

WATERFORD INSTITUTE OF TECHNOLOGY

APPLYING DATA MINING
TECHNIQUES TO MONITORING
DATA FROM A DATA CENTRE
TO IMPROVE ITS ENERGY
EFFICIENCY

Masters Dissertation

Student

John Fitzpatrick,
02323826

Supervisor

Dr. Bernard Butler

August 29, 2019

Abstract

The growing prevalence of Data Centres and their associated services to commercial and individual consumers is an underlying feature of today's technological landscape. Therefore, as the energy consumed by these Data Centres is growing exponentially, efforts must be focused to maximise energy efficiencies in Data Centres in light of an ongoing global energy crisis. Existing literature in this regard focuses on physical improvements to server, networking or energy components as well as new management support systems. This research identified a gap whereby the Data Centre is treated as an entity in its own right, combining its own operational monitoring data, that's collected on energy, weather and cooling systems and combining this with modern Data Mining techniques to uncover hidden insights that are richer and not seen from simple monitoring data. These insights could relate to predicting or explaining unexplained spikes in energy consumption or for explaining anomalies. This research specifically focused on the energy efficiency of the Chiller system in the WIT-TSSG Data Centre in Waterford. In particular, the research focused on the period of time before and after the operational Chiller alternated from Chiller 2 to Chiller 1, which saw unexpected spikes in Energy Consumption of Chiller 1. Correlation and Linear Regression analysis was carried out on the related variables (energy consumption of cabinets, weather information and cooling metrics) to identify the relationship between these variables and the operating chiller from week to week. Through this process, this research then identified certain variables that could indicate a potentially underlying issue, with the caveat that the full interpretation of these results will need input from Data Centre Engineers who understand the context of the underlying variables.

Acknowledgements

Firstly I would like to thank Dr. Bernard Butler for all his support, guidance, knowledge and crucially patience in helping me complete this research. Without the skills and genuineness that Dr. Butler has brought to this project, it would not have been successfully completed.

I would also like to thank John Ronan and Michael Kugel from the WIT-TSSG Data Centre for their time in meeting with me on multiple locations and for providing so much information on the Data Centre itself.

Finally, I would like to thank my wife Sharon and daughter Lily for supporting me over the last year while this work was being completed. Again, without your support, this project would not have been possible.

Contents

1. Introduction	1
1.1. Motivation	1
1.2. Research Problem Statement	1
1.3. Aims and Objectives	2
1.4. Research Questions	3
2. Literary Review	3
2.1. Introduction	3
2.2. Data Centre Infrastructure and the WIT-TSSG Data Centre	4
2.2.1. What is a Data Centre and How Much Energy do they Use?	4
2.2.2. The WIT-TSSG Data Centre	5
2.3. Improving Data Centre Energy Efficiency	7
2.4. Data Mining Techniques (for improving energy efficiency in buildings) . .	9
2.5. Summary	11
3. Methodology	12
4. Research Methods	14
4.1. Data Available	14
4.2. Exploratory Data Analysis	15
4.3. Research Techniques	18
4.3.1. Introduction	18
4.3.2. Import Data From Data Centre	18
4.3.3. Understanding the Data	19
4.3.4. Manipulate Data in Python to Generate Correlations for all Vari- ables	20
4.3.5. Analyse Correlations	21
4.3.6. Meet with Data Centre Administrators	21
4.3.7. Identify Micro-timelines to Analyse Chiller Performance	22
4.3.8. Identifying Variables to use in a Regression Model	23
4.3.9. Construct Regression Model	24
4.3.10. ‘Goodness of Fit’ Measures	24
5. Presentation of Findings	25
5.1. Correlation Analysis of Chiller 1 versus Chiller 2	25
5.2. Regression Analysis of Chiller 1 versus Chiller 2	26
6. Discussion of Findings	27
7. Conclusion	29
References	31

Appendices	34
A. MySQL Queries for Exploratory Data Analysis	34
B. MySQL Queries for Detailed Data Analysis	35
C. Python Code for Data Manipulation	36
D. Python Code for Regression Model	45
E. Definition of Key Variables	50
F. Detailed Correlation Analysis Results for Overall Data Centre	50
G. Detailed Correlation Analysis Results for each Chiller	52
H. Full list of Independent Variables for Regression Model	53
I. Summary of the Output of the Regression Model	54

List of Figures

1.	How Energy is Consumed in a Data Centre.	4
2.	Energy Consumption of Data Centres in the EU and US from 2000 to currently.	5
3.	The average PUE across different locations.	5
4.	The average PUE from 2000 to currently.	6
5.	The design of the WIT-TSSG Data Centre.	6
6.	The schema of the WIT-TSSG Data Centre.	14
7.	Energy consumption (in kW usage per 5 minute intervals) of one cabinet in the Data Centre over the entire year of 2017.	15
8.	Energy consumption (in kW usage per 5 minute intervals) of three different server cabinets in the Data Centre over the course of one day (2017-09-13).	16
9.	Energy consumption (in kW usage per 5 minute intervals) of both Chillers and one server cabinet in the Data Centre over the course of 2017	16
10.	Annual water flow (in litres per 5 minute intervals) to LCP 1 measured at 9am every day during 2017	17
11.	Energy consumption (in kW usage per 5 minute intervals) of the Generator before, during and after Storm Ophelia in October 15th 2017	17
12.	Schema of Data and Variables chosen for detailed Data Analysis	19
13.	The Latte Definition of Different Energy Metrics	20
14.	Apparent Energy Consumption of Chiller 1 and 2 in 2017	22
15.	Summary of Analysis from Regression Model	27
16.	Definition of key Variables from the Monitoring Data	50
17.	The Detailed Correlation Analysis Results - section 1	51
18.	The Detailed Correlation Analysis Results - section 2	51
19.	Correlation Analysis of each Chiller for 4/28 to 5/9	52
20.	Correlation Analysis of each Chiller for 6/6 to 6/20	52
21.	Correlation Analysis of each Chiller for 8/29 to 9/12	53
22.	Full list of Independent Variables used in the Regression Model	53
23.	Regression Output from Data from April 25th to May 9th	54
24.	Regression Output from Data from June 6th to June 20th	55
25.	Regression Output from Data from August 29th to September 12th	56

List of Source Codes

1.	MySQL Query for Annual Energy Consumption	34
2.	MySQL Query for Daily Energy Consumption	34
3.	MySQL Query for Annual Energy Consumption of Chillers and Server . .	34
4.	MySQL Query for Annual Water Flowr	35
5.	MySQL Query for Energy Consumption of the Generator during Storm Ophelia	35
6.	MySQL Query for Energy Consumption in 2017	35
7.	MySQL Query for All Meter ID Location	35
8.	MySQL Query for LCP information for 2017	36
9.	MySQL Query for LCP Location Information	36
10.	MySQL Query for Weather Data for 2017	36
11.	Python - Set Up and Configuration	37
12.	Python - Import and Cleanse Energy Data	38
13.	Python - Pivot Data around the meter reading time	39
14.	Python - Merge Energy Pivot Tables together	40
15.	Python - Import and Merge LCP Data	41
16.	Python - Pivot Data around the LCP reading time	42
17.	Python - Merge LCP Pivot Tables together	43
18.	Python - Import and Merge Weather Data	44
19.	Python - Merge all Data Together into one table	44
20.	Python - Get Correlations between all Variables	45
21.	Python - Analyse Correlations in Python	45
22.	Python - Set Up and Configuration for Regression	46
23.	Python - Import and Manipulate data for Regression	47
24.	Python - Create Micro-Timeline and Check that Time Series is Stationary	48
25.	Python - Construct Regression Model and Execute	48
26.	Python - Check Regression Model for Auto-Correlation and Other Tests .	49
27.	Python - Run Cochran-Orcutt Procedure (to eliminate autocorrelation)	50

1. Introduction

The demand for cloud computing services is growing exponentially, driven by the growth of internet users and applications (Mahdavi and Tschudi, 2013). As a result, within the Information and Communications Technology sector (ICT), Data Centres are the fastest growing sector in terms of energy consumption and are now responsible for over 2% of global CO₂ emissions (Avgerinou, Bertoldi, and Castellazi, 2017). With this growth in Data Centre energy consumption over the last 10 years, there arises a cost incentive, along with a general green incentive, to reduce energy consumption or to use energy more efficiently. While there is ample literature available that deals with making Data Centres more efficient by changing the hardware configuration of the Data Centre, this Dissertation documents the progress made on applying Data Mining techniques to the monitoring data, compiled from within a Data Centre, to unlock hidden information to assist system administrators to make the Data Centre more efficient or eliminate energy inefficiencies. Data from a medium sized Data Centre, the WIT-TSSG Data Centre in the Carriganore West Campus of Waterford institute of Technology is being used for this project.

1.1. Motivation

The goal behind making Data Centres more energy efficient is to reduce the energy consumption while delivering the same service to its users. Therefore the motivation that follows from this is twofold.

Firstly, there is the motivation to make the Data Centre more energy efficient for the tangible benefits this brings. If energy consumption is reduced, this will lead to cost savings, while simultaneously reducing the carbon footprint of the Data Centre. Ultimately, this could deliver savings to WIT, allowing for potential investment in other educational projects while helping the environment at the same time.

Secondly, in addition to the overarching motivation above, this research is being motivated by a desire to explore Data Mining techniques to gain insight into Data Centre energy consumption to help make them more energy efficient. While Data Centres are a part component of any cloud computing architecture, the emphasis here is to treat them as an entity in their own right, and use the vast amounts of monitoring data they generate to make them more energy efficient by highlighting energy inefficiencies. There is very little existing literature in this regard and the motivation here is that this study would contribute to the research in this field.

1.2. Research Problem Statement

As outlined in the introduction (section 1), The information and Communication Technology (ICT) sector, and Data Centres in particular, are consuming vast amounts of energy and are contributing to an ever increasing carbon footprint. Avgerinou, Bertoldi, and Castellazi (2017) highlight that the ICT sector will consume 13% of the global electricity supply by 2030.

Therefore, as the energy demand of Data Centres grows due to the changing role of technology in society, this represents a clear challenge to policymakers, researchers and Data Centre stakeholders to seek solutions to make Data Centres more efficient in terms of energy consumption.

A significant amount of research exists that focuses on making Data Centres more efficient, by concentrating on improvements to the physical hardware and infrastructure within a Data Centre. This, as detailed in the literature review, (section 2.3) specifically focused on such measures as:

1. Introducing wireless sensors to capture key metrics;
2. Following best practices and
3. Focusing on specific features within a Data Centre to gain efficiencies, such as the cooling system, server efficiencies, network modelling, or the optimisation of applications at an infrastructural level.

However, this research has a different focus. It analyses existing monitoring data, that is already stored by the Data Centre to seek hidden insights that might make a Data Centre more energy efficient.

1.3. Aims and Objectives

The ultimate aim of this research, as specified in the Research Problem Statement (section 1.2) is to use Data Mining Techniques on the monitoring data compiled at the WIT-TSSG Data Centre to gain hidden insights about energy consumption to help make the Data Centre more energy efficient. Specifically, the following are the aims and objectives of the project:

- Review all existing literature on improving Data Centre energy efficiency.
- Learn about the WIT-TSSG Data Centre and what monitoring data is compiled that can be used for analysis.
- Identify Data Mining techniques from existing literature that can be applied to the monitoring data compiled at the Data Centre.
- Apply Data Mining techniques on the monitoring data from the Data Centre to maybe gain some hidden insights about energy consumption.

If successful with these aims, one objective is to understand when there is a peak in energy demand so that the supply of energy can be better utilized. Another objective arising from ongoing discussions with staff from the Data Centre is to help identify known inefficiencies in the chiller system.

If the project can meet these objectives, it will also have contributed to the academic literature on this subject, while simultaneously providing the Data Centre in WIT-TSSG

with a predictive model to help better utilize the supply of energy at peak demand that in turn can lead to greater energy efficiencies, and assist in pinpointing specific problem areas of energy inefficiency within the Data Centre.

1.4. Research Questions

From monitoring data provided by the WIT-TSSG Data Centre and using a Data Mining based approach,

1. Can we identify features in the data that indicate where the Data Centre is not optimally consuming its supply of energy?
2. Can we identify predictive indicators from the data that forecast an unanticipated peak in energy demand at the Data Centre?
3. Specifically, can we identify energy or cooling features associated with the chiller system that are not energy efficient and could warrant remedial action?

The third Research Question was added in collaboration with system administrators in the Data Centre who had believed for some time that a potential issue was apparent in the Chiller system. While existing diagnostic techniques could not confirm any theory, feedback was provided in consultations with them, that this could be an area to concentrate on.

2. Literary Review

2.1. Introduction

The literature review can be divided into three concepts. The first is an introduction to what Data Centres are, how they work and specifically an overview of the WIT-TSSG Data Centre in Carriganore in Waterford (section 2.2).

The next concept looks at the literature around improving the energy efficiency of Data Centres (section 2.3). In most cases, the focus is on reducing energy consumption by improving the efficiency of different hardware components of a Data Centre. Specifically, that research looks at either the server, networking or cooling systems with an emphasis, by multiple authors, on implementing a wireless sensor system to capture key metrics across a Data Centre and using this to pinpoint issues in real-time.

Other work in this area presents research on optimising resources or developing new scorecards to identify efficiencies in the management of Data Centres.

The final concept looks at research on Data Mining techniques used to increase energy efficiency in buildings (section 2.4). There are numerous case studies using ‘Big Data’ from building automation systems, to be harnessed in a Data Mining approach to identify hidden insights to improve the energy efficiency of different buildings. All three concepts are explored in detail below, before a summary ties all three concepts together and identifies the research gap that this project is investigating.

2.2. Data Centre Infrastructure and the WIT-TSSG Data Centre

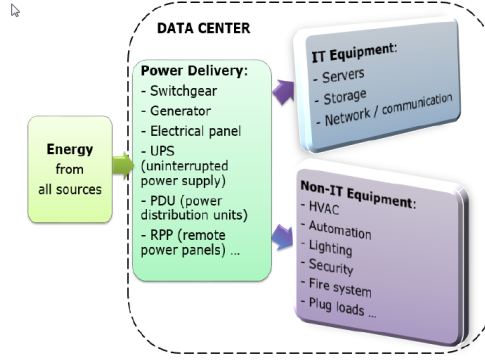
2.2.1. What is a Data Centre and How Much Energy do they Use?

In the modern economy, the demand for cloud storage and cloud computing based applications is growing exponentially as the number of internet users who rely on digital information via cloud services continues to sky-rocket across the world (Vasudevan et al., 2017). To facilitate the demand in growth for such services, it follows that Data Centres, and the facilities they provide, will also be more in demand.

Levy and Raviv (2018) define a Data Centre as "a dedicated facility with all the resources required for storage, processing and sharing digital information and its supported areas". The hardware required for such a facility comprises power distribution, environmental control systems, telecommunications, security and information technology (I.T) equipment (Levy and Raviv, 2018).

Figure 1 from (Levy and Raviv, 2018) outlines how energy is typically consumed in a Data Centre, specifically how it is channelled to IT equipment and non IT equipment.

Figure 1: How Energy is Consumed in a Data Centre.



Data Centres consume forty times more energy than conventional office buildings (Levy and Raviv, 2018) and have the fastest growing carbon footprint across the whole ICT sector (Avgerinou, Bertoldi, and Castellazi, 2017).

Figure 2, reproduced from Avgerinou, Bertoldi, and Castellazi (2017) shows the rising energy consumption of Data Centres across the EU and the US over the last 15 years as the number of Data Centres have grown.

With Data Centres consuming more energy, there is an obvious incentive, both from a cost and a green perspective, to reduce energy consumption per Data Centre. This can be done by making Data Centres more efficient for the energy they consume and the measurement that is used to measure energy efficiency is the *Power Usage Efficiency* (PUE) score. This is described as

$$\text{Power Usage Effectiveness} = \frac{\text{Total Facility Energy Usage}}{\text{IT Equipment Usage}}.$$

Lower PUE scores are better, and a PUE close to one indicates that the Data Centre is more efficient as it consumes less energy on non IT equipment such as cooling. In terms

Figure 2: Energy Consumption of Data Centres in the EU and US from 2000 to currently.

Consumption (TWh)	Reporting Year
EU consumption	
18.3	2000
41.3	2005
56	2007
72.5	2010
104	2020
US consumption	
91	2013
140	2020
Global consumption	
216	2007
269	2012

of the current trends of energy usage, research by Avgerinou, Bertoldi, and Castellazi (2017) show that Data Centres in Northern Europe have a lower PUE score than those in southern Europe, which is expected given the geographical and climate difference. Also, the same authors find that the PUE scores in general have been decreasing year after year as demonstrated by the graphs in Figures 3 and 4.

Figure 3: The average PUE across different locations.

Table 3. Geographical zoning with temperature and relative humidity average data.

Geographical Zones	Countries	Temperature Range (°C)	RH Range (%)	Average PUE	Number of Data Centres
Nordic countries	Denmark, Finland, Norway, Sweden	18–26	20–80	1.71	13
UK and Republic of Ireland	England, Scotland, Wales, Northern Ireland, Republic of Ireland.	17–30	8–80	1.83	116
Northern/Central Europe	Austria, Belgium, France, Germany, Hungary, Luxembourg, The Netherlands, Portugal, Poland, Switzerland	14–28	16–75	1.72	122
Southern Europe/Mediterranean	Gibraltar, Greece, Italy, Malta, Spain, Turkey, Monaco, Romania, Bulgaria	16–26	20–80	2.00	30
Non EU	Republic of Mauritius, US	-	-	-	5

So the challenge, as outlined in the various literature below (section 2.3), is how to make a Data Centre more energy efficient by eliminating energy inefficiencies within the IT equipment to reduce the consumption of energy overall.

