 pm, the trend decreases again. This occurs in each season but on a different scale, showing a much more aggressive change in summer than in other seasons.

From January to August is that when the summer ends, the temperature increases after this date then it begins to decrease.

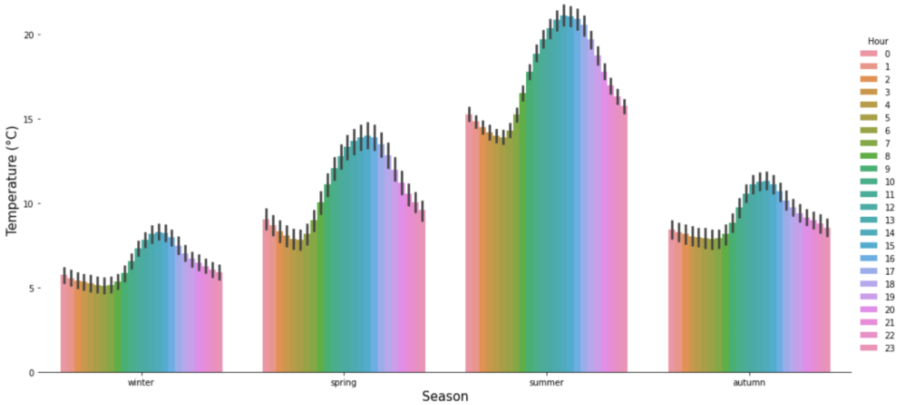


Figure 4.18: Hourly temperature per season

The effect of temperature on visibility is confirmed, on occasions where the temperature decreases, visibility also decreases, effectively on those sunny days that are more common in summer are the days where there is more visibility throughout the year.

It can be evidenced that despite the fact that the temperature is not a determining factor in visibility, since on those days when the temperature is lower in the year, less visibility is expected and it is not like that on all occasions, therefore there are other factors that they influence. The end of the year has been the period where there is less visibility and this comes down to the snowfall.

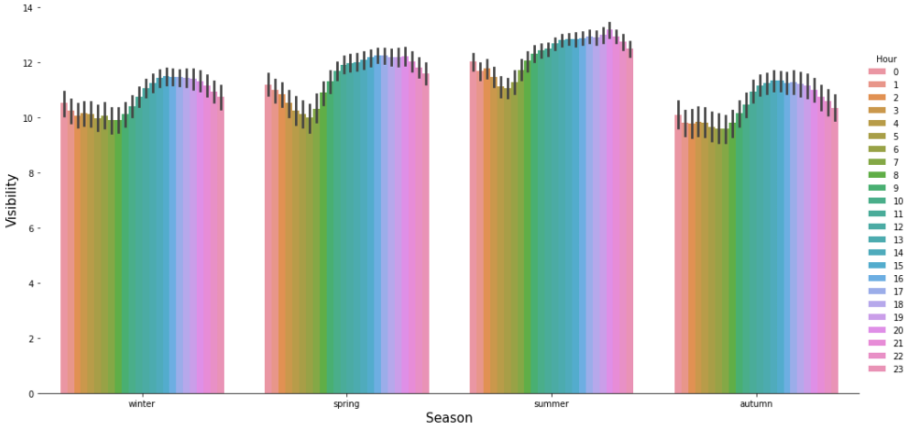
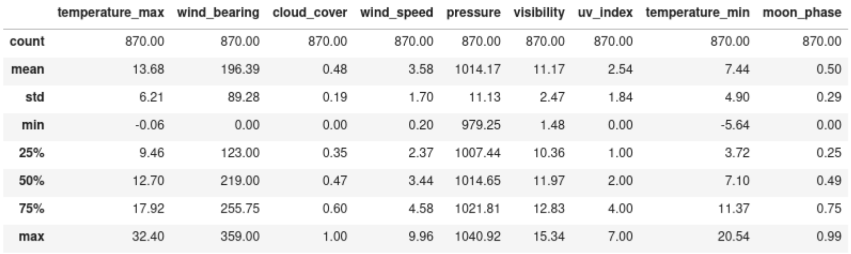


Figure 4.19: Hourly visibility per season

weather dataset In the daily weather dataset, there is meteorological information about the daily climate in London, such as minimum and maximum temperature, cloud cover, visibility, time in which the highest and lowest temperatures were registered, and wind. The dataset has 882 records, one for each day in which the study was conducted, just one record has null data in the columns cloud cover and UV index. To complete this information, an imputation of the mean values of total records from cloud cover and UV index was made. Below is a table with the summary of the most important descriptive statistics of the dataset:



	temperature_max	wind_bearing	cloud_cover	wind_speed	pressure	visibility	uv_index	temperature_min	moon_phase
count	870.00	870.00	870.00	870.00	870.00	870.00	870.00	870.00	870.00
mean	13.68	196.39	0.48	3.58	1014.17	11.17	2.54	7.44	0.50
std	6.21	89.28	0.19	1.70	11.13	2.47	1.84	4.90	0.29
min	-0.06	0.00	0.00	0.20	979.25	1.48	0.00	-5.64	0.00
25%	9.46	123.00	0.35	2.37	1007.44	10.36	1.00	3.72	0.25
50%	12.70	219.00	0.47	3.44	1014.65	11.97	2.00	7.10	0.49
75%	17.92	255.75	0.60	4.58	1021.81	12.83	4.00	11.37	0.75
max	32.40	359.00	1.00	9.96	1040.92	15.34	7.00	20.54	0.99

Figure 4.20: Basic statistics of daily weather dataset

To carry out analyses by the season of the year, an additional column was created from the date of registration, having the greatest number of records collected in the season of spring and the lowest in winter. This is aligned with the fact that the study began in November 2011 and ended in February 2014. For that reason, during the months of November and March, there are the greatest numbers of meteorological records.

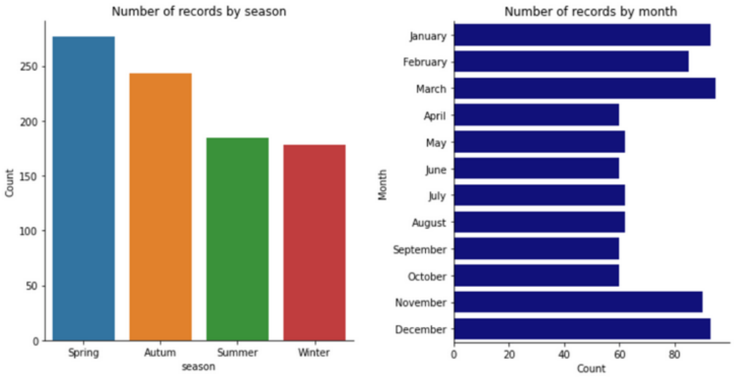
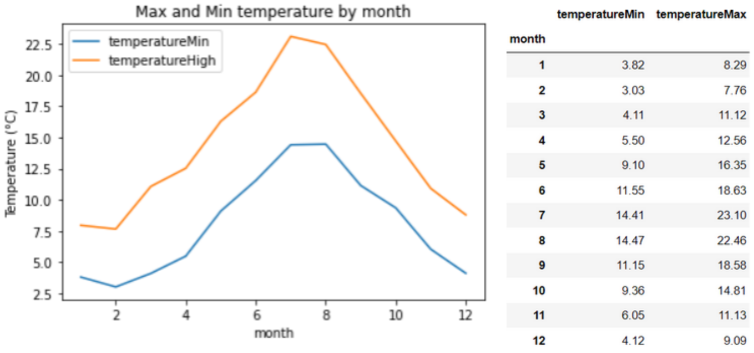


Figure 4.21: Number of records by season and month

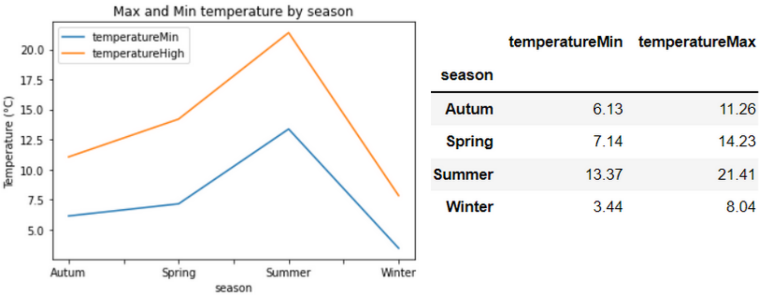
In this project, we are interested in discovering the relationship between energy consumption per household and meteorological data. The first variable to consider is temperature due to the different seasons that the London population lives in. The following figure shows the minimum and maximum temperatures per month, having the greatest temperature record between July and August and the lowest in February. The lowest temperature register is 3°C in February.



month	temperatureMin	temperatureMax
1	3.82	8.29
2	3.03	7.76
3	4.11	11.12
4	5.50	12.56
5	9.10	16.35
6	11.55	18.63
7	14.41	23.10
8	14.47	22.46
9	11.15	18.58
10	9.36	14.81
11	6.05	11.13
12	4.12	9.09

Figure 4.22: Mean maximum and minimum temperatures by month

Additionally, an analysis per season was conducted to identify the mean values of minimum and maximum temperature during each season.



season	temperatureMin	temperatureMax
Autum	6.13	11.26
Spring	7.14	14.23
Summer	13.37	21.41
Winter	3.44	8.04

Figure 4.23: Mean maximum and minimum temperatures by season

The second meteorological aspect to consider is icon used to describe the weather each day in London. According to the available data, during the months in which the study was developed, 70% of the time the weather was partly-cloudy in London.

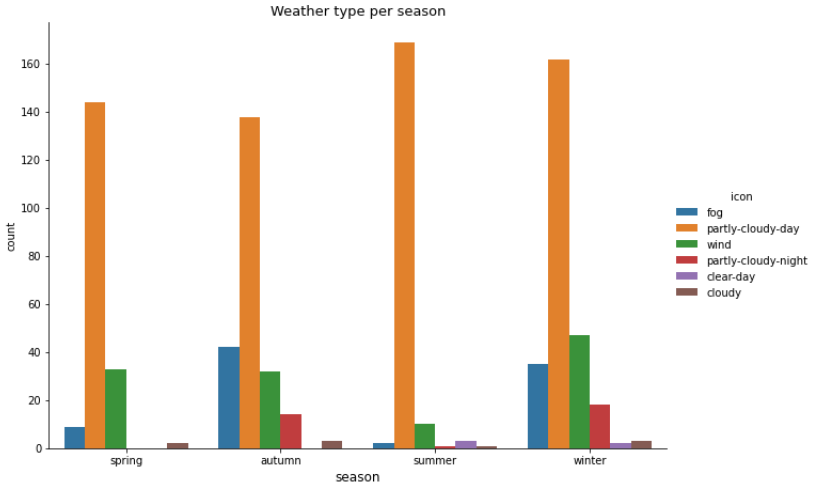


Figure 4.24: Weather type by season

The third meteorological variable is visibility, to find if there is any correlation between lower visibility and the need to consume more energy for artificial lights. According to the data, in summer (July) is the greatest visibility index.

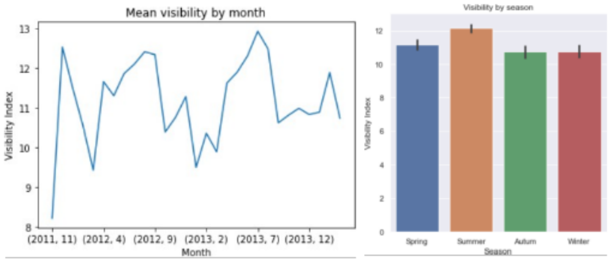


Figure 4.25: Mean visibility index per month and season

The last variable explored is the percentage of cloud cover, in summer used to be a cloud cover of almost 40%, while in winter, this measure increases to 54%. At the beginning of 2013, there was the highest cloud cover achieving more than 60%.

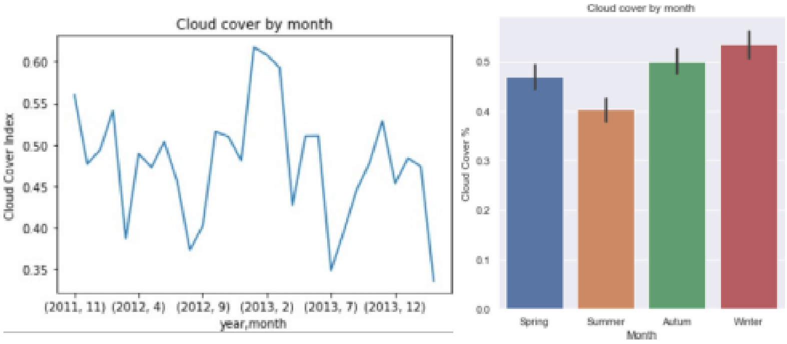


Figure 4.26: Mean cloud cover per month and season

Finally, a correlation matrix is created to detect relations between variables in the dataset. It is shown that there is a positive correlation between the min, max, and UV index. On the other hand, there is a negative relationship between temperature, humidity, and cloud cover percentage.

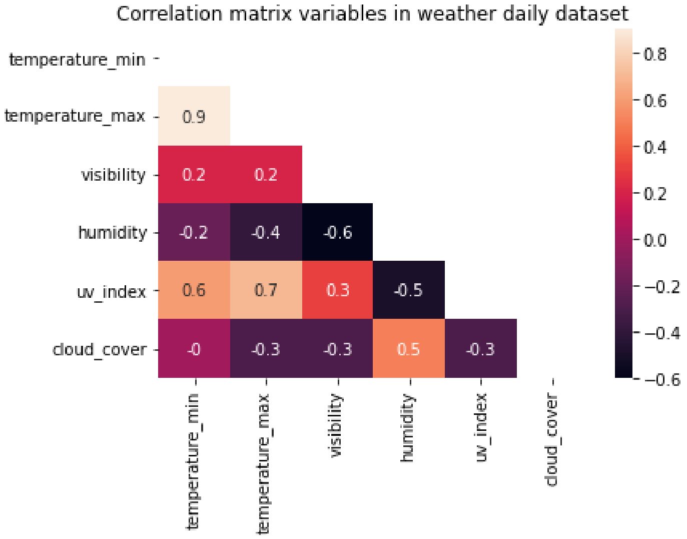


Figure 4.27: Correlation matrix variables in weather daily dataset

Additionally, it was explored the relationship between daily energy consumption in London households and the registered temperature monthly. The next figure shows graphically that if there is an increase in temperature, the energy consumption in the household will decrease.

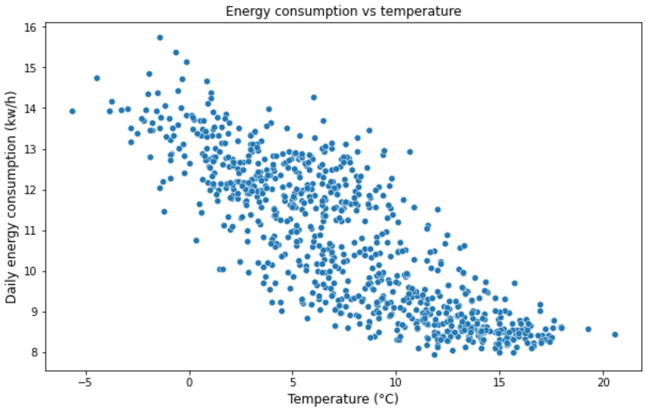


Figure 4.28: Energy consumption vs Temperature

To verify the correlation between the energy consumption and the meteorological variables explored in the weather daily report. In this graph, it is shown that the temperature has the greatest negative correlation (-0.83) with the monthly energy consumption. The other variable with a higher negative influence on energy consumption is the UV index (-0.75).

