School of Mathematics, Statistics and Physics
Newcastle University

Data analytics to support transition to zero carbon emissions

Research into the optimal electric van charger placement for Engie a large energy and services company

Maximus Johnston
Shirley Coleman
Jaume Bacardit
Daniel Walker

[2]

[10] [9]

July-2020

Contents

Abstract

This report is a summary of the initial six weeks of research into the optimal placement of electric vehicle chargers in an area. The aim of the project is to enable a service company to transition from a fleet of diesel vans to fully electric vehicles. This report describes the application of various data science techniques on four months of van data collected by the service company. Analysing this operational data we try to understand and gain insight into the usage of the fleet and how the demands of the company can be met while using electric vehicles. As data science is a combination of IT and data skills, statistical analysis and machine learning techniques, and business awareness, the report also assesses the benefits to stakeholders and the downsides of a transition from diesel to electric vehicles.

Keywords: change management, impact, frequency, cluster analysis, mapping, geolocations, electric vehicles, data science, fuel consumption.

1. Problem Statement

People are becoming ever more concerned with the irreversible changes we are making to our planet due to the rapid unsustainable consumption of fossil fuels and natural resources. As a result of this, having and owning a Battery Electric Vehicle (BEV) is becoming increasing popular as they are seen to be more sustainable. This change has meant that electric vehicle manufacturer Tesla have become the most valuable car company in the world in 2020 saying they aim to sell 500,000 cars in 2020 [8].As a result of this surge in popularity the issue of the placement of chargers around towns and cities has arisen. Managing the effect on the grid, the cost of building charging points, keeping range anxiety of driver's low are all factors to consider before placing chargers.[6],[7]

In the context of facilities management, the service company wants to reduce their carbon footprint and proposes to convert their current diesel fleet of maintenance vehicles to electric vehicles in the next few years. Hybrid vehicles are increasingly an option for personal cars and many of these do not need charging stations as they work by charging their own batteries when running under fuel. However, it is not expected that hybrid solutions will be available for vans and lorries. Therefore, a key challenge for a company changing to electric vehicles is the placing of chargers.

We apply a data science approach to the challenge. Data science consists of three spheres: IT; statistics and business awareness.We consider how the data is collected and how it is quality checked and made available for analysis. We then apply exploratory and more sophisticated statistical analysis and machine learning techniques. We also consider the business aspects such as implementation issues and change management.

We propose a greedy algorithm which takes in various parameters such as cost and power available to be supplied at different nodes to find the optimal charger placement. For example, Fredriksson [3] discusses using solving a pruned integer problem to select the optimal location of charging stations in the network. Such algorithms are called "greedy" because they offer a short-range solution which maybe a local rather than global optimal. As such the algorithms grab what is easiest and closest to hand like a greedy person.

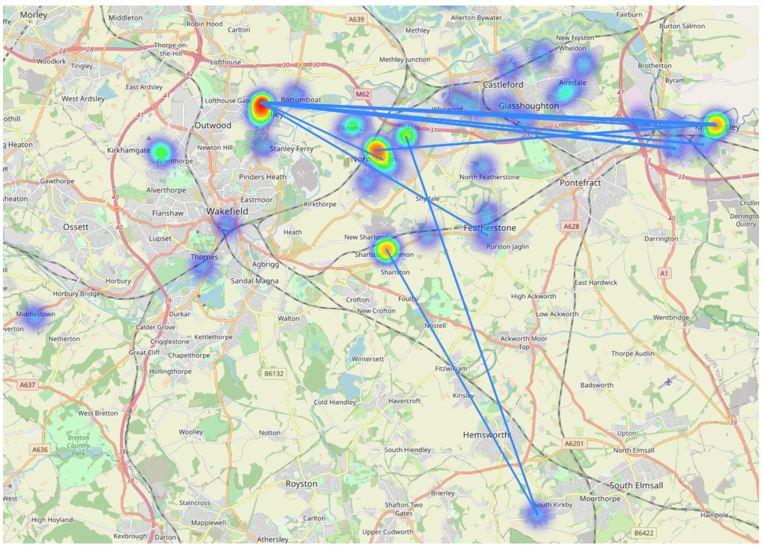
This report details the approach stage by stage with week by week activities described in the next sections. The results are gathered together with conclusions and recommendations for further work in the final section. The project is summarised in the appendix in the form of a presentation that was given to share the data science approach and results with company staff.

2. Methodology

2.1 Week One

My first week consisted of reading up on electric vehicles and the problems around using them daily as well as getting a good understanding of the data and this would be crucial for the next 6 weeks. I started out by reading Electric vehicle charging station placement[7]. This paper is the basis for the work the National Innovation Centre for Data (NICD) are doing to try and formulate the optimal placement of chargers in order to minimize disruption to the Engie Engineers using EVs whilst also minimizing the cost. I also learned about the charging time of Evs [4],[5] and how this slows dramatically as the battery reaches 80%. This gave me a real insight into what the problem is and why it is so complicated to find the optimal placement of chargers.

