University of Strathclyde Glasgow
Strathclyde Business School
MSc Data Analytics

Exploration of Energy Consumption Data with Time Series Analysis and Machine Learning

CS989: Big Data Fundamentals

Athanasios Mitselos

Glasgow, 11th November 2019

Contents

List of Figures

List of Tables

1 Introduction

The energy sector at the moment is developing at an exponential rate mainly due to the digitalisation of its infrastructure and the penetration of renewable resources. However that growth brings many new challenges that could be addressed with a balance of innovative engineering and extensive data analysis. The planning of generation is essential to meet the consumer demand and that could be dramatically improved with data analytics tools, while also consumers can also contribute to a more stable and reliable network by scheduling their consumption needs in order to lower the overall peak demand. This report aims to provide data-driven answers relevant to the industry and consumers.

2 Description of Datasets

UK Power Networks captured the consumptions of 5,567 consumers in London from November 2011 until February 2014 for the Low Carbon London project. The energy consumption dataset consisted of 168 csv files. The combined size was 10.3GB containing approximately 167 million rows and 6 columns.

The consumptions were measured in half hourly intervals and every consumer was classified by the Acorn group and type of tariff as seen on Table 2-1. To be more specific the description of the columns were:

LCLid: a unique identifier for every consumer

stdorToU: an identifier for the type of tariff charge

DateTime: the date and time

KWH/hh: the energy consumption measured in kWh per half hour

Acorn: the CACI Acorn group

Acorn_grouped: the CACI Acorn categories

LCLid	stdorToU	DateTime	KWH/hh	Acorn	Acorn_grouped
MAC000002	Std	10/12/2012 0:30	0	ACORN-A	Affluent
MAC000002	Std	10/12/2012 1:00	0	ACORN-A	Affluent
MAC000002	Std	10/12/2012 1:30	0	ACORN-A	Affluent

Table 2-1:Sample of consumption dataset

Two different types of tariffs existed in the dataset, the standard and the dynamic time of use (dToU). Consumers with standard tariffs were given fixed prices (14.23p/KWh) for their consumption, while consumers with dToU were informed one day ahead for the price signals of the next day. There were three different price signals: High (77.2p/kWh), Low (3.99 p/kWh) or Normal (11.76p/kWh). dToU charges were applied to 1100 customers for the calendar year 2013, while the remaining consumers were issued with standard tariffs [3].

TariffDateTime	Tariff
1/1/2013 0:00	Normal
1/1/2013 0:30	Normal
1/1/2013 1:00	Normal

Table 2-2:Sample of tariff dataset

Along with the energy consumption dataset UK Power Networks published the price signals of the dToU tariff. This dataset consists of 17520 rows which represented the tariff for every half hour of 2013 (Table 2-2). Only two columns existed for this dataset:

TariffDateTime: the date time

Tariff: the price signal

3 Initial Analysis

The aforementioned datasets were merged forming an even larger dataset with noisy consumptions and millions of instances. For this reason data cleansing and an initial analysis were mandatory.

3.1 Data Cleansing

There were four key challenges that were addressed during the data cleansing phase which will be described below:

Dataset size: The size of the dataset was not manageable with a personal computer. Attempts to load it as a whole resulted to memory errors and thus a decision was made to filter only data for the year 2013 which was the only year that included consumers with both standard and dToU tariffs resulting to the dataset named Raw 2013 (Table 3-1).

Duplicated values: Duplicate rows were removed because they added redundancy.

Corrupted values: By casting the correct data type to each column the corrupted values were identified and removed since they represented false time intervals.

Missing values: Another decision had to be made on whether the missing values should be neglected or extrapolated [6]. Given that the dataset was still too large to handle, the consumers with missing values were removed. Consolidated 2013 (Table 3-1) was the name of the dataset that was used further on.

File Description	Size (GB)	Entries	Consumers	Read (s)
Initial dataset	10.3	167000000	5567	NA
Raw 2013	6.17	93087837	5528	197.25
Consolidated 2013	1.55	19832640	1132	27.97

Table 3-1:Description of datasets

3.2 *Exploration of Dataset*

Before applying any predictive methods it was vital to summarise the dataset and identify its key insights that would act as reference points for further analysis. Based on the overall segmentation of consumers on Figure 3-1, consumers with dToU were considerably less than those with standard tariffs. The largest group was the Affluent.

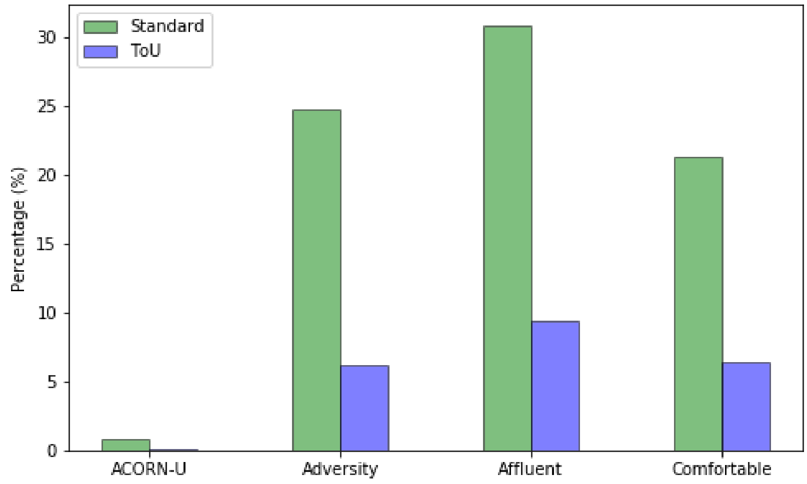

Figure 3-1: Percentage of consumers based on tariffs and Acorn categories

3.2.1 Demand Analysis

The dataset was then aggregated based on daily demand producing a vector of 365 instances. Based on the histogram on Figure 3-2 it is obvious that during the winter the daily demand was the highest and during summer the lowest with narrow range. The range and the distributions of the daily demand in fall and spring were similar.

