DATA70202 Applied Data Science

P18: EV CHARGING POINTS FOR THE OPTIBUS PLATFORM

Group 4

Student ID
11186818
11287658
7494452
11161576
10977190

Word count: 6,939 (excluding table of contents, references and appendices)

Table of Contents

1. Introduction 3

1.1. Background and Context of the Project 3

1.2. Objectives and Goals of the Project 3

1.3. Scope of the Project 3

2. Literature Review 4

3. Data Collection and Pre-processing 5

4. Methodology 12

4.1. Density Overlay Analysis 12

4.2. K-means Analysis 13

4.3. Traffic Weighted Analysis 16

4.4. Extra and Further Considerations 19

4.5. Limitation 20

5. Results 21

5.1. Density Overlay Analysis 21

5.2. K-means Analysis 23

5.3. Traffic Weighted Analysis 24

6. Discussion 26

6.1. Limitations and Challenges Encountered During the Project 26

6.2. Potential Blockers 28

6.3. Potential Obstacles to Long-Term Data Access 29

6.4. Comparison of the results with those of previous research 30

6.5. Recommendations for Future Research and Improvements to the Project 31

7. Conclusion 33

8. References 35

9. Appendices 37

Table of figures

Figure 1. 1: Bus routes GIS layer. 5
Figure 1. 2: Bus stops GIS layer. 6
Figure 1. 3: National Charging point Registry GIS layer. 7
Figure 1. 4: Electricity North West heatmap tool GIS layer. 8
Figure 1. 5: Brownfield land GIS layer. 9
Figure 1. 6: Ancient woodland GIS layer. 9
Figure 1. 7: Conservation areas and special areas of conservation GIS layer. 10
Figure 1. 8: Greenbelt GIS layer. 11
Figure 1. 9: Historic parks and gardens GIS layer. 11

Figure 2. 1: Substations heatmap distribution. 13
Figure 2. 2: Bus stops heatmap distribution. 13
Figure 2. 3: All the bus stops covered in cluster buffer zone. 14
Figure 2. 4:Bar chart for area of Great Manchester & substation buffer zone. 14
Figure 2. 5: Area after cutting out. 15
Figure 2. 6: Bus stop quantity in each buffer zone. 15
Figure 2. 7: Cluster centroid when k=44. 16
Figure 2. 8: Cluster centroid when k=96. 16
Figure 2. 9: Cluster centroid when k=136. 16
Figure 2. 10: Bus stop (orange point) and bus route (yellow line). 17
Figure 2. 11: Find the nearest line from a point. 17
Figure 2. 12: line chart for partial bus stop with traffic flow score. 18
Figure 2. 13: K-means cluster when k=2. 18
Figure 2. 14: Cluster apply weighted value (k=2). 19

Figure 3. 1: Combined heatmap 21
Figure 3. 2: High density area after cutting 22
Figure 3. 3: Connected line when k=44 23
Figure 3. 4: Connected line when k=96 23
Figure 3. 5: Connected line when k=136 Figure 23
Figure 3. 6: Intersection between bus route and line 23
Figure 3. 7: Traffic-weighted charging station distribution map 24
Figure 3. 8: Detail of traffic-weighted charging station distribution map 25
Figure 3. 9: Overlap area between substation buffer zone and cluster circle 25

1. Introduction

1.1. Background and Context of the Project

The transportation industry is experiencing a significant shift towards sustainable energy solutions. The transition from diesel to electric vehicles (EVs) is one of the key strategies to reduce vehicle emissions and promote cleaner modes of transportation. However, one of the significant challenges for EVs' widespread adoption is the lack of sufficient charging infrastructure. EVs require charging throughout the day, which diesel vehicles do not. This presents the challenge of finding suitable locations to provide electric charging points with available capacity, space to charge multiple buses simultaneously if needed, and a centralized location where multiple routes pass through regularly to ensure maximum vehicle efficiency while maintaining charge.

1.2. Objectives and Goals of the Project

The main objective of this project is to develop a geojeson data file that categorizes potential charging sites for electric buses in the Greater Manchester area. The data file will be integrated into Optibus, an end-to-end, cross-functional software platform for transportation planning, scheduling, rostering, and operations. In this project, our main task is to identify relevant data sets and processing methods to determine all existing EV charging stations, predict potential charging station locations based on bus operations, and recommend suitable locations for new charging points. Finally, we will use this data to construct a GIS layer. The interactive data file will enable clients to view potential charging locations and determine where their current routes may be able to charge EVs, thus facilitating the transition to a sustainable and greener transportation system.

1.3. Scope of the Project

The project will specifically focus on the Greater Manchester area and its surrounding areas. Key data sources will include Data.gov.uk, UK Power Networks, Bus Open Data Service, National Public Transport Access Nodes (NaPTAN), Google Maps, Land Registry data, and any other relevant data sources. The project will involve data processing and analysis, as well as the construction of a GIS layer that can be integrated into the Optibus platform. The expected outcomes for this project include a usable geojson layer, any created code to be made available through an online repository, access to the datasets used to create the GIS layer, and a written report on the methods of obtaining and processing the data as well as potential obstacles to site development and data access in the long term.

2. Literature Review

The growing concerns about greenhouse gas emissions and the urgent need for greener energy alternatives have sparked interest in transitioning from diesel to electric vehicles (EVs) in public transportation systems. One of the primary barriers to the widespread adoption of EVs is the requirement for charging infrastructure. Optibus, a leading transportation software platform, is addressing this challenge through a project aimed at developing a data file that interlinks with its platform to identify potential charging locations for electric buses. This literature review critically examines recent research on electric bus charging infrastructure, focusing on strategies and barriers to its integration.

Lopez de Briñas Gorosabel et al. (2021) highlighted the importance of considering factors such as land use, spatial constraints, and accessibility when selecting charging locations. Building on this, He et al. (2022) proposed an integrated approach to charging infrastructure planning and scheduling for battery electric bus systems, which aligns with the objectives of the Optibus project. Furthermore, efficient route planning, incorporating charging stops, and minimizing battery degradation are essential for successful integration, as emphasized by Dirks et al. (2023).

In addition to these strategies, the literature identifies various barriers to the adoption of electric buses. Aldenius et al. (2023) pointed out high upfront costs, limited driving range, and charging infrastructure challenges in England and Sweden. Similarly, Grigorieva and Nikulshin (2023) stressed the importance of a holistic approach to charging infrastructure planning in Moscow. Both studies recommend supportive policies, stakeholder collaboration, and innovative business models to overcome these barriers. Moreover, Guschinsky et al. (2021) found that strategic decisions regarding charging infrastructure significantly impact the overall performance of fast-charging city electric bus services.

The implications of these findings for the Optibus project are clear. The reviewed literature consistently underscores the need for a comprehensive and integrated approach to charging infrastructure planning. By incorporating charging infrastructure data into its platform, Optibus can facilitate better route planning, scheduling, and overall system performance for electric bus networks.

In conclusion, the critical examination of recent research on electric bus charging infrastructure reveals the need for comprehensive strategies and collaborative efforts to overcome challenges related to charging infrastructure integration.