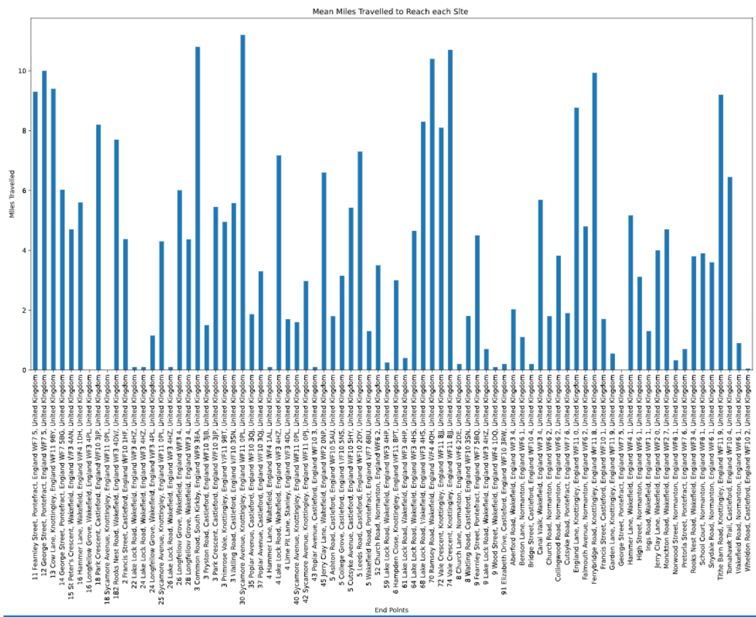
Engie have been collecting data from their vans to give this project a base to work from and eventually data to use when simulating how the electric vans would work in the real world. This data includes the journeys the vans have made and the start and stop addresses and times of all the journeys. In order to start understanding the data I needed to set up an anaconda environment in order to install various python libraries that would help sort and visualise the data. Once I had done this, I was able to start using the data. My first task was to take all the addresses of the places visited by the engineers over the four months and turn them into latitude and longitude as this is a more useful format and can be used by folium, a tool that allows you to create heat maps and easily see the most common sites visited. To try to get a grasp on the tools and and the data I plotted some simple graphs based on engineer 1's data. First I used folium to show the 90th percentile of distance of journeys travelled by the van in the four months allowing you to see the sites the engineer travelled the furthest to reach (Figure 2.1).



(a) Engineer 1 - Longest Journeys

Figure 2.1: Heat Map With Journeys

I then plotted the average distance travelled to get to each of the sites engineer 1 visited to get an idea of which sites may need a charger as the van was depleting its battery in going to them. Although these figures may not actually be used to create the model, creating them was a good exploratory data analysis exercise to help see what results you can expect when doing more formal modeling.



(a) Engineer 1 - Mean travel to reach each site

Figure 2.2: Bar Chart with distances

In Figure 2.2 the mean journey distances range from less than 0.1 miles up to 11 miles.

2.2 Week Two

In week two I started by creating a summary file of all of the data I had made in week one from the data collected by the vans. This was to allow me to produce large summaries of the data and see the outliers, then remove the clear outliers such as cases where the vans were recorded as being in Canada rather than Wakefield. With this cleaned data file I used folium to create a marker cluster map of every job the engineers had been on in the last 4 months.

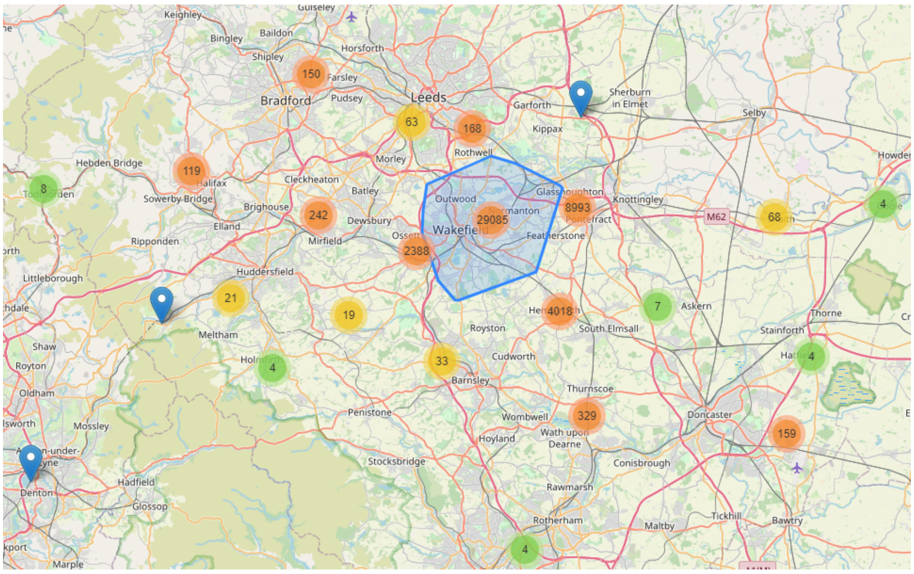
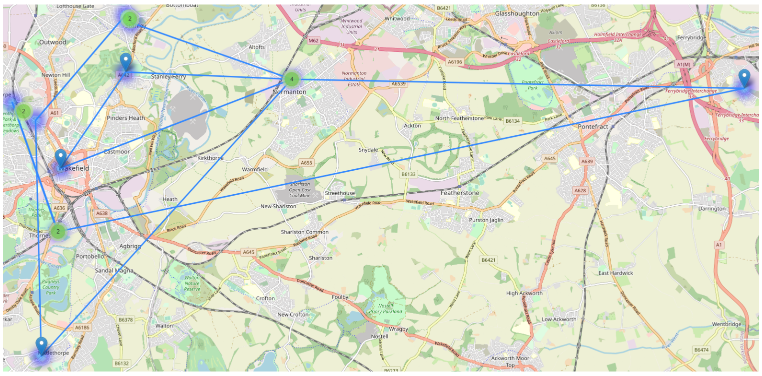