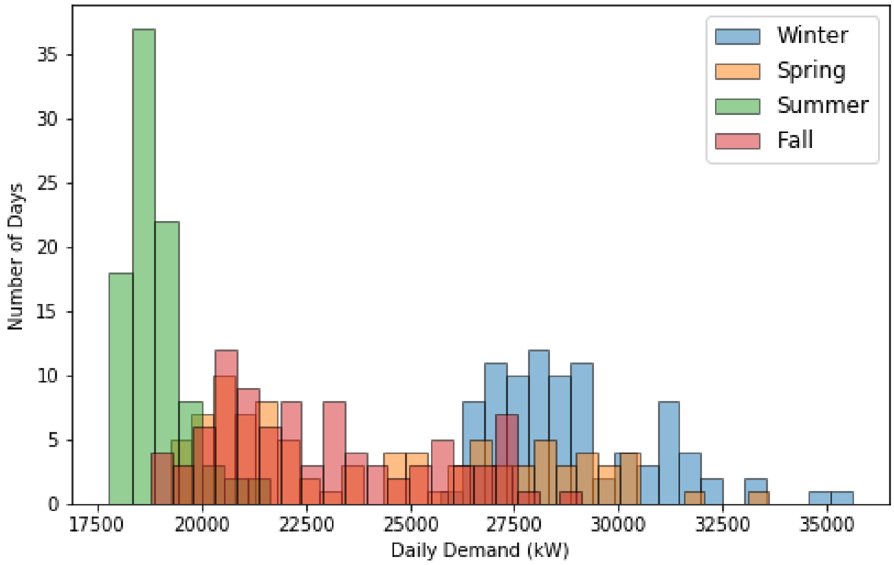

Figure 3-2: Daily demand per season

Trends also occurred on an hourly basis. The means were obtained from the hourly demand for each day of the week. The result was a *AxB* matrix with *A* being the days of the week (7) and *B* being the hours in a day (24). As shown on Figure 3-3 weekends had the highest demand overall while peak demand occurred daily from 17:00 until 21:00. As expected the lowest demand occurred during the early morning, while moderate demand during the working hours on weekdays.

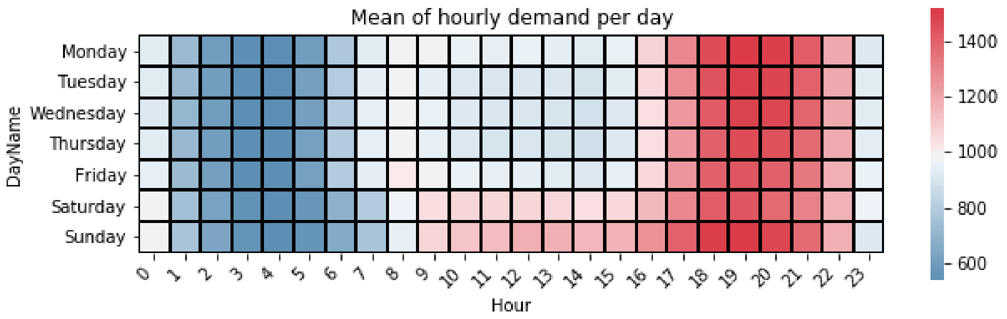

Figure 3-3: Hourly mean demand per day

3.2.2 Consumer Tariffs

Interesting insights were also discovered when consumers with different tariffs were compared. On Figure 3-4 the average hourly demand during a random day (19/01/2013) can be seen for consumers with standard and dToU tariff. The green and red background indicate the timeframe during which dToU consumers had reduced (Low) and increased (High) tariff, respectively. Consumers with dToU consumed more energy on average during the hours with decreased tariff and less energy during the hour with increased tariff compared to consumers with standard tariff.