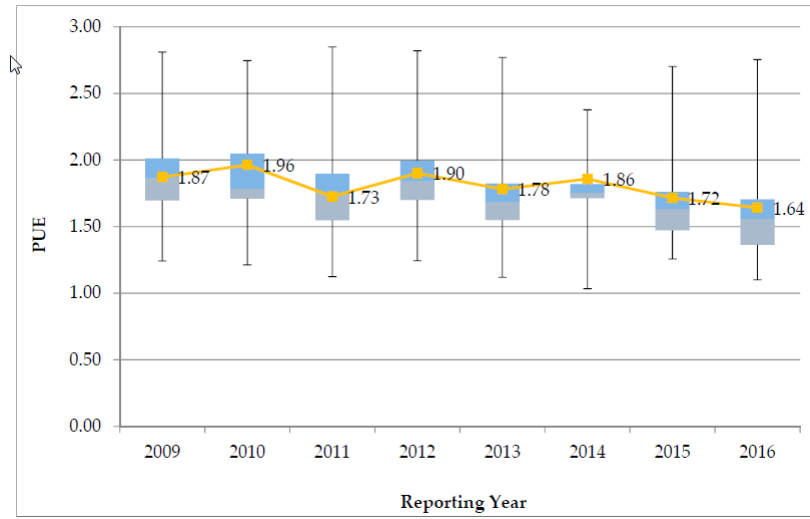
2.2.2. The WIT-TSSG Data Centre

The WIT-TSSG Data Centre is located on the Carriganore West Campus of the Waterford Institute of Technology (WIT) on the west side of Waterford city. From this location, the Data Centre supports over 50 concurrently active ICT research projects through provisioning of internet services, Cloud Computing resources, an AI cluster and project bespoke test beds (*TSSG Infrastructure Group* 2018).

In particular, the Data Centre supports TSSG projects, the Higher Education Authority, The Irish Centre for High End Computing(ICHEC), hosting its super computer Kay, Arclabs and other 3rd party commercial arrangements.

In terms of its energy usage, the Data Centre has an IT power load of 300kW, with cabinets engineered to house 30kW of IT equipment. 1+1 300kW UPSs and an 800kW Generator provide the backup power (*TSSG Infrastructure Group* 2018).

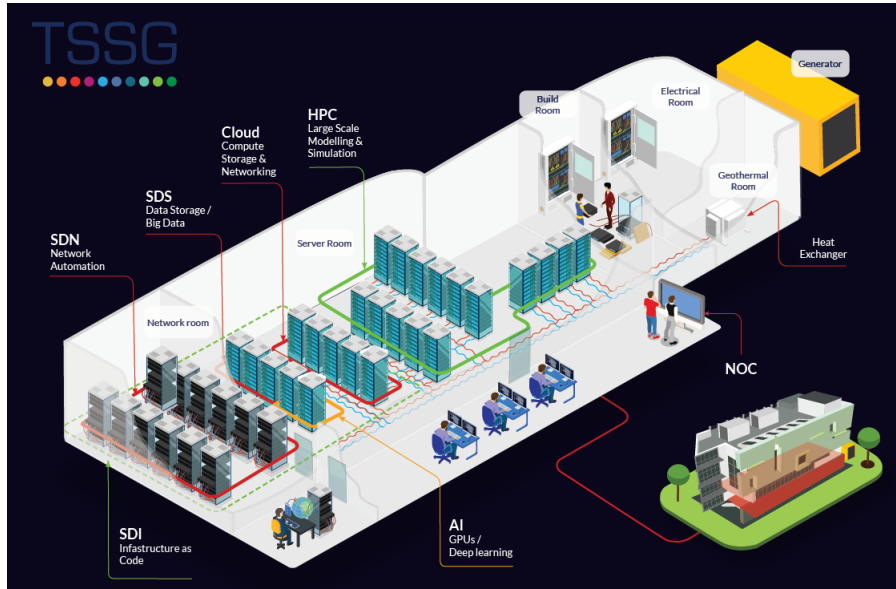
Figure 4: The average PUE from 2000 to currently.



With regards to physical hardware, the Data Centre supports over 160 physical servers, providing more than 1,000 CPU cores for processing and 400 virtual servers for cloud computing. In addition, there is over half a PB (that is 512 TB) of Data Storage, and 3,000 network port (*TSSG Infrastructure Group* 2018).

Figure 5, provided from *TSSG Infrastructure Group* (2018) demonstrates the design and infrastructure of the Data Centre building.

Figure 5: The design of the WIT-TSSG Data Centre.



This diagram illustrates that there are 26 Cabinets and these use varying degrees of KW per cabinet to house over 160 servers. In addition, the Data Centre is designed

so that there are separate rooms for servers, networking equipment, UPS and batteries with the generator and cooling equipment outside. This is to optimise the best cooling option for each room, resulting in the server room not having any hot or cold aisles. Instead, all heat and cooling is contained inside the 26 cabinets (*TSSG Infrastructure Group* 2018).

In terms of a cooling system, the Data Centre uses a water based cooling system which uses less energy than an air cooled system. There are two Chillers in the Data Centre that operate alternate weeks. For monitoring and visualisation, the Data Centre uses Nagios and the ELK software stack (*Elastic Stack* 2019); the output from that monitoring system is discussed further in section 4.1 and 4.2.

Also, as part of this research, several Data Administrators at the Data Centre were interviewed to garner further information on the Data Centre. They were able to identify specific features that might warrant investigation and outline current projects to upgrade the Data Centre while they also outlined any anomalies to expect to see in the monitoring data.

In addition to this, when conducting the correlation analysis as outlined in section 4.3.5, it became clear that further understanding was needed of the metrics involved and how energy consumption might be measured in a Data Centre. Reddy et al. (2017) provide a comprehensive set of metrics and key performance indicators that are used in Data Centres that can be used to add context to the analysis of the WIT-TSSG Data Centre.

2.3. Improving Data Centre Energy Efficiency

From the Research Problem Statement (section 1.2), the benefits for improving Data Centre energy efficiency is twofold: reducing the carbon footprint of the Data Centre, and reducing the cost of maintaining the Data Centre and these benefits should be mutually inclusive. Therefore, with such a desired outcome at stake, it is not surprising that a large amount of literature devotes itself to efforts in improving energy efficiency in Data Centres.

As outlined from the previous section (2.2), Data Centres are recent innovations, so to a degree, the literature is still catching up with the newest technologies and best practices. However, it is important to outline the earlier literature to understand why modern Data Centres, such as the WIT-TSSG Data Centre would have the ability to capture and store sensory data from its general operations. The following research summarises the need for such a sensor system and the immediate benefits this can bring to the energy efficiency of a Data Centre.

To begin, a study by Masanet et al. (2011) states that implementing a sensor system to capture data can reduce energy demand by as much as 80% through aggressive pursuit of energy efficiency measures specifically focused around servers. The authors also provide a model to estimate efficiency potentials associated with discrete efficiency measures applied to different types of IT equipment.,

While in 2013, the literature from this time also calls for the implementation of wireless sensors to begin capturing data. Here, Mahdavi and Tschudi (2013) argue that capturing

this data via wireless sensors and then using integrated software to manage and integrate this into management systems can allow for greater efficiencies in a Data Centres cooling system and power usage. Specifically, the paper suggests that by implementing such measures, Data Centres could reduce energy by as much as 17%, with specific savings within this attributed to the cooling energy.

Liu and Terzis (2012) also explain that a lack of visibility into a Data Centres operating conditions is an underlying reason for low-efficiency energy utilization. In this study, the authors argue for the introduction of sensor networks that combined with modelling, could provide insights into the distribution of energy and resolve the tension between cooling and performance, highlighting how the relationship between workload placement, performance and total energy is poorly understood.

While these outlined the immediate and obvious benefits of integrating more reliable monitoring data into management systems and decisions, which by and large has been implemented, more recent literature investigates where more marginal gains can be found.

Rygielski, Kounev, and Tran-Gia (2016) looks to the performance of network capabilities and in particular, this research develops simulation models to predict the performance of networks based on the traffic through them. It concludes that while the network performance models developed by the authors only lead to minor prediction accuracy degradation, the same models can however, accelerate the performance analysis by a much larger factor.

Arjona et al. (2014) use sensor data and a Map-Reduce computation in Hadoop to analyse server usage to maximise the efficiency of CPU power being used by each server in a Data Centre, thereby introducing some advanced data analysis of a specifically gathered sensor data. This introduces a measurement-based characterisation of energy consumption in a server and concludes that CPU consumption is neither linear or concave with the load within a server.

Marcos-Jorquera et al. (2016) moved the argument on from creating sensor networks to developing a monitoring service for energy efficiency and sustainable management in Data Centres. The approach the authors use is to collect the monitoring data, analyse the data captured and then execute based on the outcome of the analysis. This study concluded that such an approach is an adequate platform to contain a monitoring service offering a robust solution. This is further developed today in Data Centres where centres use the TICK or ELK (section 2.2.2).

Another approach used to improve Data Centre energy efficiency in the literature is advanced by Levy and Raviv (2018) who introduce a whole new set of metrics to review the efficiency of Data Centres. They argue that while the old metrics are valuable, some key metrics concerning *risk* are overlooked. They advance a new Data Centre multidimensional scorecard which gives better visualisation to identify areas of improvement and crucially, standardise a process that could eventually become best practice and allow for true comparisons Data Centres.

More advanced initiatives are based around the optimisation of physical resources within a Data Centre. Bi et al. (2015) propose a multi-tier architecture-orientated virtualised application performance model to allocate computing resources to each machine

based on SLA (service level agreement) restrictions. The authors develop a heuristic mixed optimisation algorithm to achieve this and the study finds that the proposed method is efficient in improving the overall performance and reducing the energy cost of the Data Centre.

Finally, Vasudevan et al. (2017) seek to reduce the operational costs of Data Centres and maximise energy efficiency by analysing log data to profile applications using the Data Centre. In addition, the authors profile the virtual and physical machines in the Data Centre, and using a penalty-based matching algorithm match different applications with the most compatible machines (virtual and physical) so that CPU usage is more efficient. Through experimental studies, the study concludes that such a profiling approach is feasible, scalable and energy efficient.

The literature outlined here demonstrates the changing focus of how best to increase the energy efficiency of data centres. From early research, just ten years ago, seeking the implementation of sensor systems that would lead to improved management systems and easily identified gains, to more recent and more advanced techniques looking to use algorithms to increase optimisation of the resources in a Data Centre. In all cases highlighted, the importance of capturing, understanding and using data to integrate new information into management systems and operations is stressed. However, in these instances, the approaches identify explicit data they want to capture and analyse, whether this is log data on CPU utilization or specific monitoring data. In general, few Data Mining approaches are used, in the sense of using implicit data to uncover hidden insights. The research questions in section 1.4 focuses on this concept.

2.4. Data Mining Techniques (for improving energy efficiency in buildings)

Data Mining is the process of uncovering hidden patterns from large sets of data, where the data is analysed to extract information which can be transformed into valuable knowledge for the future. Such techniques include statistical and machine learning techniques which take the form of regression analysis, classification, cluster analysis and more advanced techniques such as neural networks. These techniques analyse the relationship between variables to uncover such hidden patterns and trends.

While there is limited literature or existing research on using Data Mining techniques to improve the energy efficiency of Data Centres, there is comprehensive research on using Data Mining approaches to assess the energy efficiency of buildings in general, or of various systems within buildings. Because a Data Centre is a building in its own right, it is important to review such Data Mining techniques in the building management literature and assess if the research can be applied to Data Centres.

In general, the literature can be divided into two sections, where basic Data Mining techniques are used, as described above, and when more complex algorithms are used to gain insights. First, I explore the literature on the basic Data Mining approaches.

Xue et al. (2017) look at district heating substations in Northern China, and from operational data collected from the substations, use clustering techniques to identify

seasonal and daily operating patterns, and association rules mining to identify faults, malfunctions and inefficient operations strategies. Using data from the meter reading system, the authors conclude that the Data Mining techniques can extract potentially useful knowledge in terms of fault detection and operation optimisation.

Kuo, Lin, and Lee (2018) use regression and classification techniques on different factors in analysing the energy consumption characteristics of Taiwan’s convenience stores. The data is collected from a survey of convenience stores owners before the Data Mining techniques look at inter-correlation between energy consumption and other factors such as geography, climate and store design. The knowledge obtained through these techniques can ultimately provide the owners of the stores with design decision mechanisms in line with the findings.

Also, using monitoring data from building automation systems, Fan et al. (2018) explore how unsupervised Data Mining techniques such as clustering, association rule mining, motif discovery and unsupervised anomaly detection can be used to gain insights into energy efficiencies within buildings. Here the authors review and summarise the techniques used and elaborate on the challenges and opportunities of using each and acknowledge that this is still a developing area of study.

In another paper by the same authors, and using operational data from the construction of an education building in Hong Kong University, Fan et al. (2018) use decision trees, motif discovery and gradual pattern mining to uncover novel and valuable insights for the management of energy consumption for the building. The paper highlights how these insights on building operational characteristics which can be obtained to pursue fault detection and optimal control strategies can be developed to enhance building operational performance.

In a similar study by Zhou et al. (2018), the authors use data supplied from heat supply companies in Singapore and look at the correlation between energy consumption and building physics, the heating system and room position. The authors use the information gain ratio algorithm and decision tree classifiers to analyse energy consumption and other related factors and their main conclusion being that this a promising approach for such analysis.

Ahmad and Chen (2018) use Gaussian process regression, multiple linear regression, tree bagger, bagged tree, boosted tree and neural networks to identify abnormal behaviour and predict the future heating and cooling demands in building environments. The results from this study showed that the six models were efficient in foreseeing anomalies and predicting the future heating and cooling demand in the building environment.

These approaches will be reviewed when finalising a predictive model for energy consumption in section 4.3.9. However, in addition to those highlighted, further literature has been sourced that specifically concentrated on analysing energy consumption in a Data Centre. Makris (2017) collects data from the Data Centre processors, access memory and the network interface controllers and analyses these features using linear regression to predict energy consumption at a system-level.

A precursor to the regression analysis is correlation analysis as outlined in section 5.1, and it was important to understand the value this can bring to a data mining approach. Girma and P. Wang (2018) use multivariate correlation and data mining clustering tech-

niques for detecting distributed denial of service attacks in cloud computing. Specifically, the authors investigate the correlation among selected and ranked variables. Such an approach is used in this project to identify predictive features of energy consumption and pinpoint potential inefficient energy consumption.

However, some research also uses more advanced algorithms to gain hidden insights from the data captured.

Sönmez et al. (2018) use log data and the self-organising maps algorithm to create a decision support application for anomaly detection in IT systems. They highlight that this algorithm is used in anomaly detection in fraud cases and security attacks, but most importantly they highlight, when using the log data from systems, the self-organising maps algorithm allows for early detection of failures in critical systems in real-time.

The Random Forrest Algorithm is used by J. Wang et al. (2017) to discover critical events for event driven optimisation in building air-conditioned systems. With data collected from the building automation system, the authors make a case for using Data Mining to gain efficiencies in air conditioned systems by moving away from a time based system, to a more event driven optimisation.

The literature reviewed at this point highlights the wide and varied Data Mining approaches taken by various authors to gain hidden insights into energy efficiencies within buildings or building systems such as heating or cooling systems.

This project will adapt these approaches to the monitoring data we have gathered from WIT-TSSG Data Centre to help answer our Research Questions (section 1.4).

2.5. Summary

The research questions (section 1.4) clearly set out the goal of maximising energy efficiency within a Data Centre and the literature review sets out the framework for building the foundations for this research. Firstly, it defines what a Data Centre is and what the trends are in energy consumption by Data Centres (section 2.2.1) before introducing the WIT-TSSG Data centre in Waterford (section 2.2.2). From these foundations, the literature review focused on various computational and systematic improvements that could lead to greater energy efficiency within a Data Centre (section 2.3). The importance of wireless sensor data is a common starting point for most authors who focus on efficiency improvements to be made by concentrating on hardware improvements (network, server or cooling systems). Most of these approaches use data which is specifically captured at a system level to assist the goal of increased energy efficiency.

This is where I believe there is a research gap. Data Centres are entities in their own right, not just a collection of servers, networking and storage equipment, and store vast amounts of operational and monitoring data. This has been harvested by a monitoring system and could be useful to analyse, using Data Mining techniques, to find hidden insights around energy efficiency or to find instances of where energy is being used inefficiently.

Based on this, the final section of the literary review looked at where Data Mining techniques are used for similar applications (section 2.4) such as the energy efficiency of different building types, or heating or cooling systems within a building. Various

Data Mining techniques are explored in the literature and are assessed to see if such approaches can be adapted for use in a Data Centre.

3. Methodology

The literature review (section 2), identified a gap where data mining techniques could be used on the monitoring data compiled within a Data Centre to unlock hidden sources of information that could lead to greater energy efficiency in operating the Data Centre. The research questions (section 1.4) ask if these efficiencies could be brought about by identifying anomalies in the Data Centre where energy is not being used efficiently, with specific focus on the Chiller system and also looking to predict when energy demand peaks in order to better utilise the supply of energy to the Data Centre.

Whereas the majority of literature analyses the separate hardware components of a Data Centre to improve efficiency (section 2.3), this research proposes to concentrate on monitoring data that tracks energy demand in the Data Centre itself to gain insights to unanticipated spikes in energy demand. There are two elements to this, analysing when the demand of energy consumption spikes, and analysing how the Data Centre manages that demand.