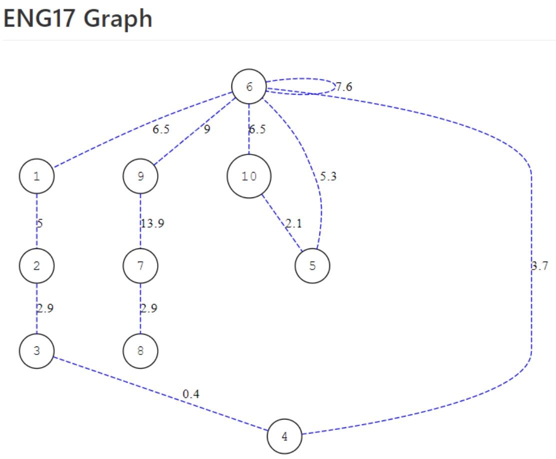
Figure 2.3: All sites visited by engineers over 6 months

On the map in Figure 2.3 you can clearly see the sites that should not be considered for EV chargers as they are well away from other sites and only visited a small number of times in the four months. In contrast you can also see the area in blue which contains over 29000 jobs and so will be suitable for placing chargers.

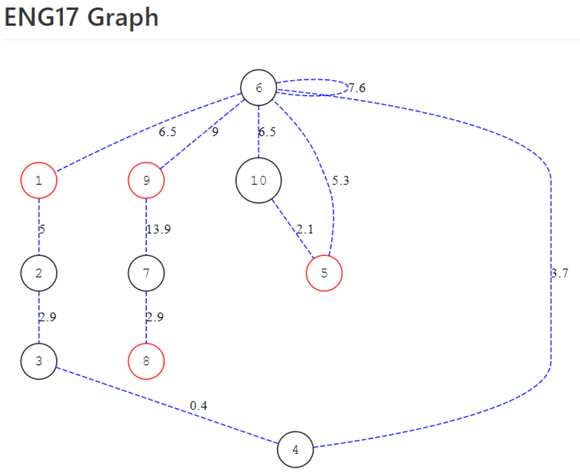
The paper on electric vehicle charging station placement[7] provides an algorithm that takes various inputs such as the range of the van, demand and supply at certain stations and uses the distance between all the sites in order to tell you which stations would need chargers in order to reach every other node. I used folium and HACKMD to represent an engineer's day.



(a) Journeys Plotted on folium



(b) Journeys represented on HackMD



(c) Journeys represented on HackMD with Chargers

Figure 2.4: Engineer 17 - 13/01/2020

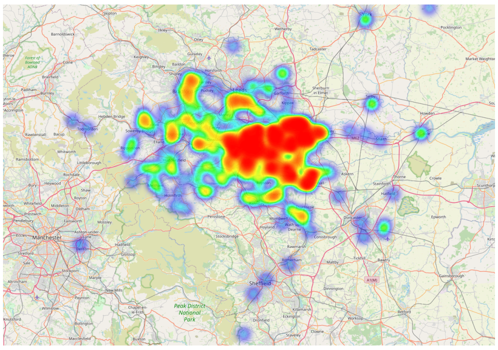
The map in Figure 2.4 (a) displays the engineer's day as an integrative HTML file which allows you to see the journeys in order, how far the next journey is and the total distance travelled so far. The graph in Figure 2.4 (b) represents the sites visited by the engineer on that day as nodes. Their day starts at node 6, goes round through nodes 1 to 4 back to 6, then 10 to 5 back to 6, travelled from 6 back to 6 then to 9 and through to 8 to finish the day.

Figure 2.4 (c) uses the algorithm. If the sites mapped in (b) were the sites visited by engineers everyday, placing a charger at all the red nodes in Figure 2.4 (c) would ensure the engineer would be able to reach every other site in the graph when using an electric van with a range of 17 miles.