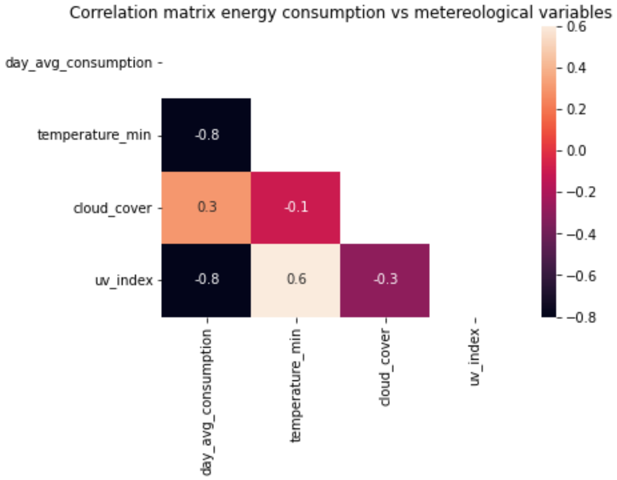


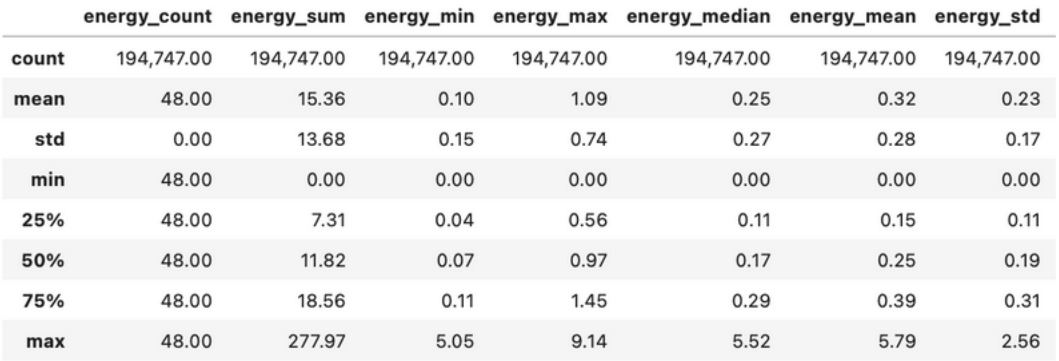
Figure 4.29: Correlation matrix energy consumption vs metereological variables

4.2 Analysis by Category

The households table contains the ACORN category and group in which each household was categorized. ACORN is a consumer classification based on demographic data, social factors, population, and consumer behavior, that segments UK postcodes into 6 main categories and 17 groups. Each category is composed of one or several groups. The dataset contains a total of 5,556 different households, classified into six ACORN categories and 17 groups. On the following graph, the number of households by category and group is presented.

4.2.1 Category 1: Affluent Achievers

The descriptive statistics for this category let us think about a low number of households, there are less than 200 thousand records (48 measures for each day for each household) as well as a high energy consumption because the maximum of the energy sum is near 280 Kw/h for a single day for a single household, too far from the average (18 times higher) which suggests the presence of outliers but it has to be analyzed in depth later to see if it is not caused for one of the groups within the category that could have a much higher mean but few households, so, few records.



	energy_count	energy_sum	energy_min	energy_max	energy_median	energy_mean	energy_std
count	194,747.00	194,747.00	194,747.00	194,747.00	194,747.00	194,747.00	194,747.00
mean	48.00	15.36	0.10	1.09	0.25	0.32	0.23
std	0.00	13.68	0.15	0.74	0.27	0.28	0.17
min	48.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	48.00	7.31	0.04	0.56	0.11	0.15	0.11
50%	48.00	11.82	0.07	0.97	0.17	0.25	0.19
75%	48.00	18.56	0.11	1.45	0.29	0.39	0.31
max	48.00	277.97	5.05	9.14	5.52	5.79	2.56

Figure 4.30: Affluent Achievers summary

To determine the behavior of each group, the next chart shows a violin plot:

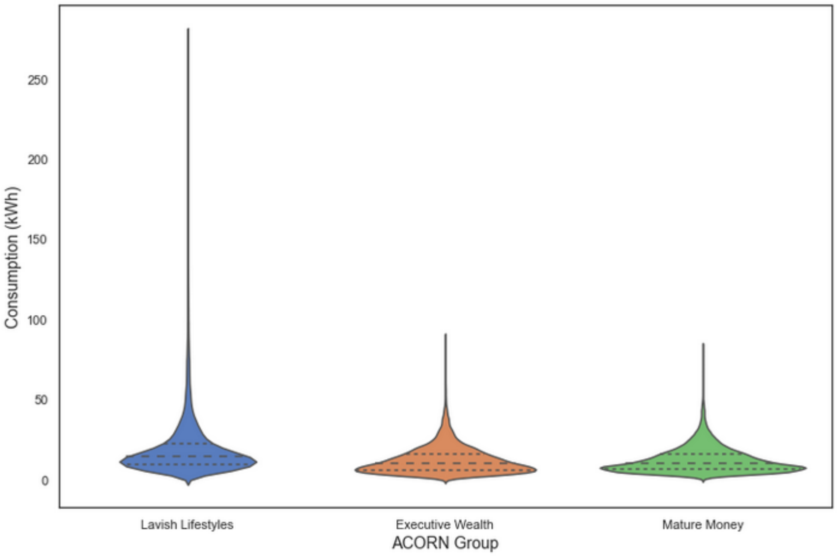


Figure 4.31: Daily energy consumption for Affluent Achievers households

The plot shows a similar consumption pattern for the households of each group, it is possible to see the outlier introduced in the previous table and it belongs to the wealthier group, not only within the category but in the entire ACORN classification, then it is interesting to look for outliers like this one in other groups.

An interesting observation is that for the 3 groups most of the records are below the mean, making the distribution skewed to the left, one possible explanation is the records of houses for vacations periods and people out from home with 0 consumption for the whole day (traveling maybe, reasonable due to the profile of the population of this category), so here rises the need for a means contrast with a subset of the groups.

To understand the behavior of the consumption across the time, a line plot is shown next:

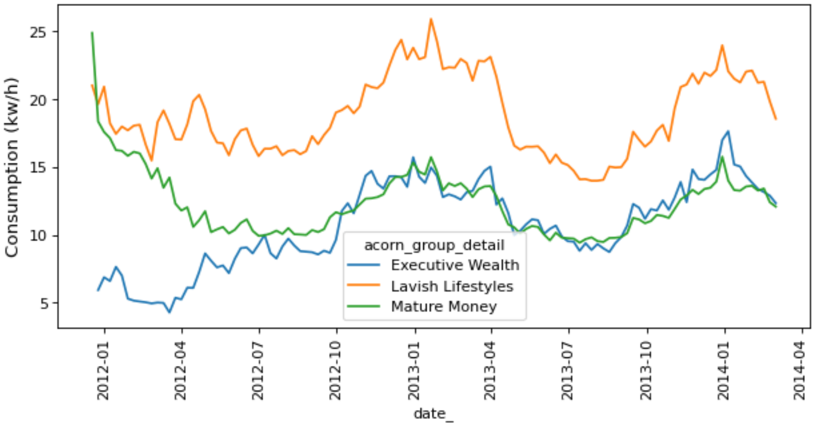


Figure 4.32: Average daily consumption for Affluent Achievers

There is a clear pattern in general for all the groups, and it is a higher consumption at the end of each year and the beginning of the next one, as well as a lower consumption for the months of the middle of the year, therefore that pattern must be analyzed in according to the average per month and the seasons of the year. Also, there is some strange noise at the beginning of the period with a large amount of volatility that needs to be seen further later, maybe just the first e months of the available data.

To continue with the behavior across the time, next there is the calculation of the average daily consumption for each one of the months of the year, separate by group to see possible different trends:

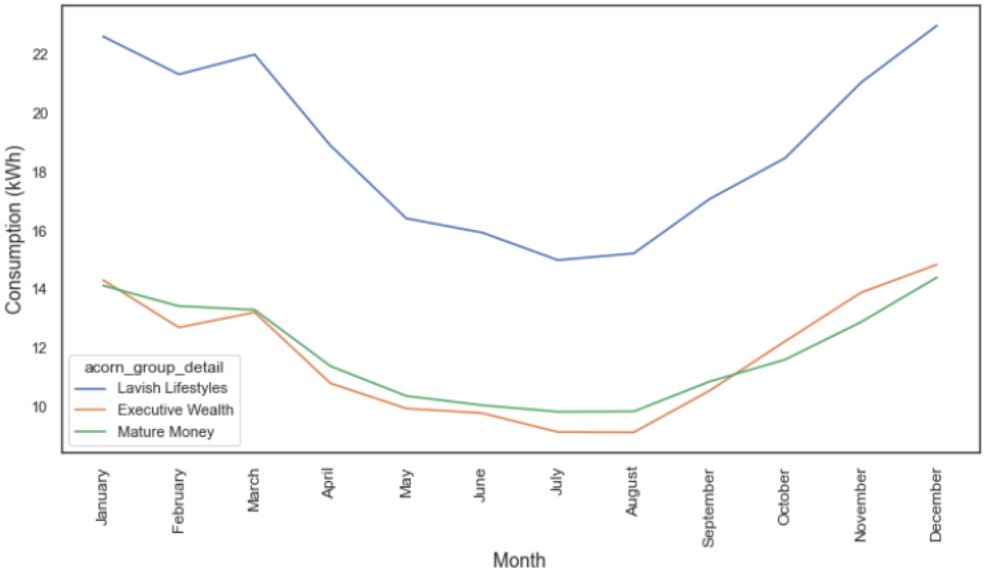


Figure 4.33: Average daily energy consumption for Affluent Achievers households per month

This chart confirms the rising trend of consumption from September to February and the decreasing trend from March to August. Although there is a clear difference between the groups, it is possible to see two of the groups: Executive Wealth and Mature Money with a similar consumption across the year, but the Lavish Lifestyles group shows a much higher consumption for every month, about 2 times higher; then a mean contrast for consumption between groups is a good approach to consider.

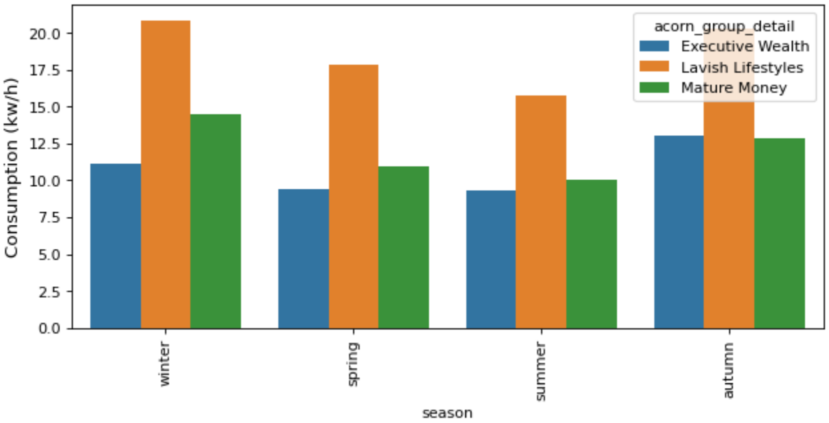
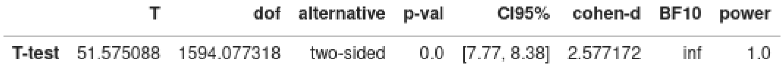


Figure 4.34: Average daily energy consumption for Affluent Achievers households per season

The previous chart shows the consumption segmented for group and season, and there is a clear rising trend beginning in the summer, growing in autumn, and reaching the highest point for the winter, no matter the group, the trend is clear, and once again there is a remarkable difference between Lavish Lifestyles and the other two groups.

To compare the means for the groups of the category, a t-test will be performed as follows:

Test 1: A subset of the data frame with the consumption per group is taken just with the energy sum (total consumption per household per day) and the group, Lavish Lifestyles in this case, and the same is performed with the group Executive Wealth, then the t-test is applied, and the result is the next:

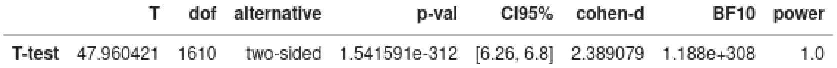


	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	51.575088	1594.077318	two-sided	0.0	[7.77, 8.38]	2.577172	inf	1.0

Figure 4.35: t-test 1

Interesting to see the p-value, which is 0.0, it means, with a 0.05 significance level that the null hypothesis must be rejected in favor of the alternative hypothesis, then it can be concluded that there is a statistical difference between the average consumption of the Lavish Lifestyles group and the Executive Wealth group.

Test 2: The same procedure is conducted again but now to compare Lavish Lifestyles and Mature Money groups:

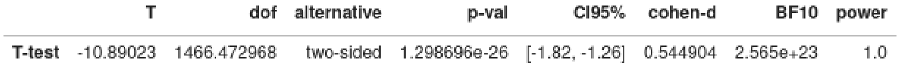


	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	47.960421	1610	two-sided	1.541591e-312	[6.26, 6.8]	2.389079	1.188e+308	1.0

Figure 4.36: t-test 2

Once again, the result is to reject the null hypothesis, it is noticeable to see the confidence interval of the two tests and how close they are. The conclusion is as expected, there is a statistical difference between the average consumption of the Lavish Lifestyles group and the Mature Money group.

Test 3: Finally, the same method is used for the Executive Wealth and Mature Money groups:



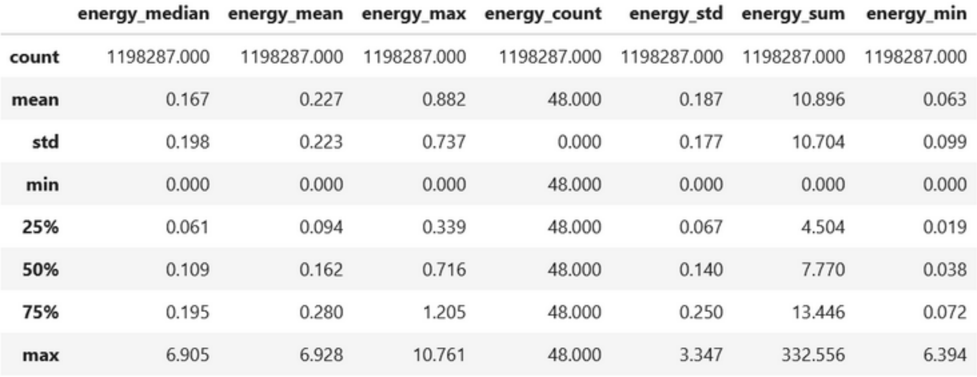
	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-10.89023	1466.472968	two-sided	1.298696e-26	[-1.82, -1.26]	0.544904	2.565e+23	1.0

Figure 4.37: t-test 3

In this case, with the same 0.05 significance level, the p-value is again lower, therefore it is possible to reject the null hypothesis and a statistical difference between the means of those groups can be infered.

4.2.2 Category 2: Rising Prosperity

The descriptive statistics for the numerical data show that the columns on the daily dataset include many statistics generated from the half-hourly dataset. In this case, the data of the Rising Prosperity category have around 1 million entries, with 7 numerical columns.



	energy_median	energy_mean	energy_max	energy_count	energy_std	energy_sum	energy_min
count	1198287.000	1198287.000	1198287.000	1198287.000	1198287.000	1198287.000	1198287.000
mean	0.167	0.227	0.882	48.000	0.187	10.896	0.063
std	0.198	0.223	0.737	0.000	0.177	10.704	0.099
min	0.000	0.000	0.000	48.000	0.000	0.000	0.000
25%	0.061	0.094	0.339	48.000	0.067	4.504	0.019
50%	0.109	0.162	0.716	48.000	0.140	7.770	0.038
75%	0.195	0.280	1.205	48.000	0.250	13.446	0.072
max	6.905	6.928	10.761	48.000	3.347	332.556	6.394

Figure 4.38: Rising Prosperity summary

We can see that there are no negative values on the data, with the minimum value of energy consumption being 0. Also, within this category, the statistics have been computed only with days with 48 half hours, so there are no missing or incomplete values.

The maximum values of all the energy-related columns are far from the 75th percentile, so surely there will be outliers. In this case, we see that there are some 0 values on the energy sum and energy min column that can be associated with periods where a specific household didn't consume energy across an entire day.

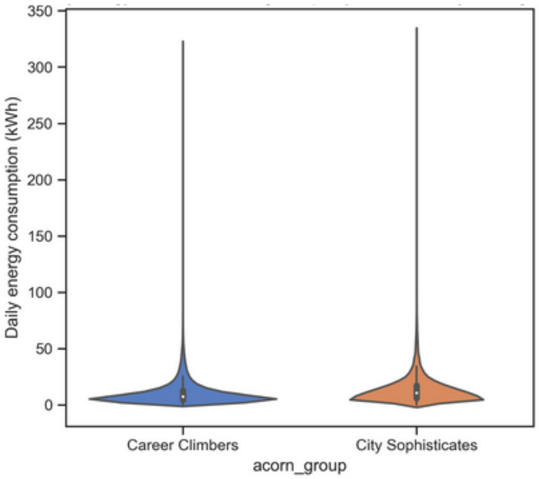


Figure 4.39: Daily energy consumption for Rising Prosperity

The violin plots of the main feature (energy sum) by group confirm that in both cases we have outliers with very large values. However, the main portion of the values shows a concentration of the observations below the mean, which can be associated with a distribution skewed to the left. This type of distribution seems to be possible because a great number of observations describe lower values, near or almost 0.

We can visualize the trend of a household's energy consumption across different time scales.

First, the daily energy consumption by each group is calculated for the entire period.

