UK Power Networks and SmartMeter Technology

Energy Consumption Report

Group Number # 02

Mohammad Hamza - 20100135
Umer Farooq Zia - 20100148
Dilawer Ahmed - 20100177
Muhamamd Hassan - 20100145

Report # 01

Exploratory Data Analysis

Data Analysis

Introduction

We were provided with the dataset of UK power networks gathered using SmartMeter Technology. The data set is gathered from 5,567 households between November 2011 and February 2014.

Dataset

The initially provided dataset contains 168 .csv files each file containing approximately 1 million rows, and the total dataset was around 10GB after unzipping. Later on due to computing issues, the numbers of files were reduced and 93 files were sampled for the analysis at this stage.

The dataset has five attributes, namely energy consumption in kWh (per half hour), unique household identifier, date and time, Acorn group, and tariff type. Meter readings were taken at half hourly intervals mostly, but there are a few exceptions in the dataset where erroneous readings were taken at irregular intervals. Such readings usually had *NaN* value and hence were replaced with ‘zero’ value in the processing stage. During the next phases of the project, we will see if these values need to be discarded or we can work to replace the null values with zeros.

Within the data set are two groups of customers. The first is a sub-group of customers who were subjected to Dynamic Time of Use (dToU) energy prices. Customers were issued following prices

1. High (67.20p/kWh)
2. Low (3.99p/kWh)
3. Normal (11.76p/kWh)

Prices were applied differently at different times of the day. The second sub-group of customers is Standard customers who were on a flat rate tariff of 14.228p/kWh.

Customers have been divided into different Acorn groups based on customer income. There are five different categories of Acorn groups, mentioned below

- Adversity
- Affluent
- Comfortable

- ACORN-U
- ACORN-

Purpose of Document

This is an open ended phase of the project where we are required to perform Exploratory Analysis on the dataset. The goal is to find patterns, get acquainted with the data, and methods to look for meaningful inferences from the data.
In this chapter, we will discuss the results of some primary data exploration and analysis techniques and will try to draw some conclusions based on these results. In later stages of the project and in next chapters, we will come back to these hypotheses and their validity.

Basic Statistics and Correlation

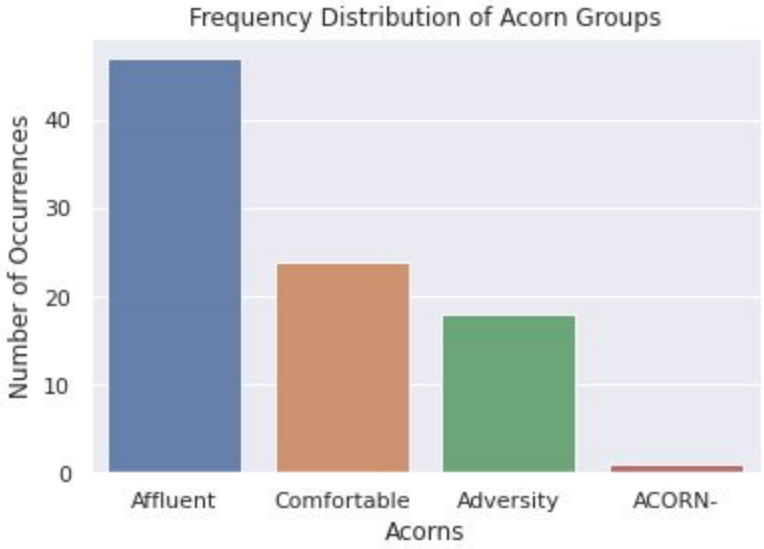
We divided the given dataset into two. One dataset of standard users and another dataset of dynamic users. Here is the summary statistic of standard user:

	KWH/hh
count	8.999696e+07
mean	2.135651e-01
std	2.922476e-01
min	0.000000e+00
25%	6.000000e-02
50%	1.200000e-01
75%	2.430000e-01
max	1.076100e+01

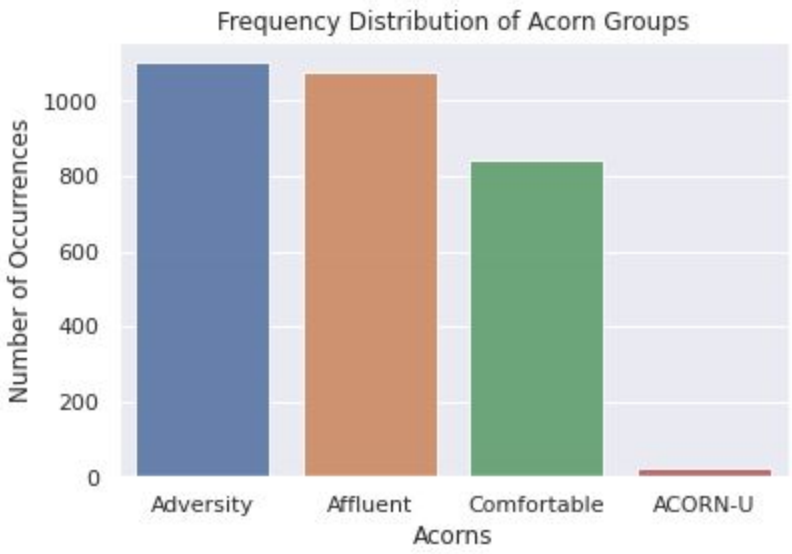
Here is the summary statistic of dynamic user:

	KWH/hh
count	2.932384e+06
mean	2.055369e-01
std	2.917551e-01
min	0.000000e+00
25%	5.300000e-02
50%	1.070000e-01
75%	2.340000e-01
max	6.162000e+00

As discussed in the introduction of the dataset, households are divided into acorn groups of 5 kinds. On further analyzing the data, we find the following division of Acorn groups among customers.
Dynamic priced customers distribution into acorn groups is shown in the graph below.