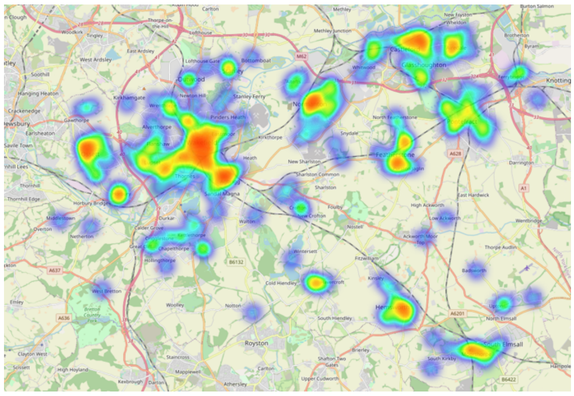
The problem with this algorithmic approach is that it does not take into consideration that engineers may not travel through nodes to get to other nodes, for example if the engineer travelled from site two they may not travel through three to get to four. A fully connected graph was too complicated for this algorithm to solve and so other approaches are being considered.

2.3 Week Three

In order to test whether the chargers have been been placed in the right place we need to know the distance from each site to every other site. With over eight hundred individual sites visited in the four months data was being collected, it was unfeasible to get the route distances between all of them. In order to get a more realistic set of sites I grouped the data by time spent at the site and the number of visits. The top three hundred of these sites were in the area Engie were looking to develop charging points and all had a significant number of visits and minutes spent at the site (Figure 2.5).



(a) All sites

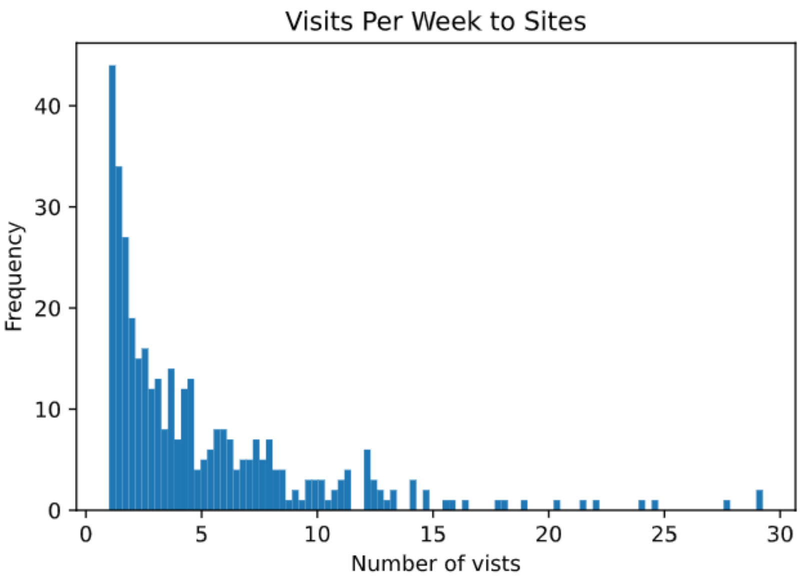


(b) Top 300

Figure 2.5: Heat Maps of Sites

As you can see the range of the sites has shrunk significantly as we have removed the outliers where engineers had made one of their trips outside of Wakefield, going as far as Sheffield. With this smaller list of sites we are able to use an API to produce a matrix of distances from each site to every other site for later use.

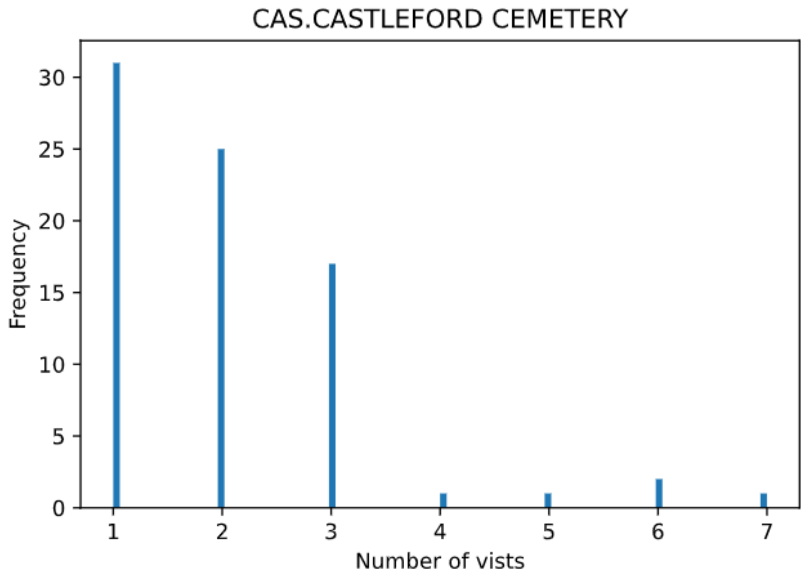
An important variable in the model NICD are aiming to produce is the demand at charging sites. To try to get an idea of what this might look like I plotted histograms of the usage of sites.



(a) All sites

Figure 2.6: Histogram of visits to sites

Figure 2.6 is a histogram plot of the visits per week to individual sites. This histogram gives us a general idea of how many times a week the sites are being visited and as expected the frequency distribution is skewed with an inverse exponential curve. There are only a few sites getting more than 10 visits per week and these would be the sites where we would expect to see the chargers placed.