3. Data Collection and Pre-processing

We needed to convert all of our datasets to geojson format after they were pre-processed so that they could be visualised as a layer in a GIS system. A GIS system is a computer system for capturing, storing, checking and displaying data related to positions on Earth's surface, this is a crucial part of spatial data infrastructure (National Geographic, 2022). The detailed visualisation for these datasets have been visualised using ArcGIS (link in references).

Figure 1. 1: Bus routes GIS layer.

We first found the bus routes dataset, so that we could visualise and determine the bus routes. “GM bus routes (1:25,000 scale map data)” published by Transport for Greater Manchester, this dataset consists of map files which show bus routes covering the Greater Manchester area. This dataset is updated around four times per year and is derived from Ordnance Survey (OS) “VectorMap District”. The routes can be used to show differences between inbound and outbound buses, full-route and part-route services, routes taken at different times, routes taken on different days. This did not require preprocessing as the dataset was already clean and ready to apply. Figure 1.1 shows the visualization of this dataset in the form of a GIS layer.

We used the National Public Transport Access Nodes (NaPTAN) dataset, which lists all public transport access points in Great Britain. It is updated when information is provided by staff based throughout local authorities around the UK. To pre-process the dataset, we needed to massively reduce the size of the dataset by only focusing on the bus stop data within the StopType variable. Then we focused only on the bus stops which are in the Greater Manchester area using the NptgLocalityCode variable. Allowing us to then filter the bus stop type, on BusStopType so that we could determine where the bus depots are. From here we were able to visualize a fixed area around these depots which will have sufficient electric charging capacity, assuming all depots will be able to facilitate the charging of buses.

Figure 1. 2: Bus stops GIS layer.

Further we used the NaPTAN dataset to determine the bus stop locations in Greater Manchester. Repeating the process stated previously but instead of looking at bus depots, we looked at active bus stops through the BusStopType variable. From here we could visualize these bus stops as a GIS layer (Figure 1.2) as well as use the data to complete the processes explained in our methodologies.

Figure 1. 3: National Charging point Registry GIS layer.

We also used the National Charging point Registry (NCR) dataset, which is a dataset regarding publicly available charging points for electric vehicles in the UK and was established in 2011. To preprocess this dataset ready for our analysis we first reduced the overall size of the dataset by limiting the available points to the Greater Manchester area, using latitudinal and longitudinal data. We then removed the charging points from the dataset which consisted of the incorrect Connector Types, Charging Methods and Charge Modes regarding our need to charge electric buses sustainably. We were then able to visualize the remaining, valuable, charging points as a GIS layer (Figure 1.3); showing us where in Greater Manchester an electric bus would be able to charge sustainably at this moment in time, without building new charging points.

Figure 1. 4: Electricity North West heatmap tool GIS layer.

Using the Electricity North West heatmap tool, which provided a dataset containing the capacity information on all the primary substations and bulk supply points. This enables us to assess the level of capacity that might be available for new connections to the network. To prepare this dataset for analysis we first needed to the Easting and Northing co-ordinates into longitude and latitude (Code 1), then we were able to focus on substations/supply points within or close to Greater Manchester. We then mapped an area around each point with a certain radius, which we used to determine all the potential areas that would have the electrical capacity necessary in order to charge multiple electric buses. This was then verified visually by inserting these points and areas as a GIS layer (Figure 1.4).

Finally, we used multiple Local Planning Authority datasets, which collect planning and household data from local planning authorities and transform them into a consistent state across England. We needed to collect the relevant Greater Manchester data, this meant joining ten datasets: Bolton, Bury, Manchester, Oldham, Rochdale, Salford, Stockport, Tameside, Trafford, Wigan. From here we were able to look at specific planning data, this included: Ancient woodland, Conservation area, Greenbelt, Historic parks and gardens, Special area of conservation, Brownfield land. This information allowed us to determine where we would be unable to place a charging point, for example areas listed as Conservation area. This allowed us to remove multiple potential locations from our search. The preprocessed data was then also converted into a GIS layer, allowing us to visualize some potential areas that could host electric vehicle charging points (Figure 1.5, 1.6, 1.7, 1.8, 1.9).

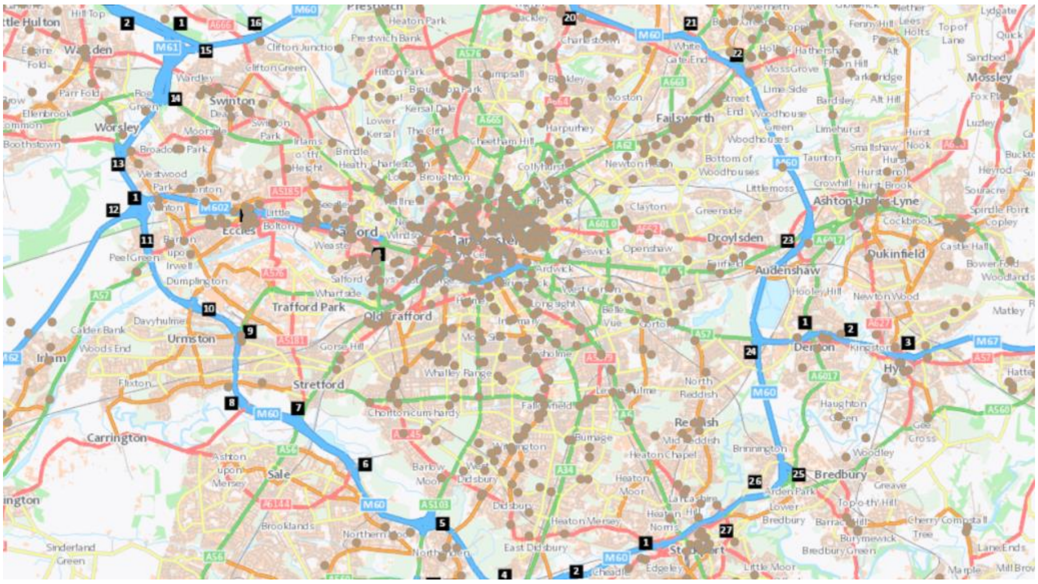

Figure 1. 5: Brownfield land GIS layer.

We included information regarding brownfield land because this refers to land which is abandoned or underutilized; therefore, this would be an ideal area to place electric vehicle charging points due to the sustainability, social and economic benefits to the surrounding areas.