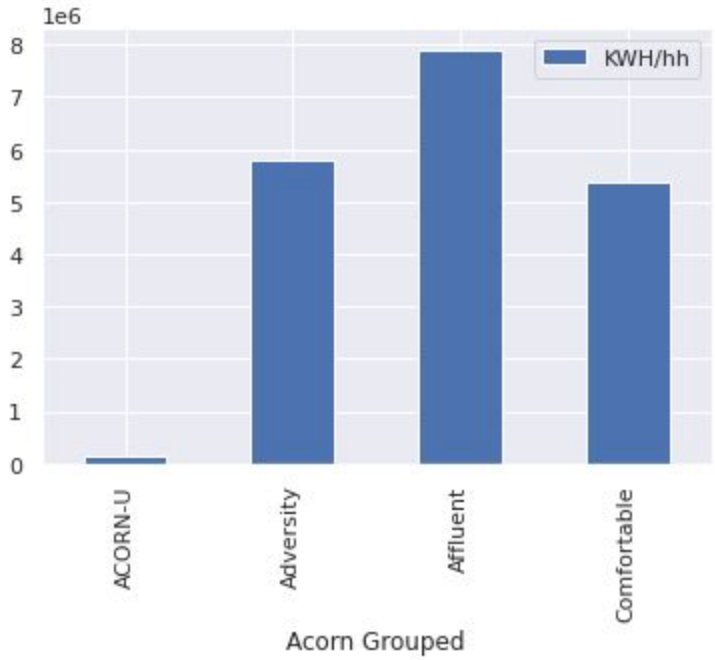


The standard priced customers had following distribution into Acorn groups as shown below.

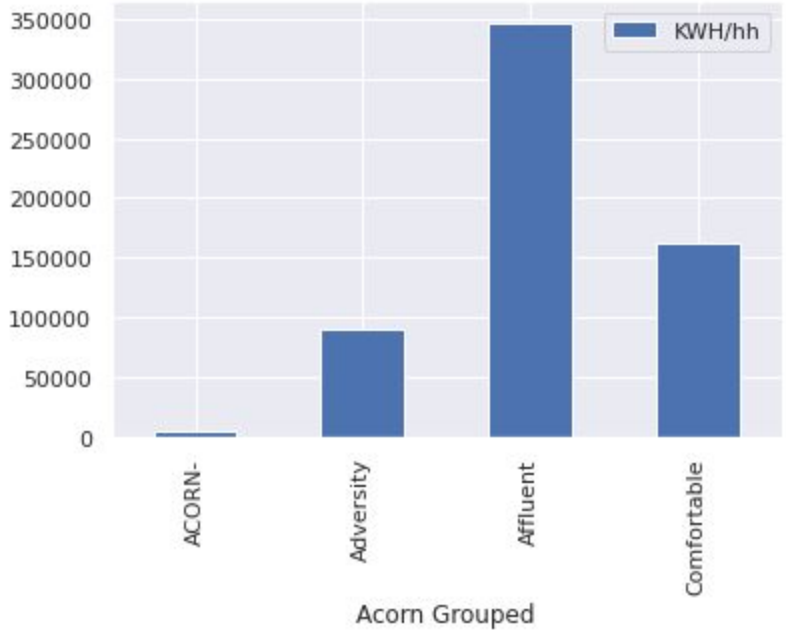


From the graphs, we can infer that the majority of customers who are asked to pay the dynamic prices belong to affluent and comfortable classes, while adversity and affluent classes are dominant in standard pricing. From a business point of view, this is an intelligent move considering that Affluent and Comfortable classes would be more willing to pay the dynamic prices, since the dynamic prices fluctuate to considerably very high value. While on the other hand, if we have a closer look at the distribution of Acorn groups into standard pricing, we find out that although ‘Adversity’ is dominant, almost ⅔ of the standard pricing customers are either ‘Affluent’ or ‘Comfortable’ households. Perhaps, in order to increase the revenue, the electricity providing company should try to shift some of these households to dynamic pricing. This can serve two purposes, reduce the load during the time when demand is high and also increase the revenue by dynamic pricing from a larger group.

If we analyze the total consumption of each pricing plan, we observe the following pattern for Standard Pricing customers. We observe that the Affluent are third most dominant in the number but their consumption is highest among all.



Below this paragraph is the total consumption graph for Dynamic Pricing plan. We realize that the total consumption of ‘Affluent’ households is greater than the combined consumption by the rest of the households in this category.



Analysis across Energy Usage:

We have analyzed the dataset and observed the trends of usage of energy on many levels. We have done the following analysis on energy usage.