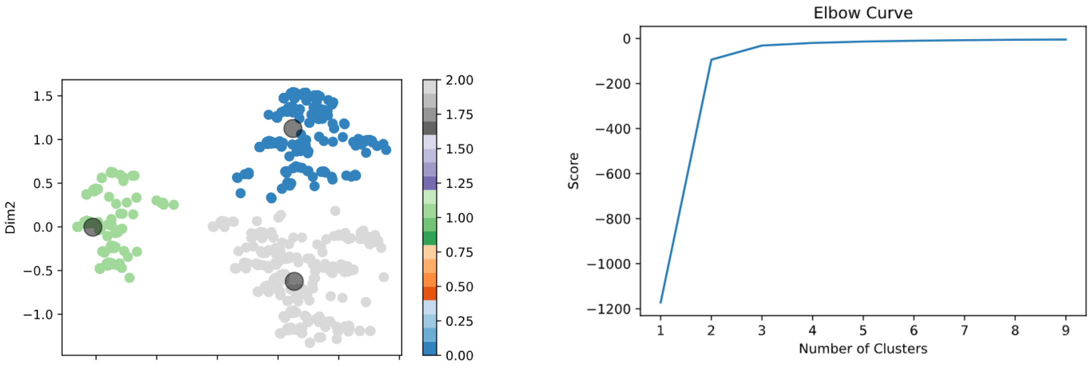
(a) All sites

Figure 2.7: Histogram of visits Sites

Figure 2.7 is a histogram of how many visits per day a site gets. These histograms can be plotted for any site in the data. This helps to give us an idea of what the demand is at that site as you may need to consider more than one charger at a site if the demand per day is often higher than one.

2.4 Week Four

To try and get the best possible understanding of the data we decided to use some unsupervised learning to find possible patterns in the data we had previously missed. To do this we used some previously unused metadata such as whether the site was active or not, the size of the site, the type of site and more. We then used a OneHotEncoder in order to turn columns where there are multiple categorical options such as site type into a usable format consisting of indicator variables (taking values 0 or 1). This produces a column for each characteristic and gives it a one if it satisfies that condition or a zero if not. This gave us an array of 1668 rows each with 114 columns. It was noticed that many of the 144 columns had only a few non-zero values. This means that the variable has very little variation. Of the 1668 cases, some columns only have 1 or 2 non-zero values. These variables therefore have little discriminating power. Removing columns with fewer than 50 non-zero cases leaves 30 variables. Some of these appeared to be duplicates so we removed the duplicates, assuming that these variables really are duplicates and not just different variables whose values just happen to coincide in this dataset. This left 27 variables. This is a useful exercise in understanding the data and suggests that the true dimension of the data is 27 rather than 144. However, using principal components analysis (PCA) has a similar effect of reducing the dimensionality of the data and further cluster analysis was carried out with the first two principals components in place of the 144 (or 27 columns). We used k means cluster analysis as this is a faster option than hierarchical methods. The analysis included constructing an elbow plot to indicate how many clusters should be looked for.



(a) Scatter Plot

(b) Elbow Plot

Figure 2.8: Kmeans Clustering

The clustering in Figure 2.8 shows what looks like three clear clusters however the elbow plot says that there are only two (Figure 2.9).

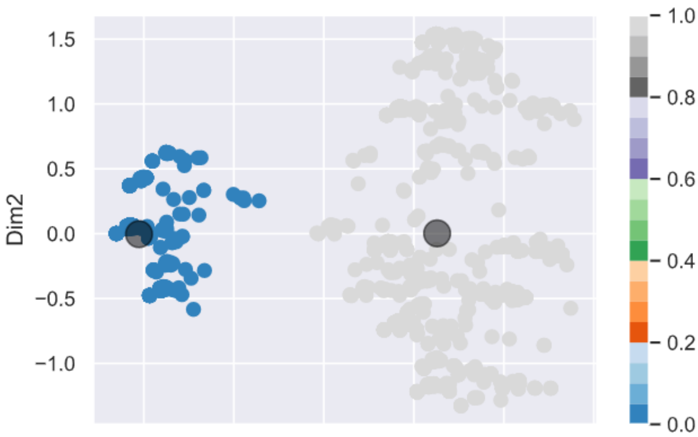
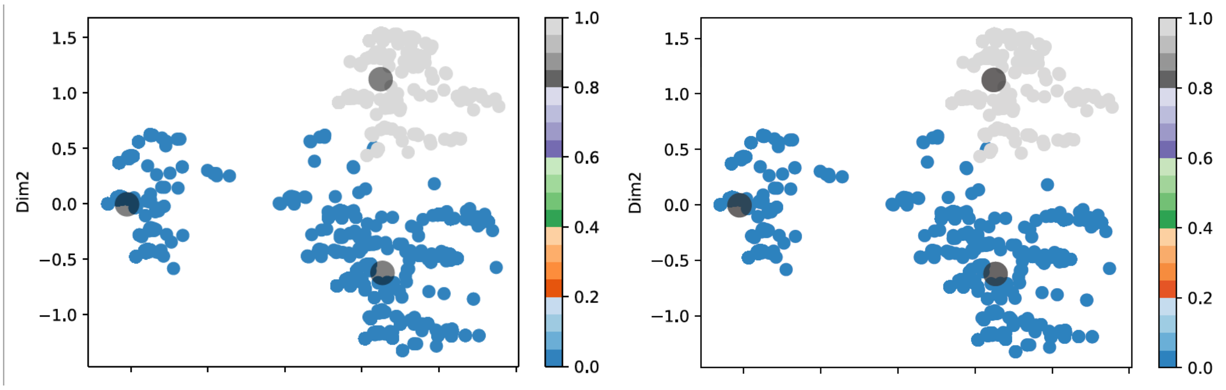


