Analysis of Energy Consumption using Smart Meters in London

Roohani Megh

Introduction

Trial lead by the European Union in order to reduce energy consumption to tackle climate change.
Aim: to install smart meters in all homes within the UK by 2020.

1. This trial was focused with Greater London. Household energy meter readings were taken every 30 mins between November 2011 - February 2014.
2. 5,567 London HHs were used in this data collection. The data from the smart meters seems associated only to the electrical consumption. 1,100 HHs were subject to smart (Dynamic Time of Use) energy meters and 4,467 kept their current standard energy meters.
3. Smart Meters use set tariff prices, which are sent to the meter home screen before the day of use.
4. Tariff Prices:
    - `High (67.20p/kWh)`
    - `Normal (11.76p/kWh)`
    - `Low (3.99p/kWh)`
5. Standard meters have a flat rate tariff of 14.228pence/kWh.

https://data.london.gov.uk, 2013

theguardian.com,2018

The Data

Many datasets were used and combined in this analysis.

1. **Information on HHs** - Each unique household identifier, along with CACI Acorn group (2010) classifier. This is a UK consumer classification dataset which analysis demographic data such as social factors, consumer financial stability and behaviour and segments them into respective groups. The customers in the trial were recruited as a balanced sample representative of the Greater London population.

2. **Daily energy reading dataset**- A snapshot of energy consumption per day for each unique HH. The data contains energy consumptions such as energy mean, median, min, max, standard deviation, count and sum. There is 826 days of data for 5,567 HHs.

3. **Acorn details** - Details about each acorn group and their profile of the people in the group.

4. **Weather daily data** - Contains information about the daily weather in London for the duration of the trial, obtained from darksky api.

How to Measure Energy Consumption

Different readings can give us different snapshots of the data.

Energy consumption varies throughout the day for any given HH. We only have daily readings which aggregate the data (mean/median/max/min/standard deviation/sum/count). We are unable to see daily variation which makes picking the measure slightly tricky.

Advantage of using mean:

- Easier to understand.
- Takes all measurements into account.
- If there are no major outliers then we get a better measure of central tendency.
- Used in statistical tests.

Disadvantages of using mean:

- If data is heavily skewed and contains many outliers, then data can be inaccurately misinterpreted.

Best way to get a sense of this is to assess how varied the data is, for this we can look at the daily standard deviation.

How to Measure Energy Consumption

Different readings can give us different snapshots of the data.

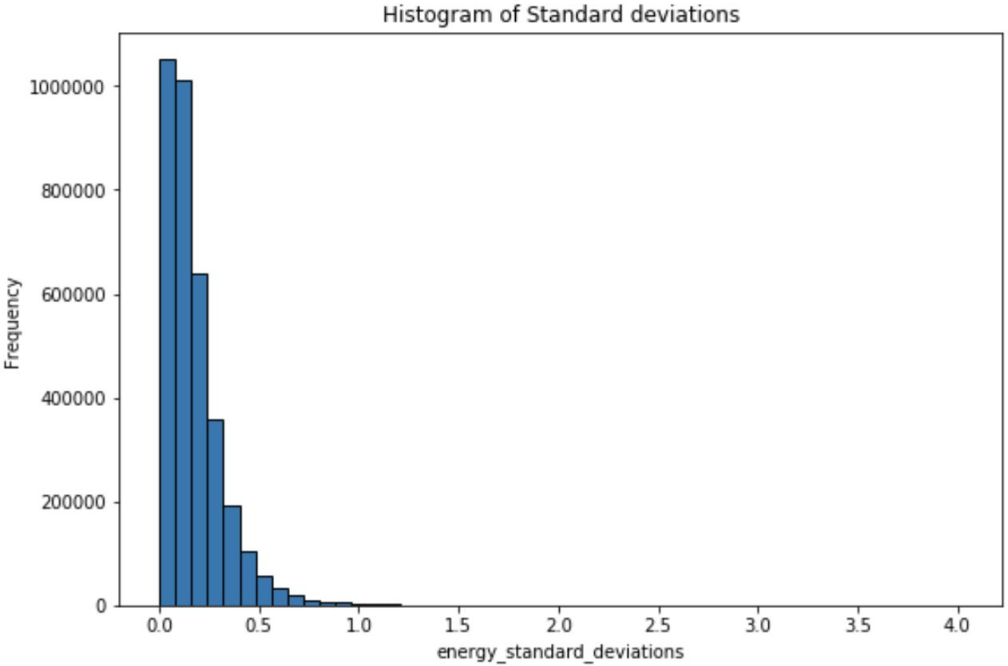
As we do not have half hourly energy readings per day for each hh, we are unable to see the skew within the data per day.