At any given time, a Data Centre will have a certain load to process, and will utilise its available resources to process the load. To maximise efficiency therefore, a Data Centre would need to be able to best match the energy demand with the supply at its disposal. The real gains in efficiency therefore, is to be able to anticipate and understand the demand to best utilise the resources at its disposal. In technical terms, this would mean moving loads between racks within a Data Centre or reducing the demand for the cooling system at certain times. Therefore, to increase the efficiency of a Data Centre by concentrating on energy demand, we need to anticipate the demand of energy usage or identify inefficiencies in the system that was causing such a demand for energy. This is encapsulated in the Research Questions (section 1.4)

However, while 'predicting' energy demand is mentioned throughout this project, it is important to explain why constructing a model to specifically solve for this in isolation is not pursued. There are numerous reasons for this. Detailed investigations into such a model revealed that such a model would be technically complex. Specifically, the monitoring data (as outlined in section 4.2) does not contain information pertaining to system loads or water pressure data. As system loads within each cabinet are not available, due to the client confidentiality of those tenants engaged with the Data Centre, the information which would be needed to predict loads and anticipate demand to a granular level required, is not available. Similarly, water pressure data, not obtainable at this time, would be needed to feed into a robust energy predicting model.

Therefore, with such data missing to construct a robust and comprehensive model to predict energy consumption in isolation, this project needed to focus on constructing a model to find anomalies or system faults that might depict when the energy supply is not being utilised inefficiently.

The approach taken to achieve this, after consultation with Data Administrators in

the Data Centre is to concentrate on the performance of both Chillers. They believe, but had no direct evidence, that the Chiller system is currently not energy efficient. Preliminary analysis (section 1.3) indicated that indeed, at certain times of the year there were unexpected spikes in energy consumption of Chiller 1. Therefore, this project would focus on those spikes, to both identify and understand the sudden spike in energy consumption and to potentially find relating factors that might be related to this. To do this, this Dissertation will use correlation analysis to identify predictive features of energy consumption and follow this with regression analysis to build a predictive model of energy demand.

Firstly, correlation analysis is used to identify the relationship between the Chiller system and other energy indicators. As there are two Chillers in the Data Centre, with each Chiller operating on alternating weeks, we can use correlation indicators to view the relationship between the Chiller in operation for one week and other measures for that week such as weather variables, the temperature of the water leaving the cooling system or the energy demand of different cabinets and servers. In theory, allowing for weather factors, the relationship between a Chiller and the other variables, which is measured as a correlation co-efficient, should be stable from week to week, regardless of the Chiller in operation. Therefore the correlation analysis will identify particular variables that are different from week to week.

From this, and using the variables identified by the correlation analysis, and controlling for other factors such as the weather, a linear regression regression model will be built that charts the relationship between those variables identified and the chiller in operation for a certain week. This is essentially a model of energy demand for a chiller for a particular week. Then, measuring the coefficients of these variables week to week, in the form of a natural experiment, will identify variables that are statistically different. The basic model takes the form of:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i \quad (1)$$

If we can find variables that vary from week to week, when the Chillers alternate, it will identify potential issues or anomalies that is causing the spikes in energy. Therefore, using Data Mining techniques to Identify potential issues for remedial action will ultimately increase the energy efficiency of the Data Centre which is our stated goal.

Therefore, to construct a model that understands how energy supply is utilised will be the missing piece of the jigsaw in helping the Data Centre become more efficient by analysing its own monitoring data to understand a surge in energy demand will allow the Data Centre to utilise its supply of energy better, making it more energy efficient and aligning with our aim of reducing the cost of running the Data Centre (section 1.3).

In conclusion, this project is treating a Data Centre as an entity in its own right. Instead of merely being an end point in a cloud computing infrastructure for many applications or physical machines, it is utilising the data created by the Data Centre itself. By doing this, a predictive model is being built to anticipate the demand of energy needed, thereby transforming the Data Centre from being an intricate part a cloud computing framework to a more Edge orientated framework in its own right.

4. Research Methods

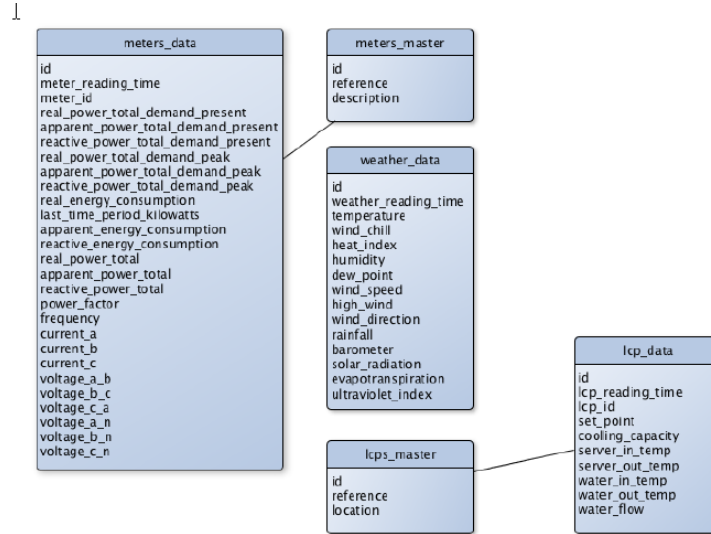
This section expands on the Methodology and firstly introduces the data available to this study before delving further into how correlation analysis and predictive models will be constructed using the data available.

4.1. Data Available

As outlined in section 2.2.2, the Data Centre examined is the WIT-TSSG (Telecommunications, Software and Systems Group) Data Centre in the Carriganore West Campus, which is part of the Waterford Institute of Technology (WIT). Construction of the Data Centre was completed in January 2010 in order to support its network-based research projects in the area of telecommunications networking and cloud computing. The Data Centre was designed with high-density, power efficiency and separation of services as its primary goals. It has a total IT power load of 300Kw with some cabinets engineered to house 30kw of IT equipment.

Similar to most modern Data Centres, sensors and the use of network devices are used to monitor and capture data on how the Data Centre is operating. As outlined in the schema model shown in Figure 6 below, there is substantive qualitative data captured that can be used for the empirical investigations to create a predictive model.

Figure 6: The schema of the WIT-TSSG Data Centre.



As we see from the schema, the data captures a multitude of metered energy usage variables, different weather data, and water usage data. A full detailed definition of the key metrics is provided in the appendices - section E. The data outlined here is available from 2014 right up to present, captured at 5 minute intervals, making this a time series data set. In the next section, the data is explored at an initial level to find any seasonal trends or predictive variables.

4.2. Exploratory Data Analysis

The data analysed for this project is a static 'data dump'. Additional data, when available, can be added for future studies if required. The data was accessed by connecting remotely into the Data Centre via PuTTY and querying the data via MySQL (See Appendices - section A). From here, the CSV output from the SQL queries was SFTP'd to my local machine and exported into Microsoft Excel where some preliminary data analysis was carried out.

The schema introduced in Figure 6 outlines the vast scale of variables available and as such data is gathered in 5 minute intervals over four years, the amount of data accumulated is vast and can be challenging to work with. While Python is used for developing the Data Mining Techniques (see sections 2.4 and C) line charts are ideal to illustrate at an exploratory level, what the time series data can show and the following charts are snippets from the data to demonstrate the range of data available.

Figure 7: Energy consumption (in kW usage per 5 minute intervals) of one cabinet in the Data Centre over the entire year of 2017.

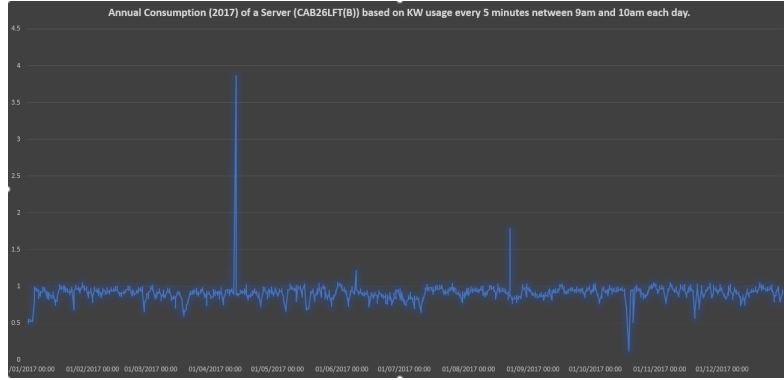


Figure 7, generated using Listing 1, illustrates the energy consumption of one particular server cabinet within the Data Centre (Cabinet 26A) over the course of 2017. The daily readings are taken from between 9am and 10am each day and while the chart may initially suggest that the energy usage has been relatively consistent, there are significant spikes in energy consumption in April and August, with some considerable drops in demand too.

However, while Figure 7 was a snapshot of energy usage over a year, the data can also be presented for a particular day.

Figure 8: Energy consumption (in kW usage per 5 minute intervals) of three different server cabinets in the Data Centre over the course of one day (2017-09-13).

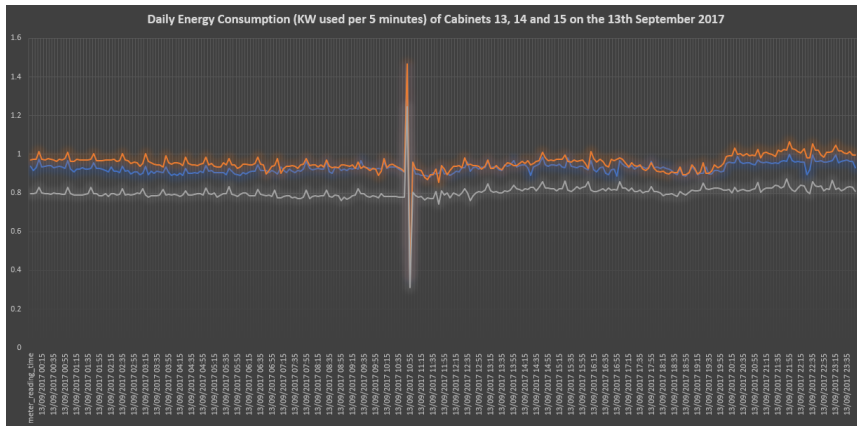
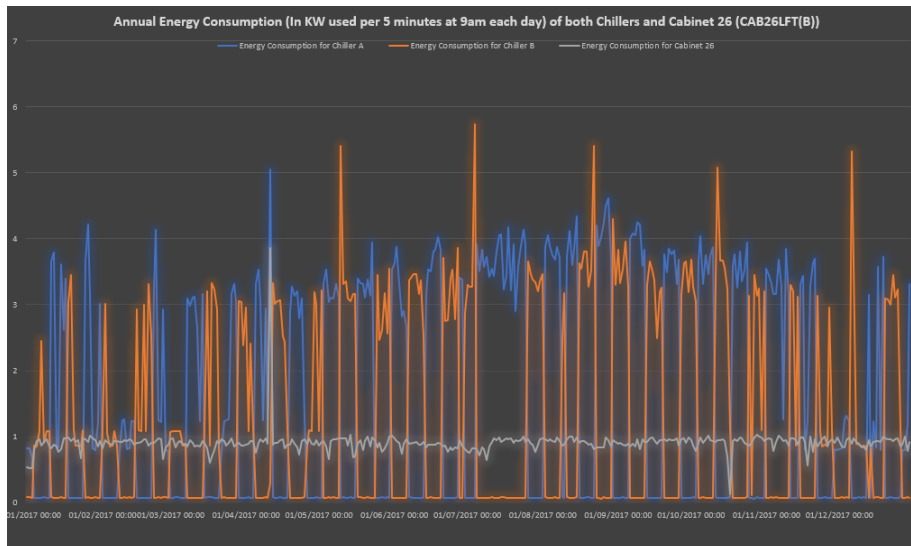


Figure 8 reports significant spikes in energy consumption in two of the cabinets around 11am. Again, this is another example of how data visualisation might be applied to assist in developing Data Mining techniques to predict when such spikes occur.

While certain spikes in energy consumption are suitable for further investigation, the preliminary data analysis should be able to detect some seasonal behaviour. Figure 9 demonstrates this. While the energy usage for the server cabinet (cabinet 26b) is relatively stable throughout the year, the energy consumption for the chillers shows a clear increase throughout the warmer summer months. Interestingly however, while each chiller works on alternate weeks, there are unexplained spikes for chiller B throughout the second half of the year that are not replicated when chiller A is running.

Figure 9: Energy consumption (in kW usage per 5 minute intervals) of both Chillers and one server cabinet in the Data Centre over the course of 2017



It is from presenting this information to Data Administrators who work in the Data Centre, where the impetus came to specifically concentrate on the inefficiency of the chiller system. It was their “hunch” that such inefficiencies existed, and it was these discussions that spawned the third research question (in section 1.4).

Other interesting preliminary analysis of the data centred around water consumption at the Data Centre (Figure 10) and the energy usage of the generator during a severe weather event, in this case being Storm Ophelia on October 2017 (Figure 11).

Figure 10: Annual water flow (in litres per 5 minute intervals) to LCP 1 measured at 9am every day during 2017

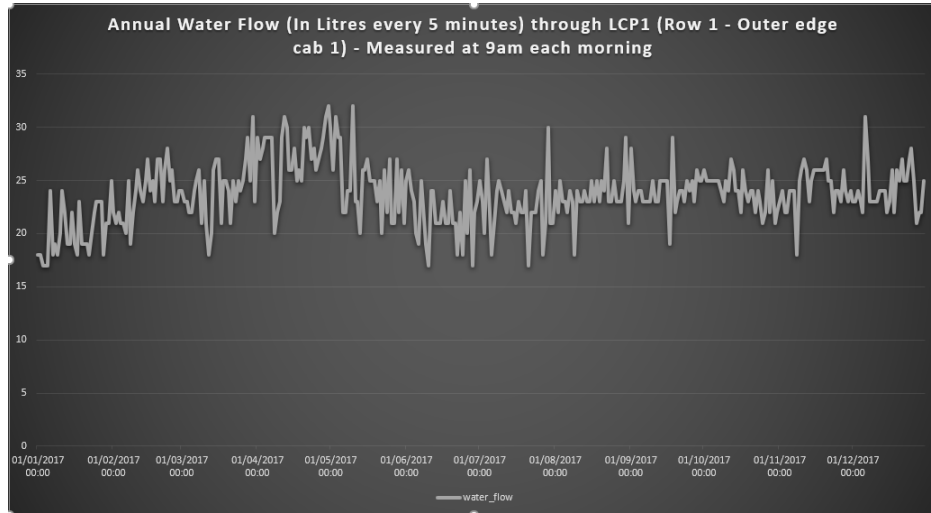
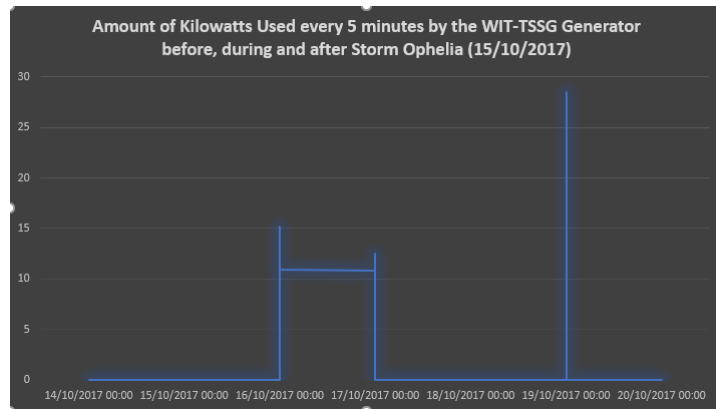


Figure 11: Energy consumption (in kW usage per 5 minute intervals) of the Generator before, during and after Storm Ophelia in October 15th 2017



All these charts are provided to simply demonstrate the range of data available. The next section discusses how this data was screened to build a model and the challenges that arose from this.

4.3. Research Techniques

4.3.1. Introduction

An advantage for this research is that the data required is already available for analysis and the previous sections (4.1 and 4.2) introduced the variables and did some exploratory data analysis. This section progresses further, and outlines how the data will be screened to build a model as outlined in the Methodology (section 3).

Figures 7, 8, 9, 10 and 11 demonstrate that the data available here is time series data. With data going back to 2014. This allowed for investigations into the energy usage per cabinet, per rack or even at a system level, whether that is the energy consumed by the network configuration, the server demand or other computational needs.

The Exploratory Data Analysis section (4.2) also illustrated seasonal effects where energy demand will increase in the summer months of the year as the warm weather precipitates the need for greater cooling within a Data Centre. However, the approach here will be to analyse these factors at a deeper level to build a more developed model that unlocks hidden insights not yet discovered.

Firstly, the literature was reviewed from section 2.4 and the various Data Mining techniques that have been used to increase energy efficiency of different building and heating systems taken into consideration. After this, the data was imported into Python for detailed Data Mining exploration to generate the correlations among all variables. The output from this is outlined in section 4.3.5. This was presented to Data Administrators in the Data Centre who provided context to these findings and suggested specific areas to investigate.

From here, regression analysis will look at the relationship between the performance of both Chillers with other variables. However, a model is only useful if it's reliable, meaning it should be accurate and precise. In attempting to develop such a model it will be important to focus on goodness of fit measures like bias, variance and auto-correlation. This will be a key factor in the project also.

Finally, as outlined in section 4.1, it is important to note that the data presented here is a static 'data dump'. This allows for the development of models on the static data only. However, an advantage of having nearly five years of static data is that it allows for the model to be tested on different subsections of the data to test its validity and reliability. Therefore data from 2017 has been used in the analysis so far. While this is useful and helps to answer our research questions (section 1.4) concerning increasing energy efficiency within a Data Centre, a more ideal model would be based on live streaming data. This would allow us to identify problems in real-time and develop models that will assist in anomaly detection or to highlight when system failures might occur in real time and assist in avoiding such failures.

4.3.2. Import Data From Data Centre

The first step was to retrieve the monitoring data for 2017 that is stored on the server of the Data Centre. Similar to section 4.2, data was exported from the Data Centre by logging on remotely via PuTTY and directly querying the data, using MySQL (See

Appendices - section B) and SFTP'ing the query results, as csv files, to my own local machine.