Figure 2.9: Cluster with 2 clusters

We looked for the variables that were important in separating the sites into 2 or 3 clusters. We did this in two ways. Firstly we looked at the loadings on the principal components and secondly we calculated chi-square statistics for the two-way frequency tables for each variable. Each table had the number of sites in cluster 1, 2 or 3 (in the rows of the table) and variable 0 or 1 (in the columns of the table). Figures 2.10, 2.14 and 2.12 show the clusters with the important variables highlighted. Site designation code "Community" characterises the two right hand clusters whereas "Land and Buildings" or SD08 which is "Investment" characterise the two lower clusters. The variable "Reactive" is spread between all three clusters.



(a) Site Designation Desc - Land and Buildings (b) Site Designation Code - SD-08

Figure 2.10: Engineer 17 - 13/01/2020

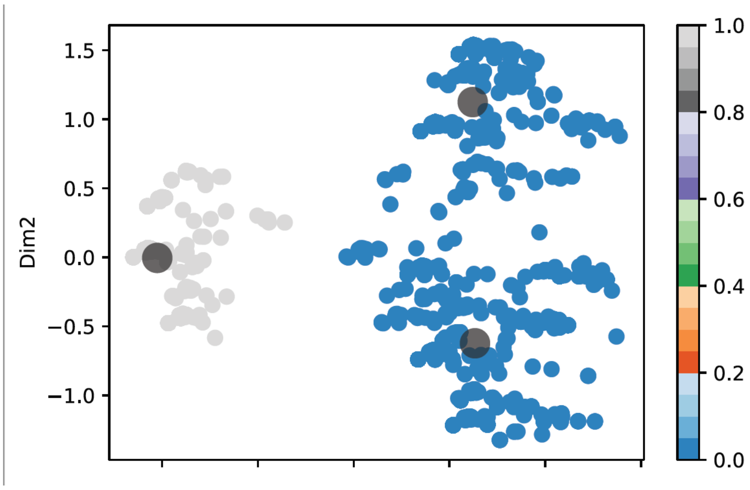


Figure 2.11: Site Designation Desc - Community

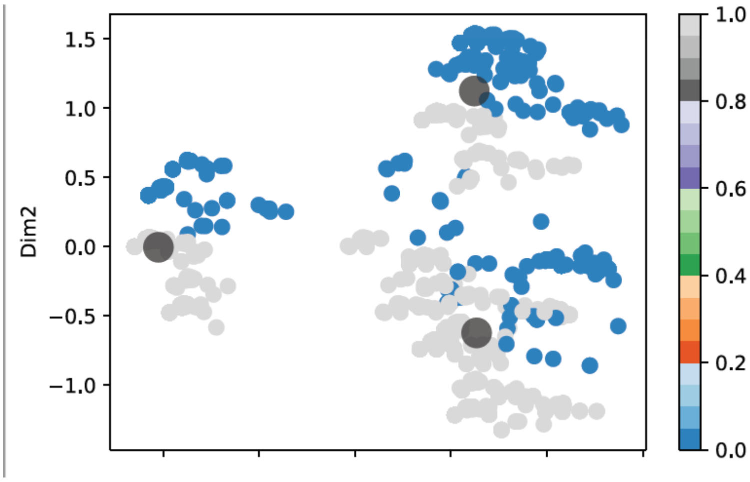


Figure 2.12: Site Service Code - Reactive Only

2.5 Week Five

I thought it was relevant to explore the importance of the project and the motivations behind switching to a fleet of electric vans from diesel. The climate is changing and this is already starting to have devastating effects such as more frequent and severe weather conditions and the sea levels rising causing flooding due to melting icecaps. One of the key contributors to climate change is the rising levels of green houses gases which are produced by various things such as volcanic eruptions, farming and burning fossil fuels like petrol and diesel. To try and combat the damage we are doing many governments and companies are now trying to reduce their carbon footprint and it is Engie's aim to have zero diesel vans by 2025.