To deal with this, I have plotted a histogram of daily standard deviations.

The 75th percentile here is 0.23, and the max SD is 4.02. The huge discrepancy between these values further proves that the vast majority of the SDs falls below 0.23.

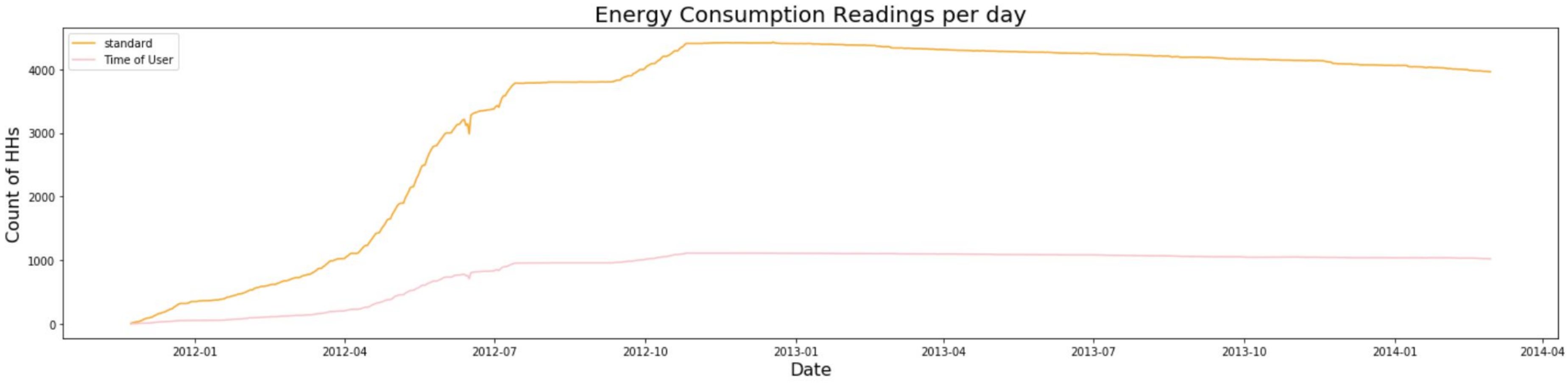
Using a SD of 0.23, the CV = 1.08, this indicates a low variation amongst the energy readings per hh per day.

If there is not much variance in the data it is fair to use the mean energy reading for my analysis.



Dealing with uneven sample sizes

Type of Meter	No. of HH in sample
Standard	4,467
Time of User	1,100



Dealing with uneven sample size

Type of Meter	Acorn Group	No. of HH in sample	As a % of Total
Standard	Affluent	1702	27
	Comfortable	1184	38
	Adversity	1518	35
Time of User	Affluent	490	44
	Comfortable	298	29
	Adversity	298	28

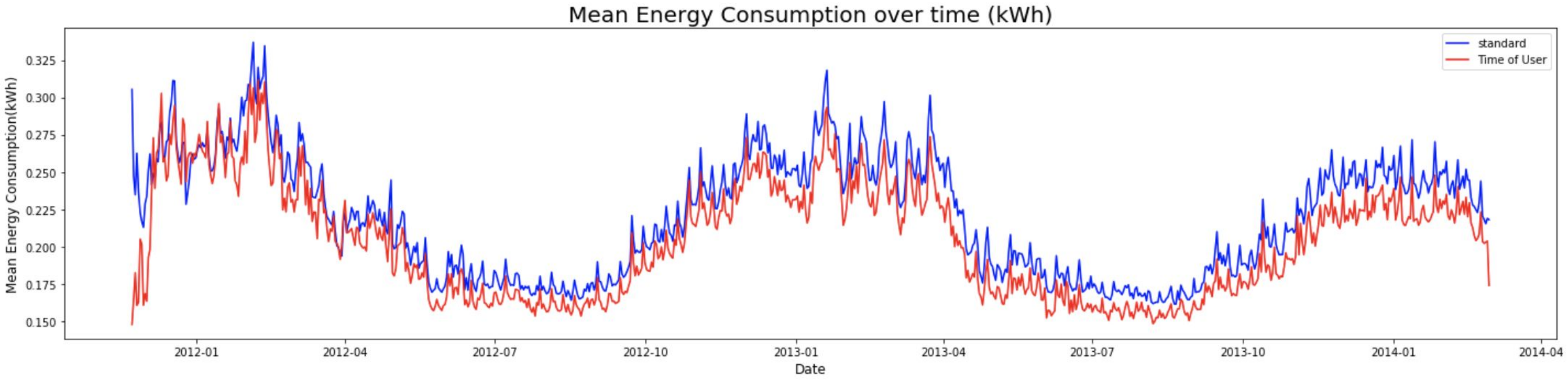
There are far more data for Standard meter readings than for Smart meter readings. As smart meters are far newer and require special installation this kind or proportion makes sense.