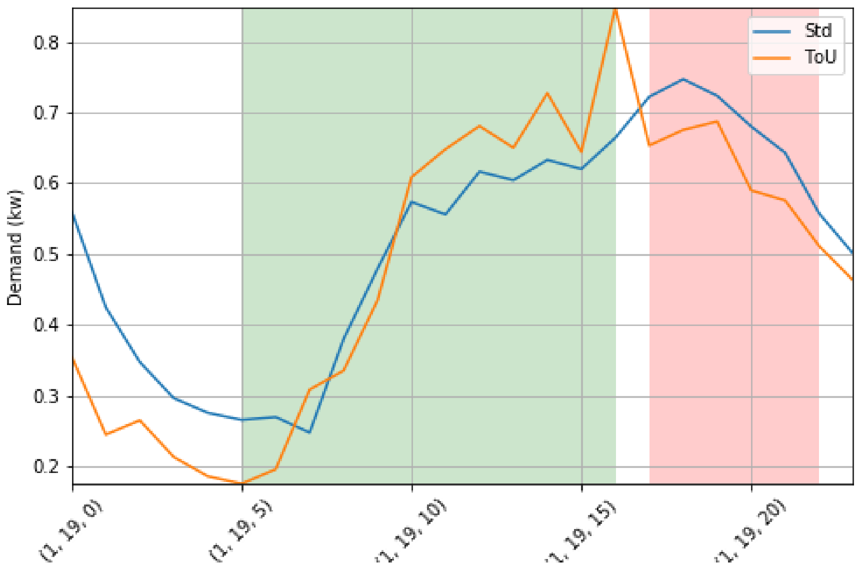

Figure 3-4: Demand for consumers with standard and dToU tariff

Overall it can be observed that consumers with standard tariffs consumed more based on the average hourly consumption (Figure 3-5). However it was clear that during High signals the difference was more sensible with lower median, range and quartiles for consumers with dToU.

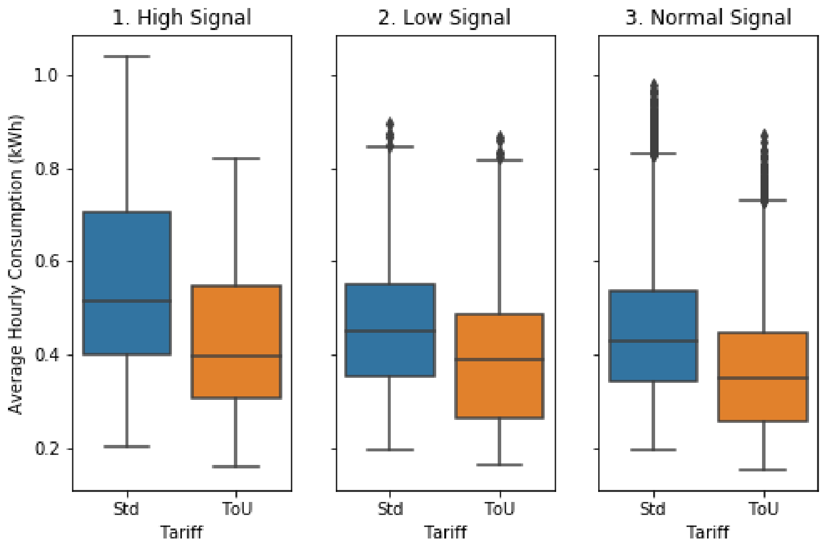

Figure 3-5: Average hourly consumptions per signal

3.3 Consumer Groups

The last part of this section was to explore average daily demand based on the Acorn group (Figure 3-6). Excluding ACORN-U which consists of unknown consumers the other groups' demand were aligned with their classification with Affluent consumers having the highest average daily demand and Adversity consumers having the lowest.

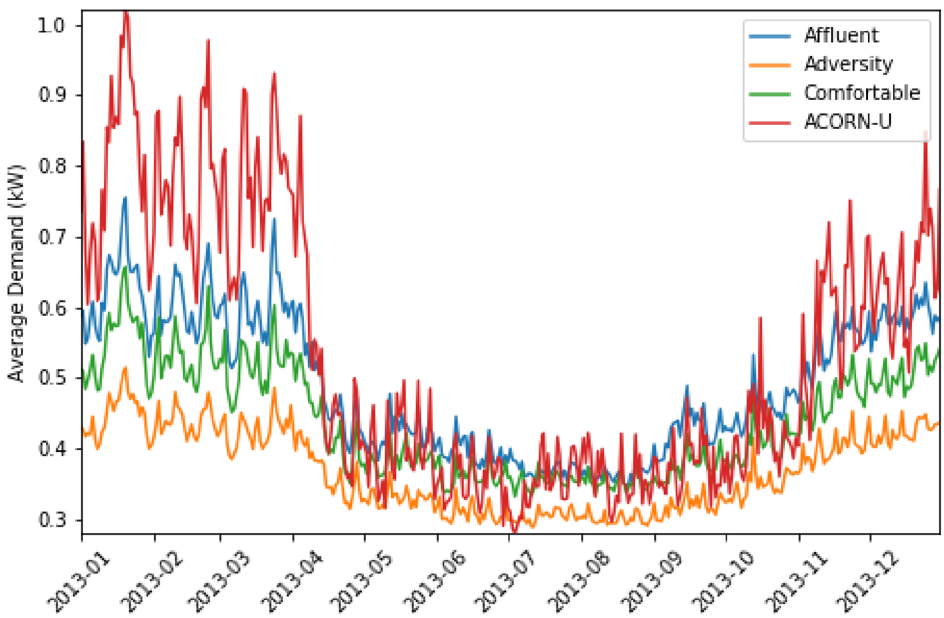

Figure 3-6: Average demand per group

4 Identification and Description of Key Challenges to be addressed

The aim of this report was to cover two different points of view by providing insights to energy suppliers and consumers. The first challenge was to forecast energy demand in short timeframes and the second was to inform consumers about the tariffs that they were offered and the impact on their bills and consumptions.

Scheduling generation in order to supply consumer needs always used to be a difficult task, but today it has become one of the crucial challenges that the energy sector has to face [7]. For this reason statistical models were utilised to explore the performance and accuracy of forecasting weekly demand, which will be covered on the 5th Section.