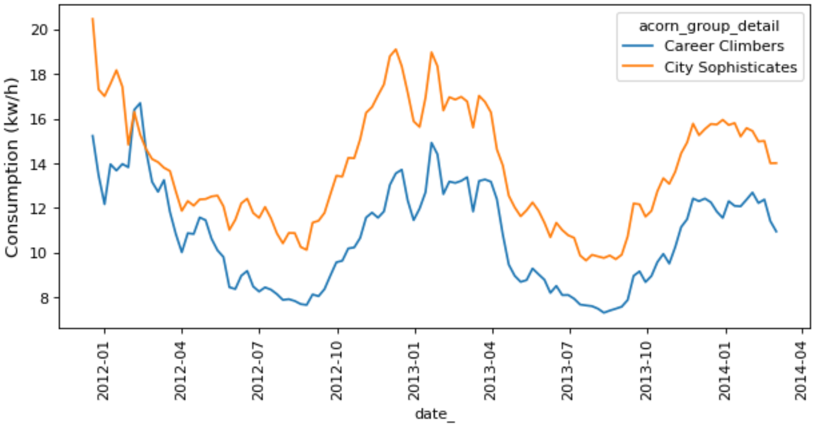


Figure 4.40: Average daily energy consumption for Rising Prosperity households

Then, the mean monthly and seasonal changes by group are calculated and plotted on the following graphs.

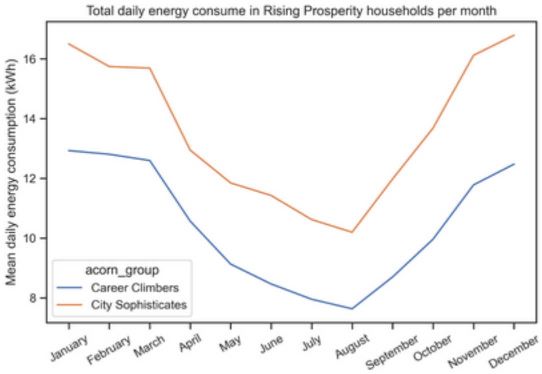


Figure 4.41: Average daily energy consumption for Rising Prosperity households per month

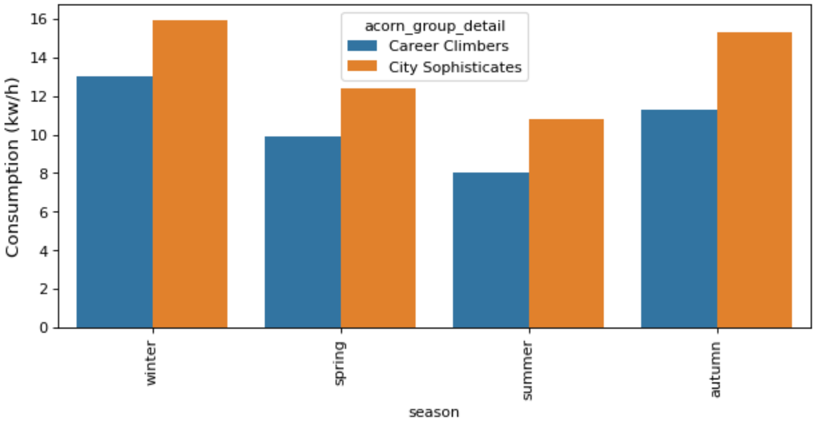


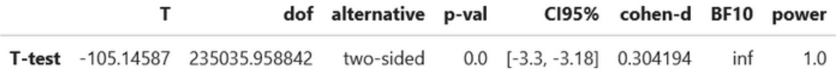
Figure 4.42: Average daily energy consumption for Rising Prosperity households per season

We can see that daily energy consumption varies with specific monthly changes within each year. In the first and last months of the year, the mean daily consumption of both groups is larger than in the rest of the months. It seems that the City Sophisticates group tends to have higher energy consumption than the Career Climbers.

The mean daily energy consumption by season shows that the described behavior can be related to the changes within each year through the seasons. Daily mean energy consumption also tends to vary across the different seasons, with a higher mean magnitude in Autumn and Winter in both groups. The violin plots show that the maximum daily energy consumption was in Autumn but in Winter the overall percentile values are the highest.

Finally, to test that there is a meaningful difference in the daily energy consumption between the two groups it is used a statistical test to quantify the difference between their arithmetic means. In this case, the t-test allows us to perform this comparison, having a null hypothesis that their means are equal and an $\alpha = 0.05$.

We can see that the p-value (0.0) $<$ α (0.05), so the null hypothesis can be rejected, concluding that there is a significant difference between the group's total energy consumption per day and their means.

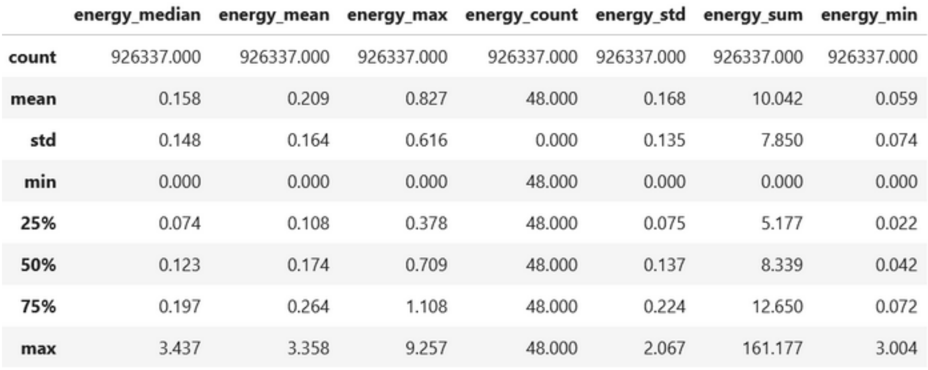


	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-105.14587	235035.958842	two-sided	0.0	[-3.3, -3.18]	0.304194	inf	1.0

Figure 4.43: t-test 1

4.2.3 Category 3: Comfortable Communities Households

The descriptive statistics for the Comfortable Communities ACORN category show that this category has almost 920 thousand observations, with 7 numerical columns. This category is composed of 5 groups, and the average daily energy consumption is 10.04 kWh.



	energy_median	energy_mean	energy_max	energy_count	energy_std	energy_sum	energy_min
count	926337.000	926337.000	926337.000	926337.000	926337.000	926337.000	926337.000
mean	0.158	0.209	0.827	48.000	0.168	10.042	0.059
std	0.148	0.164	0.616	0.000	0.135	7.850	0.074
min	0.000	0.000	0.000	48.000	0.000	0.000	0.000
25%	0.074	0.108	0.378	48.000	0.075	5.177	0.022
50%	0.123	0.174	0.709	48.000	0.137	8.339	0.042
75%	0.197	0.264	1.108	48.000	0.224	12.650	0.072
max	3.437	3.358	9.257	48.000	2.067	161.177	3.004

Figure 4.44: Comfortable Communities summary

To understand and see possible trends and differences between the groups of this category, the next figure shows a violin plot with the consumption per group and also a tree map showing the categories under analysis:

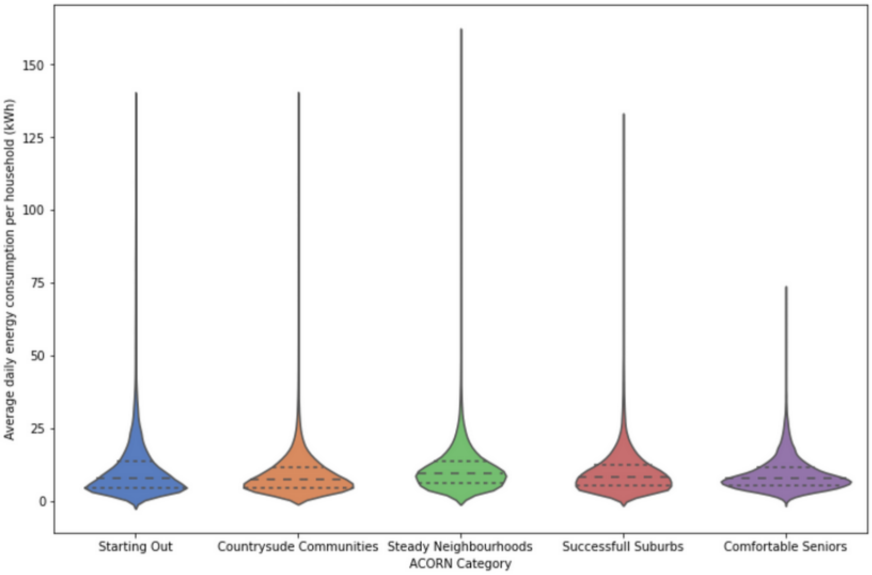


Figure 4.45: Daily energy consumption for Comfortable Communities households

It is possible to see a common behaviour relative to the average consumption, in general not higher than 25 Kw/h, although there are outliers in all the groups. Also, there is something interesting to observe and it is the distribution of the groups Steady Neighbourhoods and Successful Suburbs, because in contrast with the other groups (even versus the groups of other categories), they show a lower skewness, the concentration around the median is not as intensive as the other groups. According to this analysis, a t-test to contrast the means of the groups does not seem a good approach since the similar consumption pattern within groups.

Given the previous analysis and in line to find out if there are significant differences between the groups, a line plot to see the consumption across the time is presented below:

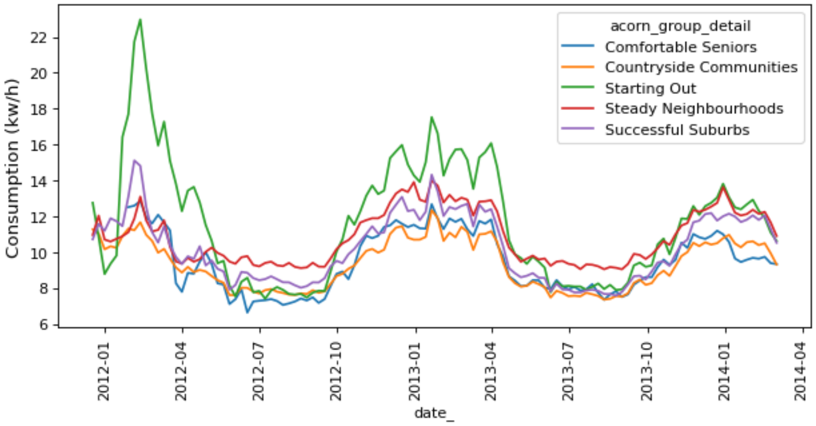


Figure 4.46: Average daily consumption for Comfortable Communities

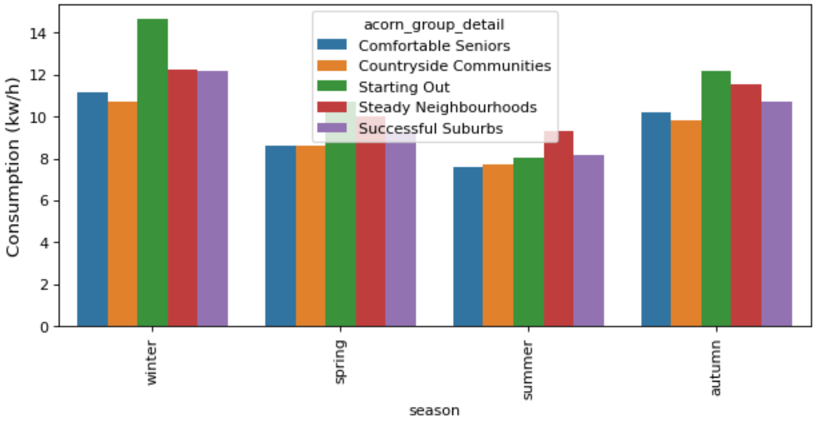


Figure 4.47: Average daily energy consumption for Comfortable Communities households per season

Once again, the consumption for all the groups seems to be really similar, except for the first months of the study (first quarter of 2012) when the chart shows a higher consumption for the green group (Starting Out). The trend is the same as usual, it is higher for the coldest months and lower for the hottest ones; but no matter the group, the amount of used energy has a common range from almost 7 Kw/h to 15 Kw/h.

Now checking the trends over the time, it really has the same behavior over the past 2 years, so the numbers match and it indicates we are going the right way with this stakeholders.

But to get further details about this segment, this is the estimation over the months, the hypothesis summer is likely to get lower numbers since it is not that necessary the warm up power. This is estimated with an possible error, and this proves the summer theory.

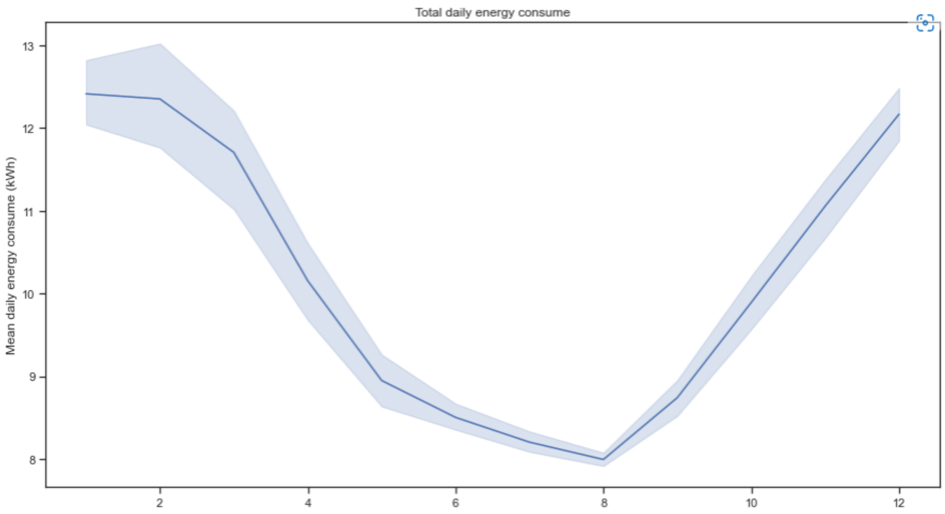
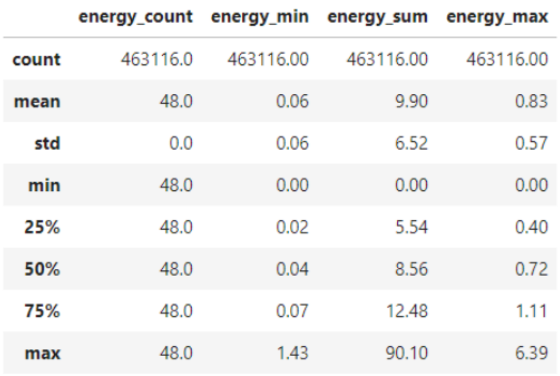


Figure 4.48: Total daily energy consumption for Comfortable Communities per month

4.2.4 Category 4: Financially Stretched Households

The descriptive statistics for the numerical data show that the columns on the daily dataset include many statistics generated from the Daily dataset. In this case, the data of the Financially Stretched category have around 460 thousand entries, with 9 numerical columns.



	energy_count	energy_min	energy_sum	energy_max
count	463116.0	463116.00	463116.00	463116.00
mean	48.0	0.06	9.90	0.83
std	0.0	0.06	6.52	0.57
min	48.0	0.00	0.00	0.00
25%	48.0	0.02	5.54	0.40
50%	48.0	0.04	8.56	0.72
75%	48.0	0.07	12.48	1.11
max	48.0	1.43	90.10	6.39

Figure 4.49: Financial Stretched summary

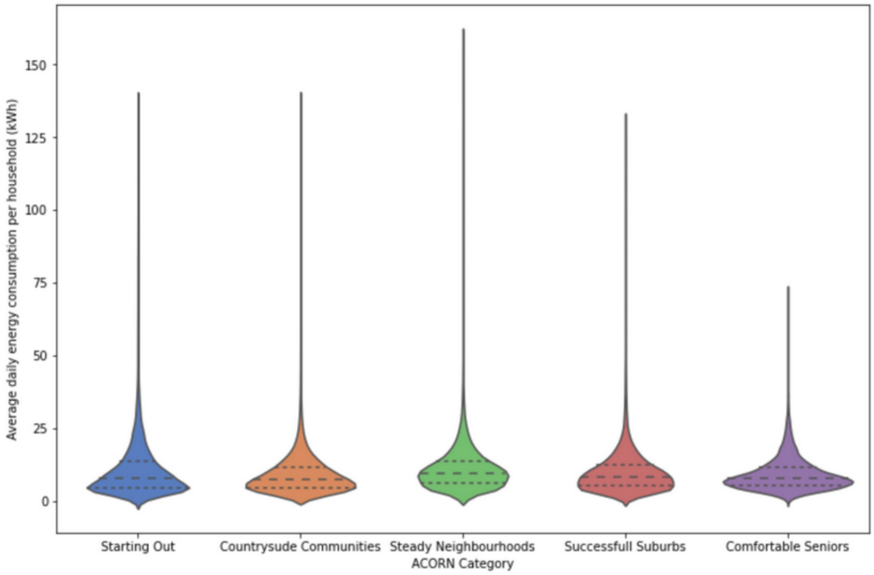


Figure 4.50: Daily energy consumption for Financial Stretched households

Financially stretched households, as well as all households considered, have an increase in energy consumption during the winter months and decrement in summer, registering the lowest daily energy consumption in August. The increment in the electric consumption could be explained by the use of electric heaters in the months when the lowest temperatures are recorded.