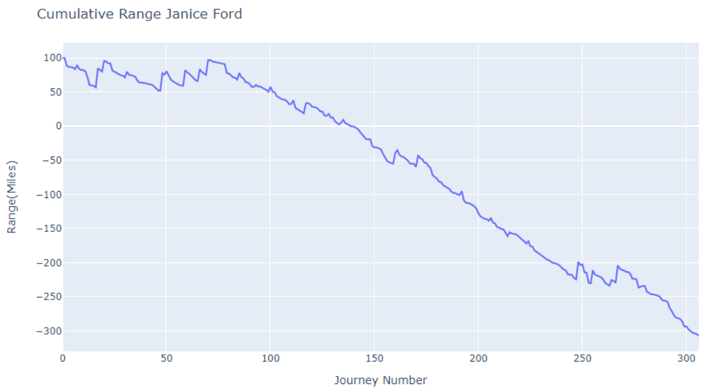
This scheme is a very ambitious task as it involves completely disregarding the infrastructure the country has built since cars became a standard means of transport and building a new infrastructure of chargers to allow their engineers to carry out their jobs as they normally would. As time moves on and more money is invested in electric vehicles and chargers become more common and faster, running out of battery will become less of an issue. Currently range anxiety is one of the key factors preventing people from switching to battery electric vehicles and so one would be able to reduce this by having a well placed set of chargers.

There are many benefits for Engie to switch to a more environmentally friendly way of working and reducing their carbon output. People are becoming environmentally conscious and as a result of this are trying to reduce their carbon footprint and make more sustainable decisions in every aspect of life. Having an environmentally friendly image is likely to attract these people as customers. Reducing carbon output can also allow you to receive compensation from the government as they have nationwide pollution targets to reach and so people and companies that contribute to meeting these targets are often compensated.

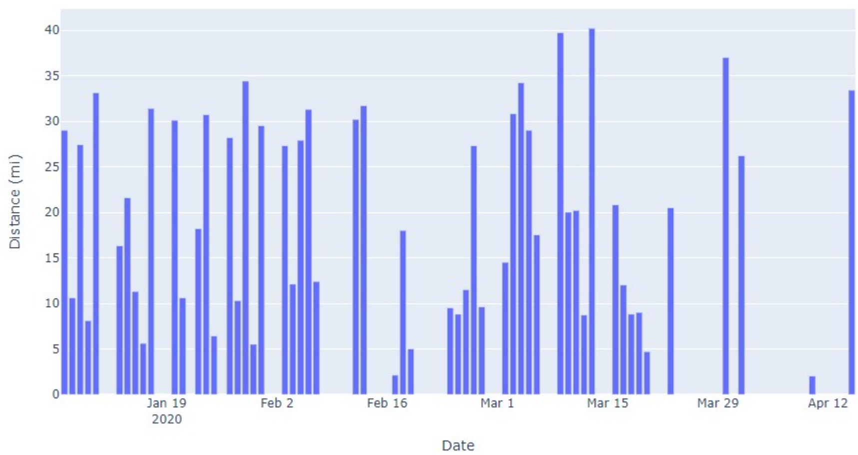
As the problem of charger placement is so complicated it will take lots of time and money to come up with a suitable effective solution. It is important to think of and explore the alternatives that may be more simple or cheaper. One solution to this would be to place chargers at engineers' houses. As this would mean they have a full charge when they start everyday. This data explores the possibility.

Stop Location	Stopped_Mins
59 Dacre Avenue, Wakefield, WF2 8, UK	4264
Wakefield Normanton Depot - Location near Collingwood Road, Normanton, WF6 2, UK	1704
36 Shepstye Rd, Horbury, Wakefield, WF4 5, UK	824
28 Cross Rd, Middlestown, Wakefield, WF4 4, UK	634
St James Way, Crigglestone, Wakefield, WF4 3, UK	604
11 Sowood Avenue, Ossett, WF5 0, UK	472
Dewsbury Rd (A638), Alverthorpe, Wakefield, WF2 9, UK	452

(a) Vans most visited locations



(b) Van battery against time



(c) Daily Mileage of electric van

Figure 2.13: Data exploring the possibility of home charging

2.13aThis is a list of the sites where this van has spent the most time stopped, over the period that this data was collected. Dacre avenue is a home address and the others appear to be schools, aside from the depot. The pattern of travel tends to be a midday to early evening shift with occasional morning visits before schools open. The van will semi-regularly attend the depot before the start of the shift, presumably to collect supplies.

2.13b If we model the range of this van as if it were an EV and with charging points at the depot, Shepstye and Cross road, we obtain the graph shown. Clearly, even with a range extender it is not possible to rely on a 'regular charging point' to sustain the range while working. This is likely due to the variety of sites visited.

2.13c We have 55 days of daily mileage data for this van. The maximum daily mileage for this van over the period of collected data is less than 41 miles. Assuming that this pattern of travel is representative of the usual travel, then this van would be suitable for home charging should it be transitioned to EV.

This van would be almost certainly suitable for transition, if we employ a home charging method. It is not very suitable, based on the data we have to rely on regular charging points at specific locations unless there was to be a wider range of sites with charging capability.