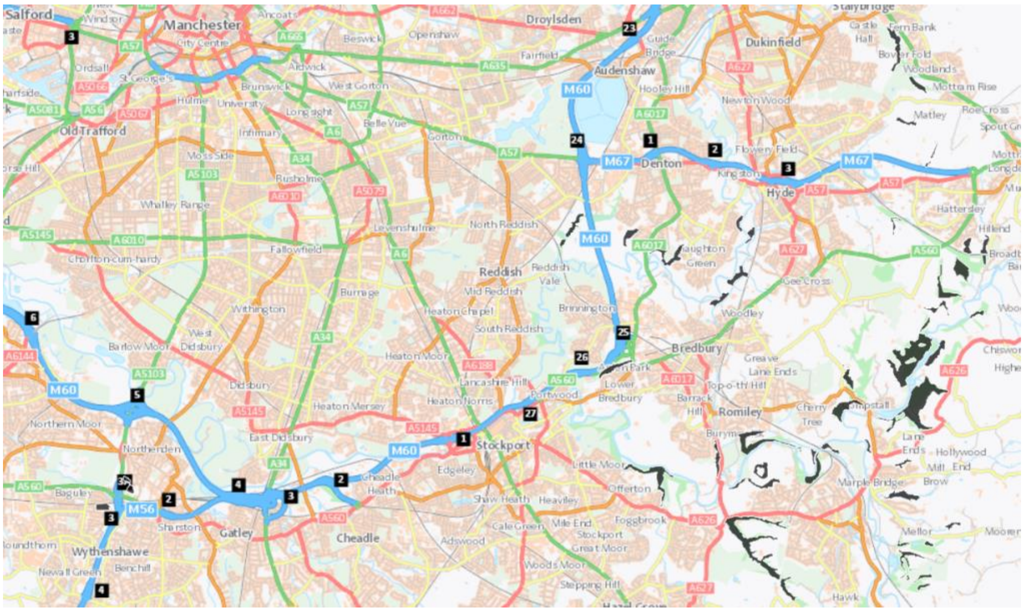

Figure 1. 6: Ancient woodland GIS layer.

We included information on ancient woodland because this represents woods which have existed since at least 1600 AD and have developed irreplaceable, complex ecosystems (Woodland Trust). Therefore, to receive permission to place an electric vehicle charging point within these ancient woodlands would be near on impossible.

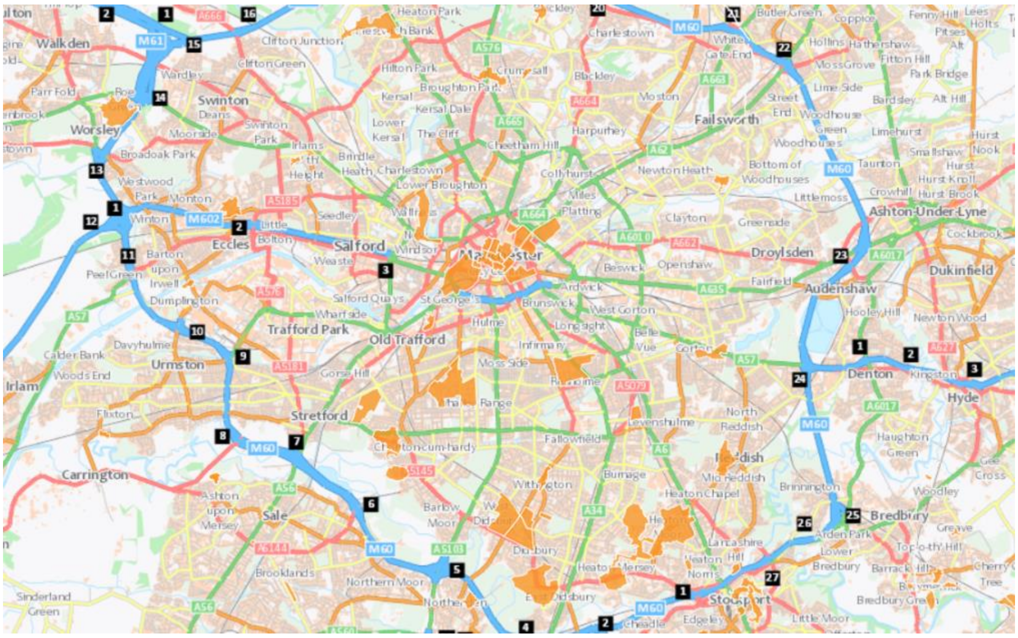

Figure 1. 7: Conservation areas and special areas of conservation GIS layer.

Information regarding conservation area and special areas of conservation refer to areas in the UK which exist to manage and protect the special architectural and historic interest of a place (Historic England[1]). These areas require have some extra planning controls and considerations in place, therefore it would be difficult to receive permission to develop electric vehicle charging points in them.

Figure 1. 8: Greenbelt GIS layer.

Information regarding greenbelt refers to specially designated areas of countryside in the UK which a protected from most forms of development, to stop urban sprawl and preserve the character of existing settlements (Oppenheimer, 2021). Therefore, to build electric vehicle charging points in these areas might lead to some backlash and may be difficult/costly to be granted access to, so we have removed these areas to minimise the chances of our proposed placements being insufficient in the long term.

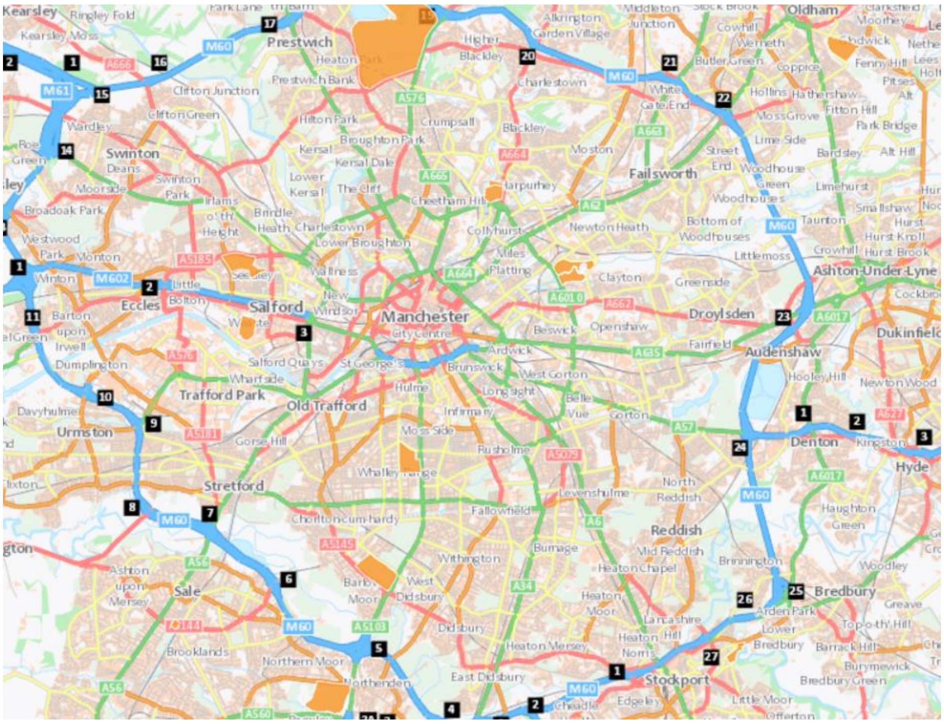

Figure 1. 9: Historic parks and gardens GIS layer.

Historic parks and gardens refer to registered sites in the UK which are not open to the public, unless advertised elsewhere (Historic England[2]). These can be areas such as the grounds of private houses, public parks and cemeteries. All areas in which planning permission to build electric vehicle charging points will not be granted or be very costly to the developers. Therefore, we have included these areas in our research to then remove these as potential electric vehicles charging point locations.