- Average usage in each month by each customer group
- Average usage by time of day
- Average usage by day of week
- Average usage by weather in DToU and STD group
- Average electricity around Christmas and New Year

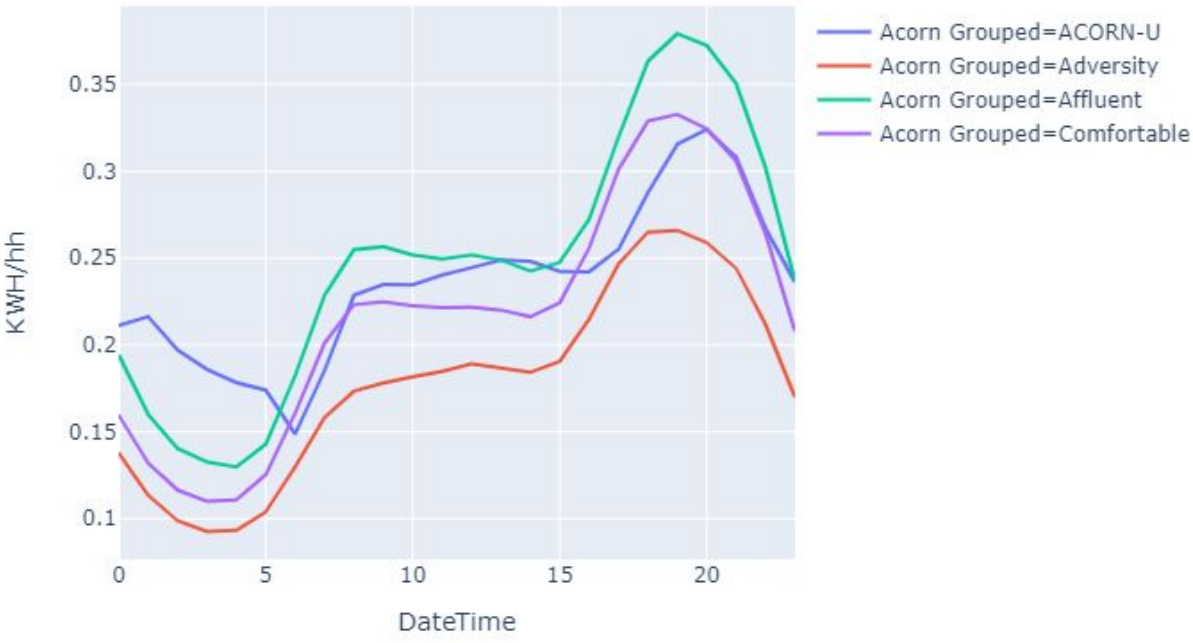
Average Usage In each month by each customer group

The following graph shows the average monthly usage of both groups. An interesting thing that can be observed is that the average usage is linked to how well off a group is. The most economically stable population i.e. Affluent have more per month usage compared to the Adversity group. This can be due to economically stable households being larger in size and have more electrical appliances than the other groups.

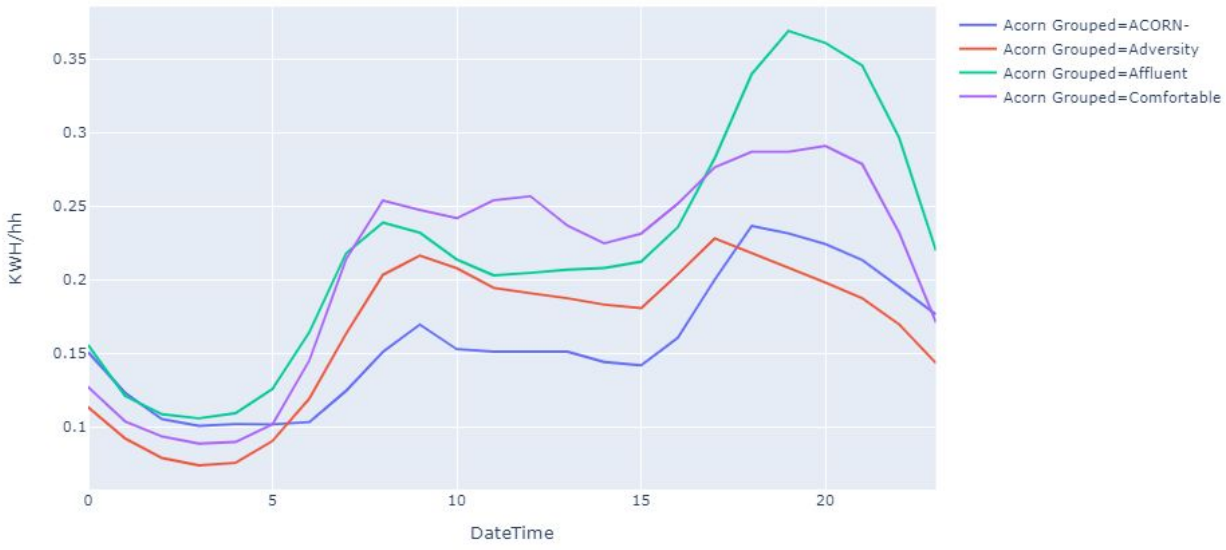
Standard Users:

Dynamic Time of Use Users:

Average Usage By Time Of Day



The graph above shows the average usage per time of day for standard users. The trend follows the expected value for household use. Usage is low during night time from 11pm to 6am. Then it increases and stays constant from 7am to 3pm. At evening time it starts to increase, reaching peak at around 7-8pm. This graph can produce important insights for electricity production. Power plants can be run and their production can be adjusted accordingly. Electricity rates can be adjusted accordingly as well to flatten the curve. The same trend is followed for dynamic users.