This may impact future analysis, when looking at these groups separately the standard meter readings hold more power, as there is more data present for this group.

However, for future statistical tests, we have already assumed equal variance, so this should not impact the validity of that.

Overall Energy Consumption trends between both Groups

Below is a plot of average energy readings (kWh) daily for Standard meter readings and Smart meter readings.

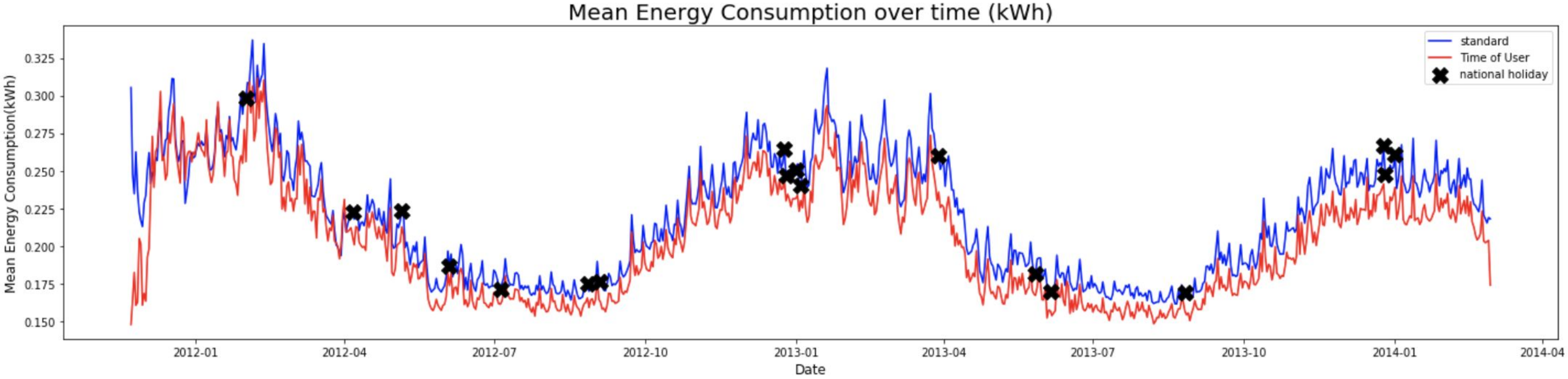


Both meters produce very similar trends. But for the vast majority of the time, Mean energy consumption is higher for standard meters than for smart meters.

It is worth noting that we have no knowledge about the accuracy of these data, whether any bugs or outages were reported. With this being said, so far, the the data seems to support my theory.