4. Methodology

Our method for identifying potential charging points for electric vehicles involves a variety of GIS tools and techniques. This includes density overlay analysis, K-means analysis, and flow-weighted analysis. Additionally, various tools such as ArcGIS online, ArcGIS software, the online site geojson.io, and the Python programming language were used to perform these analyses. Using these tools and methods, we can identify areas with high demand for charging points due to traffic flow at bus stops and proximity to substations. This helps to strategically place charging points within communities to meet the needs of electric vehicle users. Additional factors such as Brownfield land, ancient woodland etc. mentioned above are considered to ensure the accuracy of identifying potential charging point locations.

4.1. Density Overlay Analysis

This method is used to identify areas with high concentrations of features, such as bus stops or substations. By creating a heatmap from the distribution of these features, we can visually represent the areas of highest concentration. From there, we can create density circles that represent regions of similar density. These density circles can help us identify areas of overlap between bus stops and substations, which may indicate high demand for charging points.

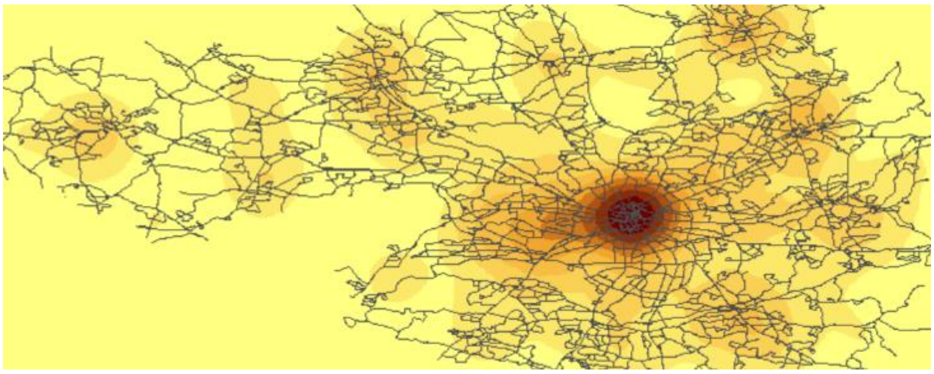

Figure 2. 1: Substations heatmap distribution.

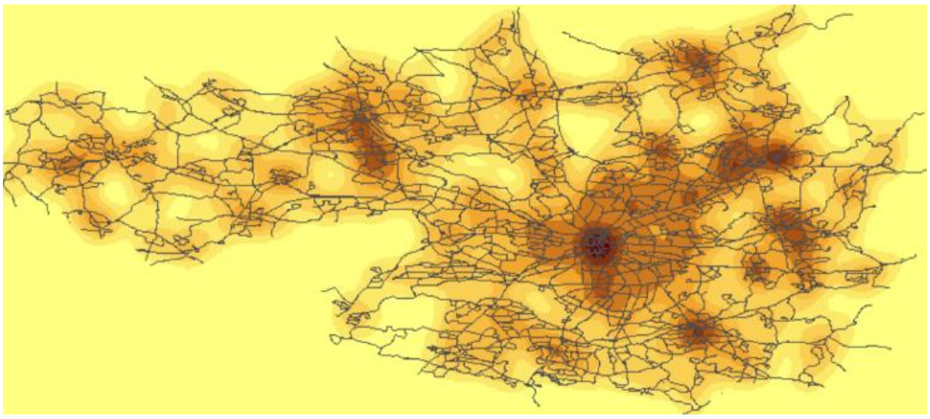

Figure 2. 2: Bus stops heatmap distribution.

To perform the density overlay analysis, we used a combination of ArcGIS software and ArcGIS online tools. We first create a layer of bus stops and substations (Figure 2.1 and Figure 2.2), then use the Generate Heatmap tool in ArcGIS Online to create a heatmap layer from the distribution of these features. This heatmap layer was then used to create density circles using the Create Buffer tool in ArcGIS software. Finally, we identified areas of overlap between these density circles to identify potential charging point locations along bus routes.

4.2. K-means Analysis

For this approach, we use the K-means clustering algorithm to group bus stops and substations based on their quantity. The K-means algorithm is a machine learning technique designed to divide data points into K clusters based on their similarity.

First, we pre-process the data to obtain the number of bus stops and substations in the matching area. We then use the K-means algorithm to bundle a specific number of bus stops together to form K clusters, where K is a predefined parameter, and the selection of K value can take a consideration by three different cases: firstly produce the buffer zone for every existing charging point and assume all the charging point has the same size buffer zone. The first case is union all the buffer zone area (Figure 2.3), then accumulate and record the number of bus stops to get average bus stops quantity in one buffer zone, k is equal to total bus stops divided by this average number. The second case is using the total transportation area in Great Manchester divided by the buffer zone area to get K value (ignore the existing charging point). The third case is extracting the area in Great Manchester without covered by buffer zone, then divided by the buffer zone area (consider the existing charging point). The detailed calculation step shown below:

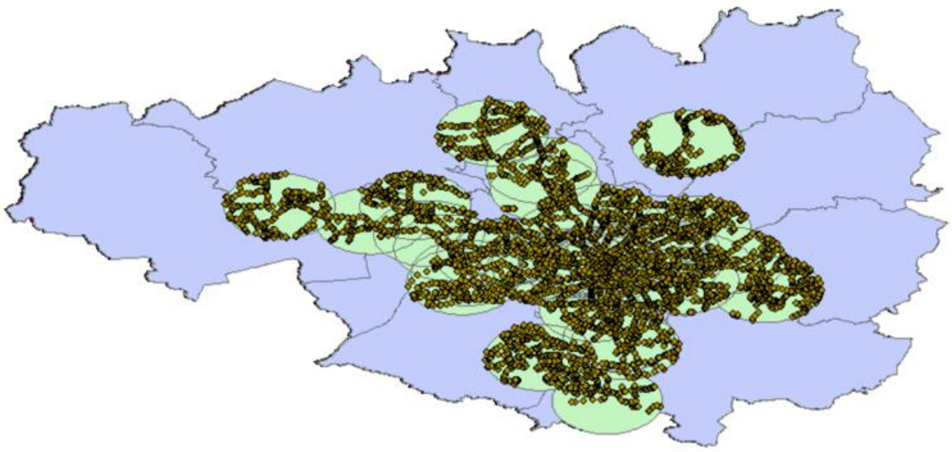

Figure 2. 3: All the bus stops covered in cluster buffer zone.