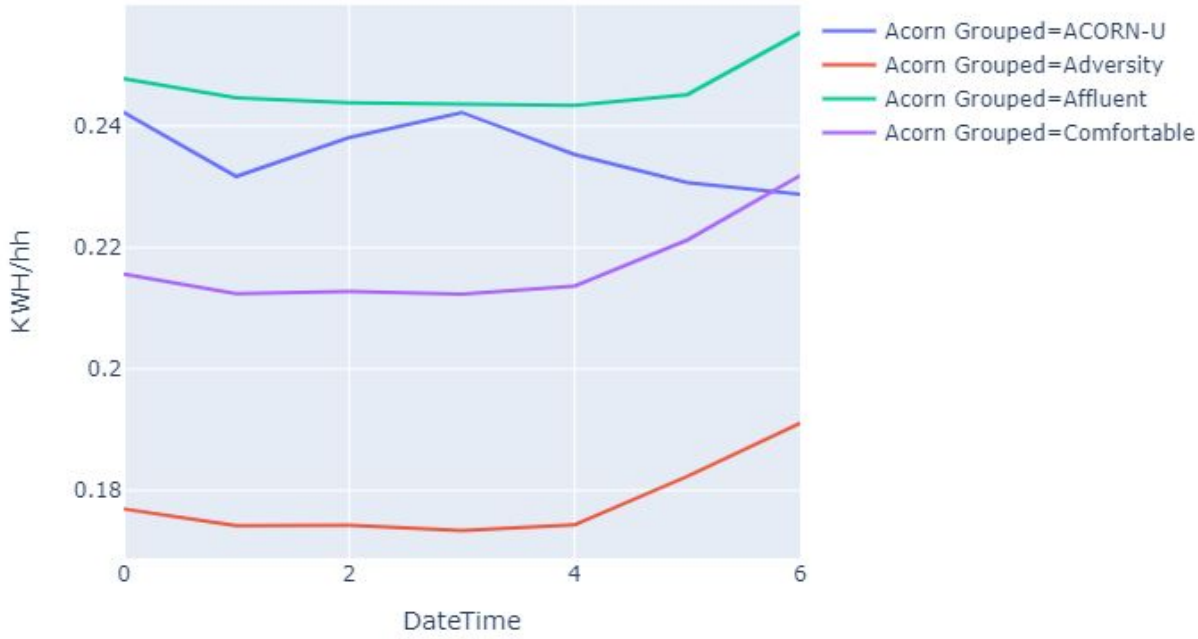


Average Usage By Day Of Week

The data was analyzed to observe the average usage with respect to weekdays. We want to see if usage is more on weekends or working days.

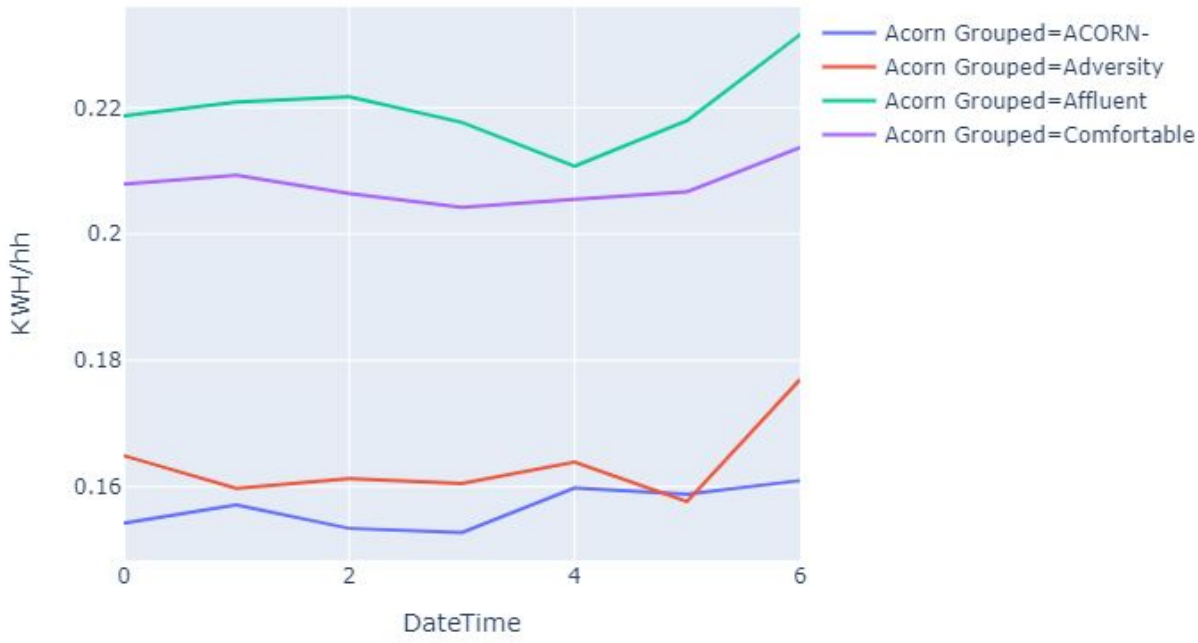
Here is the mapping of the days:
0 = Monday
1 = Tuesday
2 = Wednesday
3 = Thursday
4 = Friday
5 = Saturday
6 = Sunday
The graph for each customer group is given below. We found out that energy consumption was relatively constant throughout the day, but during the weekend there was a spike. Since this data is from SmartMeters in UK households, we can infer that the spike has been caused by people spending their weekends at home and not going to work places.
Further, we can infer that the graphs correlate with the economical question that Affluent and Comfortable households are spending more on electricity as compared to other classes.