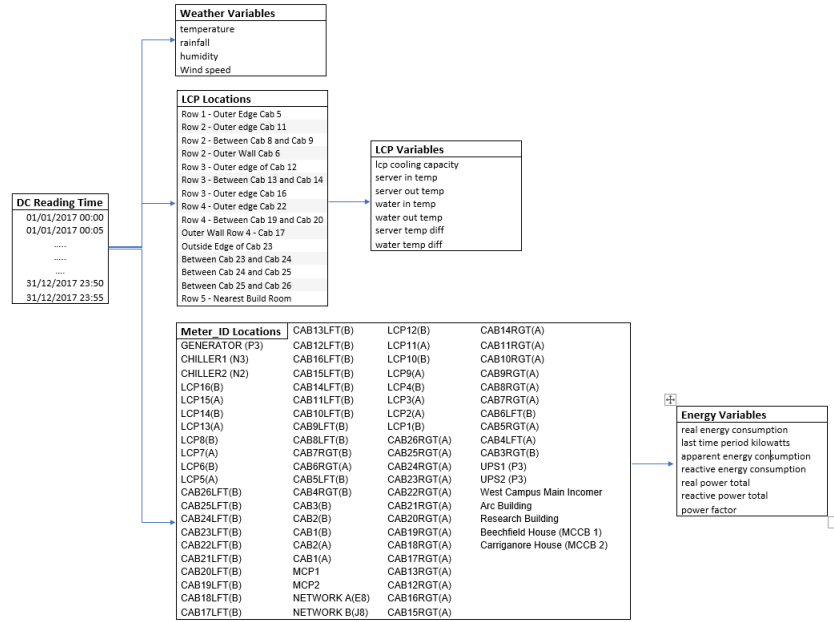
Figure 6, as previously outlined, visualises what data is available in the Data Centre. Specifically, data from 2017 was used as an initial testing sandbox for the data analysis. As the time series data is gathered in 5 minute intervals, this equates to 175,200 entries for the entire year.

In addition, specific monitoring data containing energy information from all 83 meter locations, LCP Cooling Capacity information for all 15 LCP Cooling Capacity locations and all weather variables were exported from the Data Centre. In total, for just 2017, that equated to over 210 million unique data entries, stored on three different csv files.

4.3.3. Understanding the Data

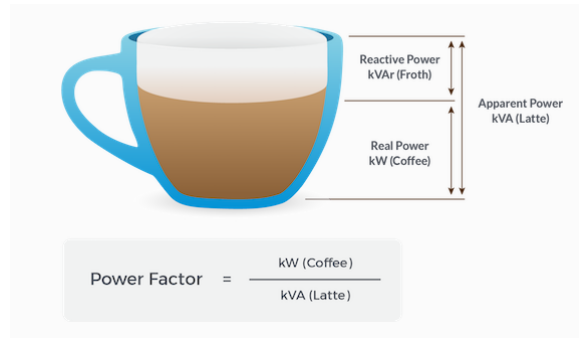
With such a vast array of information available, as earlier demonstrated in section 4.2, the next stage was to streamline the amount of variables to be considered. After consultation, the following variables, as outlined in Figure 12, were chosen.

Figure 12: Schema of Data and Variables chosen for detailed Data Analysis



While full definitions of the key variables are defined in the appendices - section E , the following diagram (Figure 13) describes and outlines the relationship between Real Power, Reactive Power, Apparent Power and Power Factor.

Figure 13: The Latte Definition of Different Energy Metrics



4.3.4. Manipulate Data in Python to Generate Correlations for all Variables

While a key requirement, as identified and prioritised during the project, was to focus on the Chiller system, it was also important to take a holistic view and unlock some insights about the Data Centre as a whole. Therefore, with the data now gathered, and the key variables identified, the next stage was to use Python to create correlations between all variables to identify some predictive features or new insights. The following steps were carried out in Python and the full code is viewable in section C of the appendices.

1. Import csv's to Python.
2. Streamline data to only include selected variables.
3. Use Pivot Tables to generate data based on the DC Reading time for every location and associated variable.
4. Merge all energy, LCP Cooling Capacity and weather information together to create a master table of all data based on the 5 minute interval values (merged on DC Reading Time).
5. Generate correlations based on the created time series for all 690 variables.
6. This created nearly 240,000 correlations. From this Python was instructed to export the top 45,000 correlations for further analysis.

Ideally, during this process efforts would be made to reduce the dimensionality of the data so that further Python based analysis could take place such as heat maps. Unfortunately, any efforts with such an array of data was not possible due to computational limitations and while further attempts to do this later in the process was pursued, at this time it was deemed that it was more important to explore the all variables identified.

4.3.5. Analyse Correlations

As Correlation analysis is the process to establish a relationship between two variables, the process to identify 'interesting' or 'useful' correlations in this instance was manual. Having received the top 45,000 correlations from Python, which had a correlation coefficient higher than ± 0.93 , this manual piece was done by importing the correlations to Microsoft Excel and using filters to trial and error different combinations.

Some correlations between variables were easily explained, such as the relationship between different energy variables but the following are high-level findings from this process:

- There is a strong correlation between Humidity (but not temperature) and Real Energy Consumption, Apparent Energy Consumption and Reactive Energy Consumption.
- There is a high correlation between Apparent Energy Consumption and Real Energy Consumption. This indicates that there is a relatively stable amount of reactive power generated at each meter.
- There is a high amount of correlation between Real power total and Last time kw.
- In terms of analysing the different performances of the chillers, out of 2103 correlations, 2089 are identical for both Chiller 1 and Chiller 2. The main difference centres around some correlations for Chiller 2 and the LCP Cooling Capacity and server temp differences of some cabinets.
- In terms of cross-correlations between meter id's and LCP locations, there is a high correlation between Real Energy Consumption of most meter id's and 5 particular LCP Cooling Capacity variables that is very consistent.

Full details of the correlation analysis, including a definition of the key variables discussed are found in section F.

4.3.6. Meet with Data Centre Administrators

Various meetings with Data Centre Administrators took place during this process. These ranged from tours of the Data Centre, to in-depth discussions around possible areas of investigation and also detailed discussions around preliminary findings.

In particular, they highlighted the inefficiencies in the Chiller system. Specifically, they explained how the two chillers alternate weeks in operation and how chiller 1 uses considerably more energy when it kicks in for its rotation.

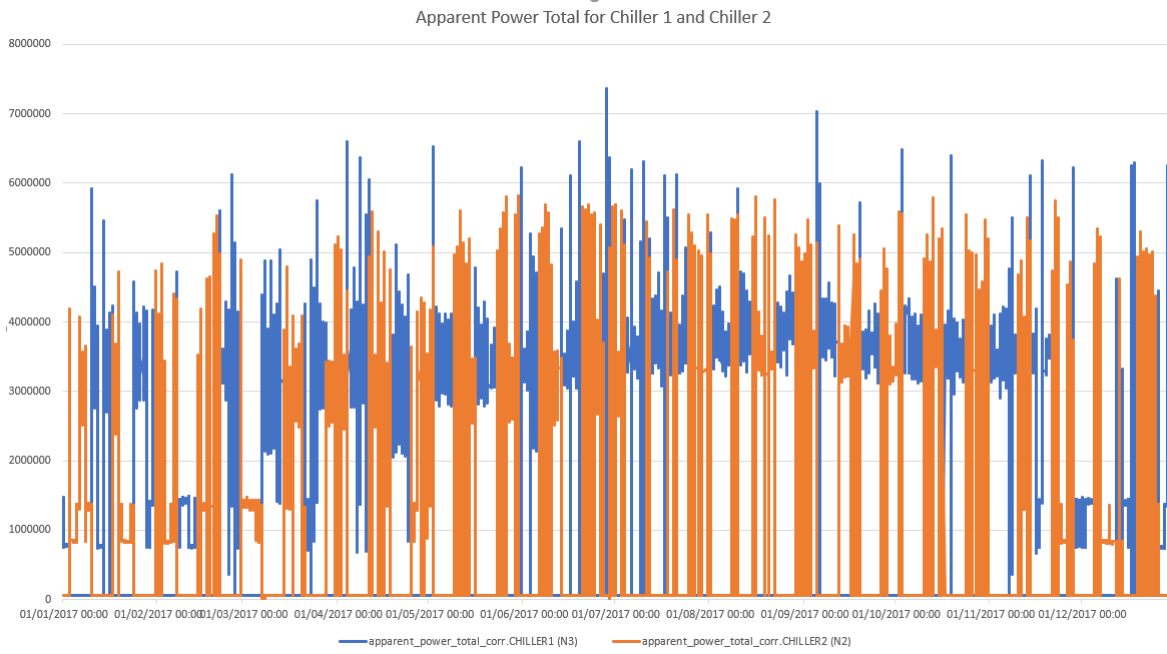
It's from these meetings, that the third research question around the energy inefficiency of the chiller system got initially prioritised. The Data Administrators have a cost incentive to correctly identify the issue at stake, so that the specific hardware component can be replaced in isolation. Because without precisely identifying the component at fault, the cost of repairing the system increases exponentially.

Therefore, from the Data Administrator's perspective their stated priority is to investigate the issues around the chiller system more so than predicting energy demand. Therefore, that's why the focus for this work narrows from here to look at the Chiller system. A predictive model is still constructed but is used in the context of comparing the performance of both chillers.

4.3.7. Identify Micro-timelines to Analyse Chiller Performance

Section 4.3.5 investigated correlations between all variables over the course of a year. However, that timescale is too broad to draw definitive conclusions about the performance of the Chiller system. Therefore, we need to discover specific windows throughout the year where there was visible anomalies in the performance of Chiller 1, and analyse those in isolation. To do this, the Apparent Power of both Chillers were charted on a line graph for the entire year to identify any such spikes.

Figure 14: Apparent Energy Consumption of Chiller 1 and 2 in 2017



The diagram clearly illustrates many spikes (in blue) that have no corresponding orange spike, highlighting the prevalence of energy consumption spiking for Chiller 1. In particular, the following cutover dates, when the chiller alternated from Chiller 2 to Chiller 1, were identified for further analysis.

1. 2nd May 2017
2. 13th June 2017
3. 5th September 2017

The next step is to take the week before the cutover, and the week after the cutover to analyse the correlations between the chiller in operation for that week, and the other variables. In theory they should be relatively stable, so any correlations that are present in one week and not the subsequent week indicates a potential issue.

4.3.8. Identifying Variables to use in a Regression Model

Ultimately, with 690 variables potentially available, a process was needed to identify some key variables to use in a regression model to analyse the performance of both Chillers. This was done by using correlation analysis, similar to section 4.3.5. However, the key difference was that this analysis was focused on just a particular week and that the correlations were focused around just the Chiller in operation for that week. The following steps were carried out in Python and the full code was similar to that as referenced in section C of the appendices, with minor filters applied for date ranges and specific variables.

1. Import saved data from previous Python code where all data is merged and ready for use (Section 4.3.4)
2. Streamline data to only include data from the week before the Cut-over (25th April to 2nd May)
3. Generate correlations based on the created time series for all 690 variables.
4. This created nearly 240,000 correlations. From this Python was instructed to export the top 30,000 correlations for further analysis.
5. Streamline data to only include data from the week after the Cut-over (2nd May to 9th May)
6. Generate correlations based on the created time series for all 690 variables.
7. This created nearly 240,000 correlations. From this Python was instructed to export the top 30,000 correlations for further analysis.
8. Repeat this step for the other 2 cut-over times identified.

This provided 2 sets of correlations for each Micro-timeline, the first for Chiller 2 and the second for Chiller 1. These were exported to excel and were manually compared to to identify, for each Micro-timeline, significant differences in correlation co-efficients between the same variables. As the Correlation co-efficients should be consistent from week to week, any particular variables that were not similar between both weeks were earmarked to be used in the regression model.

An overview of the outcome of this process is outlined in the Presentation of Findings (section 5) and the full detailed results are in the Appendices section G.

In selecting variables for the regression model, the four weather variables were included to control for external factors, while the remaining 40 variables were identified using the

correlation analysis in this section. As there are several energy consumption metrics, Apparent Power was used to bring consistency to the model.

4.3.9. Construct Regression Model

As per sections 1.4 and 3, the aim is to construct a regression model around the energy consumption of the Chiller in operation for that week (*dependent variable*) with selected other variables, identified in section 4.3.8, such as the weather, energy consumption of certain cabinets and specific water cooling features (*independent variables*).

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i \quad (2)$$

The full variable list of Independent Variables used is available in the Appendices - section H. We then compare the coefficients of this model, taking the confidence intervals into account, to the same results the next week, when the alternating Chiller is in operation.

The premise of this model is that over a two week period, the relationship of the independent variables to the Chiller in operation for each week, as measured by the model, should be relatively stable. Independent variables therefore, that are not stable from week to week, especially those are also present in analysing the other cut-over times, indicate potential issues with how the Chillers operate and possible inefficiencies in how the Data Centre manages its supply of electricity. This is how this model proposes to answer the Research Questions as outlined in section 1.4.

4.3.10. ‘Goodness of Fit’ Measures

However, there are a number of ‘Goodness of Fit’ factors to consider when constructing the regression model. As this is time series data, the data needs to be stationary, meaning that statistical properties such as *mean* and *variance* are constant over time. The Augmented Dickey-Fuller Test was used to confirm that the time series was stationary.

In addition, the model also needed to be conscious of *Auto-Correlation*. This is to check that the similarities between variables is not as a result of a time-lag between them. The Durbin-Watson statistic from the regression output quantifies this test. The result is a score between 0 and 4, with anything between 0 and 2 indicating strong positive auto-correlation and between 2 and 4 indicating strong negative auto-correlation. Ideally a regression model would have a score close to 2 which indicates no auto-correlation. However, the simple linear model as described in section 4.3.9 indicated strong auto-correlation in all six outputs of the model (for each week before and after a cut-over for the 3 cut-over times being analysed). This meant that residual coefficients would be biased for each output so they could not be compared week to week.

To overcome this, the Cochrane-Orcutt procedure was used which adjusts a linear model for auto-correlation in the error term. Once this was applied, and the model re-run, the 6 outputs of the regression model all scored a Durbin Watson score close to 2.

This allowed for the comparison of residual coefficients from week 1 and week 2, for the three micro-timelines outlined and to identify the independent variables that are causing potential energy inefficiencies.

Python was used to construct and execute the regression model and goodness of fit measures, and this code is captured in the Appendices - section D.

The results of the regression analysis are presented in the next section (5) and discussed in section 6.

5. Presentation of Findings

5.1. Correlation Analysis of Chiller 1 versus Chiller 2

The process, which analysed the Correlation of the energy consumption of an active Chiller during the week before and after a cut-over with other variables is outlined in section 4.3.8 and the full detailed results are in the Appendix - section G. The Python code for manipulating the data and generating the relevant correlations is in the Appendices - section C.

In Summary, for the first instance investigated, which covered the timeframe of April 25th through to May 9th, with the cut-over to Chiller 1 on May 2nd, out of 1,400 correlations relating to both Chillers, over 1,200 correlation coefficients match from week 1 to week 2. Out of the correlations that were not present in both weeks, The Energy Consumption of Cabinets 1, 7 and LCP units 5 and 19 along with Humidity were correlated with Chiller 2 only, while The Energy Consumption of Cabinets 3, 17 and LCP unit 7 were correlated only with Chiller 1. Interestingly, Chiller 1 was also only correlated with Water Flow and Server Temperature metrics between Cabinets 13 and 14.

Similarly, for the second instance investigated, which covered the timeframe of June 6th through to June 20th, with the cut-over to Chiller 1 on June 13th, out of 2,300 correlations relating to both Chillers, over 1,200 correlation coefficients match from week 1 to week 2. Out of the correlations that were not present in both weeks, The Energy Consumption of Cabinets 10, 22 and LCP units 9 and 17 along with LCP cooling metrics for Cabinet 22 were correlated with Chiller 2 only, while The Energy Consumption of Cabinets 3, 7, 12, 18 and LCP unit 7 were correlated only with Chiller 1. Again, Chiller 1 was also only correlated with Water Flow and Server Temperature metrics between Cabinets 13 and 14.

Finally, for the final instance investigated, which covered the timeframe of August 29th through to September 12th, with the cut-over to Chiller 1 on September 5th, out of 1,700 correlations relating to both Chillers, over 1,250 correlation coefficients match from week 1 to week 2. Out of the correlations that were not present in both weeks, The Energy Consumption of Cabinets 22, 23, 24, 26 and LCP units 12 and 13 along with LCP cooling metrics for Cabinet 22, 23, 24 and 25 were correlated with Chiller 2 only, while The Energy Consumption of Cabinets 11 and the LCP units of 2, 5 and 11 were correlated only with Chiller 1. In addition, Chiller 1 was also only correlated with

Water Temperature metrics between Cabinets 23 and 25 and Windspeed.

The variables highlighted above, whose correlation coefficient with the energy consumption of the operating Chiller from week to week varied significantly after the cut-over in operating Chillers, potentially provides diagnostic information on features in the Data Centre where energy is being consumed inefficiently. These variables therefore form the basis of the regression model outlined in section 4.3.9. The full list of variables used is outlined in the Appendices - section H.

5.2. Regression Analysis of Chiller 1 versus Chiller 2

In developing the model, the full process to construct the model and then to test that the time series is stationary is outlined in section 4.3.9. It was confirmed that the time series was stationary. In all 6 instances of running the model, the Augmented Dickey-Fuller value is less than the 1% critical value so we can reject the null hypothesis - confirming the presence of a stationary time series.

In testing for Auto-correlation, whereas the original Linear regression models had a Durbin-Watson score of between 0 and 1, confirming the presence of positive auto-correlation, which rendered the output biased. Using the Cochrane-Orcutt procedure to eliminate the bias, the generalised least squares method produced Durbin-Watson scores, for the 6 instances of the regression model of between 1.89 and 2.02. As a score close to 2 indicates no auto-correlation, the residual coefficients of each independent variable can be compared week to week, to identify where the relationship between the Chiller in operation and the independent variables differed significantly after the cut-over of the Chiller.