Given that the industry is changing it made sense to explore how that may affect consumers as well. Some consumers were offered dToU tariffs, which encouraged them to decrease their peak demand by using electricity on off-peak times (noon, morning). Two questions were instantly raised: Did consumers change their consumption patterns based on their tariff and if yes did they eventually pay less? These questions will be answered on the 6th Section.

5 Walk Forward Forecasting with SARIMA

Seasonal Autoregressive Integrated Moving Average (SARIMA) was utilized to forecast demand because it was known to perform well with seasonal data.

5.1 Configuration of Parameters and Evaluation of Training

The first part of the process was to fine-tune two different sets of parameters for SARIMA. The first defined the trend elements (p, d, q) and the second the seasonal elements (P, D, Q, m). The degree of differencing (d) was estimated with ADF test, while the degree of seasonal differencing (D) was estimated with a Canova-Hansen test to make sure that each training set was stationare [8]. The periodicity was expected to be capture on a daily basis and for this reason the number of instances for each seasonal cycle (m) was set to 24 corresponding to the hours of a day. The remaining parameters were configured automatically with stepwise searches with respect to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) metrics in order to minimise the mean square error of prediction without disregarding the sample size [1].

5.2 Evaluation of Training

The consumptions were aggregated to get the hourly demand for a year. Walking forward validation was used instead of k fold validation since it is considered as the appropriate validation technique for time-series forecasting [2]. Walking forward validation was implemented with a fixed size sliding window of 3 weeks for each training to forecast the next week’s consumption. More specifically on every iteration a vector of 504 values was trained to forecast a vector of 168 values (75% training, 25% testing). A full scan was completed after 50 iterations resulting in 50 trained models and 50 tested sets.

From Figure 5-1 it can be observed that the curves of the evaluation metrics follow a positive quadratic trend. The background colors are segmenting the different seasons starting from January to December. From Table 5-1 it is clear that sample size the metrics had similar values and that the accuracy on the training was consistent with relatively small standard deviation compared to the mean.

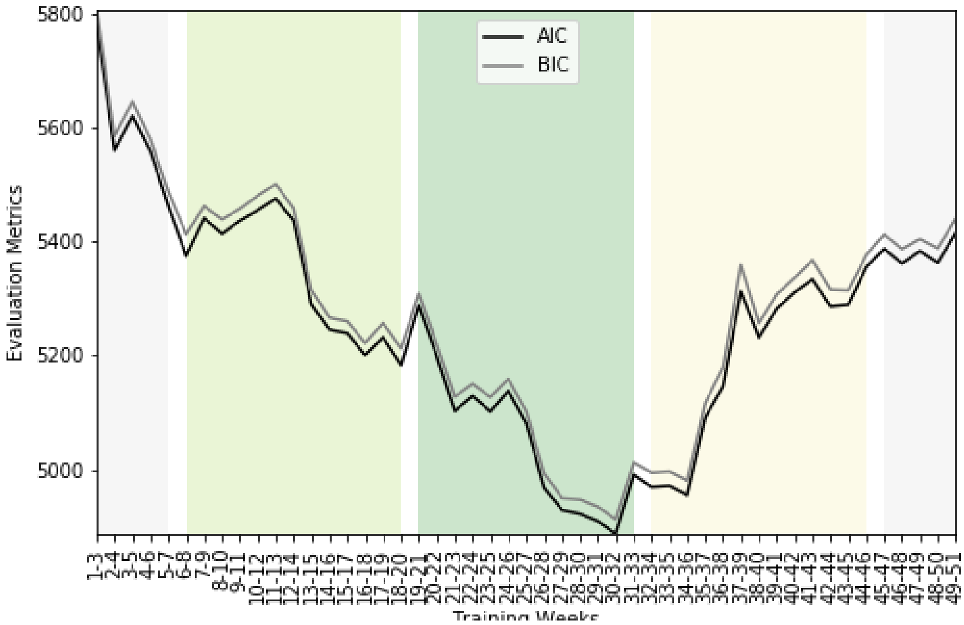

Figure 5-1: Evaluation curves for SARIMA models

Evalualtion	Mean	Standard Deviation
AIC	5255.399	205.2159
BIC	5280.568	205.5316

Table 5-1: Evaluation statistics for SARIMA models

5.3 *Validation of Testing*

As soon as the models were obtained the next step was to test their accuracy on the test sets. The objective was to obtain not only the overall accuracy of weekly demand but also accuracy of detecting the level and time of the daily peak. On the Figure 5-2 the actual and forecasted demand can be seen as well as the confidence interval for the fourth week. As seen on Figure 5-2 the model2 had a difficulty on predicting the demand during the noon.