Overall Energy Consumption trends between both Groups

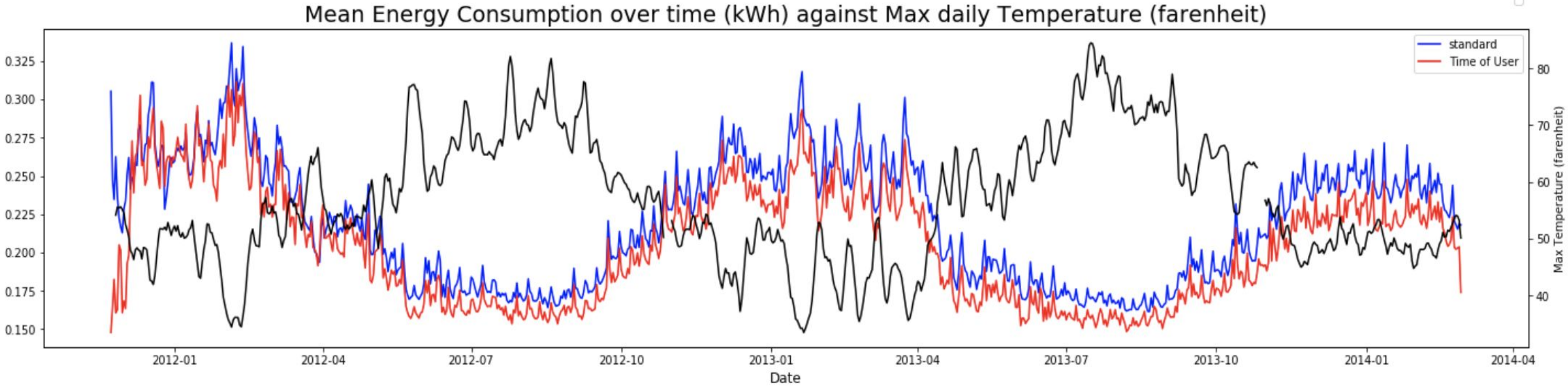
In order to omit any variation, I wanted to assess whether national holidays (bank holidays) made a difference to energy consumption. This would impact the total sample, people are likely to be out of the house enjoying the holidays, energy consumption may be significantly less for these days, these would count as outliers.



National holidays don't seem to have any impact on the energy consumption for that day. Both meters don't deter from the trend that is created by succeeding and preceding data. These data points will be kept in.

Energy Consumption vs Weather

Below is a plot of average energy readings (kWh) daily for Standard meter readings and Smart meter readings, plotted against the max temperature (farenheit) in London for that day.



Energy consumption is highly correlated with the daily temperature. The temperature patterns are mimicked by both energy reading consumption patterns. A correlation test further proved this with a high correlation of 0.83.

As the temperature increases, the energy consumption decreases. As temperature decreases, the energy consumption increases.

This makes sense, the majority of temperature regulation done within UK households is through the use of heatings. London does not generally get warm enough to apply air conditioning.

Conducting an Independent T-test

To test the hypothesis, I am going to conduct an independent t-test to determine whether the two energy consumptions are in fact statistically significant or not.

Independent T-tests:

The T - test is a powerful statistical test which can help establish whether there is a statistical difference within two independent groups of data.

These groups are independent of each other - one HH can not both have standard AND a smart meter.

Independent T-tests require the data to follow a normal distribution. We have seen that the that energy consumption is largely dictated by the daily temperature, and this feature is the same for every HH in London. With this being the case, I assume that most houses will largely use the same amount of energy so the energy means will be closely related to the overall daily mean and hence follow a normal distribution.

Another assumption made to carry out valid independent t- tests are that the two groups have a similar variance. This has been proved earlier on.

Conducting an Independent T-test

To test the hypothesis, I am going to conduct an independent t-test to determine whether the two energy consumptions are in fact statistically significant or not.

Independent t -tests were taken for the 829 days that this trial was run. This omits the variation in energy consumption that occurs with seasonality during the trial period.

Below are the results:

No. of days	T-test result
754	Reject null hypothesis (the two groups are statistically different)
72	Fail to reject null hypothesis (the two groups are the same)