Standard Users:



Dynamic Users:

For the dynamic user data, All data points that we have were recorded on Thursday. So, no can't compare the usage with other week days for dynamic user's data points.



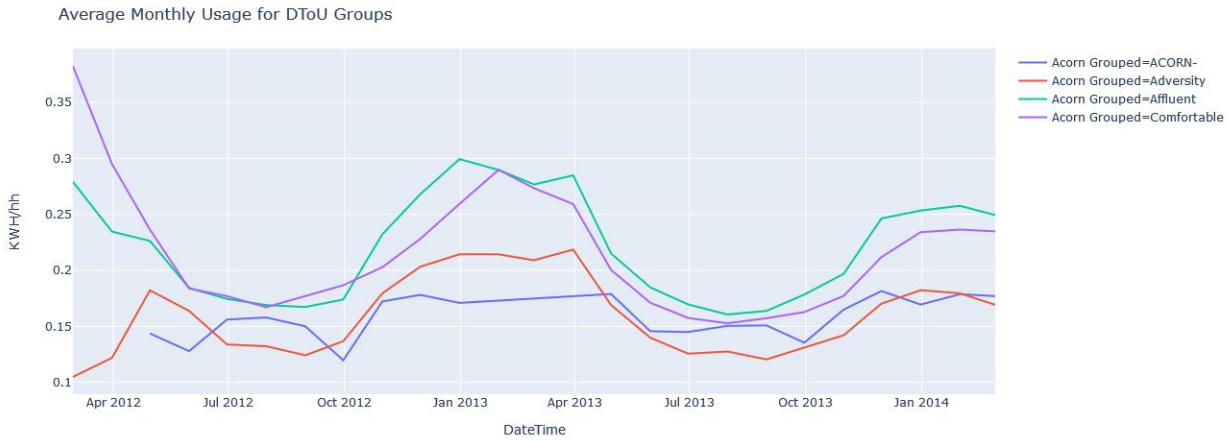
Average usage by weather in DToU and STD group

There are four seasons in UK (source):

- Spring: March - May
- Summer: June to August
- Autumn: September - November
- Winter: December to February

Referring to Figure 20, that graph is for average “KWH/hh” usage per “LCLid” for a month (STD Group). This group has four “Accron_grouped” in it, namely: “ACORN-U”, “Adversity”, “Affluent”, “Comfortable”. As expected, “KWH/hh” follows a trend of seasonality for all “Accron_grouped”. During winters, the average usage zeniths and steadily decreases as moving towards summers. During summers, the average usage is the lowest. During spring and autumn, the average usage is similar. This pattern repeats for all the years. One conclusion that can be drawn from this is that electricity is used for heating purposes more during the winters. Correlating this with winter and summer conditions in the UK, where winters are very cold and harsh, and summers are very pleasant.

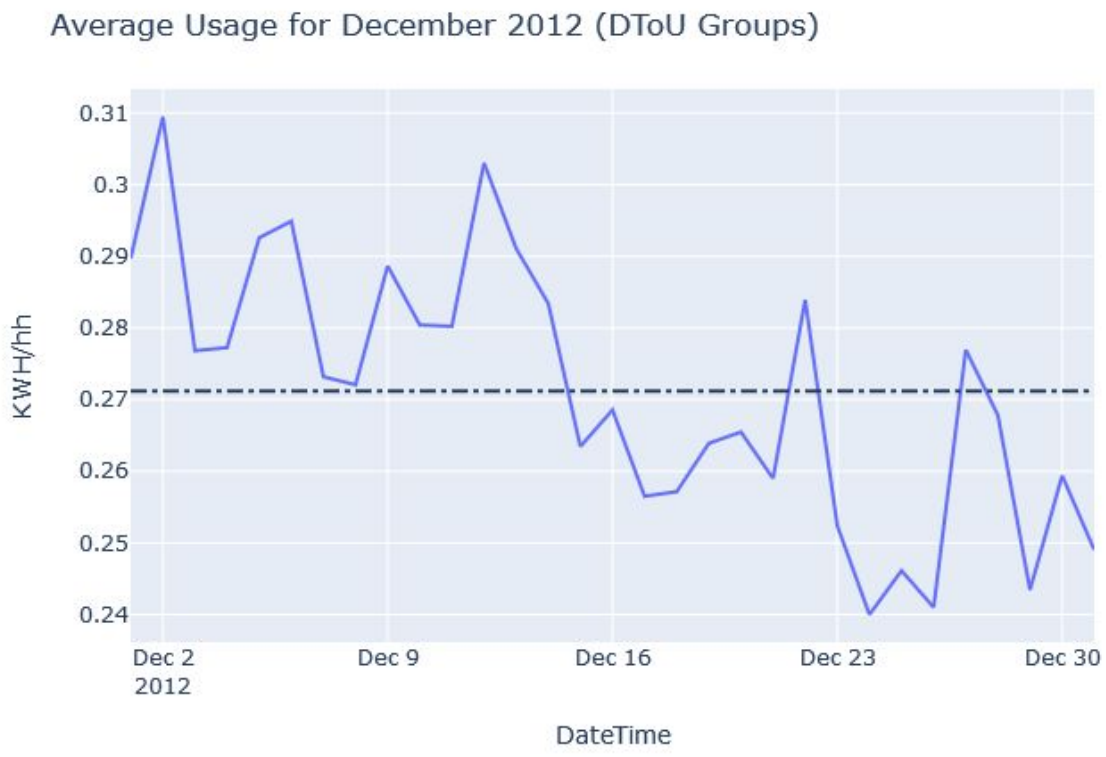
Up next is the graph for average “KWH/hh” usage per “LCLid” for a month (DToU Group). This group has four “Accron_grouped” in it, namely: “ACORN-”, “Adversity”, “Affluent”, “Comfortable”.