This was done by matching the coefficients of the independent variables from week 1 to week 2, allowing for the 0.025 and 0.975 confidence intervals. This means that the coefficient value from week 1, would have to record a value between the confidence levels of the same variable for Week 2 and visa versa. The full breakdown, including the coefficient values for all 3 micro-timelines are in the Appendices - section I while the Python code is also in the Appendices - section D. A summary of the variables, whose coefficients do not match from week 1 to week 2, for the 3 micro-timelines is provided in Figure 15.

Figure 15: Summary of Analysis from Regression Model

Independent Variables from Regression Model whose co-efficients do NOT overlap from week 1 to week 2		
Micro-timeline 1 - 25th April to 9th May	Micro-timeline 2 - 6th June to 20th June	Micro-timeline 1 - 28th August to 12th September
Cut-over: 2nd May 11am	Cut-over: 13th June 11am	Cut-over: 5th September 11am
temperature rainfall humidity wind_speed apparent_power_total_corr.CAB13LFT(B) apparent_power_total_corr.CAB13RGT(A) apparent_power_total_corr.CAB17LFT(B) apparent_power_total_corr.CAB17RGT(A) Apparent_power_total_corr.CAB18RGT(A) apparent_power_total_corr.CAB19LFT(B) apparent_power_total_corr.CAB19RGT(A) apparent_power_total_corr.CAB23LFT(B) apparent_power_total_corr.LCP11(A) apparent_power_total_corr.LCP13(A) server_temp_diff_Between_Cab_23_and_Cab_24 server_temp_diff_Between_Cab_24_and_Cab_25 server_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14 server_temp_diff_Row_3_-_Outer_edge_of_Cab_12 water_temp_diff_Between_Cab_23_and_Cab_24 water_temp_diff_Between_Cab_24_and_Cab_25 water_temp_diff_Between_Cab_25_and_Cab_26 water_temp_diff_Outside_Edge_of_Cab_23 water_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14 water_temp_diff_Row_3_-_Outer_edge_of_Cab_12 water_temp_diff_Row_4_-_Outer_edge_Cab_22	temperature wind_speed apparent_power_total_corr.CAB12LFT(B) apparent_power_total_corr.CAB12RGT(A) apparent_power_total_corr.CAB13RGT(A) apparent_power_total_corr.CAB19RGT(A) apparent_power_total_corr.CAB22LFT(B) apparent_power_total_corr.CAB23LFT(B) apparent_power_total_corr.CAB23RGT(A) apparent_power_total_corr.CAB3(B) apparent_power_total_corr.LCP12(B) apparent_power_total_corr.LCP13(A) apparent_power_total_corr.LCP7(A) server_temp_diff_Between_Cab_23_and_Cab_24 server_temp_diff_Between_Cab_24_and_Cab_25 server_temp_diff_Between_Cab_25_and_Cab_26 server_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14 server_temp_diff_Row_4_-_Outer_edge_Cab_22 water_temp_diff_Between_Cab_24_and_Cab_25 water_temp_diff_Between_Cab_25_and_Cab_26 water_temp_diff_Outside_Edge_of_Cab_23 water_temp_diff_Row_3_-_Outer_edge_of_Cab_12 water_temp_diff_Row_4_-_Between_Cab_19_and_Cab_20 water_temp_diff_Row_4_-_Outer_edge_Cab_22	temperature wind_speed apparent_power_total_corr.CAB17RGT(A) apparent_power_total_corr.CAB18LFT(B) Apparent_power_total_corr.CAB18RGT(A) apparent_power_total_corr.CAB19RGT(A) apparent_power_total_corr.CAB22LFT(B) apparent_power_total_corr.CAB24RGT(A) apparent_power_total_corr.CAB26RGT(A) apparent_power_total_corr.CAB3RGT(B) apparent_power_total_corr.CAB9LFT(B) apparent_power_total_corr.LCP12(B) apparent_power_total_corr.LCP2(A) server_temp_diff_Between_Cab_23_and_Cab_24 server_temp_diff_Between_Cab_24_and_Cab_25 server_temp_diff_Between_Cab_25_and_Cab_26 server_temp_diff_Outside_Edge_of_Cab_23 server_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14 server_temp_diff_Row_3_-_Outer_edge_of_Cab_12 server_temp_diff_Row_4_-_Between_Cab_19_and_Cab_20 server_temp_diff_Row_4_-_Outer_edge_Cab_22 water_temp_diff_Between_Cab_23_and_Cab_24 water_temp_diff_Between_Cab_25_and_Cab_26 water_temp_diff_Outside_Edge_of_Cab_23 water_temp_diff_Row_3_-_Outer_edge_of_Cab_12 water_temp_diff_Row_4_-_Between_Cab_19_and_Cab_20 water_temp_diff_Row_4_-_Outer_edge_Cab_22

While the significance of these findings is discussed below in section 6, this highlights that while the majority of independent variables have a stable and consistent relationship to the dependent variable (the operational Chiller) from week 1 to week 2, the variables listed in Figure 15 do not. It is also worth noting that some variables are consistently present across multiple micro-timelines in Figure 15 which indicates the presence of a pattern.

6. Discussion of Findings

From the outset, the exploratory data analysis (section 4.2) carried out on simple monitoring data, (Figures 9 and 14) demonstrated that there was unexplained peaks in energy consumption of Chiller 1, which was further corroborated by staff working in the Data Centre. Therefore, the aim of this project as outlined in the Research Questions (section 1.4) and the Aims and Objectives (section 1.3) was to use monitoring data combined with Data Mining techniques to identify hidden insights as to why the Chiller system, and subsequently the Data Centre as a whole is consuming energy inefficiently.

The correlation analysis, which looked at the Data Centre as a whole (section 4.3.4) and specifically for correlations between the operating Chiller and all other variables (section 4.3.8) also provided some interesting findings. Specifically it emerged that energy consumption is more correlated with humidity than temperature.

That process also discovered variables that were not correlated to the same degree with the operational Chiller when the cut-over from Chiller 2 to Chiller 1 happened. These are interesting, as they are the exception with over 75% of the variables recording a correlation coefficient that is stable and consistent across both weeks when the cut-over to

Chiller 1 happens. Significantly again, some variables were present in the analysis across multiple micro-timelines indicating a potential underlying issue with their performance vis a vis the operating Chiller.

However, as richer data is achieved from regression analysis, these variables then formed the basis for the regression model to quantify the relationship between the operating Chiller for a particular week and these selected other *independent variables*. When the results for week 1 were compared with week 2, this again highlighted particular variables whose relationship to the operating Chiller was not consistent across both weeks. In analysing Figure 15, there are some assumptions which must be stated.

1. It would be expected that the 'weather' variables are not consistent from week 1 to week 2. This is Ireland after-all. These were included so that their relationship to the Chiller could still be measured (even if the weather varied wildly from week to week).
2. The demand for energy consumption for any given cabinet *could* vary from week to week also. It is not unforeseeable that a particular client of the Data Centre has differing requirements from week to week. To mitigate this, that's why data was gathered for a week prior to the cut-over and not just 2-3 days. However, if the same cabinet had a consistently different relationship to the operating Chiller from week to week it would be more alarming.
3. The water and server temperature variables could be related to the weather. If the humidity raised considerably in week 2, that would have knock on effects on these variables. However, again, if they consistently had a different relationship to the operating Chiller from week 1 to week 2 it would raise alarm. The weather in Ireland doesn't fluctuate that much in such short spaces of time. Also to mitigate this the variable used in the model is water/server temperature difference. This is a contrived variable (developed in Python as demonstrated in the Appendices - section C) to eliminate the fluctuations in the actual rise in the temperature and simply record the in/out difference in the temperature.

Taking this into account, some inferences can still be drawn from the results as outlined in Figure 15. The energy consumption of Cabinets 3, 12, 13, 17, 18, 22 and 23 as well as the LCP units of 12 and 13 appear on the list more than once. What this translates as is that more than once during our analysis, when the operating Chiller switched from Chiller 2 to Chiller 1, these cabinets and LCP units suddenly had a different relationship to the energy consumption of Chiller 1 than it had with Chiller 2. Furthermore, the energy consumption of Cabinet 19 has a different relationship to the operating Chiller across both weeks for all Micro-timelines. These results are also consistent with the correlation analysis of both Chillers.

In addition, and maybe more significantly, the relationship between the operating chiller and the water temperature difference (*the difference in water temperature entering and leaving the unit*) and server temperature difference (*the difference in server temperature following the water being applied to cool the servers*) is far from stable from

week to week. Particularly this is true of the cooling operations between cabinets 23, 24, 25 and 26 and this is consistent over all 3 micro-timelines. What this indicates, is that when Chiller 1 cut-in to be the operating chiller, it had a consistently different relationship to these cooling variables than Chiller 2 did. Again, this was broadly consistent with the correlation analysis of both Chillers.

Turning these raw findings into an interpretable set of rules or potential warnings for Data Centre engineers is open to correction and would need further collaboration but the general narrative from these findings centre around what happens when the Chiller system alternates to Chiller 1 from Chiller 2. The findings here outline that when this cut-over happens, the newly operating Chiller (Chiller 1) has a significantly different relationship with cooling operations of Cabinets 23-26 to when Chiller 2 was operating. This would seem to indicate that Chiller 1 is working harder to cool these cabinets than Chiller 2 was. The underlying issue could be an issue with the water flow after the cut-over or an undetected issue with Chiller 2. To implement a potential warning system for the Data Centre Administrators, monitoring data from both Chillers and the impacted cooling systems for the 48 hours before and after a cut-over could be garnered to highlight when spikes in energy consumption of Chiller 1 is not within normal expectations.

In conclusion, while the analysis carried out here is consistent and highlights certain variables that may be responsible for inefficient energy consumption in the Chiller system and a general narrative is offered, in terms of drawing definitive conclusions from the findings, that is not for me to say. I am not a field expert and all I can do is pass this information onto the Data Administrators and Engineers in the Data Centre for further analysis on their side. To add to this, this project, with all the constraints of a taught masters programme, only looked to use correlation and regression techniques on 3 micro-timelines to garner these results and findings. It has to be stated that there is much scope for further analysis using both extended Data Mining techniques and additional data. Richer data around water pressure was not made available in time for this project but does exist. This along with deeper analysis into other years could yield potentially richer results.

However, to answer the 3 Research Questions as stated in section 1.4, a model was constructed, using monitoring data from a Data Centre and combining this with Data Mining techniques to identify hidden features which might find anomalies or inefficiencies in energy consumption. We identified that energy consumption of Chiller 1 spikes when it takes over from Chiller 2 on a weekly rotation, and the regression results here show a link to certain water and server temperature variables of specific cabinets when this happens. These are the hidden features that are not evidently visible from normal monitoring data and this information will be passed to the relevant people in the Data Centre to fully understand these findings.

7. Conclusion

This project concentrates on the growing prevalence that Data Centres play in the ICT sector and how much energy they currently consume. The literature review introduces

Data Centres in more detail, including the WIT-TSSG Data Centre, before surveying the existing literature that focused on increasing the energy efficiency of Data Centres. The literature review also looks at how Data Mining Techniques are used in increasing energy efficiency of other building types before a gap was established where Data Mining techniques could be used to help Data Centres become more energy efficient.

The Methodology section expanded on this gap and set the framework for the research methods section which firstly explains the data available before outlining the Data Mining techniques used to develop a model that will analyse the energy consumption of both Chillers. Specifically this model focused around when we see an unexplained spike in the energy consumption of Chiller 1 and the model and subsequent analysis outlines potential features that may be responsible for such a spike in energy consumption.

These results are presented and discussed with the caveat that it would require a Data Centre Engineer to interpret the findings fully and that there is much scope for further research to back up these findings. However, during the duration of finalising this document, it has been confirmed by staff in the Data Centre, that the Chiller system did indeed require remedial action and that is currently ongoing. Unfortunately at the time of completing this research it is too soon to say if the underlying physical issues uncovered acquiesce with my findings, but it does validate the decision to concentrate on energy inefficiencies in the Chiller system.

Therefore in terms of the generalisability of this research, while all Data Centres will have similar monitoring data that consist of metered energy usage, weather information and other operational data, not all Data Centres will have the same geographical constraints or energy demands driven by its individual clients. So while there could be no direct inference made to other specific Data Centres, any findings could be relate-able in a broader sense if a common set of features or attributes can be found. This would be the contribution of this proposed study to the existing literature.

Finally, in terms of the reasoning applied, this project with Data Mining at its core, straddles both inductive and deductive reasoning. When Data Mining is used in an exploratory mode, it is inductive while when it is used to build predictions, it is deductive. This project has elements of both as the quantitative data is already compiled and available (sections 4.1 and 4.2). It also exhibits abductive reasoning by finding an explanation for an apparent behaviour discovered in data.

References

- Ahmad, Tanveer and Huanxin Chen (2018).
“Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches”.
In: *Energy and Buildings* 166, pp. 460–476. ISSN: 0378-7788.
- Arjona, Jordi et al. (2014). “A Measurement-based Analysis of the Energy Consumption of Data Center Servers.” In: *CoRR* abs/1402.0804. ISSN: 14020804.
- Avgerinou, Maria, Paolo Bertoldi, and Luca Castellazi (Oct. 2017).
“Trends in Data Centre Energy Consumption under the European Code of Conduct for Data Centre Energy Efficiency.” In: *Energies* 10.1470. ISSN: 1996-1073.
- Bi, Jing et al. (2015). “SLA-based optimisation of virtualised resource for multi-tier web applications in cloud data centres.”
In: *Enterprise Information Systems* 9.7, pp. 743–767. ISSN: 17517575.
- Elastic Stack (2019). <https://www.elastic.co/elk-stack>. Accessed: 2019-04-29.
- Fan, Cheng et al. (2018). “Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review”.
In: *Energy and Buildings* 159, pp. 296–308. ISSN: 0378-7788.
- Girma, Anteneh and Ping Wang (2018). “An efficient hybrid model for detecting distributed denial of service (DDOS) attacks in cloud computing using multivariate correlation and data mining clustering techniques.”
In: *Issues in Information Systems* 19.2, p. 1. ISSN: 15297314.
- Kuo, Chung-Feng Jeffrey, Chieh-Hung Lin, and Ming-Hao Lee (2018).
“Analyze the energy consumption characteristics and affecting factors of Taiwan’s convenience stores-using the big data mining approach”.
In: *Energy and Buildings* 168, pp. 120–136. ISSN: 0378-7788.
- Levy, Moses and Daniel Raviv (2018). “An Overview of Data Center Metrics and a Novel Approach for a New Family of Metrics.” In: *Advances in Science, Technology and Engineering Systems, Vol 3, Iss 2, Pp 238-251 (2018)* 2, p. 238. ISSN: 2415-6698.
- Liu, Jie and Andreas Terzis (2012). “Sensing data centres for energy efficiency.”
In: *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 1958, p. 136. ISSN: 1364503X.
- Mahdavi, Rod and William Tschudi (2013).
“Wireless Sensor Network for Improving the Energy Efficiency of Data Centers.”
In: *Lawrence Berkely National Library*.
- Makris, Theodosios (2017).
“Measuring and Analyzing Energy Consumption of the Data Center”.
MA thesis. Aalto University.
- Marcos-Jorquera, Diego et al. (Sept. 2016). “Smart Monitoring Embedded Service for Energy-Efficient and Sustainable Management in Data Centers.”
In: *Energies* 9.515. ISSN: 9070515.
- Estimating the Energy Use and Efficiency Potential of U.S. Data Centers.* (2011).
Proceedings of the IEEE.

- Reddy, V. Dinesh et al. (2017). “Metrics for Sustainable Data Centers.”
In: *IEEE Transactions on Sustainable Computing, Sustainable Computing, IEEE Transactions on, IEEE Trans. Sustain. Comput* 3, p. 290. ISSN: 2377-3782.
- Rygielski, Piotr, Samuel Kounev, and Phuoc Tran-Gia (2016).
“Flexible Performance Prediction of Data Center Networks using Automatically Generated Simulation Models.” In: *EAI Endorsed Transactions on Scalable Information Systems, Vol 3, Iss 9, Pp 1-10 (2016)* 9, p. 1. ISSN: 2032-9407.
- Sönmez, Ferdi et al. (2018). “Anomaly Detection Using Data Mining Methods in IT Systems: A Decision Support Application.”
In: *Sakarya University Journal of Science* 22.4, p. 1109. ISSN: 13014048.
- TSSG Infrastructure Group (2018). <http://www.tssg.org/research/research-supports/infrastructure-group-2/>. Accessed: 2018-12-12.
- Vasudevan, Meera et al. (2017). “Profile-based application assignment for greener and more energy-efficient data centers”.
In: *Future Generation Computer Systems* 67, pp. 94–108. ISSN: 0167-739X.
- Wang, Junqi et al. (2017). “A Data Mining Approach to Discover Critical Events for Event-Driven Optimization in Building Air Conditioning Systems”.
In: *Energy Procedia* 143. Leveraging Energy Technologies and Policy Options for Low Carbon Cities, pp. 251–257. ISSN: 1876-6102.
- Xue, Puning et al. (2017). “Fault detection and operation optimization in district heating substations based on data mining techniques”.
In: *Applied Energy* 205, pp. 926–940. ISSN: 0306-2619.
- Zhou, Hao et al. (2018).
“Analysis of correlation between actual heating energy consumption and building physics, heating system, and room position using data mining approach”.
In: *Energy and Buildings* 166, pp. 73–82. ISSN: 0378-7788.