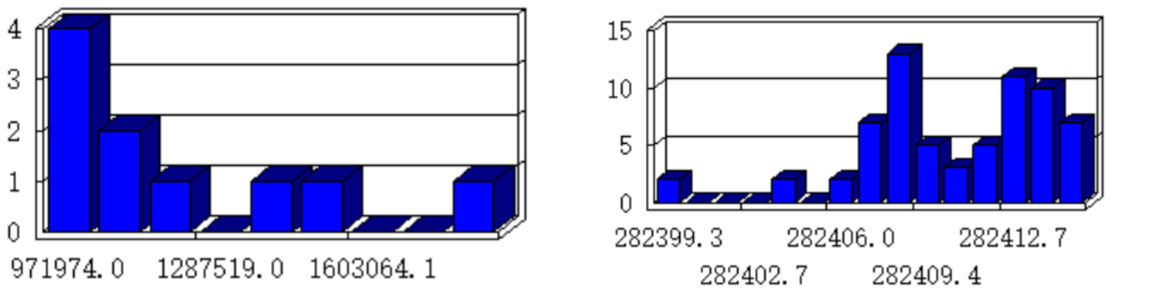

Figure 2. 4:Bar chart for area of Great Manchester & substation buffer zone.

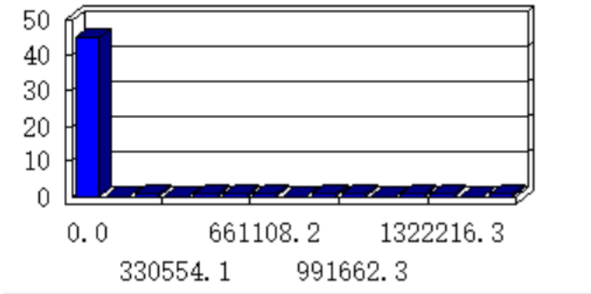

Figure 2. 5: Area after cutting out.

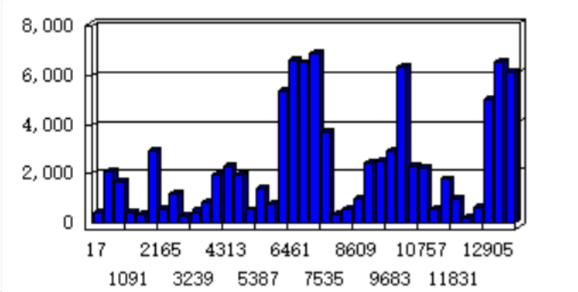

Figure 2. 6: Bus stop quantity in each buffer zone.

	the great Manchester area	the substation buffer area	the area after the buffer is cut out
COUNT	10	67	54
MINIMUM	971973.9	282399.32	0.01
MAXIMUM	1881710.08	282416.04	1570286.7
TOTAL	12452734.23	**18921533.17**	8304073.83
AVERAGE	1245273.42	**282410.94**	153779.14
STANDARD	279882.75	3.5791	375477.98

Table 1: Area integration table.

To determine the number of clusters in the K-means method, several calculations are required. Firstly, we need to calculate the size of each buffer, which is known as the buffer area. In this example, the total buffer area is 18921533 square meters, resulting in an average buffer size of 282410.9403 square meters. Secondly, we need to identify the number of bus stops within the buffer zone using the spatial correlation tool. In this case, there are a total of 91458 bus stops within the buffer zone, leading to an average of 1365.044776 bus stops per buffer zone. We then calculate the clipped area of each buffer by subtracting the area after clipping the buffer zone from the TGM area, which is 8304073 square meters.

Finally, we calculate three k values to determine the number of clusters. The first k value is the ratio of the clipped area of the buffers to the size of each buffer, which equals 29.4 and adds the number of existing charging points equal to 96. The second k value is the ratio of the clipped area of the buffer to the total area of the buffer, resulting in 44. The third and final k value is the ratio of the total number of bus stops to the average number of bus stops per buffer zone, which is 136.

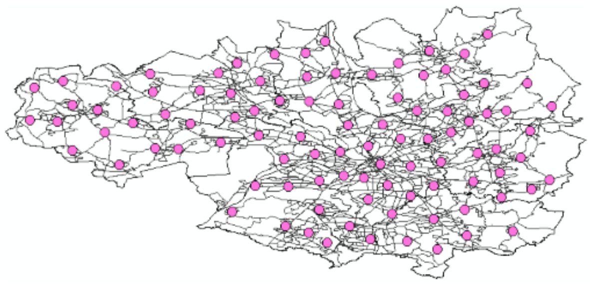

Figure 2. 7: Cluster centroid when k=44. *Figure 2. 8: Cluster centroid when k=96.*

Figure 2. 9: Cluster centroid when k=136.

Finally, each cluster forms a circle and finds the centroid coordinates of this circle. Next, we connect the centroid of each bus station cluster to the nearest substation. Areas close to this line represent a high concentration of traffic and sufficient power supply. We also decided and found the intersection of this line with the bus route to ensure that the location of the charging point is convenient for public transport users.

4.3. Traffic Weighted Analysis

This method is the improved version for the second method which involves clustering bus stops based on traffic flow values and combining this information with the location of substations to identify areas that need charging points. Firstly, we give each bus stop a traffic flow value and cluster them into different blocks according to this value, which can objectively show the traffic congestion near a bus stop and thus indirectly explain the demand for charging piles under the circumstances.

There are two different ways to calculate the traffic flow value for each transit stop: one is to add up the number of bus lines that pass through a particular bus stop. From (Figure

2.10) it is easy to see that the location of the bus stop is not on the bus line, so we decided to use the nearest line-point algorithm between the bus line (a line) and the bus stop (a point) to move the bus stop to the nearest of bus lanes, and then use python to accumulate the number of bus stops for each route. Another way is to use bus schedules to see how many buses pass by a particular bus stop during the day. After this step, it is equivalent to that all bus stops have their own weight scores.

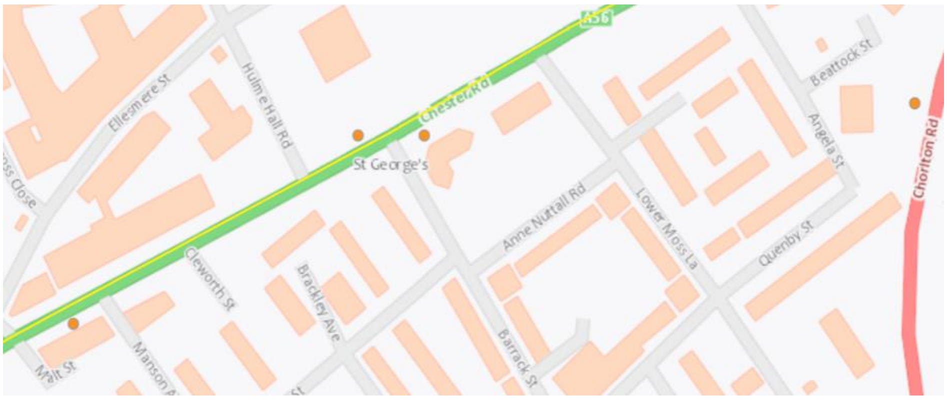

Figure 2. 10: Bus stop (orange point) and bus route (yellow line).