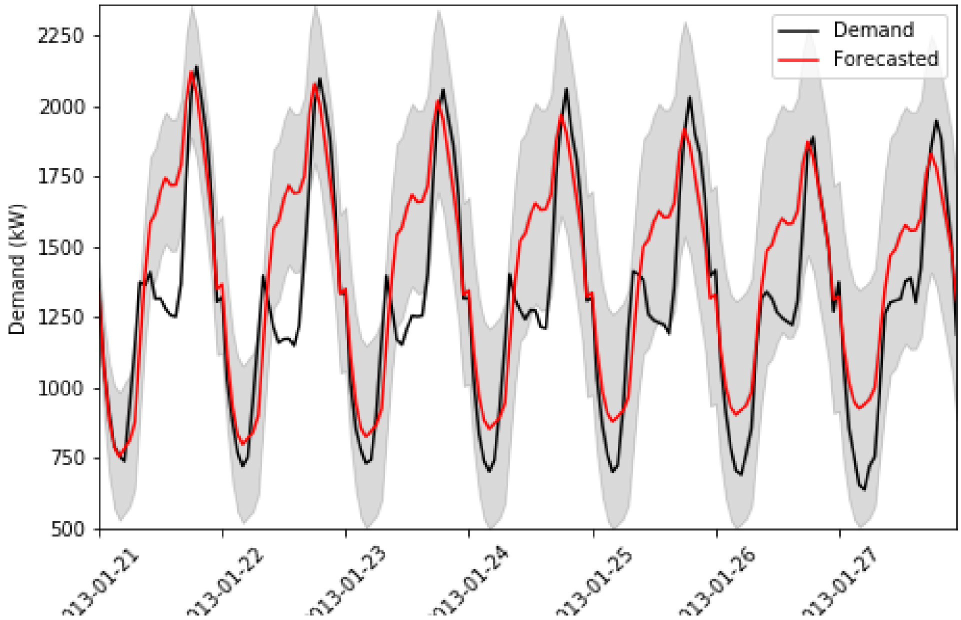

Figure 5-2: Actual vs forecasted demand for the one test week

For the overall accuracy of the models Mean Absolute Error (MAPE) and R-Squared were used in order to understand both the percentage errors and the goodness of fit [9]. Based on the Table 5-2 it can concluded that the average R-Squared and MAPE reached 75% and 12.28% respectively proving that generally the training was relatively successful with a balanced tradeoff between bias and variance. An interesting fact was that the R-Squared had a positive trend as the testing continued (Figure 5-3).

Overall	Mean	Standard Deviation
MAPE	12.28	3.07
R-Squared	75.21	8.73

Table 5-2: Validation statistics for overall forecast

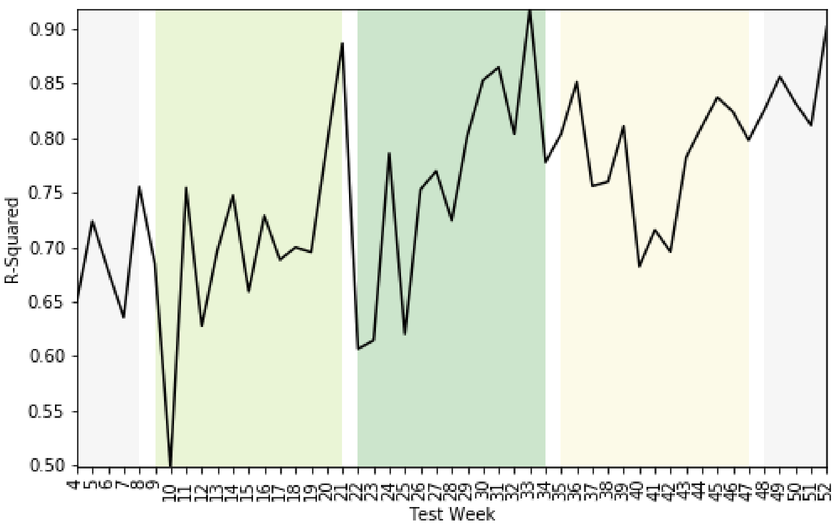

Figure 5-3: R-Squared curve

Apart from exploring the overall accuracy it made sense to look deeper into the level and the hourly offset of the actual and forecasted peak measured with MAPE and (Mean Absolute Error) MAE respectively. The reason behind using MAE was that MAPE is known to result to errors when the results approach zero. Based on Table 5-3 and Figure 5-4 it can be seen that the MAPE for the peak is smaller than the MAPE for the weekly forecast, which means that the model captured the peaks with higher accuracy. If these results are combined with the results shown at Table 5-4 it is clear that the average hourly offset was approximately half an hour on average.

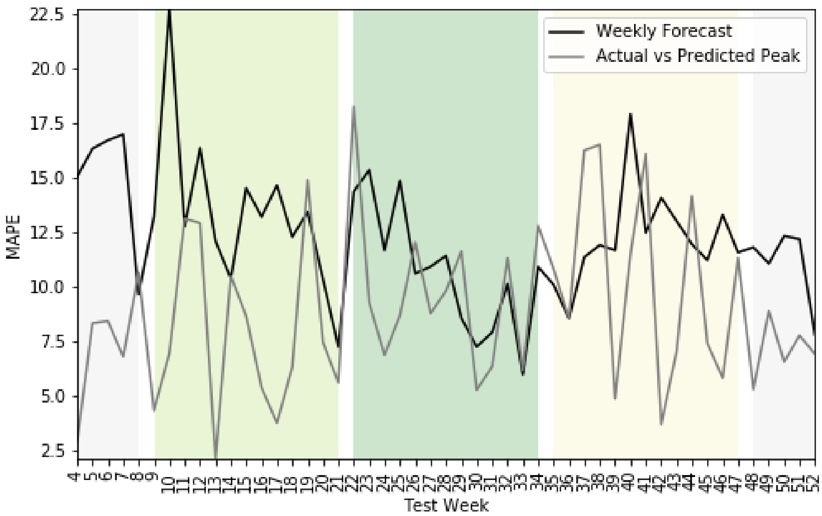

Figure 5-4: Weekly vs peak forecast

Actual vs Forecasted Peak	MAPE
Mean	8.9
Standard Deviation	3.78

Table 5-3: Validation statistics for actual vs forecasted peak demand

Peak Offset	MAE
Mean	0.47
Standard Deviation	0.48

Table 5-4: Validation statistics for hourly peak offset

6 Comparison of Consumptions among Consumers with Different Tariffs

Given that the signals for different tariff charges were provided it made sense to explore if consumers with different tariffs had different consumption patterns and whether consumers with dToU tariff paid less than the consumers with standard tariff.