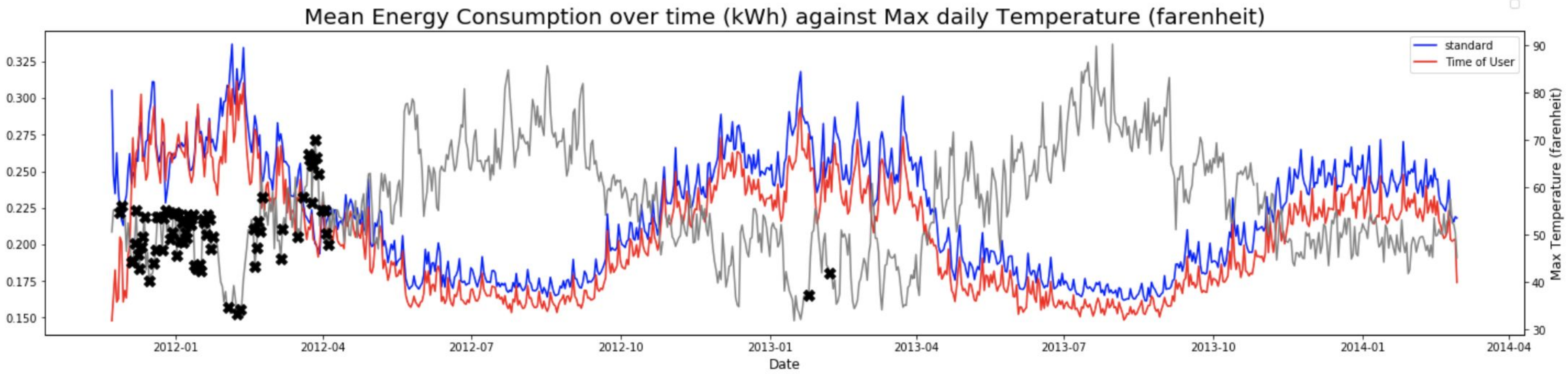
91% of the time, the test shows that there is a difference between the energy consumption means between those HHs that have a smart meter and those who have a standard meter.

8% of the time, the test shows no difference between the energy mean consumptions between the two meters.

Let's look at the days that those occur.

Conducting an Independent T-test

Here is a plot, annotating those days where there is no statistical difference between the energy consumption between the two groups.

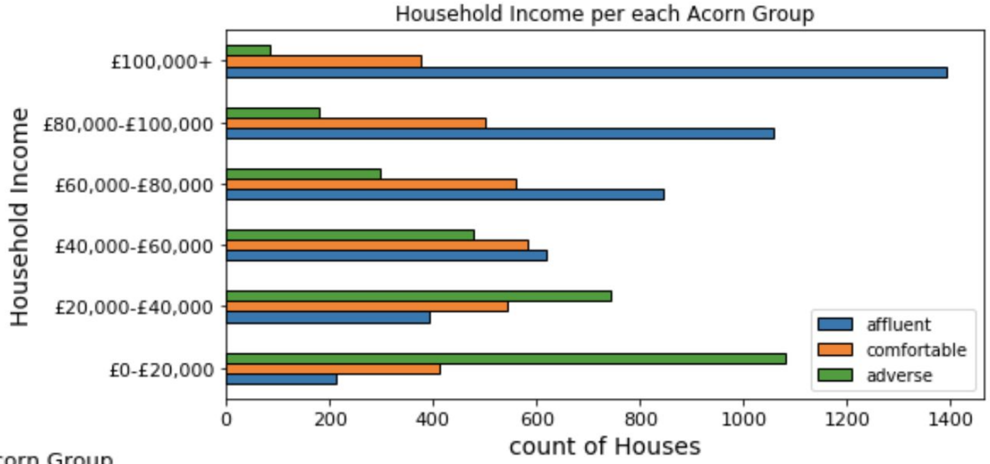


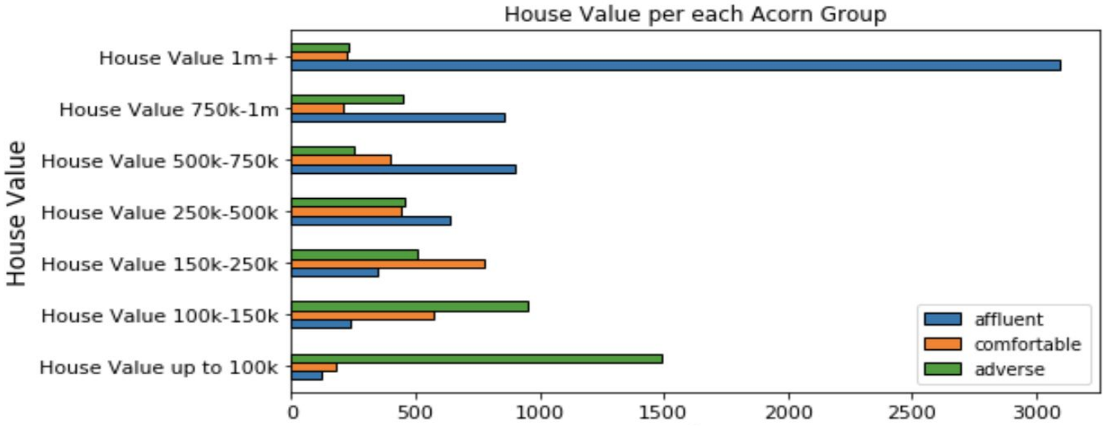
Majority of these instances occur 11/2011-02/12. It's interesting that these instances occur right in the beginning of the trial. This could be attributed to HHs getting used to a Time of User meter. These meters work differently and often its reviewing past data usage that gives users an understanding of how much energy they save which drives further improvement.