As you can see the data for this van shows it would ideally be put on home charging as it never exceeds the standard range of an electric van and so would never run out of battery as it would start the day with a full charge everyday. In an ideal world the home charging would be very effective however there are real world problems which prevent this from being the solution. Installing an electric charger in an engineer's house would first require that this was possible with the electrics and location of their home i.e would it be safe and secure. Installing a home charger would also be a large investment for Engie and so they would have to make sure the engineer was not renting the house or going to move soon otherwise they would have spent large amounts on a charger and have to repeat this each time a person moves home. Another problem with this is that the engineers would be paying for the the electricity to charge the vans with their own money which Engie would then have to reimburse. This would require Engie to set up a fast and easy to use reimbursement system. Have to do this extra admin with bills at home may put engineers off having the home charger. Placing chargers in peoples homes may be very effective in terms of maintaining charge despite this it is not a totally practical way to solve this problem but may become more practical in the future as more homes get EV charging stations and more people switch to EVs for their day to day cars.

2.6 Week 6 Overview and next steps

Exploratory data analysis was carried out using python and also the seaborn library for graphical displays. The modelling tools used in this project include folium, geopy, matplotlib and HACKMD. Unsupervised data mining and data science made use of the scikit learn python package. [1].

The next steps for the project will be to start to incorporate more real world elements into the model, we will then scale the model up for the full data set and this will give us the model we would hope to use this model.

Before large amounts of money are invested in installing the chargers it is important that we validate that the placement of the chargers would in fact work in the real world. To do this two approaches are being used. The first will simulate the journeys the vans are likely to take in the future, this will give us an idea of weather the model will work and you will be measure how may times would fail and address this. This second will run the last four months of data with the chargers placed where the model placed them and test that it would work for the past data.

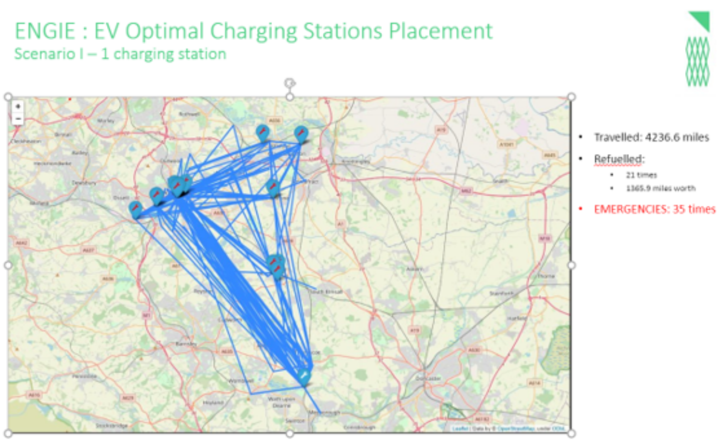


Figure 2.14: Map of past journeys for validation

This method will produce a map that will show all of the breakdowns, emergency charges of the vans and the time spent wasted by the engineers. Once these have been completed and the validation supports the model the chargers will be ready to be installed and the model applied to other areas for EV transition.

3. Conclusions

The project has been very successful in collecting and preparing a useful set of data on van journeys that can be analysed.

The project has carried out exploratory analysis using graphical and geolocational displays, frequency distributions, principal components analysis, cluster analysis and chi-square analysis. We have demonstrated the use of an algorithm aimed at determining the optimal position of charging points but have not carried out the analysis on the full set of real data.

The cluster analysis showed that the sites can be grouped into distinct subsets and this could be useful in combining assets of reorganising the journey routes.

When the full analysis is carried out, it will be important to establish metrics to assess the baseline situation, for example in terms of current fuel consumption and consequent emissions then consider the impact of the solution and evaluate the achievement of aims.

Bibliography

[1] A good overview related to data mining in python, https://scikit-learn.org/stable/modules/clustering.html. 2020.

[2] Logos Download. newcastle-university-logo, 2020. Last accessed 08 July 2020.

[3] Henrik Fredriksson, Mattias Dahl, and Johan Holmgren. Optimal placement of charging stations for electric vehicles in large-scale transportation networks. *Procedia Computer Science*, 160:77 – 84, 2019. The 10th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2019) / The 9th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2019) / Affiliated Workshops.

[4] C Herron and Coleman S. Making way for ultra-low-emission vehicles. *Focus*, June:48–49, 2018.

[5] C Herron and Coleman S. A methodology for calculating the required number of rapid chargers for a population, https://zerocarbonfutures.co.uk/insights/. 2018.

[6] A. Y. S. Lam, Y. Leung, and X. Chu. Electric vehicle charging station placement: Formulation, complexity, and solutions. *IEEE Transactions on Smart Grid*, 5(6):2846–2856, 2014.

[7] Albert Y.S. Lam, Yiu-Wing Leung, and Xiaowen Chu. Electric vehicle charging station placement. In *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, oct 2013.

[8] BBC News. Tesla overtakes toyota to become world's most valuable carmaker - bbc news.

[9] ENGIE Resources. Engie-logo-solid-white-on-blue-full, 2020. Last accessed 08 July 2020.

[10] www.ncl.ac.uk. Nicd%20logo, 2020. Last accessed 08 July 2020.