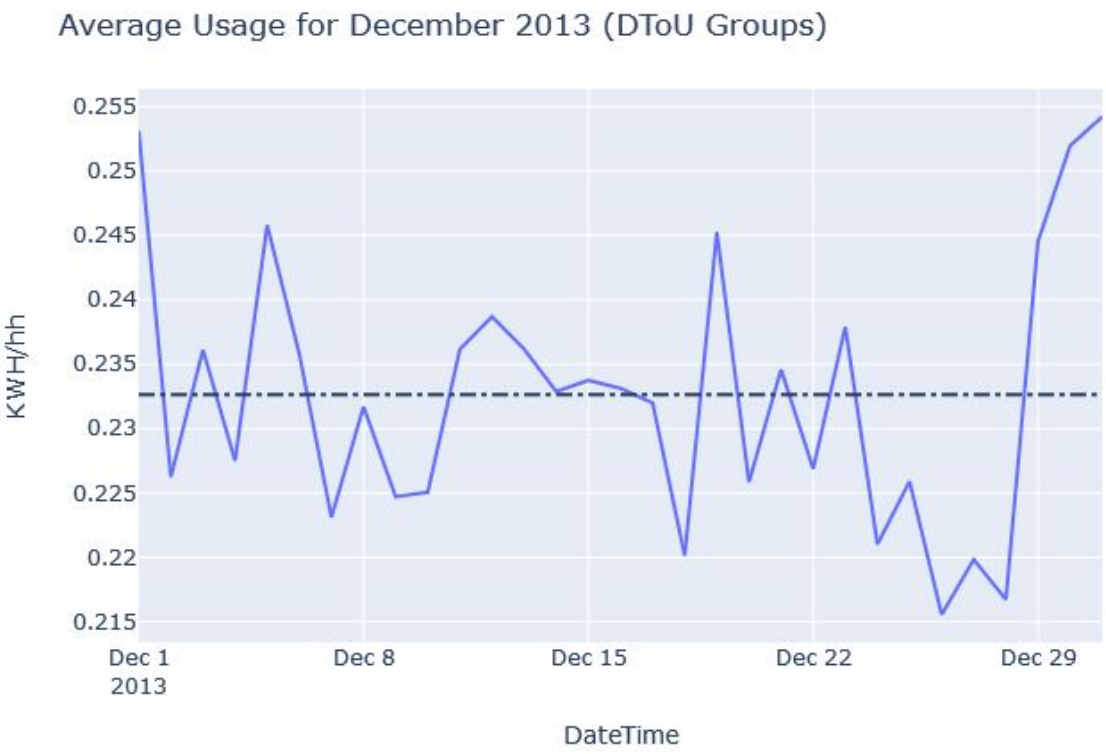


DToU also follows a similar usage pattern to STD but around June 2012 there are some anomalies, but these can be attributed to the lower number of input data (only 3 files from DToU were used).