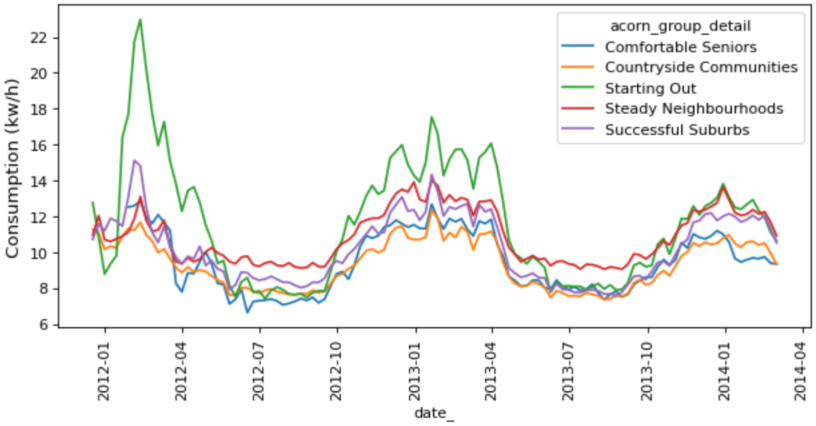
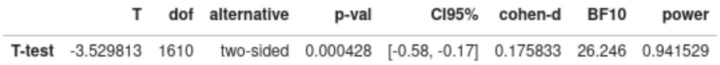


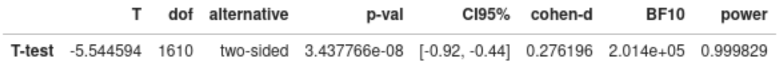
Figure 4.51: Average daily consumption for Financially Stretched

According to the violin graphic plotted below, there was explored the existence of a difference in the mean of the daily energy consume by financially stretched households in seasons. In the autumn and winter, there 25% of daily energy records are greater than 10 kWh. While 25% of daily energy collected in summer and spring is greater than 8 kWh.



	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-3.529813	1610	two-sided	0.000428	[-0.58, -0.17]	0.175833	26.246	0.941529

Figure 4.53: t-test 1



	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-5.544594	1610	two-sided	3.437766e-08	[-0.92, -0.44]	0.276196	2.014e+05	0.999829

Figure 4.54: t-test 2

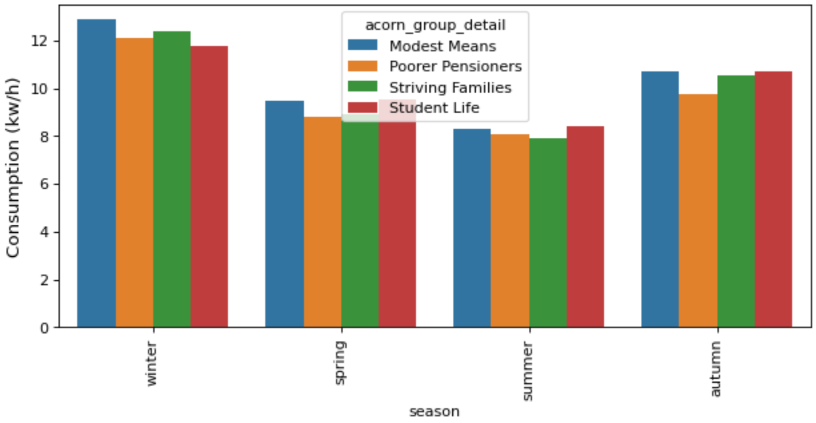


Figure 4.52: Average daily energy consumption for Financially Stretched households per season

To compare the means for the groups of the category, a t-test will be performed as follows:

Test 1: A subset of the data frame with the consumption per group is taken just with the energy sum (total consumption per household per day) and the group, Poorer Pensioner in this case, and the same is performed with the group Student Life, then the t-test is applied, and the result is the next:

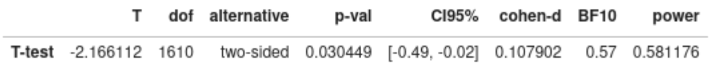
It's interesting to see the p-value, which is near 0, it means, with a 0.05 significance level that the null hypothesis must be rejected in favor of the alternative hypothesis, then it can be concluded that there is no statistical difference between the average consumption of the Poorest Pensioners group and the Student Life group.

Test 2: The same procedure is conducted again but now to compare Poorest Pensioners and Modest Means groups:

Once again, the result is to reject the null hypothesis, it is noticeable to see the confidence interval of the two tests and how close they are. The conclusion is as expected, there is not a statistical difference between the average consumption of the Poorest Pensioners group and the Modest Means group.

Test 3: Finally, the same method is used for the Poorest Pensioners and Strivings Families groups:

In this case, with the same 0.05 significance level, the p-value is near 0, the confidence interval includes negative and positive values, and there is no statistical difference between the average consumption of the Poorest Pensioners group and the Strivings Families group.

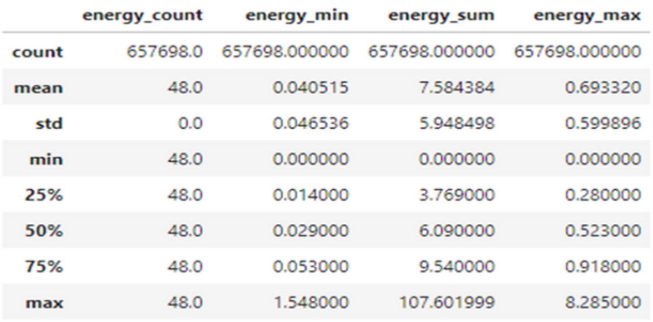


	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-2.166112	1610	two-sided	0.030449	[-0.49, -0.02]	0.107902	0.57	0.581176

Figure 4.55: t-test 3

4.2.5 Category 5: Urban Adversity

This category includes 3 subcategories, classified as: Acorn O= Young hardship, Acorn P= Struggling Estates, and Acorn Q= Difficult Circumstances.



	energy_count	energy_min	energy_sum	energy_max
count	657698.0	657698.000000	657698.000000	657698.000000
mean	48.0	0.040515	7.584384	0.693320
std	0.0	0.046536	5.948498	0.599896
min	48.0	0.000000	0.000000	0.000000
25%	48.0	0.014000	3.769000	0.280000
50%	48.0	0.029000	6.090000	0.523000
75%	48.0	0.053000	9.540000	0.918000
max	48.0	1.548000	107.601999	8.285000

Figure 4.56: Urban Adversity summary

The descriptive statistics show us that we have 48 counts, a record each half-hourly, while the 24 hours of the day. The mean consumption by day in this group is 7,58 kw/h and the standard deviation shows us that the data are scattered.

Now, we can see the information in a violin plot:

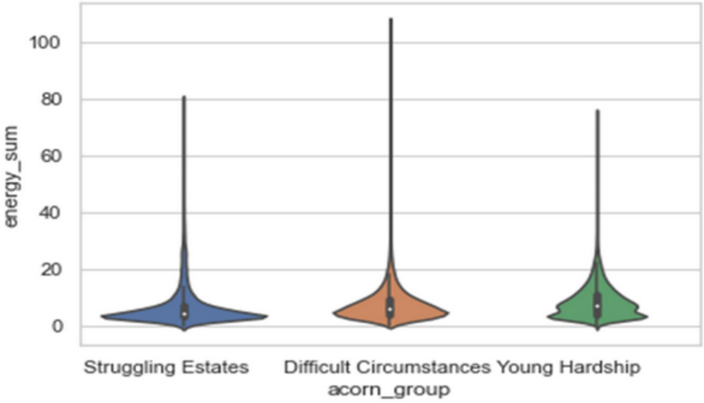


Figure 4.57: Daily energy consumption for Urban Adversity households

The plot shows the distribution in each one of the groups inside Urban Adversity. The median seems to be similar and the major difference is in the maximum values. Difficult circumstances show more consumption than Struggling Estates and the least more consumption than Young Hardship.

We're going to see the behavior of the daily consumption of energy through time.

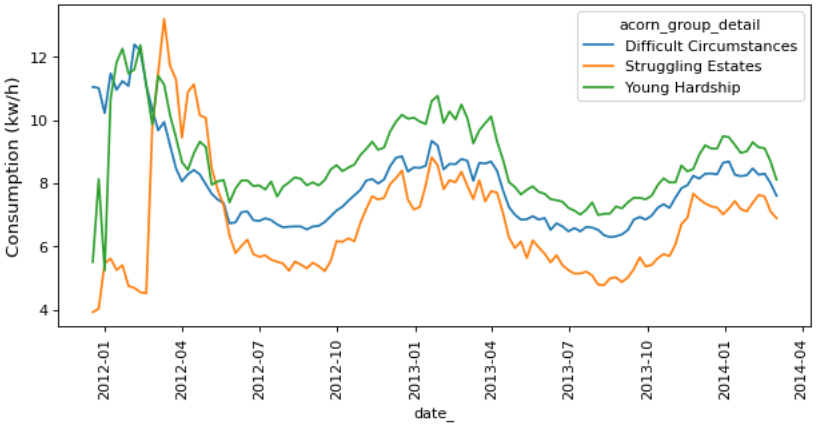


Figure 4.58: Average daily consumption for Urban Adversity

The demand for energy shows a deep since 2012 and their behavior is seasonal.

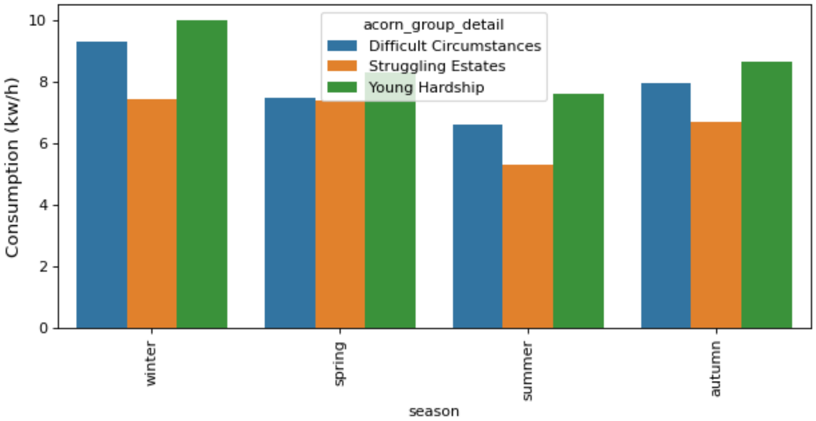
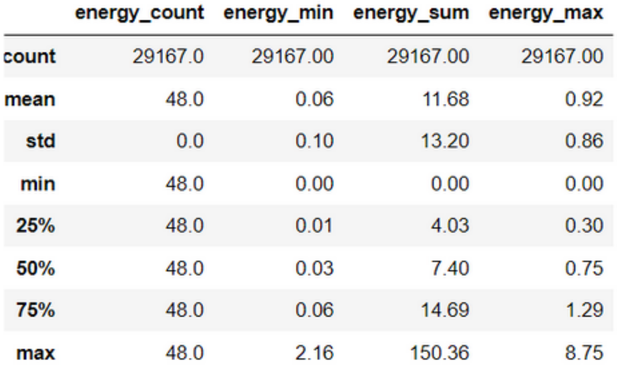


Figure 4.59: Average daily energy consumption for Urban Adversity households per season

We can see that Winter is the season when the energy demand is the highest, followed by autumn, spring and summer show the lowest consumption.

4.2.6 Category 6: Not Private Households

In the category "not private households", all communal, business, and non-residential areas are grouped. During the development of the study, this group was considered to collect their daily and half-hourly energy consumption. Below, you will find a summary chart with the basic statistics of energy consumption.



	energy_count	energy_min	energy_sum	energy_max
count	29167.0	29167.00	29167.00	29167.00
mean	48.0	0.06	11.68	0.92
std	0.0	0.10	13.20	0.86
min	48.0	0.00	0.00	0.00
25%	48.0	0.01	4.03	0.30
50%	48.0	0.03	7.40	0.75
75%	48.0	0.06	14.69	1.29
max	48.0	2.16	150.36	8.75

Figure 4.60: Not private households summary

The following graphic shows the historical behavior of daily energy consumption in the group "not private households" during the time the study was conducted.

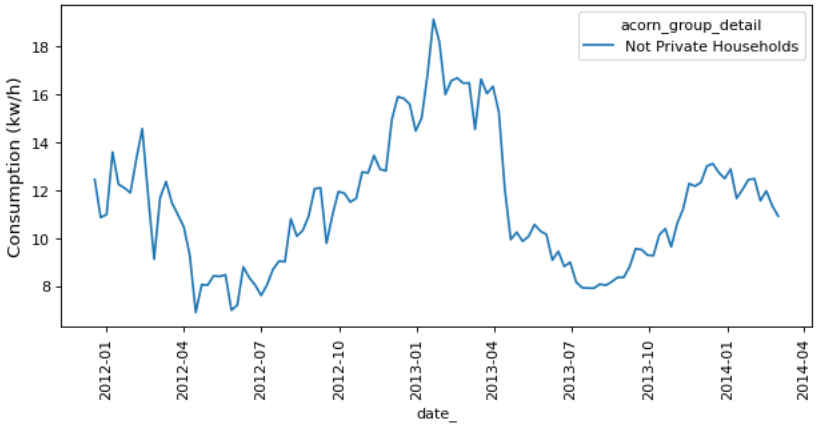


Figure 4.61: Average daily consumption for Not Private Households

Not private households, as well as all households considered, have an increase in energy consumption during the winter months and a decrease in summer, registering the lowest daily energy consumption in July. The increment in the electric consumption could be explained by the use of electric heaters in the months when the lowest temperatures are recorded.

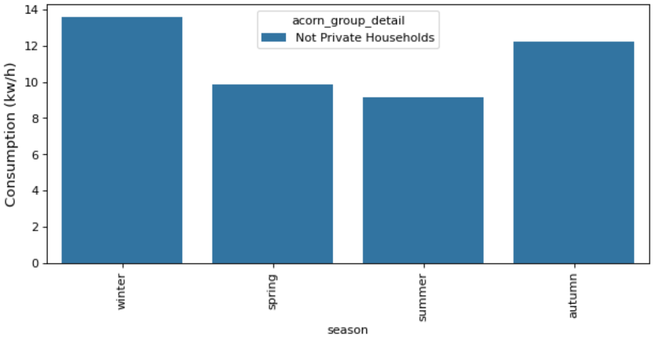


Figure 4.62: Average daily energy consumption for Not Private households per season

According to the violin graphic plotted below, there was explored the existence of difference in the mean of the daily energy consume by not private households in seasons. In the autumn and winter, there 25% of daily energy records are greater than 17 kWh. While 25% of daily energy collected in summer and spring is greater than 14 kWh.