Appendices

A. MySQL Queries for Exploratory Data Analysis

This section provides the MySQL queries used to source the data for the line charts in Figures 7, 8, 9, 10 and 11.

```
1 use dc;
2 select *
3 from meters_data
4 where meter_reading_time > '2017-01-01'
5 and meter_reading_time < '2017-12-31'
6 and HOUR(`meter_reading_time`) = 9
7 and meter_id not in ('1','2','3','4','5','6','7','8','9','10','11',
8 '12','77','78');
```

Listing 1: MySQL Query for Annual Energy Consumption

```
1 use dc;
2 select *
3 from meters_data
4 where meter_reading_time > '2017-09-12'
5 and meter_reading_time < '2017-09-14'
6 where meter_id in ('13','14','15');
```

Listing 2: MySQL Query for Daily Energy Consumption

```
1 use dc;
2 select *
3 from meters_data
4 where meter_reading_time > '2017-01-01'
5 and meter_reading_time < '2017-12-31'
6 and HOUR(`meter_reading_time`) = 9
7 and MINUTE(`meter_reading_time`) = 0
8 and meter_id in ('3','4','13');
```

Listing 3: MySQL Query for Annual Energy Consumption of Chillers and Server

```
1 use dc;
2 select *
3 from lcps_data
4 where lcp_reading_time > '2017-01-01'
5 and lcp_reading_time < '2017-12-31'
6 and HOUR(`lcp_reading_time`) = 9
7 and MINUTE(`lcp_reading_time`) = 0;
```

Listing 4: MySQL Query for Annual Water Flowr

```
1 use dc;
2 select *
3 from meters_data
4 where meter_reading_time > '2017-10-10'
5 and meter_reading_time < '2017-10-24'
6 and meter_id = '2';
```

Listing 5: MySQL Query for Energy Consumption of the Generator during Storm Ophelia

B. MySQL Queries for Detailed Data Analysis

This section provides the MySQL queries used to source the data for section 4.3.2

```
1 use dc;
2 select *
3 from meters_data
4 where meter_reading_time > '2017-01-01'
5 and meter_reading_time < '2017-12-31';
```

Listing 6: MySQL Query for Energy Consumption in 2017

```
1 use dc;
2 select *
3 from meters_master;
```

Listing 7: MySQL Query for All Meter ID Location

```
1 use dc;
2 select *
3 from lcps_data
4 where lcp_reading_time > '2017-01-01'
5 and lcp_reading_time < '2017-12-31';
```

Listing 8: MySQL Query for LCP information for 2017

```
1 use dc;
2 select *
3 from lcps_master;
```

Listing 9: MySQL Query for LCP Location Information

```
1 use dc;
2 select *
3 from weather_data
4 where weather_reading_time > '2017-01-01'
5 and weather_reading_time < '2017-12-31';
```

Listing 10: MySQL Query for Weather Data for 2017

C. Python Code for Data Manipulation

This section provides Python code to cleanse, manipulate and derive correlations the data in section 4.3.5

```
1 import numpy as np
2 import pandas as pd
3 pd.options.display.max_rows = 1000
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6 sns.set_style("darkgrid")
7 sns.set_context("paper")
8 from itertools import combinations, groupby
9 from collections import Counter
10
11 import sys, os, glob
12 def size(obj):
13     """Return size of object in MB"""
14     return "{0:.2f} MB".format(sys.getsizeof(obj) / (1000 * 1000))
```

Listing 11: Python - Set Up and Configuration

```
1 energyusagepython = pd.read_csv("data/energyusagepython.csv")
2
3 metersmaster = pd.read_csv("data/metersmaster.csv")
4
5 energy_info = energyusagepython.merge(metersmaster, on="meter_id")
6
7 energy_regression = energy_info[['meter_reading_time',
8 'description',
9 'real_energy_consumption',
10 'last_time_period_kilowatts',
11 'apparent_energy_consumption',
12 'reactive_energy_consumption',
13 'real_power_total',
14 'apparent_power_total',
15 'reactive_power_total',
16 'power_factor']]
17
18 energy_regression.columns = ['dc_reading_time', 'meter_location',
19 'real_energy_consumption',
20 'last_time_period_kilowatts',
21 'apparent_energy_consumption',
22 'reactive_energy_consumption',
23 'real_power_total',
24 'apparent_power_total',
25 'reactive_power_total',
26 'power_factor']
```

Listing 12: Python - Import and Cleanse Energy Data

```
1 last_time_period_kilowatts_corr = energy_regression.pivot_table
2 (index='dc_reading_time',
3  columns='meter_location',
4  values='last_time_period_kilowatts')
5
6 last_time_period_kilowatts_corr.columns =
7  ['last_time_period_kilowatts.Arc Building',
8   'last_time_period_kilowatts.Beechfield House (MCCB 1)',
9   .....
10  'last_time_period_kilowatts.West Campus Main Incomer']
11
12 The same process was followed for all energy variables.
```

Listing 13: Python - Pivot Data around the meter reading time

```
1 last_time_period_kilowatts_corr.reset_index( inplace=True)
2 real_energy_consumption_corr.reset_index( inplace=True)
3 total_energy_info_1 = last_time_period_kilowatts_corr.merge
4 (real_energy_consumption_corr, on="dc_reading_time")
5
6 apparent_energy_consumption_corr.reset_index( inplace=True)
7 total_energy_info_2 = apparent_energy_consumption_corr.merge
8 (total_energy_info_1, on="dc_reading_time")
9
10 reactive_energy_consumption_corr.reset_index( inplace=True)
11 total_energy_info_3 = reactive_energy_consumption_corr.merge
12 (total_energy_info_2, on="dc_reading_time")
13
14 real_power_total_corr.reset_index( inplace=True)
15 total_energy_info_4 = real_power_total_corr.merge
16 (total_energy_info_3, on="dc_reading_time")
17
18 apparent_power_total_corr.reset_index( inplace=True)
19 total_energy_info_5 = apparent_power_total_corr.merge
20 (total_energy_info_4, on="dc_reading_time")
21
22 reactive_power_total_corr.reset_index( inplace=True)
23 total_energy_info_6 = reactive_power_total_corr.merge
24 (total_energy_info_5, on="dc_reading_time")
25
26 power_factor_corr.reset_index( inplace=True)
27 total_energy_info_final = power_factor_corr.merge
28 (total_energy_info_6, on="dc_reading_time")
```

Listing 14: Python - Merge Energy Pivot Tables together

```
1 lcpdata2017 = pd.read_csv("data/lcpdata2017.csv")
2 print("lcpdata2017 -- dimensions: {0}; size: {1}".format
3 (lcpdata2017.shape, size(lcpdata2017)))
4
5 lcpdmaster = pd.read_csv("data/lcpdmaster.csv")
6
7 lcp_info = lcpdata2017.merge(lcpdmaster, on="lcp_id")
8
9 lcp_regression = lcp_info[['lcp_reading_time',
10 'location',
11 'cooling_capacity',
12 'server_in_temp',
13 'server_out_temp',
14 'water_in_temp',
15 'water_out_temp',
16 'water_flow']]
17
18 lcp_regression['server_temp_difference']
19 = lcp_regression['server_out_temp'] -
20 lcp_regression['server_in_temp']
21 lcp_regression['water_temp_difference']
22 = lcp_regression['water_out_temp']
23 - lcp_regression['water_in_temp']
24
25 lcp_regression.columns = ['dc_reading_time',
26 'lcp_location',
27 '_lcp_cooling_capacity',
28 'server_in_temp',
29 'server_out_temp'
30 , 'water_in_temp',
31 'water_out_temp',
32 'water_flow',
33 'server_temp_diff',
34 'water_temp_diff']
```

Listing 15: Python - Import and Merge LCP Data

```
1 _lcp_cooling_capacity = lcp_regression.pivot_table
2 (index='dc_reading_time',columns='lcp_location',values=
3 '_lcp_cooling_capacity')
4
5 _lcp_cooling_capacity.columns =
6 ['lcp_cooling_capacity_Row 1 - Outer edge Cab 1',
7  'lcp_cooling_capacity_Row 1 - Outer edge Cab 5',
8  .....
9  'lcp_cooling_capacity_Row 5 - Nearest Build Room']
10
11 The same process was followed for all LCP variables.
```

Listing 16: Python - Pivot Data around the LCP reading time

```
1 _lcp_cooling_capacity.reset_index( inplace=True)
2 server_in_temp.reset_index( inplace=True)
3 lcp_info_1 = _lcp_cooling_capacity.merge(server_in_temp,
4 on="dc_reading_time")
5
6 server_out_temp.reset_index( inplace=True)
7 lcp_info_2 = server_out_temp.merge(lcp_info_1,
8 on="dc_reading_time")
9
10 water_in_temp.reset_index( inplace=True)
11 lcp_info_3 = water_in_temp.merge(lcp_info_2,
12 on="dc_reading_time")
13
14 water_out_temp.reset_index( inplace=True)
15 lcp_info_4 = water_out_temp.merge(lcp_info_3,
16 on="dc_reading_time")
17
18 water_flow.reset_index( inplace=True)
19 lcp_info_5 = water_flow.merge(lcp_info_4,
20 on="dc_reading_time")
21
22 server_temp_diff.reset_index( inplace=True)
23 lcp_info_6 = server_temp_diff.merge(lcp_info_5,
24 on="dc_reading_time")
25
26 water_temp_diff.reset_index( inplace=True)
27 lcp_info_final = water_temp_diff.merge(lcp_info_6,
28 on="dc_reading_time")
```

Listing 17: Python - Merge LCP Pivot Tables together

```
1 weatherdata = pd.read_csv("data/weatherdata.csv")
2
3 weatherdata.columns = ['weather_id',
4 'dc_reading_time',
5 'temperature',
6 'wind_chill'
7 , 'heat_index'
8 , 'humidity'
9 , 'dew_point'
10 , 'wind_speed',
11 'high_wind',
12 'wind_direction',
13 'rainfall'
14 , 'barometer',
15 'solar_radiation',
16 'evapotranspiration',
17 'ultraviolet_index']
18
19 revised_weatherdata = weatherdata[['dc_reading_time',
20 'temperature',
21 'rainfall',
22 'humidity',
23 'wind_speed']]
```

Listing 18: Python - Import and Merge Weather Data

```
1 final_data_merge_1
2 = total_energy_info_final.merge
3 (lcp_info_final, on="dc_reading_time")
4
5 final_data
6 = revised_weatherdata.merge
7 (final_data_merge_1, on="dc_reading_time")
8
9 final_data.to_csv("final_data.csv",
10 index=False, header=True)
```

Listing 19: Python - Merge all Data Together into one table

```

1 final_data_correlations = final_data.corr()
2
3 final_data_correlations.to_csv
4 (\ "final_data_correlations.csv\ ",
5 index=False, header=True)

```

Listing 20: Python - Get Correlations between all Variables

```

1 def get_redundant_pairs(final_data_correlations):
2     '''Get diagonal and lower triangular pairs of correlation matrix'''
3     pairs_to_drop = set()
4     cols = final_data_correlations.columns
5     for i in range(0, final_data_correlations.shape[1]):
6         for j in range(0, i+1):
7             pairs_to_drop.add((cols[i], cols[j]))
8     return pairs_to_drop
9
10 def get_top_abs_correlations(final_data_correlations, n=5):
11     au_corr = final_data_correlations.corr().abs().unstack()
12     labels_to_drop = get_redundant_pairs
13     (final_data_correlations)
14     au_corr = au_corr.drop
15     (labels=labels_to_drop).
16     sort_values(ascending=False)
17     return au_corr[0:n]
18
19 Top_correlations =
20 (get_top_abs_correlations
21 (final_data_correlations, 45000))
22
23 Top_correlations.to_csv
24 ("Top_Correlation_45000.csv",
25 index=True, header=True)

```

Listing 21: Python - Analyse Correlations in Python

D. Python Code for Regression Model

This section provides Python code to construct the regression model and goodness of fit measures as outlined in sections 4.3.9 and 4.3.10

```

1 import numpy as np
2 import pandas as pd
3 pd.options.display.max_rows = 1000
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6 sns.set_style("darkgrid")
7 sns.set_context("paper")
8 from itertools import combinations, groupby
9 from collections import Counter
10 import scipy.stats as stats
11 import sys, os, glob
12 from pandas import Series
13 from statsmodels.tsa.stattools import adfuller
14 from sklearn import linear_model
15 import statsmodels.api as sm
16 import statsmodels.tsa.api as smt
17 from scipy.stats.stats import pearsonr
18 from scipy import stats
19 import statsmodels.formula.api as smf
20 import statsmodels.stats as sm_stats
21 from sklearn import linear_model
22
23 def size(obj):
24     """Return size of object in MB"""
25     return "{0:.2f} MB".format(sys.getsizeof(obj) / (1000 * 1000))

```

Listing 22: Python - Set Up and Configuration for Regression

```

1 final_data = pd.read_csv("final_data_revised_2.csv")
2 print("final_data -- dimensions: {0}; size: {1}".format(final_data.shape, size(final_data)))
3 display(final_data.head())
4
5 final_data_revised = final_data[['dc_reading_time', 'temperature', 'rainfall', 'humidity', 'wind_speed',
6     'apparent_power_total_corr.CAB10LFT(B)',
7     'apparent_power_total_corr.CAB10RGT(A)',
8     'apparent_power_total_corr.CAB12LFT(B)',
9     'apparent_power_total_corr.CAB12RGT(A)',
10    'apparent_power_total_corr.CAB13LFT(B)',
11    'apparent_power_total_corr.CAB13RGT(A)',
12    'apparent_power_total_corr.CAB17LFT(B)',
13    'apparent_power_total_corr.CAB17RGT(A)',
14    'apparent_power_total_corr.CAB18LFT(B)',
15    'apparent_power_total_corr.CAB18RGT(A)',
16    'apparent_power_total_corr.CAB19LFT(B)',
17    'apparent_power_total_corr.CAB19RGT(A)',
18    'apparent_power_total_corr.CAB22LFT(B)',
19
20     -----
21
22     'server_temp_diff_Row 4 - Outer edge Cab 22',
23     'water_temp_diff_Between Cab 23 and Cab 24',
24     'water_temp_diff_Between Cab 24 and Cab 25',
25     'water_temp_diff_Between Cab 25 and Cab 26',
26     'water_temp_diff_Outside Edge of Cab 23',
27     'water_temp_diff_Row 3 - Between Cab 13 and Cab 14',
28     'water_temp_diff_Row 3 - Outer edge of Cab 12',
29     'water_temp_diff_Row 4 - Between Cab 19 and Cab 20',
30     'water_temp_diff_Row 4 - Outer edge Cab 22']]
31 display(final_data_revised.head())

```

Listing 23: Python - Import and Manipulate data for Regression

```

1 Peak3_Week1 = final_data_revised[(final_data_revised['dc_reading_time'] > '2017-08-29 11:00:00') & (
2
3 Peak3_Week1.drop(['dc_reading_time'], axis=1, inplace=True)
4
5 Peak3_Week1.to_csv("Peak3_Week1.csv", index=True, header=True)
6
7 from pandas import Series
8 from statsmodels.tsa.stattools import adfuller
9 series = Series.from_csv('Peak3_Week1.csv', header=0)
10 X = series.values
11 result = adfuller(X)
12 print('ADF Statistic: %f' % result[0])
13 print('p-value: %f' % result[1])
14 print('Critical Values:')
15 for key, value in result[4].items():
16     print('\t%s: %.3f' % (key, value))

```

Listing 24: Python - Create Micro-Timeline and Check that Time Series is Stationary

```

1 Peak3_Week1.drop(['apparent_power_total_corr.CHILLER1 (N3)'], axis=1, inplace=True)
2 y = Peak3_Week1['apparent_power_total_corr.CHILLER2 (N2)'].copy()
3 X = Peak3_Week1.copy()
4 X.drop(['apparent_power_total_corr.CHILLER2 (N2)'], axis=1, inplace=True)
5
6 model = linear_model.LinearRegression()
7
8 X_constant = sm.add_constant(X)
9 lin_reg_Peak3_Week1 = sm.OLS(y,X_constant).fit()
10 lin_reg_Peak3_Week1.summary()

```

Listing 25: Python - Construct Regression Model and Execute

```

1 acf = smt.graphics.plot_acf(lin_reg_Peak3_Week1.resid, lags=40 , alpha=0.05)
2 acf.show()
3
4 for column in X.columns:
5     corr_test = pearsonr(X[column], lin_reg_Peak3_Week1.resid)
6     print(f'Variable: {column} --- correlation: {corr_test[0]:.4f}, p-value: {corr_test[1]:.4f}')
7
8 X.apply(np.var, axis=0)
9
10 from scipy import stats
11
12 def normality_of_residuals_test(model):
13     '''
14     Function for drawing the normal QQ-plot of the residuals and running 4 statistical tests to
15     investigate the normality of residuals.
16
17     Arg:
18     * model - fitted OLS models from statsmodels
19     '''
20     sm.ProbPlot(model.resid).qqplot(line='s');
21     plt.title('Q-Q plot');
22
23     jb = stats.jarque_bera(model.resid)
24     sw = stats.shapiro(model.resid)
25     ad = stats.anderson(model.resid, dist='norm')
26     ks = stats.kstest(model.resid, 'norm')
27
28     print(f'Jarque-Bera test ---- statistic: {jb[0]:.4f}, p-value: {jb[1]}')
29     print(f'Shapiro-Wilk test ---- statistic: {sw[0]:.4f}, p-value: {sw[1]:.4f}')
30     print(f'Kolmogorov-Smirnov test ---- statistic: {ks.statistic:.4f}, p-value: {ks.pvalue:.4f}')
31     print(f'Anderson-Darling test ---- statistic: {ad.statistic:.4f}, 5% critical value: {ad.critical}')
32     print('If the returned AD statistic is larger than the critical value, then for the 5% significance level, the residuals are not normally distributed.')
33
34 normality_of_residuals_test(lin_reg_Peak3_Week1)

```

Listing 26: Python - Check Regression Model for Auto-Correlation and Other Tests

```

1 lin_reg_Peak3_Week1_CORC = smf.GLSAR(y,X)
2 lin_reg_Peak3_Week1_CORC_fit = lin_reg_Peak3_Week1_CORC.iterative_fit(maxiter = 100)
3 lin_reg_Peak3_Week1_CORC_fit.summary()
4
5 print(lin_reg_Peak3_Week1_CORC.rho)