This proves that the introduction of smart meters does infact cause a change HH energy consumption.

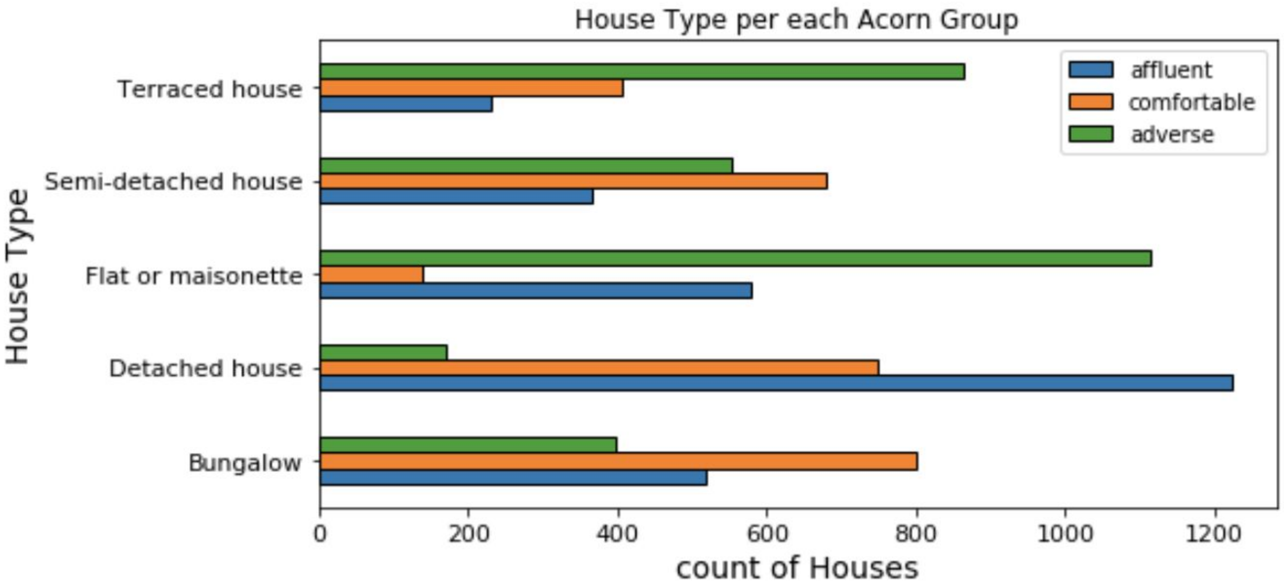
Looking into Acorn Groups

The energy dataset provided had grouped the several acorn classes into 3 main groups; Affluent, Comfortable and Adverse. Let's look into these classes.