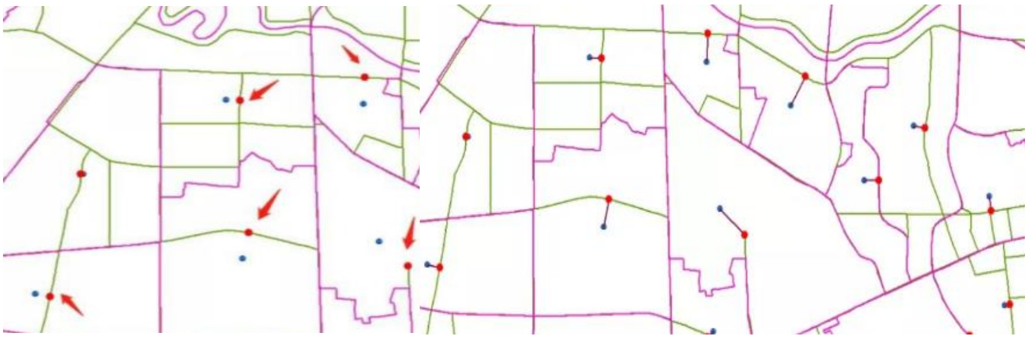

Figure 2. 11: Find the nearest line from a point.

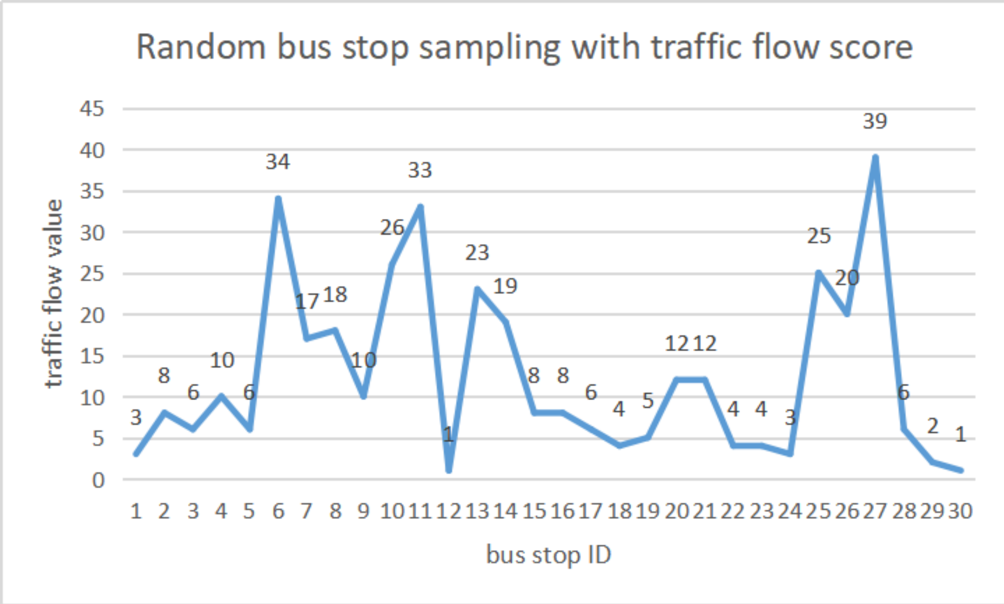

Figure 2. 12: line chart for partial bus stop with traffic flow score.

Next, we divide the bus stops into several blocks by a clustering method according to the weight scores. Specifically, each block should have a similar block score (total weighted score for the block's bus stops). Clustering using the traffic flow values of bus stops instead of the number of bus stops is a better way to reflect the traffic congestion in the area, because the traffic flow values can more accurately represent the actual traffic flow. For example, if we have 4 bus stops: blue and red with 100 traffic flow score, black and yellow with 6 traffic flow score, the result of k-means cluster when k=2 in Figure 2.14 and the result of this method when k keep same in Figure 2.14, it is clear to show that two clusters in the Figure 2.14 has balanced allocation.

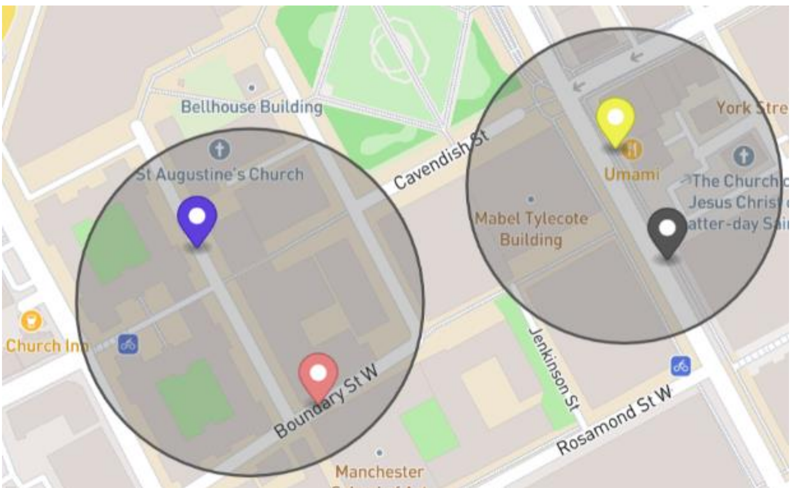

Figure 2. 13: K-means cluster when k=2.

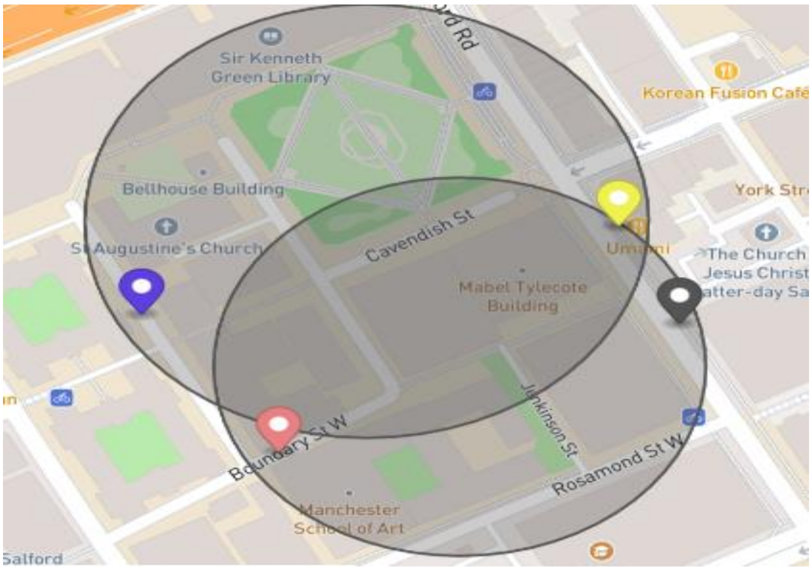

Figure 2. 14: Cluster apply weighted value (k=2).

Then set the substation buffer zone, we create a circular area for each substation to simulate its coverage area, the buffer radius is set to 1km by default.

Finally, the potential charging points are identified by finding the overlapping area of the bus station group circle and the substation buffer zone. This area represents the locations where there is high demand for charging points due to high traffic flow at bus stops and proximity to substations. Charging points can be placed within this area to serve the needs of EV in the community.