6.1 Segmentation of consumer behavior with K-Means and TSNE

To further explore this subject K-Means was used to group consumers based on their hourly load profiles in order to compare consumers with similar consumption patterns (shape). Every instance of the dataset (1132) included the hourly demand of a consumer and every feature (8760) the consumption of all consumers during a specific hour. Each instance was scaled to [0, 1] to increase the efficiency of the algorithm and to reduce the numeric differences among consumers [4]. Silhouette score and Calinski-Harabasz index were obtained to decide on the number of clusters and by observing the curves on Figure 6-1 it is clear that three clusters had the highest silhouette score and considerably high Calinski-Harabasz index. Given that the dataset is expected to have some outliers three clusters were chosen in order to capture some anomalies.

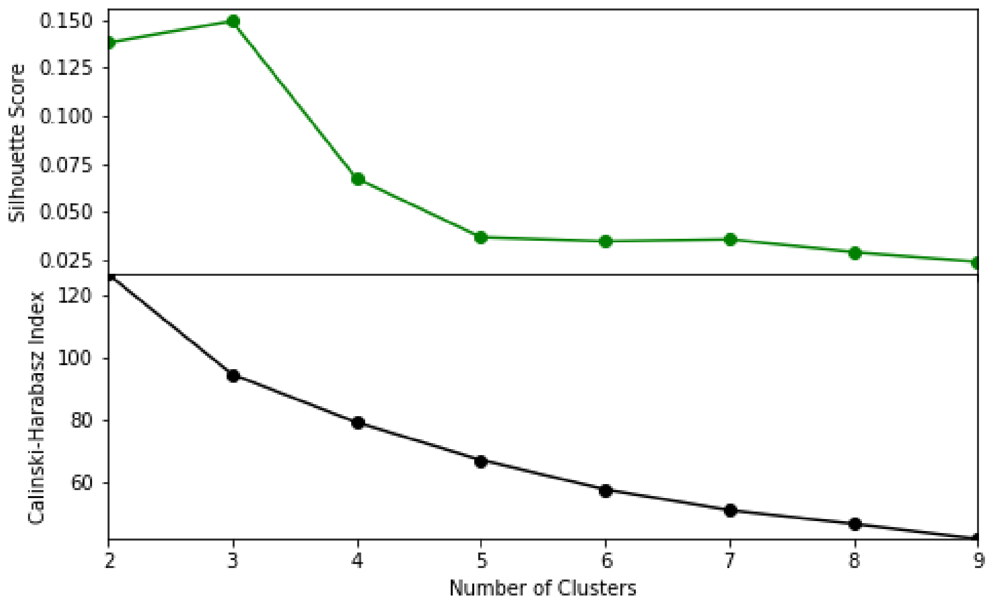

Figure 6-1: Silhouette and Calinski-Harabasz curve

The scaled dataset was the input for the K-Means algorithm and TSNE was used to reduce the dimensions of the dataset to visualise the results. TSNE was preferred over PCA due to its probabilistic nature and its performance on high dimensional data [5]. The results of the K-Means and TSNE algorithm are shown on Figure 6-2 and some statistics extracted from the hourly mean consumption of each cluster can be observed on Table 6-1. From the Figure 6-3 combined with Table 6-1 it can be seen that Cluster 2 consists only of consumers with standard tariffs and increased consumptions. For this reason they were excluded for the remaining analysis as outliers.