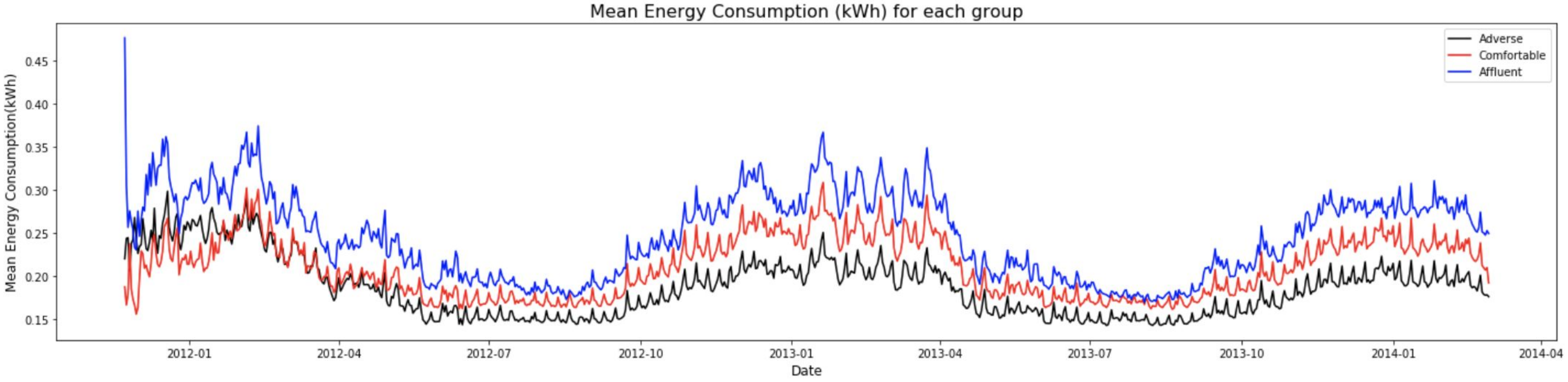
Looking into Acorn Groups



Detached houses are a lot less insulated so require a lot more heating to maintain and regulate a desired internal temperature, This will be a big contributing factor to energy consumption along with just the financial circumstances of each group.

Looking into Acorn Groups

Below is a plot of overall average energy readings (kWh) readings, split by the acorn groups.



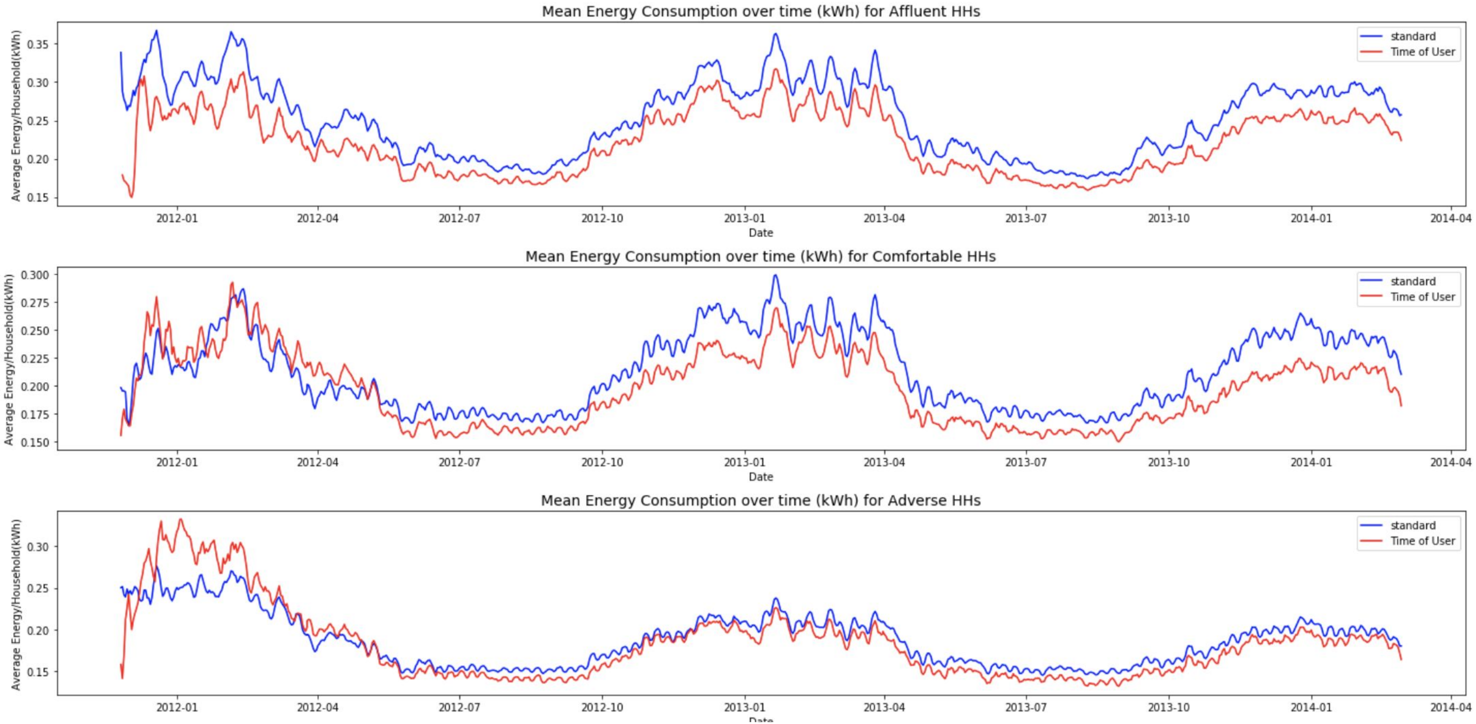
Expected trends can be seen here. Affluent groups consume the most energy throughout the year, then comfortable groups. Adverse groups consume the least.