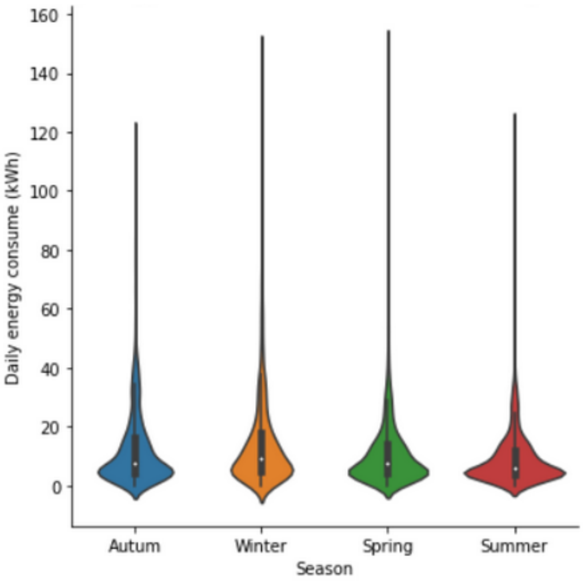
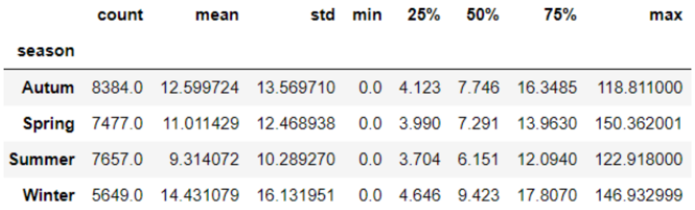


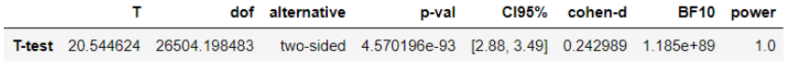
Figure 4.63: Energy consumption for Not Private households per season



	count	mean	std	min	25%	50%	75%	max
season								
Autum	8384.0	12.599724	13.569710	0.0	4.123	7.746	16.3485	118.811000
Spring	7477.0	11.011429	12.468938	0.0	3.990	7.291	13.9630	150.362001
Summer	7657.0	9.314072	10.289270	0.0	3.704	6.151	12.0940	122.918000
Winter	5649.0	14.431079	16.131951	0.0	4.646	9.423	17.8070	146.932999

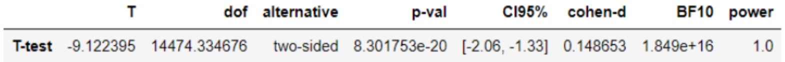
Figure 4.64: Energy consumption summary for Not Private households per season

To determine if there is a statistically significant difference between non-private households' daily



	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	20.544624	26504.198483	two-sided	4.570196e-93	[2.88, 3.49]	0.242989	1.185e+89	1.0

Figure 4.65: t-test mean difference between first and second half of the year



	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-9.122395	14474.334676	two-sided	8.301753e-20	[-2.06, -1.33]	0.148653	1.849e+16	1.0

Figure 4.66: t-test mean difference between spring and summer

energy consumption in the first and second half of the year, a T-test was conducted with a significant level of 5% obtaining a p-value of 4.57e-93. This result proves that there is a difference between the means of both periods.

Exploring the difference in the seasons of the first half, spring and summer, an additional mean difference T-test was conducted with a significant level of 5 %. The obtained p-value is lower than the significant level, then the nule hypothesis of mean equality is rejected proving than the mean daily consumption of spring and summer are different.

Finally, the difference between the means of consumption of the autumn and winter seasons was verified with a significance level of 5%. Below are the results of the T-test developed.

5 MODELING

5.1 Implementation of Prophet forecasting model

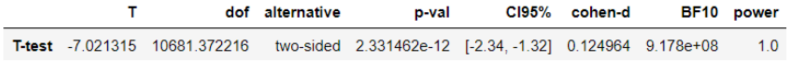
The Prophet forecasting model was selected to model the behavior of the daily consumption time series of the London households, because it shows the better mean absolute percentage error (MAPE) in comparison with the rest of the tested models, with MAPE = 0.98%. Additionally, this model considers the existence of differences between workdays and holidays, which was one of the insights discovered during the EDA phase of the project. As well as the existence of other frequencies in the seasonality, monthly and yearly. In the table below, you will find a comparison of the MAPE metric for each model, according to their cross validation:

Forecast Model	MAPE (Mean Absolute Percentage Error)
Prophet	0.98%
Exponential Smoothing	6.27%
SARIMAX	39.51%

Table 3: MAPE results

The modeling process can be divided into three main steps: data preparation, hyper parameter tuning and fitting of the model, and cross-validation and forecasting.

In this case, the model was implemented using the aggregated daily energy consumption data and the national UK holidays data. For the hyper parameter tuning and the cross-validation, the dataset was automatically split into training and testing periods on a rolling basis, according to a defined train



	T	dof	alternative	p-val	CI95%	cohen-d	BF10	power
T-test	-7.021315	10681.372216	two-sided	2.331462e-12	[-2.34, -1.32]	0.124964	9.178e+08	1.0

Figure 4.67: t-test mean difference between autumn and winter

period and a forecasting horizon, which were established as 540 and 180 days. For that reason, the data used to perform the forecasting later will be included into the training set, since random samples cannot be used in time series.

Fitting the model is a very straightforward process but some key hyper parameters were adjusted to optimize the model performance. We perform an iterative process to select which of all the hyper parameters were most likely to be tuned by comparing the MAPE obtained by adjusting each individual hyper parameter with a baseline MAPE with a standard fitted model. The most relevant hyper parameters were the type of trend, its flexibility or the seasonality, and its strength, so its values were optimized using the grid search method.

After the hyper parameter tuning and the after cross-validation we obtained the best performing model, which exhibits a MAPE of 1.357%. This model will be used for forecasting and comparison with the other time series models.

The figure 5.1 shows the behavior of the mean daily household energy consumption in London (black dots), and the fitted series obtained with the model with the best performance (blue line) with its confidence interval (light blue). This figure also includes the forecasting of the next 90 days, with a confidence interval of 95%.

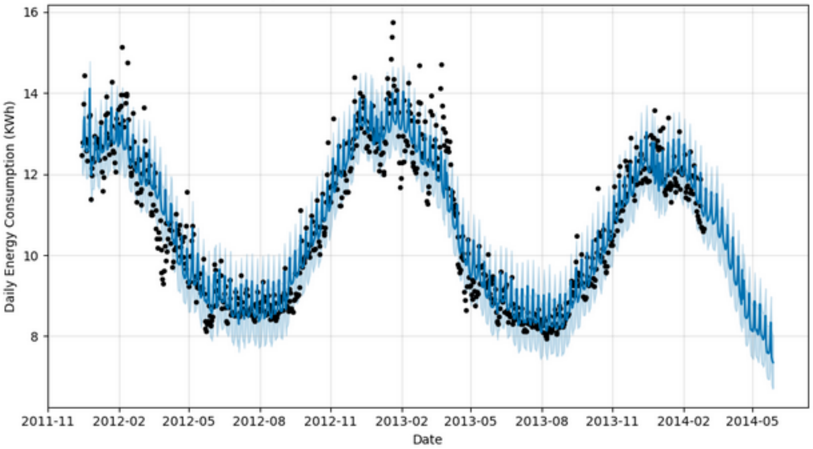


Figure 5.1: Forecasting for next 90 days

The Prophet model represents the overall seasonality changes, in each time scale. However, there are some outliers that the model cannot accurately take into account, so these anomalies affect the predictions and the model performance.

Also, the prophet package allows users to see the forecast components, showing the trend, yearly seasonality, holidays, and weekly seasonality of the time series, wich is on the figure 5.2.

The component graph shows us the distinct behavior that energy consumption has on different time scales, with a clear downward yearly trend, and some information on weekly and monthly patterns. We can see that the effect of energy consumption by day of the week is very clear, increasing its levels on Mondays and weekends. There is also an intra-annual trend, where at the beginning and the end of the year the energy consumption rises, most likely due to the weather and the low temperatures.

Cross-validation with historical data was done to evaluate the model performance and measure the forecast error. This procedure can be done automatically using the function implemented on the package, obtaining a data frame with the metrics computed for the prediction performance. The low values of the MAPE indicate that we can explain some of the variability of the household energy consumption behavior, considering the different seasonal effects that have been described by the data.

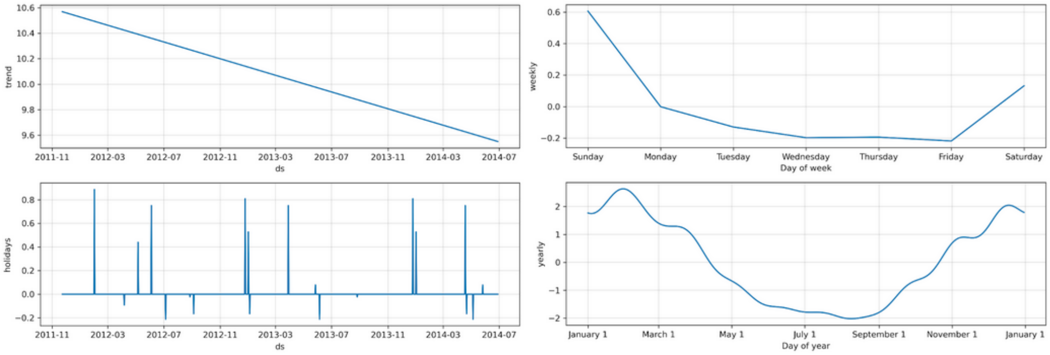


Figure 5.2: Forecasting components

5.2 Prophet model by Category

For a more in-depth analysis, the same procedure was applied to the aggregated data by ACORN categories, obtaining the corresponding metrics and forecast. This gave us insights of the behavior that the daily energy consumption has across the distinct ACORN groups and its impact on the performance of the model. Some of the fitted models are presented to compare their predicted values to the observations.

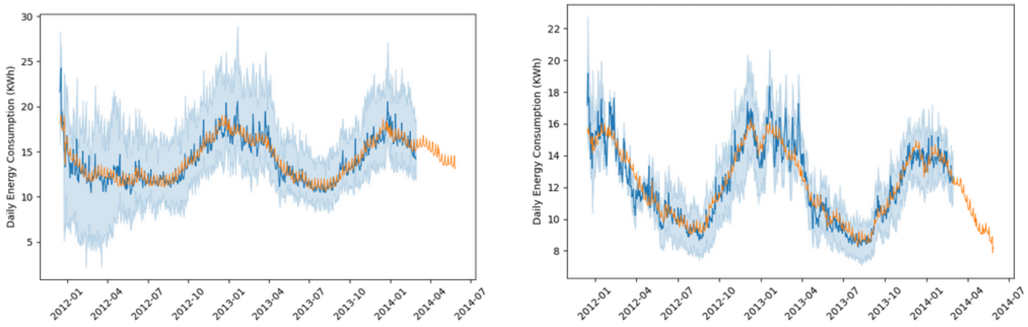


Figure 5.3: Forecast for Affluent Achievers (left) and Rising Prosperity (right)

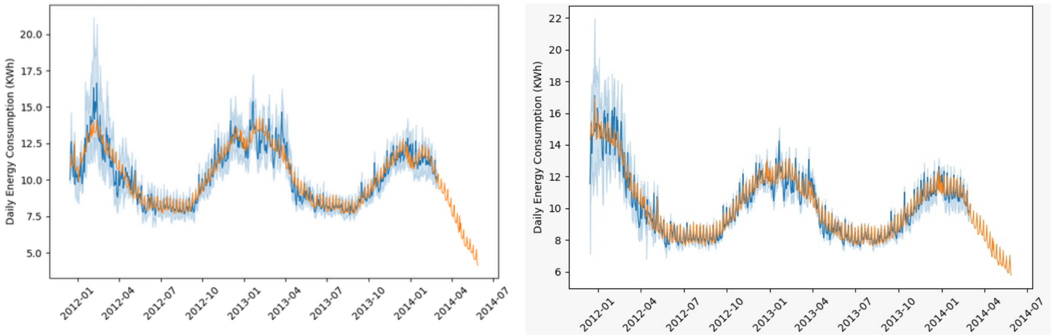


Figure 5.4: Forecast for Comfortable Communities (left) and Financial Stretched (right)

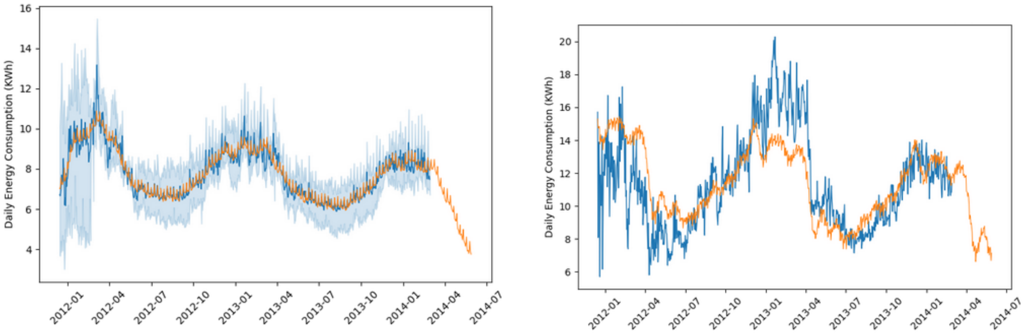
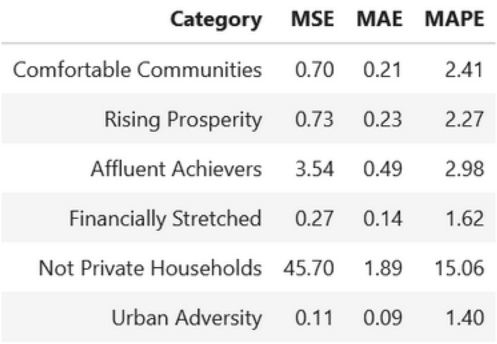


Figure 5.5: Forecast for Urban Adversity (left) and Not Private Households (right)

The described process of fitting the model was performed to the dataset of each category, obtaining the following metrics:



Category	MSE	MAE	MAPE
Comfortable Communities	0.70	0.21	2.41
Rising Prosperity	0.73	0.23	2.27
Affluent Achievers	3.54	0.49	2.98
Financially Stretched	0.27	0.14	1.62
Not Private Households	45.70	1.89	15.06
Urban Adversity	0.11	0.09	1.40

Figure 5.6: Error summary

Both metrics and the plots show us a generally good response of the model across the different categories, with a MAPE in a range of 1% - 2.5%. However, in particular the Not Private Households category shows a poor performance due to the high variation across the period, it makes harder to take the accuracy of the predictions.

It's possible to see that the energy demand will increase in the upcoming years, and die to the average energy growth demand will increase at the seasons stands approximately 4% (according to the prophet model), compared with previous years and it won't be significantly different, so the number of departments, categories and commercial growth over the median will be more.

Finally, computing the variable's importance for doing the classification in the model, we identified that the most important variables were: season, and population. We'd like to clarify that all the information was summarized just to have a general overview and take to most of the performance out of the model, to clear the bigger picture, and to stay tuned with the changes. It was also summed up to prevent the model to be over fitted.

6 AWS Database

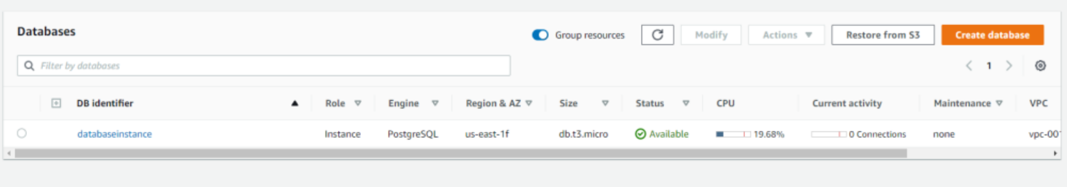
The server that provides the information was uploaded to a Postgres instance in Amazon Relational Database Service (RDS). The connection can be made using the following URL:

postgresql://london:as3fgd6@databaseinstance.c8611n47i7sr.us-east-1.rds.amazonaws.com:5432/database_ds4a

The link of the instance is: <http://instanceds4a.c8611n47i7sr.us-east-1.rds.amazonaws.com/> and the details are next:

Database	database_ds4a
User	london
Password	as3fgd6
Instance	databaseinstance.c8611n47i7sr.us-east-1.rds.amazonaws.com
Port	5432

Table 4: Database details

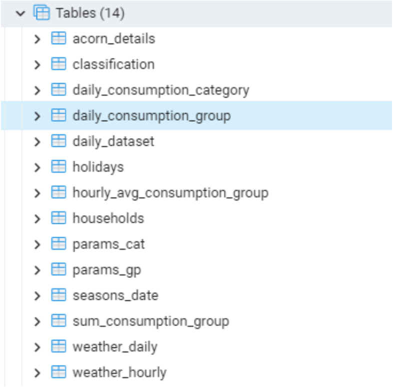


	DB identifier	Role	Engine	Region & AZ	Size	Status	CPU	Current activity	Maintenance	VPC
○	databaseinstance	Instance	PostgreSQL	us-east-1f	db.t3.micro	Available	19.68%	0 Connections	none	vpc-00

Figure 6.1: Database

For the correct operation of the dashboard, the instance must be turned on. Due to the storage limits of the instance, it was necessary to update the database tables with the information transformed into the database. In the staging scheme.