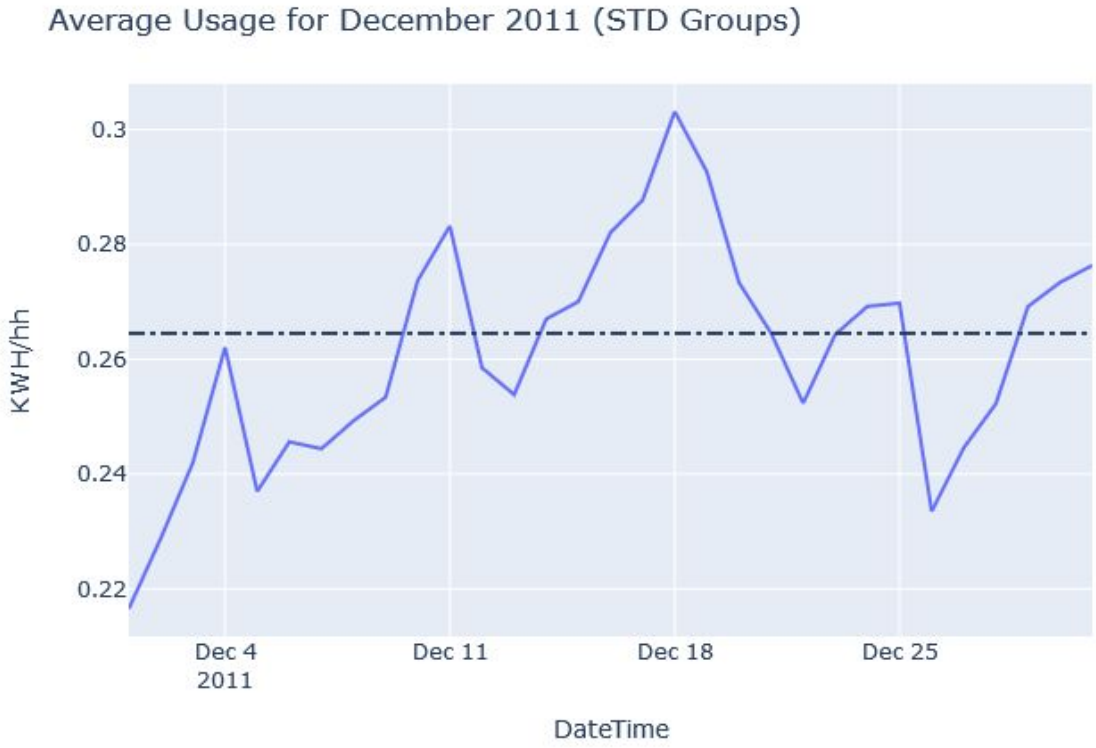
Average Electricity around Christmas and New Year

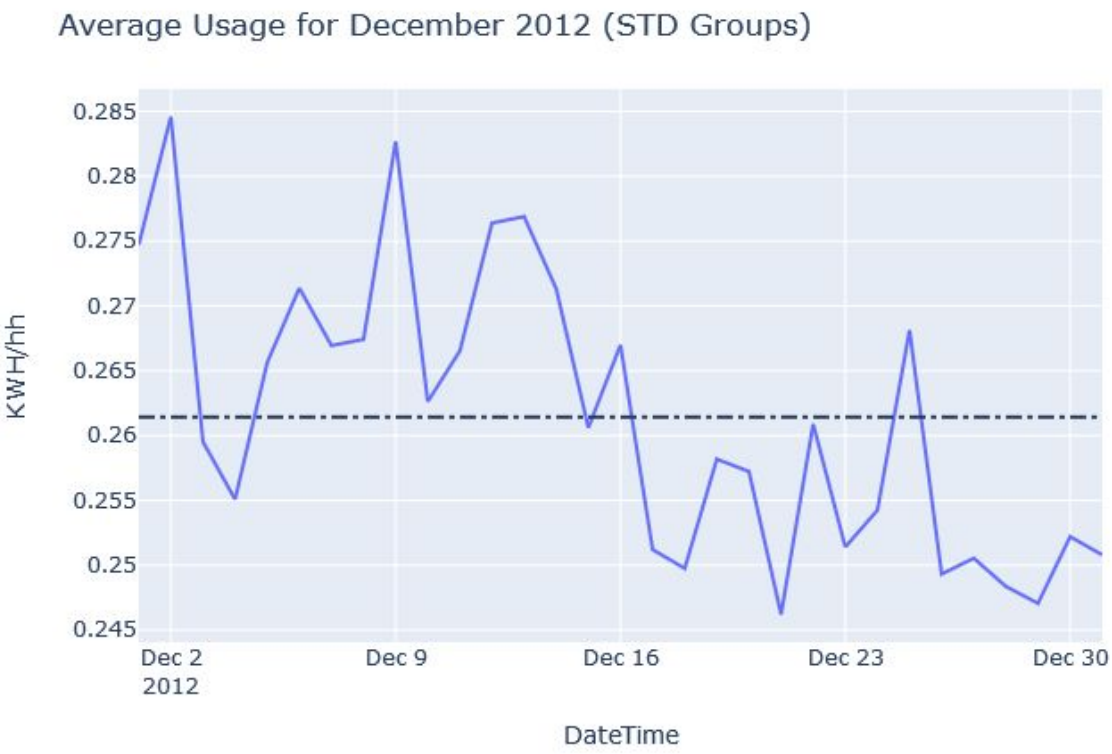
Following are the graphs for daily average and monthly average in December for DToU group. The dashed line represents the mean electricity usage for the month.