4.4. Extra and Further Considerations

In the future, spatial accessibility analysis can be utilized to calculate the 15-minute accessibility of existing charging stations to the bus stops, in order to optimize the locations of charging points. This analysis can be conducted by using GIS tools to generate a network dataset, incorporating information such as road networks, public transportation routes, and the location of charging points. By calculating the travel time between each bus stop and charging point, the accessibility of charging stations can be determined based on the maximum travel time of 20 minutes.

And this analysis can be further extended to include other factors, such as population density and EV ownership rates, in order to identify areas with the greatest potential demand for electric vehicle charging infrastructure. The results of such analysis can aid decision-makers in making informed decisions about the optimal locations for future charging points, ensuring efficient and effective deployment of electric vehicle charging infrastructure.

In addition, Machine learning model applying for improving the model accuracy, extracting the suitable features from existing charging point dataset like the location including the geographic coordinates (latitude and longitude) of the charging point, availability of a charging point (e.g., whether it is occupied or available), the type of charger used at the charge point and distance to nearest public transportation and amenities. By analysing the relationship between the extracted features and the demand for charging points, machine learning model can be developed to predict where new charging points should be installed to help us double check the previous charging point location.

4.5. Limitation

For the traffic weighted analysis, we did not achieve to use bus schedules to calculate the traffic flow value, because there is lack of data in bus time schedule and related data are difficult to integrate. This method has higher accuracy than using the quantity of bus routes. In addition, some factors are not considered like population density, road capacity, and proximity to amenities.

5. Results

5.1. Density Overlay Analysis

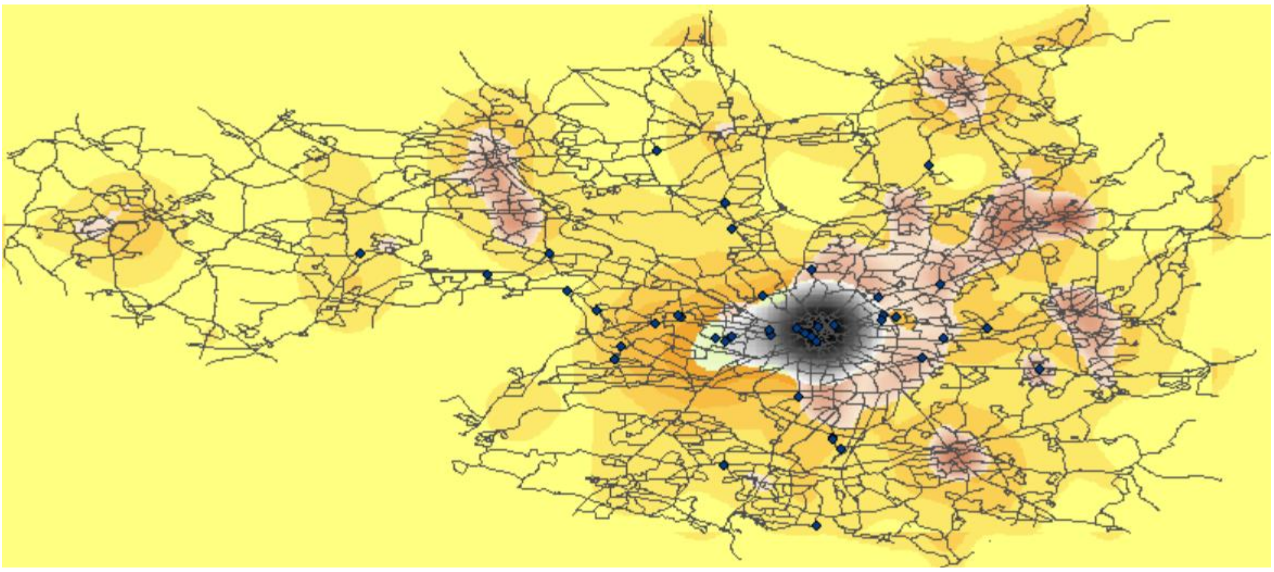

Figure 3. 1: Combined heatmap.

The combination of bus stops and substations heatmaps allowed for the identification of high value areas that are suitable for electric vehicle charging points. The red and black areas denote the high value ranges of the bus stops and substations, respectively, while the overlap of these ranges is represented by the black area. Further analysis of these high value areas revealed two key findings. Firstly, combining selected high value areas would be beneficial to maximize the potential coverage of electric vehicle charging points. Secondly, the distribution of these high value areas was found to be reasonable and consistent with the existing charging point locations, which provides confidence in the selection of these areas for future charging point installations. The intuitive nature of the similarities and consistency between the distributions further supports the potential for successful implementation of new charging points in these high value areas.

Figure 3. 2: High density area after cutting.

To identify suitable locations for electric vehicle charging points, we needed to consider certain limitations. Areas such as ancient woodland, Conservation area, Greenbelt, Historic parks and gardens, Special area of conservation, and Brownfield land were excluded as unsuitable for charging point installation. It is worth noting that the limitations were not only based on areas marked in black but also those marked in red and green. However, it is important to acknowledge the limitations of the analysis. While the results can be explained, the accuracy may not represent the actual traffic flow. Therefore, we recommend installing charging points at crossroads in black areas to increase the likelihood of usage. Additionally, we suggest adding charging points to the red areas where no charging points currently exist as there is demand but no facilities are in place.

5.2. K-means Analysis

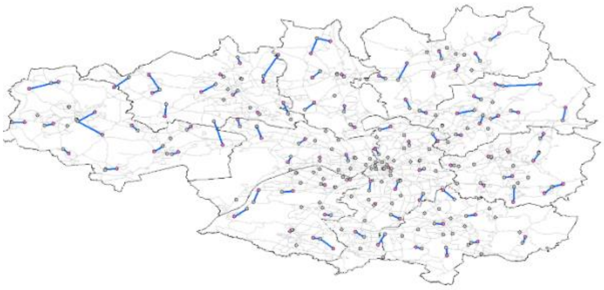

Figure 3. 3: Connected line when k=44.

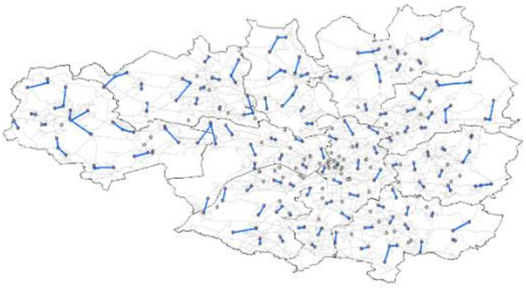

Figure 3. 4: Connected line when k=96.

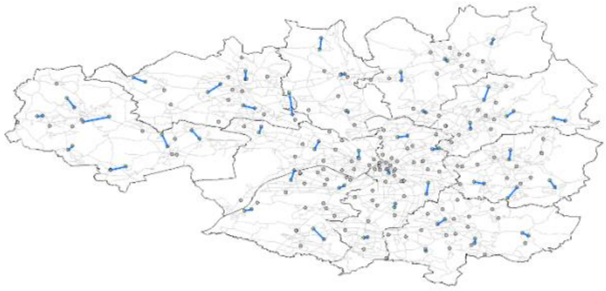

Figure 3. 5: Connected line when k=136.