The database is made up of:

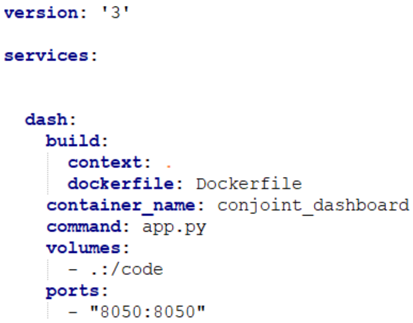


Tables (14)
- acorn_details
- classification
- daily_consumption_category
- daily_consumption_group
- daily_dataset
- holidays
- hourly_avg_consumption_group
- households
- params_cat
- params_gp
- seasons_date
- sum_consumption_group
- weather_daily
- weather_hourly

Figure 6.2: Tables

6.1 Deployment

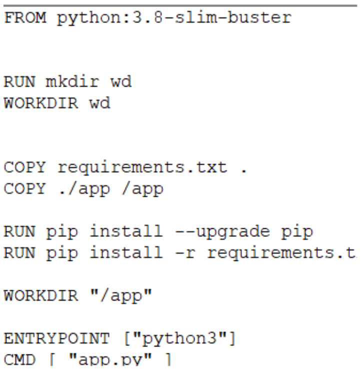
Once the application is finished, it is necessary to use Docker to create an image and from this its respective container. To create the image, and container and run the app, it was done using docker-compose, which instantiates the name of the Docker file, the name of the container and the application to open once the container is deployed. As well as instantiating port 8050. As follows:



```
version: '3'

services:


  dash:
    build:
      context: .
      dockerfile: Dockerfile
    container_name: conjoint_dashboard
    command: app.py
    volumes:
      - .:/code
    ports:
      - "8050:8050"
```

Figure 6.3: Docker file

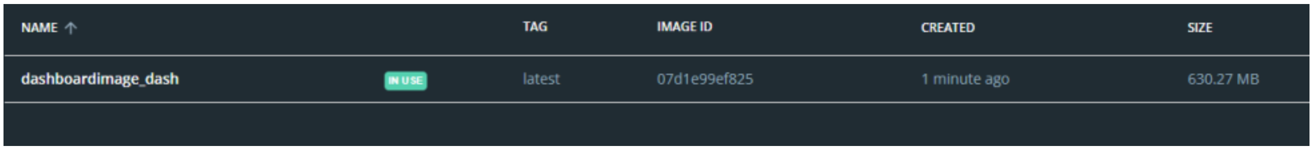
The package from which the libraries will be obtained is instantiated in the Docker file. Then, on the container server, the mkdir path is opened, the requirements and the files that make up the application are copied to the app folder. After doing the installation and update of pip for the installation of packages contained in the text file requirements. Finally, in the container in the app folder, the Python file app.py is opened, as follows:



```
FROM python:3.8-slim-buster


RUN mkdir wd
WORKDIR wd


COPY requirements.txt .
COPY ./app /app

RUN pip install --upgrade pip
RUN pip install -r requirements.t

WORKDIR "/app"

ENTRYPOINT ["python3"]
CMD [ "app.py" ]
```

Figure 6.4: Container

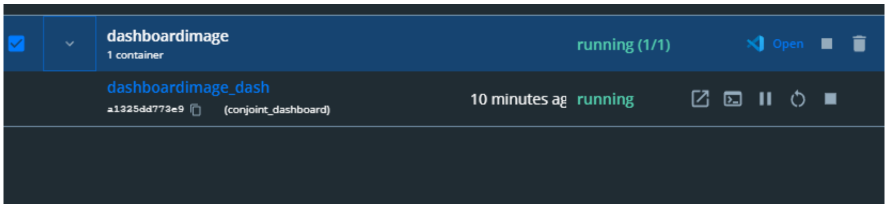
Once docker-compose is run, the image and container are created, and the app is run. In Docker Desktop the image is displayed like this:



NAME ↑		TAG	IMAGE ID	CREATED	SIZE
dashboardimage_dash	IN USE	latest	07d1e99ef825	1 minute ago	630.27 MB

Figure 6.5: Docker desktop

Docker Desktop shows the creation of the container and the evidence that it runs.



dashboardimage
1 container
running (1/1)

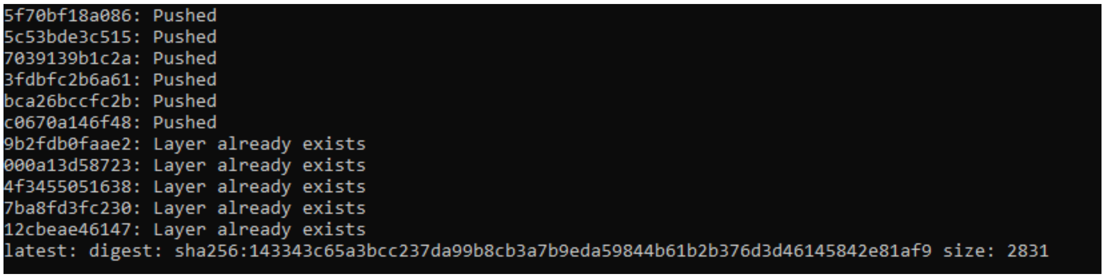
dashboardimage_dash
a1325dd773e9 (conjoint_dashboard)
10 minutes ag running

Figure 6.6: Dashboard running

The connection to the user who has the resources to use is proceeded through the AWS Management Console:

- Elastic Container Registry: Amazon repository where the code that was uploaded to the container is uploaded.
- Elastic Container Service: Amazon service where the cluster is created where the service (server) will be located that will allow the visualization of the dashboard in any part of the world.

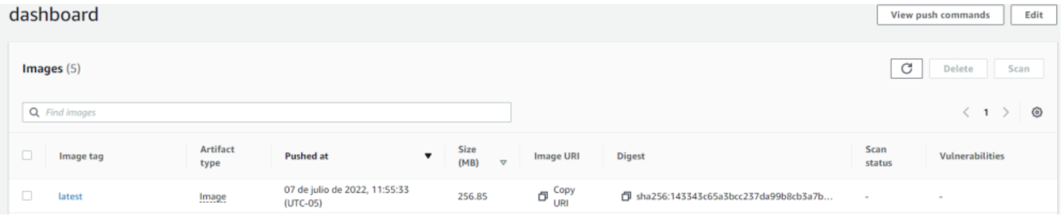
The connection to the user from the AWS management Console is necessary to have the access key id and secret access key available, once the connection is created, it is started in the Elastic Container Registry to then tag the container with the path of the repository previously created in AWS. Finally, the image is uploaded to the repository, as follows:



```
5f70bf18a086: Pushed
5c53bde3c515: Pushed
7039139b1c2a: Pushed
3fdbfc2b6a61: Pushed
bca26bccfc2b: Pushed
c0670a146f48: Pushed
9b2fdb0faae2: Layer already exists
000a13d58723: Layer already exists
4f3455051638: Layer already exists
7ba8fd3fc230: Layer already exists
12cbeae46147: Layer already exists
latest: digest: sha256:143343c65a3bcc237da99b8cb3a7b9eda59844b61b2b376d3d46145842e81af9 size: 2831
```

Figure 6.7: Image

On AWS it looks like this:



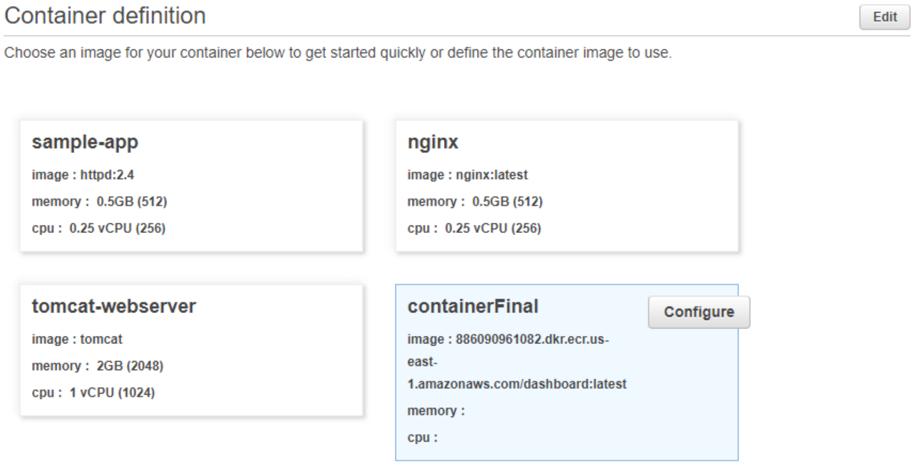
dashboard

Images (5)

□	Image tag	Artifact type	Pushed at	Size (MB)	Image URI	Digest	Scan status	Vulnerabilities
□	latest	Image	07 de julio de 2022, 11:55:33 (UTC-05)	256.85	Copy URI	sha256:143343c65a3bcc237da99b8cb3a7b...	-	-

Figure 6.8: ASW view

In Elastic Container Registry, a cluster is created where the definition of the container in AWS is specified:



Container definition

Choose an image for your container below to get started quickly or define the container image to use.

sample-app
image : httpd:2.4
memory : 0.5GB (512)
cpu : 0.25 vCPU (256)

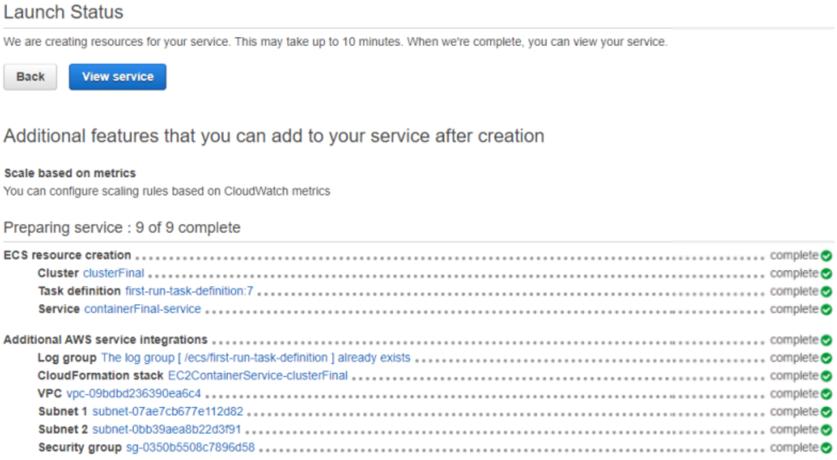
nginx
image : nginx:latest
memory : 0.5GB (512)
cpu : 0.25 vCPU (256)

tomcat-webserver
image : tomcat
memory : 2GB (2048)
cpu : 1 vCPU (1024)

containerFinal
image : 886090961082.dkr.ecr.us-east-1.amazonaws.com/dashboard:latest
memory :
cpu :

Figure 6.9: Container definition

Finally, the deployment is done, giving a successful status:



Launch Status

We are creating resources for your service. This may take up to 10 minutes. When we're complete, you can view your service.

Additional features that you can add to your service after creation

Scale based on metrics
You can configure scaling rules based on CloudWatch metrics

Preparing service : 9 of 9 complete

```
ECS resource creation ........................................ complete ✓
    Cluster clusterFinal ..................................... complete ✓
    Task definition first-run-task-definition:7 .............. complete ✓
    Service containerFinal-service ........................... complete ✓

Additional AWS service integrations .......................... complete ✓
    Log group The log group [ /ecs/first-run-task-definition ] already exists ... complete ✓
    CloudFormation stack EC2ContainerService-clusterFinal .... complete ✓
    VPC vpc-09bdbd236390ea6c4 ................................ complete ✓
    Subnet 1 subnet-07ae7cb677e112d82 ........................ complete ✓
    Subnet 2 subnet-0bb39aea8b22d3f91 ........................ complete ✓
    Security group sg-0350b5508c7896d58 ...................... complete ✓
```

Figure 6.10: Application running on the public IP

7 DASHBOARD

The dashboard and backend implementation was driven by our talented team with prior knowledge on programming and SQL skills, they took care of everything in further detail. The backend is a set of web services (API) that allow the front-end to access the data to build and present charts. The instance is stored in an AWS which stands for Amazon Web Services cloud, all the tables mentioned before are a PostgreSQL with summarized data to make it more powerful and quicker to be processed. Less time more information, quicker insights, and smarter decisions. The Dashboard can be consulted by clicking here <http://54.147.102.232:8050/>. It was built with the open-source tool named Dash which uses the Python program to execute, carry out and show all the different plots set up by our team.

7.1 About the tool

SmartEnerx is a visualization tool designed to provide a graphical representation of the historical daily energy consumption of households in London from November 2011 to February 2014. This application is very flexible and allows the user to display information about various ACORN categories and groups, within a specific period or season of interest. SmartEnerx integrates a high-accuracy forecasting model that can be used to predict the daily energy consumption of one category and its groups of interest, in a defined forecast horizon. It gives the user the possibility to compare the performance of different models by changing the values of the most relevant hyperparameters.

Overview of the application and layout
This section gives a brief description of the Overview and Application Layout

7.2 Header and top navigation bar

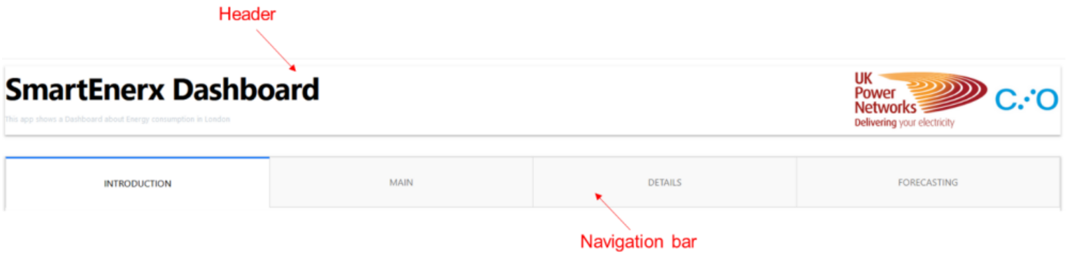


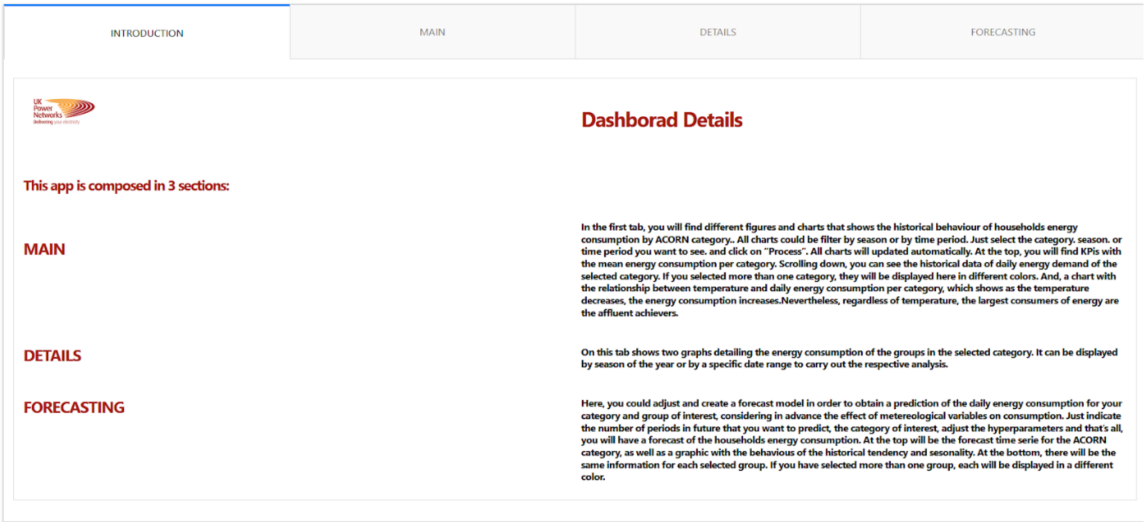
Figure 7.1: Navigation bar

The application header includes the corresponding title, its main purpose, and the main institutions related to its development. The navigation tabs below the header can be used to navigate across the different tabs or panels of the application, changing the content area and its corresponding filters.

SmartEnerx is composed of four panels: the introduction to the tool, the main dashboard panel, a panel with specific details of the household energy consumption by the ACORN group, and the forecasting panel.

7.3 Introduction tab

The introduction tab includes a brief description of the content and functionalities of each tab.



Dashborad Details

This app is composed in 3 sections:

MAIN

In the first tab, you will find different figures and charts that shows the historical behaviour of households energy consumption by ACORN category.. All charts could be filter by season or by time period. Just select the category. season. or time period you want to see. and click on "Process". All charts will updated automatically. At the top, you will find KPis with the mean energy consumption per category. Scrolling down, you can see the historical data of daily energy demand of the selected category. If you selected more than one category, they will be displayed here in different colors. And, a chart with the relationship between temperature and daily energy consumption per category, which shows as the temperature decreases, the energy consumption increases.Nevertheless, regardless of temperature, the largest consumers of energy are the affluent achievers.