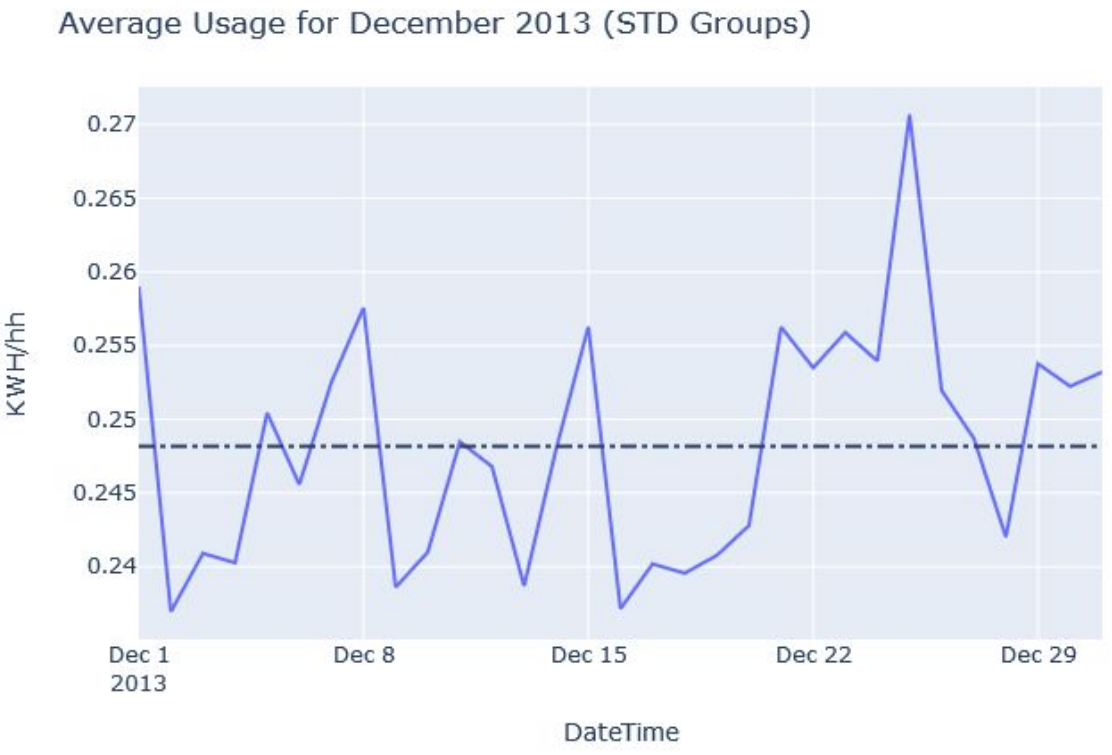




Following are the graphs for average daily and monthly usage for STD group.



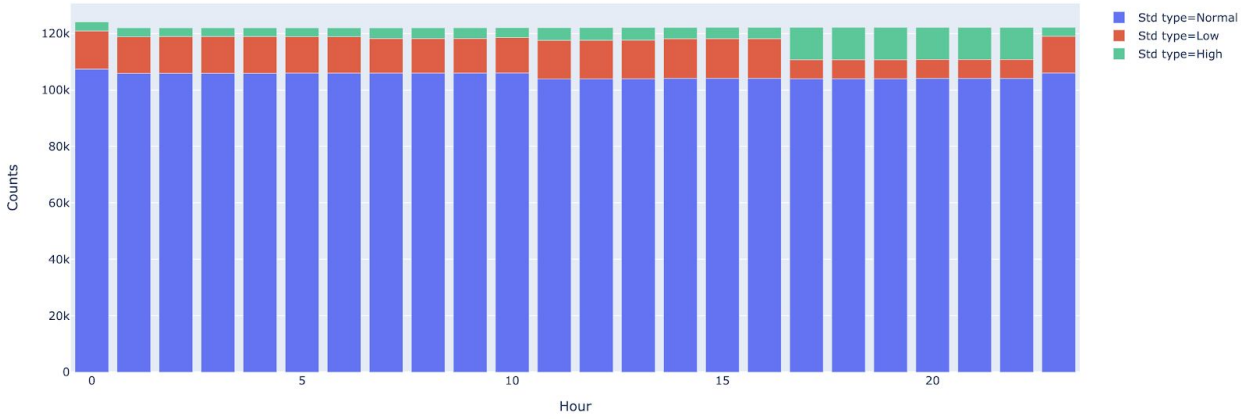




Looking at these 5 graphs, it is difficult to say that the average daily electricity usage strictly follows some trend in the month of December. Although in some cases, the

usage was reduced in the second half of December, i.e. from 16th December to 31st December. This could be a result of people getting together (electricity usage falling for some houses) for Christmas and
New Year events.

Tariff types of dynamic customer by time of day



The stacked bar graph above shows the tariff types with respect to time of day. As can be seen mostly normal tariffs are in place. Around 6-10pm high tariffs are relatively more charged as compared to the rest of the day.

Comparison of average cost of standard customer and Dynamic customer

The graph does the average monthly cost per customer in UK Pounds for Standard tariff and Dynamic Time of Use Tariff. There is no clear indication of if one is better than the other but overall Dynamic tariffs are ever so slightly lower compared to standard.

Summary

- Affluent and Comfortable were spending more in terms of electricity cost as compared to others.
- Affluent were spending more in standard connections as well, the company can transfer them to Dynamic usage and try to maximize the profits. This is one of the hypotheses drawn from current inferences, we can only verify this in later stages where we will have better evidence for such business decisions.
- We tried to correlate categorical data through label encoding of categorical features, further we employed Pearson and Kendall methods in Python, but we could not find any strong correlation. Given this scenario, we referred to graphing our findings to correlate different attributes.
- The energy consumption was greater in winters as compared to summers, and is possible due to weather conditions in the UK.
- The energy consumption was greater on weekends as compared to weekdays
- The energy consumption was greater from 7 pm to 11 pm than the rest of the day.
- During holidays like New year and Christmas, the energy consumption has decreased. Maybe it's due to the fact that people go and visit their friends and relatives during this time period.
- Dynamic customers are being charged at Normal for most of the time but around 6pm to 10 pm high tariffs are relatively more charged as compared to the rest of the day.
- The average cost of Dynamic and Standard customer per month was relatively similar, we could not find any interesting points from it.