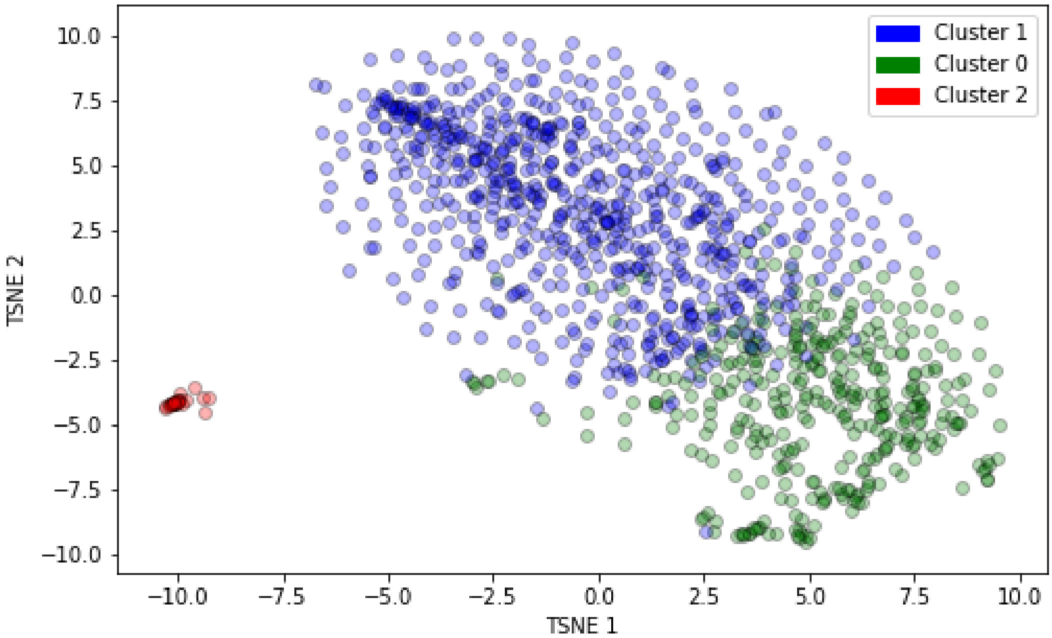

Figure 6-2: K-Means results with TSNE

Cluster	Tariff	Count	Average Consumption	Average Bill
0	Std	315	5587.699	795.13
	dToU	57	4601.593	663.543
1	Std	538	2900.401	412.727
	dToU	194	2840.719	405.776
2	Std	28	6980.668	993.349

Table 6-1: Statistics for clusters

Cluster 1 was the largest cluster and its consumers had the lowest average consumption. At first glance it looks like consumers with dToU paid less (yearly basis) but considering that they also consumed less the percentage difference was obtained. For the bill it was almost 1%, but for the consumption it was 2% which means that there is no significant reason to believe that they paid less on a yearly scale.

Cluster 0 consisted of consumers with higher consumptions than Cluster 1 but not as high as Cluster's 2. Again consumers paid less but they also consumed less.

The consumption percentage difference was 19% while the bill percentage difference was 18%, which also doesn't prove that they had significant differences.

6.2 Hypothesis testing

Given that there wasn't enough evidence in place to support that consumers with dToU paid less hypothesis testing was introduced to make sure that the average consumption of each group came from different distributions. Wilcoxon non-parametric test was used because the samples were paired and non-normal (Shapiro test). The P-Value indicated that the sample distributions were not the same, thus the mean consumption of each cluster and tariff was indeed different.

6.3 Comparison of bills and consumptions

At this point it would be helpful to observe the average bills and consumptions based on different tariffs for the existing tariff signals. The summation of the average consumption of consumers with dToU is smaller for high signals and larger for low signals for Cluster 1 (Figure 6.3.3). On the other hand consumers of Cluster 0 didn't have the same patters, which show that the clustering was successful. The impact of the pricing can observe on Figure 6-3.2 and 6-3.4, which illustrates how differently the consumers were priced. That shows that the consumers took seriously into account the pricing but on average no significance discount was achieved.

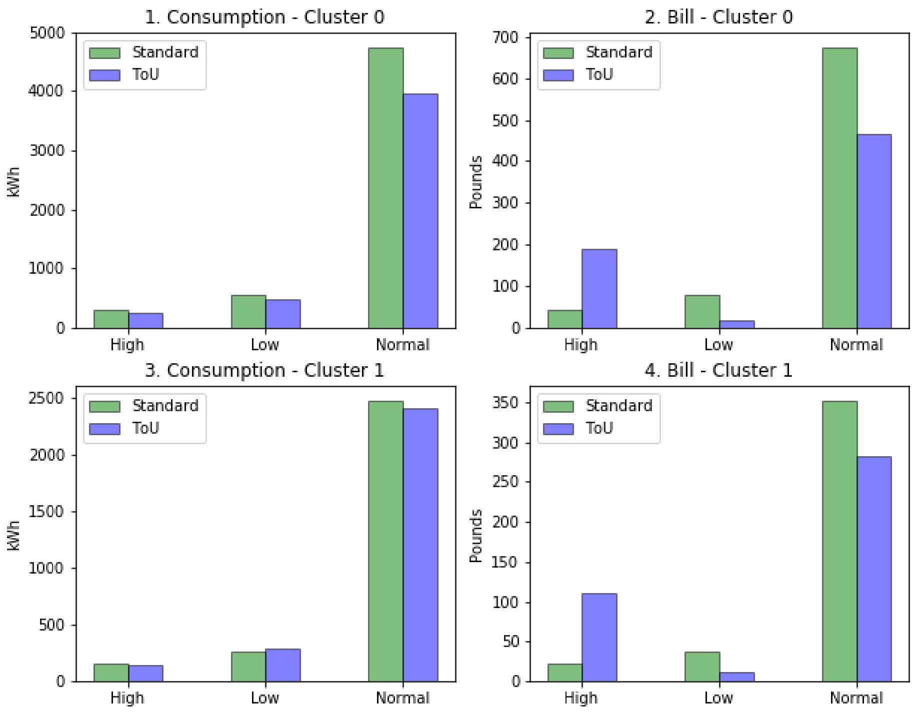

Figure 6-3: Sums of demand per cluster and tariff

7 Reflection on methods

Due the size of dataset the analysis was rather time consuming and resource hungry. Syntax, logic and runtime errors had a significantly negative impact on the process of the report.