```

Listing 27: Python - Run Cochrane-Orcutt Procedure (to eliminate autocorrelation)

E. Definition of Key Variables

This section provides definitions of key variables as discussed in section 4.3.5

Figure 16: Definition of key Variables from the Monitoring Data

Definition of Key Variables	
Real Energy Consumption	Real Energy is the actual power consumed by the equipment to do useful work. It is distinguished from apparent power by eliminating the reactive power component that may be present
Reactive Energy Consumption	Reactive Energy is the part of complex power that corresponds to storage and retrieval of energy rather than consumption
Apparent Energy Consumption	The combination of reactive power and real power is called apparent power, and it is the product of a circuit's voltage and current, without reference to phase angle. Apparent power is measured in the unit of Volt-Amps (VA) and is symbolized by the capital letter S
Power Factor	Power Factor (PF) is the ratio between 0 and 1 of real power (kW) to apparent power (kVA)
LCP Cooling Capacity	The capacity of the Liquid Cooling Package to cool a rack in a Data Centre

F. Detailed Correlation Analysis Results for Overall Data Centre

This section provides the detailed correlation analysis as outlined in section 4.3.5

Figure 17: The Detailed Correlation Analysis Results - section 1

Variable	Highly Correlated with	Other Comments
Humidity	Strong Correlation between Humidity and (Real energy consumption, apparent energy consumption and reactive energy consumption)	Humidity is only strongly correlated with the LCP use of one area - row 3 outer area row 12
	Out of the 83 meter ID's - 77 have a strong correlation between Apparent Energy consumption and Humidity (>0.93)	
Apparent Energy Consumption	real_energy consumption This indicates that there is a relatively stable amount of reactive power generated at each meter_id All 83 meter id's are correlated with each other in terms of apparent_energy_consumption apart from apparent_energy_consumption_corr.CAB15LFT(A)	For example, the apparent energy in CAB1(A) is highly correlated to the real energy consumption in 79/93 other meter I'd's
Real_Power_Total	High amount of correlation between real_power_total and last_time_kw 38 electrical meters have a correlation higher than 0.93	
Chiller 1 (following variables) apparent energy consumption of chiller 1 is correlated to power factor of chiller 1 is correlated to power factor of chiller 1 is correlated to power factor of chiller 1 is correlated to power factor of chiller 1 is correlated to reactive energy consumption of chiller 1 is correlated to reactive power total of chiller 1 is correlated to reactive power total of chiller 1 is correlated to reactive power total of chiller 1 is correlated to real energy consumption of chiller 1 is correlated to	Out of 2103 correlations, 2089 are identical for both Chiller 1 and chiller 2, the following 14 are the only differences between chiller 1 and chiller 2 (might explain variances in energy consumption) apparent_energy_consumption_corr.CHILLER2 (N2) reactive_power_total_corr.Carriganore House (MCCB 2) apparent_power_total_corr.Carriganore House (MCCB 2) real_power_total_corr.Carriganore House (MCCB 2) last_time_period_kilowatts.Carriganore House (MCCB 2) reactive_energy_consumption_corr.CHILLER2 (N2) apparent_power_total_corr.CHILLER1 (N3) real_power_total_corr.CHILLER1 (N3) last_time_period_kilowatts.CHILLER1 (N3) real_energy_consumption_corr.CHILLER2 (N2)	
Chiller 2 (following variables) apparent energy consumption of chiller 2 is correlated to reactive energy consumption of chiller 2 is correlated to real energy consumption of chiller 2 is correlated to real energy consumption of chiller 2 is correlated to	water_flow_Between Cab 23 and Cab 24 server_out_temp_Between Cab 24 and Cab 25 lcp_cooling_capacity_Between Cab 23 and Cab 24 water_flow_Between Cab 23 and Cab 24	

Figure 18: The Detailed Correlation Analysis Results - section 2

Variable	Highly Correlated with	Other Comments
Server in temp/Water in temp of the following locations: Between Cab 23 and 24 Between Cab 24 and 25 Between Cab 25 and 26	lcp_cooling_capacity_Row 2 - Between Cab 8 and Cab 9 Row 1 - Outer Edge Cab 5 server_in_temp_Row 2 - Between Cab 8 and Cab 9 server_in_temp_Row 3 - Between Cab 13 and Cab 14 server_in_temp_Row 3 - Outer edge Cab 16 server_in_temp_Row 3 - Outer edge of Cab 12 server_in_temp_Row 4 - Between Cab 19 and Cab 20 server_in_temp_Row 4 - Outer edge Cab 22 water_in_temp_Row 1 - Outer edge Cab 1 water_in_temp_Row 2 - Between Cab 8 and Cab 9 water_in_temp_Row 2 - Outer edge Cab 11 water_in_temp_Row 2 - Outer Wall Cab 6 water_in_temp_Row 3 - Between Cab 13 and Cab 14 water_in_temp_Row 3 - Outer edge Cab 16 water_in_temp_Row 3 - Outer edge of Cab 12 water_in_temp_Row 4 - Between Cab 19 and Cab 20 water_in_temp_Row 4 - Outer edge Cab 22	
Real Energy Consumption/Apparent Energy Consumption of most Meter ID's is highly correlated to:	server_temp_diff_Row 3 - Outer edge of Cab 12 server_out_temp_Row 3 - Outer edge of Cab 12 lcp_cooling_capacity_Between Cab 24 and Cab 25 water_flow_Between Cab 23 and Cab 24 lcp_cooling_capacity_Between Cab 23 and Cab 24	In terms of cross-checking meter-id energy usage and LCP info - this is extraordinary consistent and the only meaningful insight Why are these 5 LCP variables to be linked to the real energy consumption of the majority of meter_id's

G. Detailed Correlation Analysis Results for each Chiller

Figure 19: Correlation Analysis of each Chiller for 4/28 to 5/9

	Chiller 2	Chiller 1
Peak 1:	4/25 to 5/2	5/2 to 5/9
	30,000 top Correlations taken among 759 metrics	30,000 top Correlations taken among 759 metrics
	715 relate to Chiller 2	701 relate to Chiller 1
	559 Correlations match between both Chillers	
	Average Difference of these 559 = 0.001905	
	Chiller 2 Correlation > Chiller 1 Correlation = 26 Instances	Chiller 1 Correlation > Chiller 2 Correlation = 533 Instances
	Unique Correlations for each Chiller for the week its in operation	
	155	136
	Energy Consumption C2 is UNIQUELY correlated to	Energy Consumption C1 is UNIQUELY correlated to
	Energy consumption of Cab 1 (7 different metrics) Energy Consumption of Cab 17 (15 different metrics) Energy Consumption of LCP 19 (12 different metrics) LCP 5 (6 different metrics)	Energy consumption of Cab 13 (27 different metrics) Energy Consumption of Cab 3 (9 different metrics) Energy Consumption of LCP 7 (12 different metrics) LCP 7 (12 different metrics)
	Humidity	LCP Cooling Capacity of Row 3 - Between Cab 13 & Cab 14 Water Flow of Row 3 - Between Cab 13 & Cab 14 Water Temp Diff of Row 3 - Between Cab 13 & Cab 14 Server Temp Difference of Row 3 - Between Cab 13 & Cab 14 Server Out Temp of Row 3 - Between Cab 13 & Cab 15

Figure 20: Correlation Analysis of each Chiller for 6/6 to 6/20

	Chiller 2	Chiller 1
Peak 2:	6/6 to 6/13	6/13 to 6/20
	30,000 top Correlations taken among 759 metrics	30,000 top Correlations taken among 759 metrics
	824 relate to Chiller 2	1525 relate to Chiller 1
	559 Correlations match between both Chillers (501 of these match Peak 1)	
	Average Difference of these 559 = 0.001579	
	Chiller 2 Correlation > Chiller 1 Correlation = 211 Instances	Chiller 1 Correlation > Chiller 2 Correlation = 348 Instances
	Unique Correlations for each Chiller for the week its in operation	
	265	940
	Energy Consumption C2 is UNIQUELY correlated to	Energy Consumption C1 is UNIQUELY correlated to
	Energy consumption of Cab 22 (43 different metrics) Energy Consumption of Cab 10 (21 different metrics) Energy Consumption of LCP 17 (24 different metrics) LCP 1 (5 different metrics) LCP 9 (5 different metrics)	Energy consumption of Cab 3 (39 different metrics) Energy Consumption of LCP 7 (20 different metrics) Energy Consumption of Cab 12 (23 different metrics) Energy Consumption of Cab 18 (27 different metrics) LCP 5 (15 different metrics) LCP 7 (12 different metrics)
	LCP Cooling Capacity of Row 4 - Outer Edge Cab 22 (4 different metrics) Water Flow of Row 4 - Outer Edge Cab 22 (7 different metrics)	Water Flow of Row 3 - Between Cab 13 & Cab 14 Temperature (3 different metrics) Wind Speed (2 different metrics)

Figure 21: Correlation Analysis of each Chiller for 8/29 to 9/12

	Chiller 2	Chiller 1
Peak 3:	8/29 to 9/5	9/5 to 9/12
	30,000 top Correlations taken among 759 metrics	30,000 top Correlations taken among 759 metrics
	947 relate to Chiller 2	822 relate to Chiller 1
	625 Correlations match between both Chillers	
	Average Difference (in correlation) of these 625 = 0.003988	
	Chiller 2 Correlation > Chiller 1 Correlation = 412 Instances	Chiller 1 Correlation > Chiller 2 Correlation = 213 Instances
	Unique Correlations for each Chiller for the week its in operation	
	322	173
	Energy Consumption C2 is UNIQUELY correlated to	Energy Consumption C1 is UNIQUELY correlated to
	Energy consumption of Cab 22 (39 different metrics) Energy consumption of Cab 23 (40 different metrics) Energy consumption of Cab 24 (56 different metrics) Energy consumption of Cab 26 (24 different metrics) LCP 12 (13 different metrics) LCP 13 (8 different metrics) LCP Cooling Capacity Between Cab 24 and 25 LCP Cooling Capacity Outside Edge of Cab 23 LCP Cooling Capacity Row 4 - Outer edge Cab 22 Server Temp Diff Between Cab 24 and 25 Server Temp Diff Outside Edge of Cab 23 Server Temp Diff - Outer edge Cab 22 Server Temp Diff Between Cab 19 and 20 Row 5 - Nearest Build Room (10 electrical and chiller metrics)	Energy consumption of Cab 11 (24 different metrics) LCP 2 (12 different metrics) LCP 11 (12 different metrics) LCP 5 (9 different metrics) Water Temp Diff of Row 3 - Between Cab 24 & Cab 25 Wind Speed (3 different metrics)

H. Full list of Independent Variables for Regression Model

Figure 22: Full list of Independent Variables used in the Regression Model

Final List of Independent Variables used in Regression Model		
temperature rainfall humidity wind_speed apparent_power_total_corr.CAB10LFT(B) apparent_power_total_corr.CAB10RGT(A) apparent_power_total_corr.CAB12LFT(B) apparent_power_total_corr.CAB12RGT(A) apparent_power_total_corr.CAB13LFT(B) apparent_power_total_corr.CAB13RGT(A) apparent_power_total_corr.CAB17LFT(B) apparent_power_total_corr.CAB17RGT(A) apparent_power_total_corr.CAB18LFT(B) Apparent_power_total_corr.CAB18RGT(A) apparent_power_total_corr.CAB19LFT(B) apparent_power_total_corr.CAB19RGT(A) apparent_power_total_corr.CAB22LFT(B)	apparent_power_total_corr.CAB22RGT(A) apparent_power_total_corr.CAB23LFT(B) apparent_power_total_corr.CAB23RGT(A) apparent_power_total_corr.CAB24LFT(B) apparent_power_total_corr.CAB24RGT(A) apparent_power_total_corr.CAB26LFT(B) apparent_power_total_corr.CAB26RGT(A) apparent_power_total_corr.CAB3(B) apparent_power_total_corr.CAB3RGT(B) apparent_power_total_corr.CAB9LFT(B) apparent_power_total_corr.CAB9RGT(A) apparent_power_total_corr.GENERATOR (P3) apparent_power_total_corr.LCP11(A) apparent_power_total_corr.LCP12(B) apparent_power_total_corr.LCP13(A) apparent_power_total_corr.LCP2(A) apparent_power_total_corr.LCP5(A)	apparent_power_total_corr.LCP7(A) server_temp_diff_Between Cab 23 and Cab 24 server_temp_diff_Between Cab 24 and Cab 25 server_temp_diff_Between Cab 25 and Cab 26 server_temp_diff_Outside Edge of Cab 23 server_temp_diff_Row 3 - Between Cab 13 and Cab 14 server_temp_diff_Row 3 - Outer edge of Cab 12 server_temp_diff_Row 4 - Between Cab 19 and Cab 20 server_temp_diff_Row 4 - Outer edge Cab 22 water_temp_diff_Between Cab 23 and Cab 24 water_temp_diff_Between Cab 24 and Cab 25 water_temp_diff_Between Cab 25 and Cab 26 water_temp_diff_Outside Edge of Cab 23 water_temp_diff_Row 3 - Between Cab 13 and Cab 14 water_temp_diff_Row 3 - Outer edge of Cab 12 water_temp_diff_Row 4 - Between Cab 19 and Cab 20 water_temp_diff_Row 4 - Outer edge Cab 22

I. Summary of the Output of the Regression Model

Figure 23: Regression Output from Data from April 25th to May 9th

Regression Output from Micro-timeline 1 - 25th April through to 9th May							
Variable	Week 1 Output			Week 2 Output			Result:
	Durbin Watson Value: 2.157			Durbin Watson Value: 1.915			
	coef	[0.025	0.975]	coef	[0.025	0.975]	
temperature	179800.000	145000.000	214000.000	17020.000	4408.691	29600.000	Variables Don't Match
rainfall	81550.000	-137000.000	300000.000	0.000	0.000	0.000	Variables Don't Match
humidity	16140.000	8394.396	23900.000	1302.530	-1280.638	3885.698	Variables Don't Match
wind_speed	16110.000	7258.062	25000.000	-15.049	-2856.438	2826.339	Variables Don't Match
apparent_power_total_corr.CAB10LFT(B)	-0.273	-3.863	3.317	-0.762	-3.379	1.856	Variables Match
apparent_power_total_corr.CAB10RGT(A)	-2.084	-7.120	2.951	-2.579	-6.029	0.871	Variables Match
apparent_power_total_corr.CAB12LFT(B)	-2.440	-12.440	7.559	-4.508	-12.378	3.361	Variables Match
apparent_power_total_corr.CAB12RGT(A)	0.445	-9.658	10.547	3.412	-4.435	11.259	Variables Match
apparent_power_total_corr.CAB13LFT(B)	7.035	1.932	12.137	-2.954	-7.353	1.444	Variables Don't Match
apparent_power_total_corr.CAB13RGT(A)	-3.953	-9.501	1.596	1.337	-3.165	5.838	Variables Don't Match
apparent_power_total_corr.CAB17LFT(B)	-1.941	-6.961	3.079	5.089	1.238	8.941	Variables Don't Match
apparent_power_total_corr.CAB17RGT(A)	-5.890	-11.220	-0.559	-1.816	-5.822	2.189	Variables Don't Match
apparent_power_total_corr.CAB18LFT(B)	-7.338	-15.680	1.004	-3.014	-7.618	1.590	Variables Match
Apparent_power_total_corr.CAB18RGT(A)	1.954	-9.701	13.609	13.246	5.055	21.437	Variables Don't Match
apparent_power_total_corr.CAB19LFT(B)	8.587	-0.501	17.675	-2.404	-4.828	0.021	Variables Don't Match
apparent_power_total_corr.CAB19RGT(A)	4.247	-4.240	12.735	-1.946	-6.033	2.142	Variables Don't Match
apparent_power_total_corr.CAB22LFT(B)	0.878	-0.734	2.490	-0.153	-1.475	1.170	Variables Match
apparent_power_total_corr.CAB22RGT(A)	-0.122	-1.754	1.511	-1.085	-2.391	0.220	Variables Match
apparent_power_total_corr.CAB23LFT(B)	-0.022	-0.882	0.837	-0.722	-1.417	-0.026	Variables Don't Match
apparent_power_total_corr.CAB23RGT(A)	0.438	-0.419	1.294	0.007	-0.651	0.666	Variables Match
apparent_power_total_corr.CAB24LFT(B)	0.105	-0.979	1.190	0.266	-0.670	1.201	Variables Match
apparent_power_total_corr.CAB24RGT(A)	0.163	-0.715	1.042	0.038	-0.702	0.777	Variables Match
apparent_power_total_corr.CAB26LFT(B)	0.567	-0.527	1.662	0.172	-0.528	0.872	Variables Match
apparent_power_total_corr.CAB26RGT(A)	0.391	-0.727	1.509	0.408	-0.283	1.098	Variables Match
apparent_power_total_corr.CAB3(B)	2.109	-6.143	10.361	-1.984	-8.463	4.495	Variables Match
apparent_power_total_corr.CAB3RGT(B)	1.920	-6.794	10.632	-0.926	-7.333	5.480	Variables Match
apparent_power_total_corr.CAB9LFT(B)	-3.173	-7.117	0.771	-0.456	-3.357	2.444	Variables Match
apparent_power_total_corr.CAB9RGT(A)	8.956	-5.726	23.638	-0.521	-12.616	11.574	Variables Match
apparent_power_total_corr.GENERATOR_(P3)	0.000	0.000	0.000	0.000	0.000	0.000	Variables Don't Match
apparent_power_total_corr.LCP11(A)	-47.296	-90.764	-3.828	6.173	-23.335	35.680	Variables Don't Match
apparent_power_total_corr.LCP12(B)	-1.652	-4.756	1.451	-0.509	-2.850	1.833	Variables Match
apparent_power_total_corr.LCP13(A)	8.381	4.556	12.207	2.690	0.286	5.095	Variables Don't Match
apparent_power_total_corr.LCP2(A)	9.710	-26.506	45.925	13.817	-16.816	44.451	Variables Match
apparent_power_total_corr.LCP5(A)	-16.076	-51.937	19.786	3.661	-21.566	28.888	Variables Match
apparent_power_total_corr.LCP7(A)	0.821	-2.439	4.081	-1.592	-4.468	1.285	Variables Match
server_temp_diff_Between_Cab_23_and_Cab_24	59040.000	33100.000	85000.000	7909.898	-12700.000	28500.000	Variables Don't Match
server_temp_diff_Between_Cab_24_and_Cab_25	40860.000	13300.000	68400.000	17710.000	-3496.799	38900.000	Variables Don't Match
server_temp_diff_Between_Cab_25_and_Cab_26	23410.000	-3848.968	50700.000	9652.619	-17500.000	36800.000	Variables Match
server_temp_diff_Outside_Edge_of_Cab_23	33050.000	10500.000	55600.000	24540.000	4927.764	44200.000	Variables Match
server_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14	-3666.556	-29300.000	22000.000	33160.000	6125.203	60200.000	Variables Don't Match
server_temp_diff_Row_3_-_Outer_edge_of_Cab_12	-982.996	-30800.000	28800.000	103800.000	58200.000	149000.000	Variables Don't Match
server_temp_diff_Row_4_-_Between_Cab_19_and_Cab_20	81620.000	53100.000	110000.000	60890.000	25900.000	95900.000	Variables Match
server_temp_diff_Row_4_-_Outer_edge_Cab_22	26500.000	2316.059	50700.000	17630.000	-8403.842	43700.000	Variables Match
water_temp_diff_Between_Cab_23_and_Cab_24	-12880.000	-42800.000	17100.000	-51720.000	-77300.000	-26100.000	Variables Don't Match
water_temp_diff_Between_Cab_24_and_Cab_25	-22850.000	-49900.000	4187.739	-105900.000	-129000.000	-82300.000	Variables Don't Match
water_temp_diff_Between_Cab_25_and_Cab_26	21030.000	-8675.505	50700.000	-23740.000	-48600.000	1115.162	Variables Don't Match
water_temp_diff_Outside_Edge_of_Cab_23	38380.000	9887.560	66900.000	79000.000	54200.000	104000.000	Variables Don't Match
water_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14	-66.556	-24400.000	24300.000	29970.000	5140.763	54800.000	Variables Don't Match
water_temp_diff_Row_3_-_Outer_edge_of_Cab_12	-61980.000	-88300.000	-35600.000	-11750.000	-34800.000	11300.000	Variables Don't Match
water_temp_diff_Row_4_-_Between_Cab_19_and_Cab_20	33700.000	-2345.698	69800.000	10240.000	-22200.000	42700.000	Variables Match
water_temp_diff_Row_4_-_Outer_edge_Cab_22	-15530.000	-40500.000	9444.721	58180.000	32500.000	83800.000	Variables Don't Match