DETAILS

On this tab shows two graphs detailing the energy consumption of the groups in the selected category. It can be displayed by season of the year or by a specific date range to carry out the respective analysis.

FORECASTING

Here, you could adjust and create a forecast model in order to obtain a prediction of the daily energy consumption for your category and group of interest, considering in advance the effect of metereological variables on consumption. Just indicate the number of periods in future that you want to predict, the category of interest, adjust the hyperparameters and that's all, you will have a forecast of the households energy consumption. At the top will be the forecast time serie for the ACORN category, as well as a graphic with the behavious of the historical tendency and sesonality. At the bottom, there will be the same information for each selected group. If you have selected more than one group, each will be displayed in a different color.

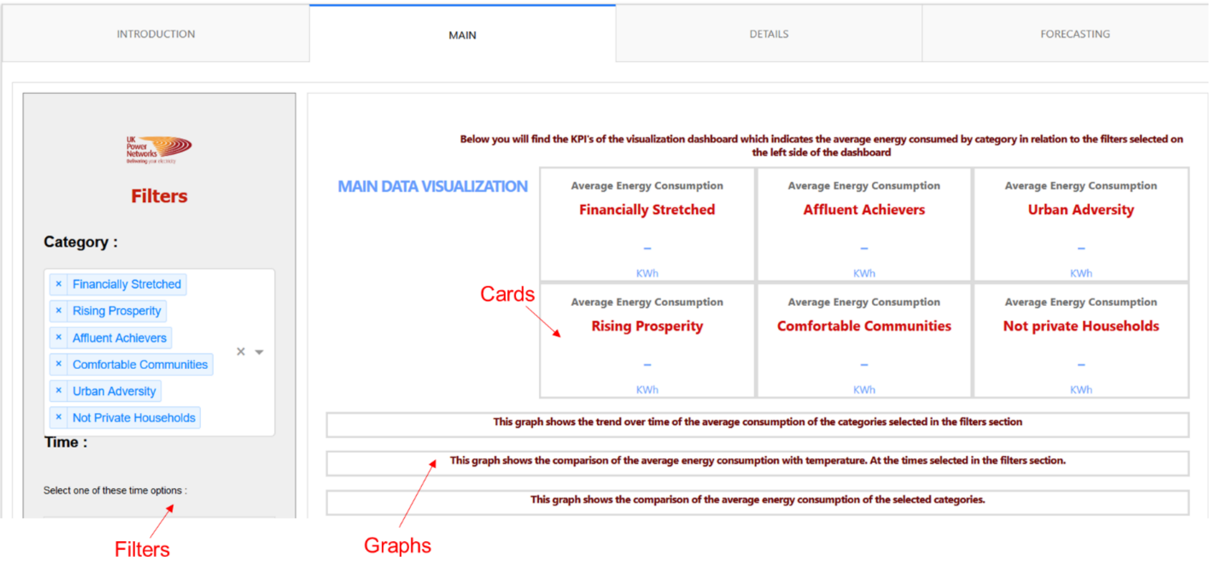
Figure 7.2: Dashboard - INTRODUCTION

This tab is loaded by default by the application to give the user a piece of additional information on how to use it and what to expect from each tab.

7.4 Main tab

On the main tab, there are three main elements:

1. Filters: this allows the user to display the information by the categories and period of interest.
2. Cards: display the most important Key Performance Indicators (KPIs) of the household energy consumption by category.
3. Graphs: display the filtered information in three graphs:
    (a) Trend over time of the daily average consumption of the selected categories.
    (b) Comparison of the average daily energy consumption with the mean temperature.
    (c) Comparison of the mean daily energy consumption between the selected categories on the period.



Below you will find the KPI's of the visualization dashboard which indicates the average energy consumed by category in relation to the filters selected on the left side of the dashboard

This graph shows the trend over time of the average consumption of the categories selected in the filters section

This graph shows the comparison of the average energy consumption with temperature. At the times selected in the filters section.

This graph shows the comparison of the average energy consumption of the selected categories.

Figure 7.3: Dashboard - MAIN

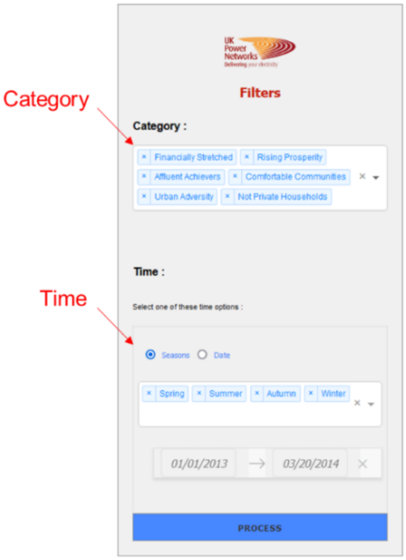
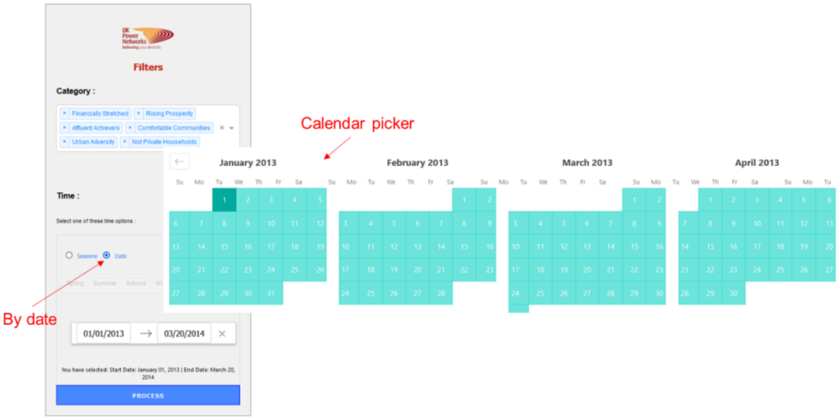


Figure 7.4: Dashboard - MAIN Filters

The filter options are divided into two main elements:

1. Category: where the user selects the categories of interest
2. Time: which can be filtered by season or by date:
   (a) By season the user is allowed to select the corresponding seasons of interest.
   (b) By date, a calendar picker is displayed, and the user can select between two specific dates.



Category :

× Financially Stretched × Rising Prosperity × Affluent Achievers × Comfortable Communities × Urban Adversity × Not Private Households

Time :

Select one of these time options :

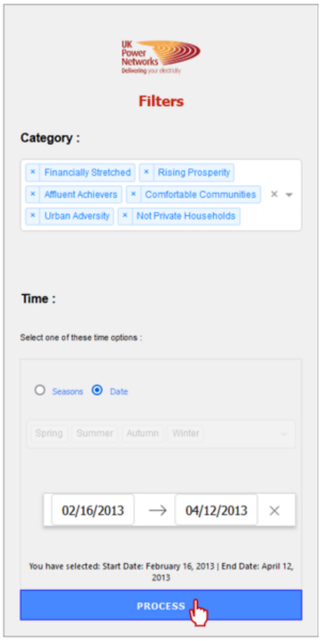
○ Seasons ● Date

01/01/2013 → 03/20/2014

You have selected: Start Date: January 01, 2013 | End Date: March 20, 2014

Figure 7.5: Dashboard - Filters

Finally, to show the graphs the application needs user authorization, thus is necessary that the process button is clicked.



Category :

× Financially Stretched × Rising Prosperity × Affluent Achievers × Comfortable Communities × Urban Adversity × Not Private Households

Time :

Select one of these time options :

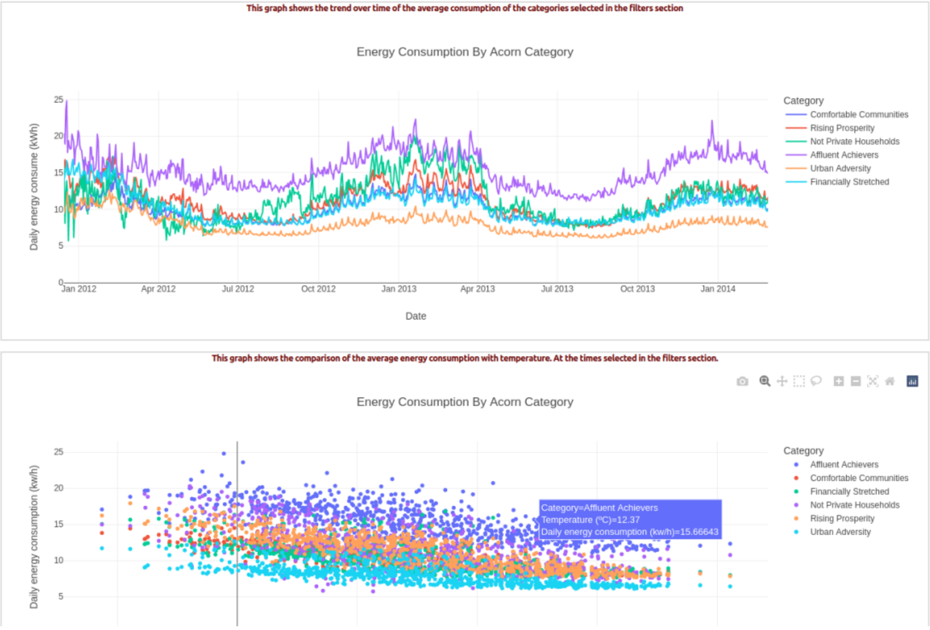
○ Seasons ● Date

02/16/2013 → 04/12/2013

You have selected: Start Date: February 16, 2013 | End Date: April 12, 2013

Figure 7.6: Dashboard - Main Process

Some of the output graphs displayed on the tab:



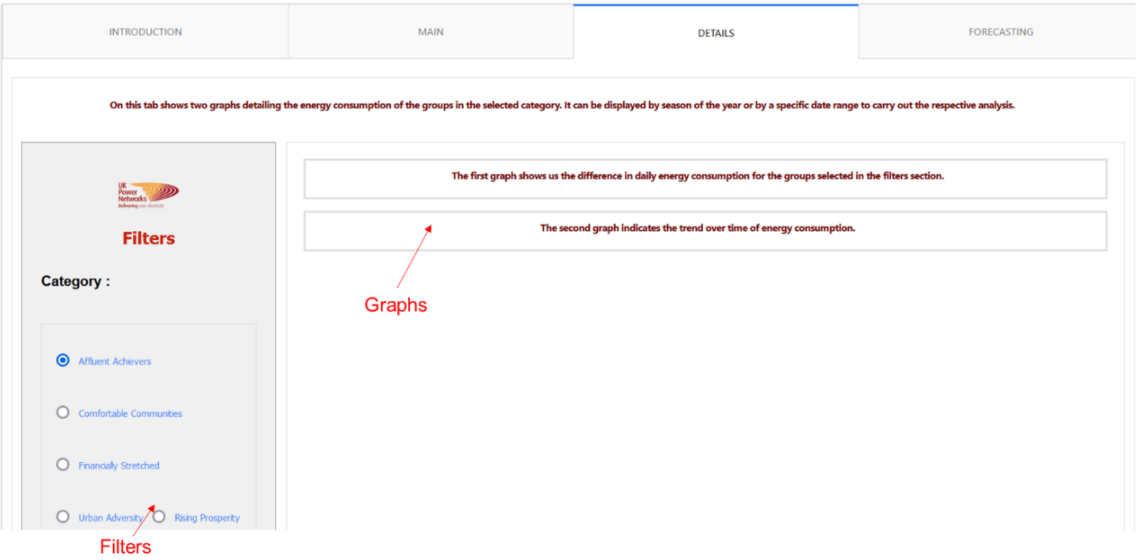
This graph shows the trend over time of the average consumption of the categories selected in the filters section

This graph shows the comparison of the average energy consumption with temperature. At the times selected in the filters section.

Figure 7.7: Dashboard - MAIN

7.5 Details tab

On the detail tab, two graphs are displayed with the detailed behavior of the daily energy consumption by each group. Although the structure is like the main tab, the filters and graphs are different. In this case for each ACORN category, its corresponding groups are available to select.



On this tab shows two graphs detailing the energy consumption of the groups in the selected category. It can be displayed by season of the year or by a specific date range to carry out the respective analysis.

Filters

Category :

● Affluent Achievers

○ Comfortable Communities

○ Financialy Stretched

○ Urban Adversity ○ Rising Prosperity

The first graph shows us the difference in daily energy consumption for the groups selected in the filters section.

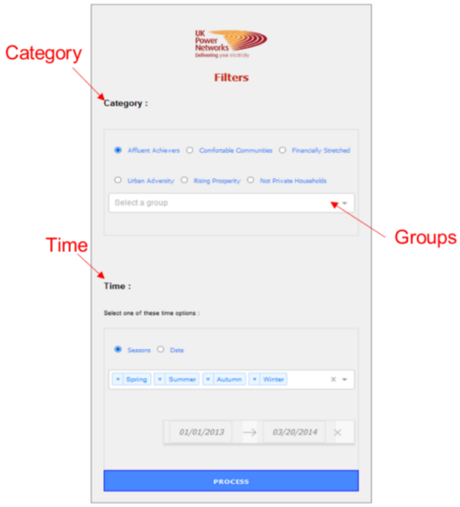
The second graph indicates the trend over time of energy consumption.

Figure 7.8: Dashboard - DETAILS

1. Filters: to select the groups and period of interest.
2. Graphs: display the filtered information in two graphs:

(a) Difference in daily energy consumption of the groups

(b) Daily trend of the energy consumption by group



Category :

● Affluent Achievers ○ Comfortable Communities ○ Financially Stretched

○ Urban Adversity ○ Rising Prosperity ○ Not Private Households

Select a group

Time :

Select one of these time options :

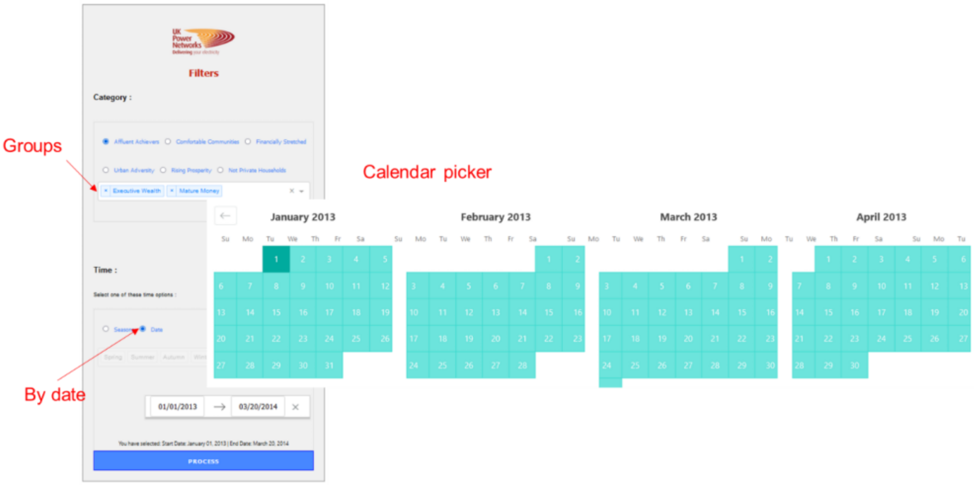
● Seasons ○ Date

× Spring × Summer × Autumn × Winter

01/01/2013 → 03/20/2014

Figure 7.9: Dashboard - Filter DETAILS

In this case, the user can select between the different groups that belong to each ACORN category. The time selection works like the previous tab, with the opportunity to select between seasons of interest or specific dates.