Time series analysis packages were explored and standard packages like linear regression and ARIMA failed to capture the seasonality of data. The accuracy of the selected package (SARIMA) could be improved by adding exogenous features such as the temperature or the time. Temperature would adjust the levels of consumption while time would adjust the seasonality. Apart from that it would be reasonable to test different seasonal cycles to focus on the weekly demand rather than the daily. Unfortunately attempts to change the seasonal cycle resulted to memory errors. The reason behind using different seasonal cycles lies on the fact that the model was influenced by weekends resulting on forecasting higher consumption for the noon.

K-Means was used mainly because it is a fast and straightforward algorithm. More sophisticated clustering algorithms (DBSCAN, HDBSCAN) would need considerably more time to train and test but they would address the challenges of selecting number of clusters and identifying outliers. Although the clustering performed well on segmenting consumers with different patterns, there was no way to evaluate it because it was qualitative in nature (shape of consumptions) and labels existed. PCA and TSNE were tested to reduce the dimensionality of data with TSNE providing more insightful results that led on identifying outlying consumers.

If there was more time and resources it would be interesting to further explore the existing questions with different algorithms and approaches and also to test whether classification algorithms could identify the Acorn groups.

8 Conclusions

Concluding the SARIMA models performed relatively well on the test set although there was a difficulty on forecasting the hours during the noun. That may be caused because the interval of seasonality was set to 24 (day) instead of 168 (week). Although different challenges were addressed the results indicate that forecasting with statistical packages is a viable option to solve the problem of short time energy consumption forecasting.

The clustering with the dimensionality reduction algorithm allowed to group consumers based on their consumption patterns, to compare them and to identify anomalies. Consumers with different tariffs had similar consumption and bills, but with different underlying distributions. Consumers from only one cluster consumed with respect to signals. However their bills were not significantly lower, concluding that on average consumers were not enough to lower their bills.

9 References

[1] Brewer, M.J., Butler, A. and Cooksley, S.L., 2016. *The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity*. *Methods in Ecology and Evolution*, *7*(6), pp.679-692.

[2] Brownlee, J. (2016). *How To Backtest Machine Learning Models for Time Series Forecasting*. [online] Medium. Available at: https://towardsdatascience.com/scale-standardize-or-normalize-with-scikit-learn-6ccc7d176a02 [Accessed 17 Oct. 2019]

[3] Data.london.gov.uk. (2019). *SmartMeter Energy Consumption Data in London Households – London Datastore*. [online] Available at: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households [Accessed 14 Oct. 2019].

[4] Hale, J. (2019). *Scale, Standardize, or Normalize with Scikit-Learn*. [online] Medium. Available at: https://towardsdatascience.com/scale-standardize-or-normalize-with-scikit-learn-6ccc7d176a02 [Accessed 17 Oct. 2019].

[5] Maaten, L.V.D. and Hinton, G., 2008. *Visualizing data using t-SNE*. *Journal of machine learning research*, *9*(Nov), pp.2579-2605.

[6] Peppanen, J., Zhang, X., Grijalva, S. and Reno, M.J., 2016, September. *Handling bad or missing smart meter data through advanced data imputation*. In *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)* (pp. 1-5). IEEE.

[7] Shcherbakov, M.V., Brebels, A., Shcherbakova, N.L., Tyukov, A.P., Janovsky, T.A. and Kamaev, V.A.E., 2013. *A survey of forecast error measures*. *World Applied Sciences Journal*, *24*(24), pp.171-176.

[8] Smith, T. (2018). *6. Tips to using auto_arima — pmdarima 1.4.0 documentation*. [online] Alkaline-ml.com. Available at: https://www.alkaline-ml.com/pmdarima/tips_and_tricks.html [Accessed 9 Nov. 2019].

[9] Tugend, A. (2017). *The Challenges for the Energy Industry*. [online] Nytimes.com. Available at: https://www.nytimes.com/2017/10/15/business/energy-environment/challenges-for-the-energy-industry.html [Accessed 4 Nov. 2019].

10 Appendix

10.1 Development Details

Overall Specifications

- **Processor:** Inter(R) Core(TM) i5-6300 CPU @ 2.40-2.50GHz
- **RAM:** 8GB
- **Operating System:** Windows 10 Pro
- **Python version:** 3.7.2

Python Packages

- **IDE:** jupyter lab
- **Aggregations:** pandas
- **Mathematic calculations:** numpy
- **Machine learning:** sklearn
- **Statistical models:** pmdarima and statsmodel
- **Hypothesis Testing:** scipy
- **Model format:** pickle
- **Date Manipulation:** datetime and calendar
- **Visualisations:** seaborn and matplotlib

10.2 Hypothesis Testing Tables

cluster	stdorToU	Tariff	Shapiro F	Shapiro P
0	Std	High	0.977	0
0	Std	Low	0.954	0
0	Std	Normal	0.953	0
0	ToU	High	0.971	0
0	ToU	Low	0.958	0
0	ToU	Normal	0.955	0
1	Std	High	0.47	0
1	Std	Low	0.604	0
1	Std	Normal	0.524	0
2	Std	High	0.963	0
2	Std	Low	0.962	0
2	Std	Normal	0.95	0
2	ToU	High	0.971	0
2	ToU	Low	0.963	0

Table 10-1:Shapiro test

Cluster	High		Low		Normal	
	Statistic	P value	Statistic	P value	Statistic	P value
0	59	0	18922	0	413746	0
1	4655	0	81884	0	8788755	0

Table 10-2:Wilcoxon test