Figure 24: Regression Output from Data from June 6th to June 20th

Regression Output from Micro-timeline 2 - 6th June through to 20th June							
Variable	Week 1 Output			Week 2 Output			Result:
	Durbin Watson Value: 1.847			Durbin Watson Value: 1.828			
	coef	[0.025	0.975]	coef	[0.025	0.975]	
temperature	35370.000	12300.000	58500.000	62980.000	45700.000	80300.000	Variables Don't Match
rainfall	-80050.000	-151000.000	-9229.746	175100.000	-332000.000	682000.000	Variables Match
humidity	1573.058	-2792.322	5938.437	-1998.487	-7193.584	3196.610	Variables Match
wind_speed	-2262.594	-6064.235	1539.047	2820.330	-1801.782	7442.442	Variables Don't Match
apparent_power_total_corr.CAB10LFT(B)	1.404	-6.010	8.817	-2.567	-8.098	2.963	Variables Match
apparent_power_total_corr.CAB10RGT(A)	4.855	-7.911	17.621	0.824	-9.604	11.253	Variables Match
apparent_power_total_corr.CAB12LFT(B)	-14.915	-31.080	1.250	8.282	-2.126	18.690	Variables Don't Match
apparent_power_total_corr.CAB12RGT(A)	14.384	-1.256	30.024	-4.595	-14.382	5.191	Variables Don't Match
apparent_power_total_corr.CAB13LFT(B)	-2.725	-17.368	11.919	-0.926	-3.799	1.946	Variables Match
apparent_power_total_corr.CAB13RGT(A)	6.038	-3.499	15.575	-0.256	-3.160	2.648	Variables Don't Match
apparent_power_total_corr.CAB17LFT(B)	1.316	-5.921	8.552	0.095	-4.557	4.747	Variables Match
apparent_power_total_corr.CAB17RGT(A)	-2.103	-9.910	5.704	0.524	-4.293	5.341	Variables Match
apparent_power_total_corr.CAB18LFT(B)	-0.706	-9.084	7.671	3.633	-2.803	10.068	Variables Match
Apparent_power_total_corr.CAB18RGT(A)	-3.858	-18.547	10.831	-6.363	-16.092	3.367	Variables Match
apparent_power_total_corr.CAB19LFT(B)	-2.365	-12.709	7.980	1.245	-3.637	6.128	Variables Match
apparent_power_total_corr.CAB19RGT(A)	-6.379	-16.036	3.278	5.344	-1.046	11.734	Variables Don't Match
apparent_power_total_corr.CAB22LFT(B)	-3.137	-5.837	-0.438	-0.855	-2.415	0.705	Variables Don't Match
apparent_power_total_corr.CAB22RGT(A)	-0.360	-3.067	2.348	0.085	-1.458	1.629	Variables Match
apparent_power_total_corr.CAB23LFT(B)	0.025	-1.403	1.453	-2.015	-2.908	-1.122	Variables Don't Match
apparent_power_total_corr.CAB23RGT(A)	-0.981	-2.365	0.403	1.077	0.198	1.955	Variables Don't Match
apparent_power_total_corr.CAB24LFT(B)	-1.704	-4.278	0.870	-0.956	-2.641	0.729	Variables Match
apparent_power_total_corr.CAB24RGT(A)	-0.138	-2.212	1.937	0.309	-1.038	1.656	Variables Match
apparent_power_total_corr.CAB26LFT(B)	-0.490	-2.566	1.586	0.247	-1.233	1.727	Variables Match
apparent_power_total_corr.CAB26RGT(A)	-0.729	-2.778	1.321	-0.480	-1.961	1.001	Variables Match
apparent_power_total_corr.CAB3(B)	-8.540	-20.809	3.730	3.348	-5.402	12.098	Variables Don't Match
apparent_power_total_corr.CAB3RGT(B)	-3.314	-15.024	8.395	-3.311	-12.034	5.413	Variables Match
apparent_power_total_corr.CAB9LFT(B)	3.798	-2.406	10.002	3.247	-0.427	6.921	Variables Match
apparent_power_total_corr.CAB9RGT(A)	-1.294	-9.645	7.058	-1.160	-10.953	8.634	Variables Match
apparent_power_total_corr.GENERATOR_(P3)	0.000	0.000	0.000	0.000	0.000	0.000	Variables Match
apparent_power_total_corr.LCP11(A)	-8.098	-111.740	95.544	-34.181	-79.047	10.686	Variables Match
apparent_power_total_corr.LCP12(B)	4.115	-0.460	8.689	-1.510	-4.283	1.263	Variables Don't Match
apparent_power_total_corr.LCP13(A)	26.460	20.399	32.521	11.665	6.250	17.079	Variables Don't Match
apparent_power_total_corr.LCP2(A)	-33.316	-128.964	62.332	-0.570	-43.141	42.002	Variables Match
apparent_power_total_corr.LCP5(A)	5.925	-66.626	78.476	-6.851	-38.145	24.443	Variables Match
apparent_power_total_corr.LCP7(A)	12.544	3.981	21.107	5.499	1.829	9.170	Variables Don't Match
server_temp_diff_Between_Cab_23_and_Cab_24	72670.000	37700.000	108000.000	26940.000	-2246.459	56100.000	Variables Don't Match
server_temp_diff_Between_Cab_24_and_Cab_25	69950.000	38900.000	101000.000	41250.000	15600.000	66900.000	Variables Don't Match
server_temp_diff_Between_Cab_25_and_Cab_26	67150.000	29800.000	105000.000	30250.000	-1952.295	62500.000	Variables Don't Match
server_temp_diff_Outside_Edge_of_Cab_23	66870.000	34700.000	99100.000	45410.000	19500.000	71300.000	Variables Match
server_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14	-10340.000	-48600.000	27900.000	55810.000	23800.000	87800.000	Variables Don't Match
server_temp_diff_Row_3_-_Outer_edge_of_Cab_12	-51190.000	-90300.000	-12100.000	-46260.000	-90500.000	-2047.629	Variables Match
server_temp_diff_Row_4_-_Between_Cab_19_and_Cab_20	195500.000	155000.000	236000.000	220500.000	170000.000	271000.000	Variables Match
server_temp_diff_Row_4_-_Outer_edge_Cab_22	139700.000	104000.000	175000.000	51200.000	22700.000	79800.000	Variables Don't Match
water_temp_diff_Between_Cab_23_and_Cab_24	61640.000	19200.000	104000.000	71810.000	31400.000	112000.000	Variables Match
water_temp_diff_Between_Cab_24_and_Cab_25	-58010.000	-94600.000	-21400.000	46130.000	12000.000	80200.000	Variables Don't Match
water_temp_diff_Between_Cab_25_and_Cab_26	141200.000	100000.000	182000.000	68990.000	25900.000	112000.000	Variables Don't Match
water_temp_diff_Outside_Edge_of_Cab_23	44300.000	3580.716	85000.000	157400.000	120000.000	195000.000	Variables Don't Match
water_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14	-52060.000	-84500.000	-19600.000	-32860.000	-61600.000	-4103.620	Variables Match
water_temp_diff_Row_3_-_Outer_edge_of_Cab_12	-63480.000	-99100.000	-27900.000	-113300.000	-144000.000	-82900.000	Variables Don't Match
water_temp_diff_Row_4_-_Between_Cab_19_and_Cab_20	-6791.959	-60200.000	46600.000	119200.000	58200.000	180000.000	Variables Don't Match
water_temp_diff_Row_4_-_Outer_edge_Cab_22	136600.000	92500.000	181000.000	2856.472	-25300.000	31000.000	Variables Don't Match

Figure 25: Regression Output from Data from August 29th to September 12th

Regression Output from Micro-timeline 3 - 29th August through to 12th September							
Variable	Week 1 Output			Week 2 Output			Result:
	Durbin Watson Value: 1.906			Durbin Watson Value: 1.861			
	coef	[0.025	0.975]	coef	[0.025	0.975]	
temperature	65900.000	47500.000	84300.000	99300.000	85200.000	113000.000	Variables Don't Match
rainfall	-6964.283	-86700.000	72800.000	-154900.000	-410000.000	100000.000	Variables Match
humidity	3092.437	-1510.545	7695.418	3043.463	-488.969	6575.896	Variables Match
wind_speed	-437.004	-6667.340	5793.333	3949.452	1117.075	6781.830	Variables Don't Match
apparent_power_total_corr.CAB10LFT(B)	3.051	-8.495	14.597	2.051	-2.155	6.257	Variables Match
apparent_power_total_corr.CAB10RGT(A)	4.620	-9.374	18.615	0.383	-4.676	5.442	Variables Match
apparent_power_total_corr.CAB12LFT(B)	1.414	-16.479	19.307	1.618	-5.978	9.215	Variables Match
apparent_power_total_corr.CAB12RGT(A)	-5.390	-19.473	8.692	-1.942	-9.148	5.264	Variables Match
apparent_power_total_corr.CAB13LFT(B)	-1.065	-3.554	1.424	-1.119	-2.131	-0.107	Variables Match
apparent_power_total_corr.CAB13RGT(A)	1.683	-1.220	4.587	0.591	-0.622	1.804	Variables Match
apparent_power_total_corr.CAB17LFT(B)	0.268	-9.093	9.629	2.210	-0.925	5.346	Variables Match
apparent_power_total_corr.CAB17RGT(A)	8.650	-0.735	18.036	0.013	-3.247	3.274	Variables Don't Match
apparent_power_total_corr.CAB18LFT(B)	-1.079	-10.572	8.413	4.479	0.432	8.527	Variables Don't Match
Apparent_power_total_corr.CAB18RGT(A)	5.514	-12.562	23.589	-2.731	-9.316	3.855	Variables Don't Match
apparent_power_total_corr.CAB19LFT(B)	-2.287	-6.676	2.102	-2.306	-4.963	0.352	Variables Match
apparent_power_total_corr.CAB19RGT(A)	2.695	-1.614	7.005	-1.372	-4.055	1.311	Variables Don't Match
apparent_power_total_corr.CAB22LFT(B)	-3.329	-7.497	0.839	-0.550	-2.612	1.513	Variables Don't Match
apparent_power_total_corr.CAB22RGT(A)	-3.763	-7.878	0.352	-2.297	-4.299	-0.295	Variables Match
apparent_power_total_corr.CAB23LFT(B)	-0.730	-2.445	0.985	-0.499	-1.129	0.131	Variables Match
apparent_power_total_corr.CAB23RGT(A)	-0.499	-2.171	1.173	-0.698	-1.323	-0.072	Variables Match
apparent_power_total_corr.CAB24LFT(B)	-0.307	-1.078	0.464	-0.086	-0.413	0.241	Variables Match
apparent_power_total_corr.CAB24RGT(A)	-0.464	-1.082	0.154	-0.069	-0.336	0.198	Variables Don't Match
apparent_power_total_corr.CAB26LFT(B)	-0.029	-1.884	1.827	-0.548	-1.153	0.058	Variables Match
apparent_power_total_corr.CAB26RGT(A)	0.242	-1.619	2.103	-0.549	-1.153	0.055	Variables Don't Match
apparent_power_total_corr.CAB3(B)	6.996	-7.638	21.630	6.289	0.729	11.848	Variables Match
apparent_power_total_corr.CAB3RGT(B)	6.017	-8.845	20.878	-16.747	-22.641	-10.853	Variables Don't Match
apparent_power_total_corr.CAB9LFT(B)	2.856	-4.336	10.048	-0.411	-3.042	2.219	Variables Don't Match
apparent_power_total_corr.CAB9RGT(A)	-2.716	-32.669	27.237	-6.910	-17.706	3.886	Variables Match
apparent_power_total_corr.GENERATOR_P3)	0.000	0.000	0.000	0.000	0.000	0.000	Variables Match
apparent_power_total_corr.LCP11(A)	-44.776	-141.682	52.131	-19.022	-45.788	7.744	Variables Match
apparent_power_total_corr.LCP12(B)	-10.372	-17.567	-3.177	18.975	11.199	26.751	Variables Don't Match
apparent_power_total_corr.LCP13(A)	4.795	-11.327	20.917	0.573	-4.287	5.434	Variables Match
apparent_power_total_corr.LCP2(A)	-80.431	-174.203	13.342	-2.291	-27.373	22.792	Variables Don't Match
apparent_power_total_corr.LCP5(A)	55.648	-43.682	154.979	27.207	-6.704	61.118	Variables Match
apparent_power_total_corr.LCP7(A)	3.791	-7.422	15.003	-9.237	-34.178	15.705	Variables Match
server_temp_diff_Between_Cab_23_and_Cab_24	116700.000	67100.000	166000.000	38750.000	15500.000	62000.000	Variables Don't Match
server_temp_diff_Between_Cab_24_and_Cab_25	145600.000	96000.000	195000.000	33640.000	12200.000	55000.000	Variables Don't Match
server_temp_diff_Between_Cab_25_and_Cab_26	82440.000	29700.000	135000.000	27660.000	3342.492	52000.000	Variables Don't Match
server_temp_diff_Outside_Edge_of_Cab_23	141400.000	89500.000	193000.000	18480.000	-8733.118	45700.000	Variables Don't Match
server_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14	-56770.000	-109000.000	-4145.047	-4568.994	-25500.000	16300.000	Variables Don't Match
server_temp_diff_Row_3_-_Outer_edge_of_Cab_12	-91420.000	-146000.000	-37000.000	38360.000	12200.000	64500.000	Variables Don't Match
server_temp_diff_Row_4_-_Between_Cab_19_and_Cab_20	160600.000	98900.000	222000.000	29140.000	6211.007	52100.000	Variables Don't Match
server_temp_diff_Row_4_-_Outer_edge_Cab_22	271900.000	214000.000	330000.000	94750.000	61900.000	128000.000	Variables Don't Match
water_temp_diff_Between_Cab_23_and_Cab_24	112100.000	50800.000	173000.000	168900.000	114000.000	223000.000	Variables Don't Match
water_temp_diff_Between_Cab_24_and_Cab_25	-58290.000	-111000.000	-5934.079	-86690.000	-145000.000	-27900.000	Variables Match
water_temp_diff_Between_Cab_25_and_Cab_26	184100.000	127000.000	242000.000	97150.000	46400.000	148000.000	Variables Don't Match
water_temp_diff_Outside_Edge_of_Cab_23	150600.000	85100.000	216000.000	71500.000	41000.000	102000.000	Variables Don't Match
water_temp_diff_Row_3_-_Between_Cab_13_and_Cab_14	-113300.000	-175000.000	-51500.000	-127600.000	-209000.000	-45900.000	Variables Match
water_temp_diff_Row_3_-_Outer_edge_of_Cab_12	-22130.000	-65700.000	21400.000	12250.000	-11100.000	35600.000	Variables Don't Match
water_temp_diff_Row_4_-_Between_Cab_19_and_Cab_20	-211300.000	-274000.000	-149000.000	9258.121	-23600.000	42100.000	Variables Don't Match
water_temp_diff_Row_4_-_Outer_edge_Cab_22	134300.000	66200.000	202000.000	37270.000	13900.000	60700.000	Variables Don't Match