Category :

● Affluent Achievers ○ Comfortable Communities ○ Financially Stretched

○ Urban Adversity ○ Rising Prosperity ○ Not Private Households

× Executive Wealth × Mature Money

Time :

Select one of these time options :

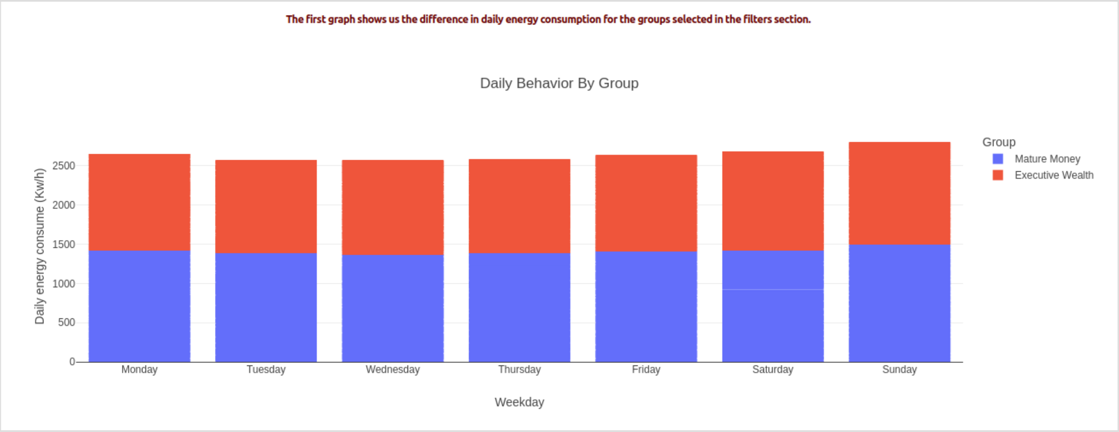
○ Seasons ● Date

01/01/2013 → 03/20/2014

You have selected: Start Date: January 01, 2013 | End Date: March 20, 2014

Figure 7.10: Dashboard - Calendar picker

Similarly, to the previous tab, it is necessary to click on the process button to display the graphs. Some of the output graphs displayed on the tab:

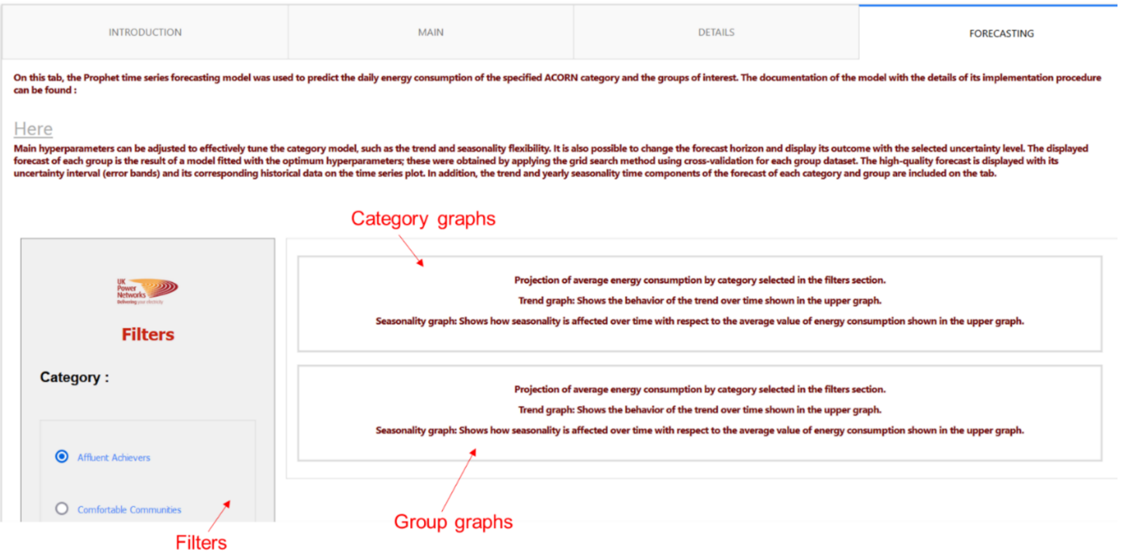


The first graph shows us the difference in daily energy consumption for the groups selected in the filters section.

Figure 7.11: Dashboard - DETAILS

7.6 Forecasting tab

On the forecast tab, the Prophet time series forecasting model was used to predict the daily energy consumption of the specified ACORN category and the groups of interest. The historical data with the obtained forecast is displayed for both the selected category and its groups, having selected the custom hyperparameters and the forecast horizon.



On this tab, the Prophet time series forecasting model was used to predict the daily energy consumption of the specified ACORN category and the groups of interest. The documentation of the model with the details of its implementation procedure can be found :

Here

Main hyperparameters can be adjusted to effectively tune the category model, such as the trend and seasonality flexibility. It is also possible to change the forecast horizon and display its outcome with the selected uncertainty level. The displayed forecast of each group is the result of a model fitted with the optimum hyperparameters; these were obtained by applying the grid search method using cross-validation for each group dataset. The high-quality forecast is displayed with its uncertainty interval (error bands) and its corresponding historical data on the time series plot. In addition, the trend and yearly seasonality time components of the forecast of each category and group are included on the tab.

Filters

Category :

● Affluent Achievers

○ Comfortable Communities

Projection of average energy consumption by category selected in the filters section.

Trend graph: Shows the behavior of the trend over time shown in the upper graph.

Seasonality graph: Shows how seasonality is affected over time with respect to the average value of energy consumption shown in the upper graph.

Projection of average energy consumption by category selected in the filters section.

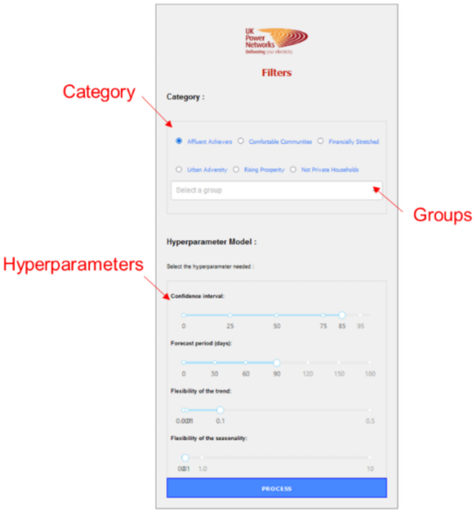
Trend graph: Shows the behavior of the trend over time shown in the upper graph.

Seasonality graph: Shows how seasonality is affected over time with respect to the average value of energy consumption shown in the upper graph.

Figure 7.12: Dashboard - FORECASTING

1. Filters: allows selecting a category and groups of interest with custom hyperparameters.

2. Graphs: Two types of graphs are displayed, historical data with the forecast and the trend and yearly seasonality time components of the forecast.



Filters

Category :

● Affluent Achievers ○ Comfortable Communities ○ Financially Stretched

○ Urban Adversity ○ Rising Property ○ Not Private Households

Select a group

Hyperparameter Model :

Select the hyperparameter needed :

Confidence interval:

Forecast period (days):

Flexibility of the trend:

Flexibility of the seasonality:

Figure 7.13: Dashboard - FORECASTING Filters

There are four Hyperparameters that can be adjusted:

1. Confidence interval: uncertainty range (both for group and category models).
2. Forecast period: days to be forecasted (both for group and category models).
3. Flexibility of the trend: hyperparameter that affects the trend of the category model.
4. Flexibility of the seasonality: hyperparameter that affects the seasonality of the category model.

All the hyperparameters have a predetermined value that is loaded with the application, but the user has a defined range of values in which these hyperparameters can be settled to the category model. **It is important to consider that the forecasts made by the group models had the optimum parameters previously found for each model.**

The forecast is displayed right after the end of the historical data, to clearly distinguish the forecasted values the uncertainty range is included in the graph. Similarly, to the previous tab, it is necessary to click on the process button to display the graphs.

Some of the output graphs displayed on the tab:

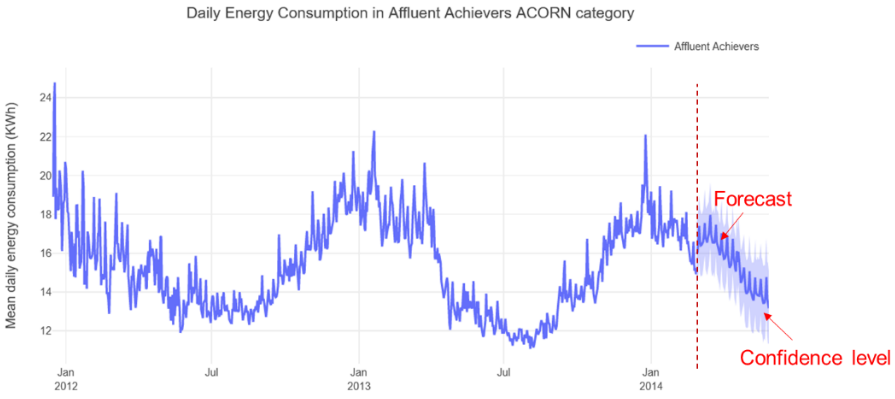


Figure 7.14: Dashboard - FORECASTING

The layout of the graph is similar between the category and the groups models. The forecast is displayed with the corresponding uncertainty level

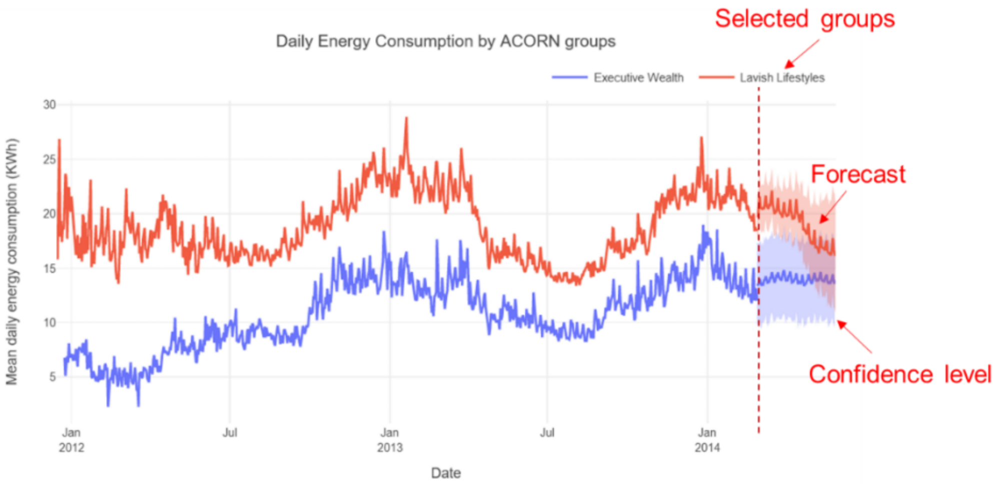


Figure 7.15: Dashboard - FORECASTING

Also, the time components of the forecast are presented both for the selected category and the groups of interest. Below is presented the graph of the trend and yearly seasonality that will be displayed for the selected category.

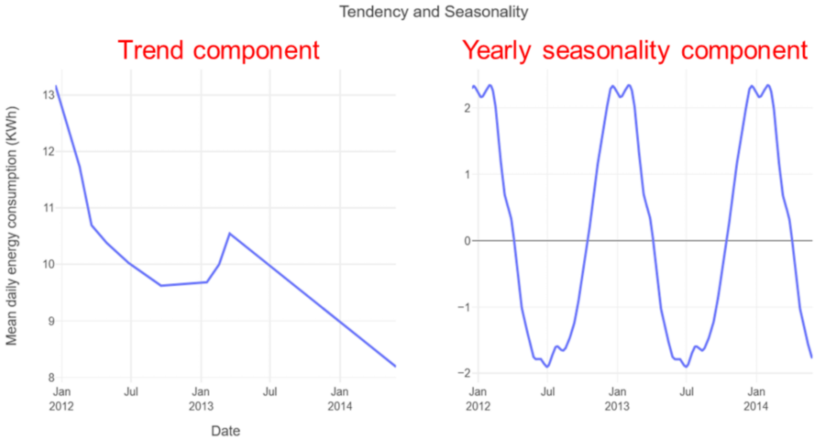


Figure 7.16: Dashboard - FORECASTING

7.7 Graphs

Since all the graphs are built with plotly graphing library, which has integrated some tools to interact with the visualization, allowing to zoom in it or exporting it in a .png format.

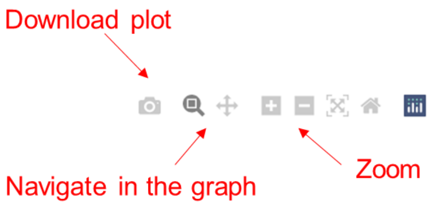


Figure 7.17: Dashboard - FORECASTING

For example, it is possible to zoom in a specific region of the graph.

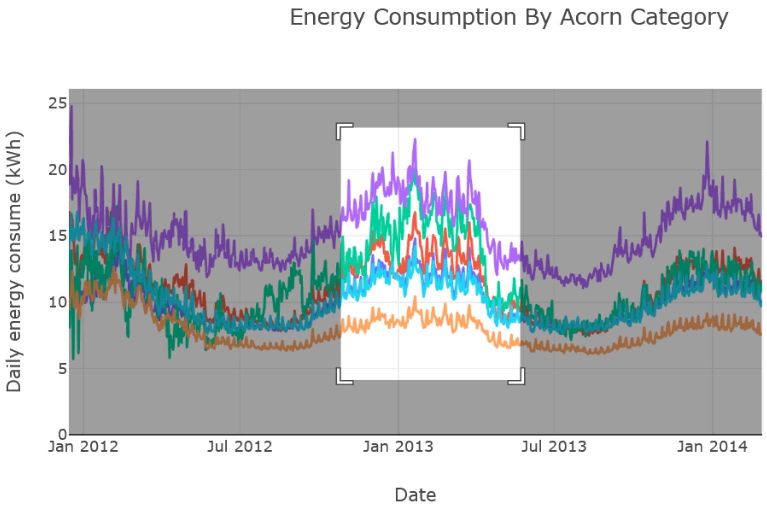


Figure 7.18: Dashboard - FORECASTING

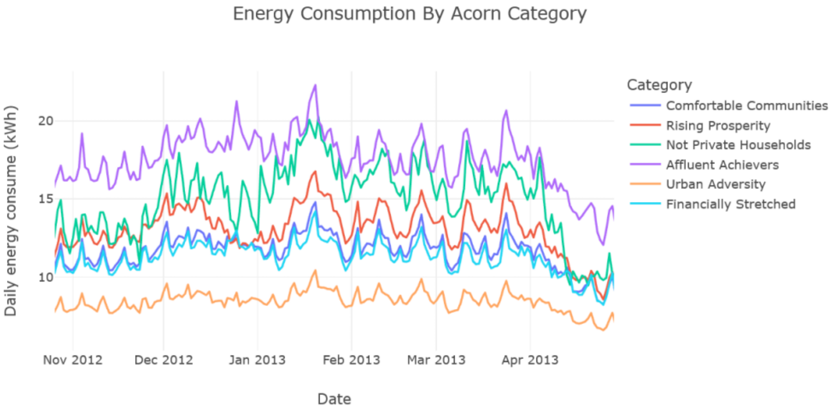


Figure 7.19: Dashboard - FORECASTING

8 CONCLUSIONS

The model performance is related to the amount of data available as input, the model is great in the short, and medium term. It's really difficult to predict 6 or more months in advance but with the data, we can see there is a really strong correlation between the demand for energy and the actual season.

We can definitely say the demand will continue to grow even in the low demand months, it depends on how the economy will grow, how the population will behave, and always according to the statistics, hypothesis tests, and graph plotting. This is just London, but we could translate those insights to other main cities in England, for example, Manchester, Liverpool, and Bristol. Something that could generate a more accurate concept is taking into account the surroundings, the outer cities around London, it's something that could give us more insights and a general grasp on how London could provide even smaller, and closets cities, this could drive more profit, infrastructure and investments from companies, one of the items checked on this research was the commercial stakeholders, they are classified and the investigation could head that way if needed as a potential investor.

On the other hand, this sort of model represents most of the time curves, the challenges we have when taking time series with average, maximum, minimum, and inputting data to try to get the sum of all the factors having an impact and trying to forecast with the most accurate tool possible. A long-term forecast will need more data even not from the same company, but from a similar economy like Paris, Frankfurt, Copenhagen, or any other main rich city.

As we had information up to 2014 mid-year, we forecasted the data just for the upcoming 6 months and it will decrease the energy demand but it will be bigger compared to the previous year's season. It will be likely to keep a positive (increase) trend for the next months, this could be expected but something that could take the research to a next level would be how climate change, clean energy sources, and other factors not explored here will impact the metrics for the forecasts and how accurate makes this model with those conceptions.

References

[1] Menculini, L., Marini, A., Proietti, M., Garinei, A., Bozza, A., Moretti, C., & Marconi, M. (2021). Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices. Forecasting, 3(3), 644–662. https://doi.org/10.3390/forecast3030040

[2] Taylor, S. J., & Letham, B. (2017). Forecasting at scale. PeerJPreprints. https://doi.org/10.7287/peerj.preprints.3190v2

[3] Mejía, E. & Gonzales, S. (2019). Prediction of residential electric power consumption in the Cajamarca Region through Holt -Winters models. https://www.redalyc.org/journal/3291/329160723002/html/

[4] Banda, H. & Garza, R. (2014). Aplicación teórica del método HoltWinters al problema de credit scoring de las instituciones de microfinanzas. https://dialnet.unirioja.es/descarga/articulo/5811252.pdf

[5] Internet open source, article: ARIMA vs Prophet vs LSTM for Time Series Prediction - neptune.ai.