As mentioned before, the nature of the houses these groups are another influencing factor, along with the financial situation of each groups.

However as adverse groups are already price sensitive, it would be interested to see which group would be most impacted by the introduction of smart meters.

Looking into Acorn Groups

Below is a plot of average energy readings (kWh) daily for Standard meter readings and Smart meter readings, split by the acorn groups.



Conducting an Independent T-test- grouped by Acorn groups

The same independent t-test was conducted for the datasets, this time split between Affluent, Comfortable and Adverse groups.

To further prove this, it would be best to conduct another series of independent t-tests for each acorn group. Using the acorn groupings, I can get a much better insight into energy consumption behaviour. This can also be a useful feature when trying to forecast and model energy consumption in the UK.

We have seen how the 3 different splits have been created, here are the T-test results:

Acorn Group	**% of total Days**	**T-test result**
Affluent	2	Fail to reject null hypothesis (the two groups are the same)
	98	Reject null hypothesis (the two groups are statistically different)
Comfortable	12	Fail to reject null hypothesis (the two groups are the same)
	88	Reject null hypothesis (the two groups are statistically different)
Adverse	17	Fail to reject null hypothesis (the two groups are the same)
	83	Reject null hypothesis (the two groups are statistically different)

Conclusions from Analysis

- We can reject the null hypothesis and safely say that there is a significant change in consumption between HHs using standard meters and smart meters. This change adds validity to the use and reason of smart meters.
- Energy consumption seems to be largely dependent on the outside temperature, this can be seen as the trends between energy consumption between both smart meters and standard meters mimics London's temperature trend during the trial. A correlation test further proved this with a high correlation of 0.83. This further illustrated my research stating that heating is a large energy suck for UK households
- As expected, affluent groups are the largest consumers of energy. This is due to a mixture of their financial circumstances, along with the types of houses they tend to occupy.
- Being the largest consumers of energy however, affluent groups show the least daily variation in energy consumption. This may be due to the use of smart and more efficient appliances.
- Even between each group, there is a significant change in consumption between HHs using standard meters and smart meters. The adverse group shows the smallest absolute change, this may be due to energy usage already being low with standard meters as they are the most price sensitive group.
- The greatest changes can be seen during the high ToU periods, likely having prior knowledge of the cost of electricity during this time encouraged people to reduce consumption. Given the absolute change in consumption is smallest during the low ToU periods, suggests households may be shifting their consumption from normal and high periods to this time.

Recommendations

- The boxplots show a significant change during high ToU periods However there is still a margin for improvement for the other periods. The normal and low periods correspond to 'less peak' times (i.e when the majority of the members are out of the house). This means energy consumption should already be lowered, so incentives can be given to the public to further lower consumption during these less peak time.
- Energy consumption is largely dependent on weather and seasonality. Maybe there could be different price tariffs for winters and summers.
- The dataset is ideal for forecasting energy consumption. Creating a model can help predict user behaviour which can help manage future renewable energy operations (as this will be rolled out).

Next Steps

- Show the magnitude of differences between the two groups.
- Calculate actual cost savings occurred when using ToU meters, so that this significant change can be quantified into something the general public can comprehend.
- Further explore the half hourly readings dataset to see how energy consumption was affected by the day of week to gain any further insight.
- Cluster the acorn groups myself to get better defined groupings.
- There is a lot of data (many households recording up to 9 energy readings per day for 829 days). This would make for a great interactive dashboard.
- Forecast energy consumption for Standard vs Time of User meters using random forest or linear regression modelling.