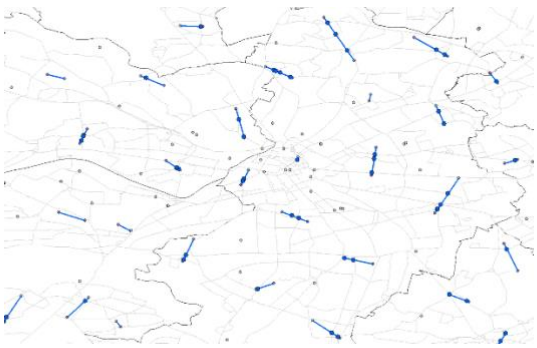

Figure 3. 6: Intersection between bus route and line.

After generating clustering points using the three methods, we conducted a 3km neighbourhood analysis on both the clustering points and substation points in ARCGIS. This enabled us to determine the nearest substation point for each clustering point and connect them with lines. We then overlaid these lines with the bus station routes to obtain intersection points, which we identified as recommended potential charging station construction sites.

Our selection of these areas was based on the fact that the points on the lines were located at a moderate distance from the clustering points and charging stations. This approach ensured a reliable source of electricity for charging and helped to reduce construction costs. Additionally, given the high demand for charging in the areas near the clustering points, our selection was effective in meeting the bus charging needs throughout Greater Manchester.

To optimize the use of available space, we compared and analysed council planning plans within the regional scope and avoided areas that could not be utilized. Furthermore, when selecting specific locations, we prioritized crossroads due to their high traffic flow, which would increase the visibility and accessibility of the charging station.

5.3. Traffic Weighted Analysis

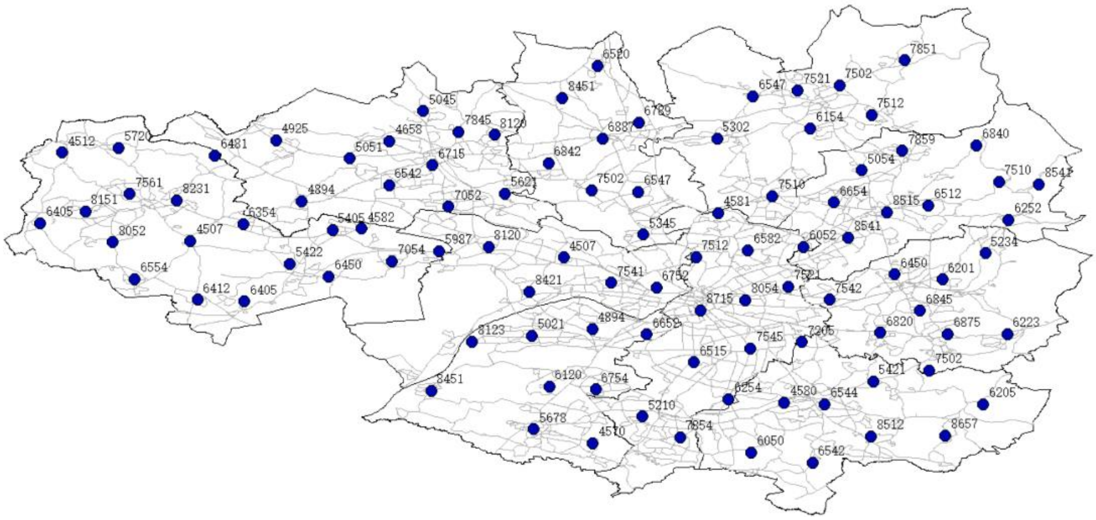

Figure 3. 7: Traffic-weighted charging station distribution map.

Figure 3. 8: Detail of traffic-weighted charging station distribution map.

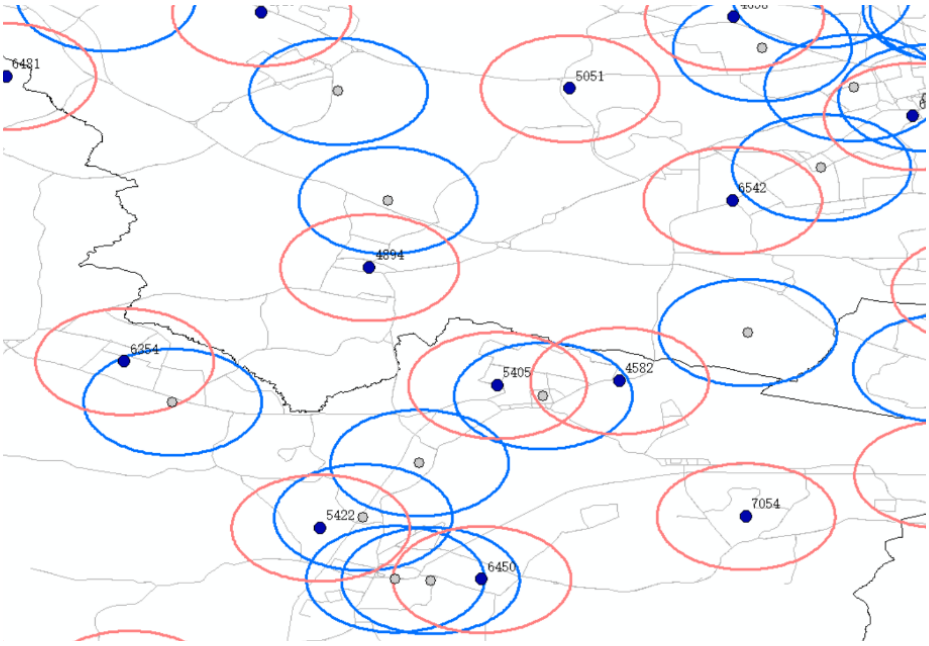

Figure 3. 9: Overlap area between substation buffer zone and cluster circle.

The results of the Traffic Weighted Analysis are displayed in Figure 3.9. This analysis method guarantees that each charging point can serve approximately the same bus flow, as shown by the numerical values on the figure. This clustering approach ensures a uniform distribution of charging resources, making the allocation of space charging resources more equitable while also meeting the charging demand.

Additionally, we were able to avoid areas where the construction of charging stations is not permitted in the planning, and we chose locations which overlap the area between substation buffer zone (blue circle) and cluster circle (red circle). This approach ensures that the charging stations are strategically placed in areas that will have the most impact in meeting the charging points needs of the buses.

6. Discussion

6.1. Limitations and Challenges Encountered During the Project

During the execution of our project, we encountered several limitations and challenges, which we outline below:

Data Quality and Completeness: The quality and completeness of data was a significant challenge. Not all datasets were complete, and some contained outdated or inaccurate information. The National Public Transport Access Nodes (NaPTAN) dataset, despite being comprehensive, suffers from potential time lags in data updates, inaccuracies in geospatial coordinates, and lack of up-to-date information on the status of bus stops. Additionally, it lacks data on electricity supply at bus stops, crucial for determining their potential as charging points. The National Charge Point Registry (NCR) dataset, the key source for existing EV charging points, also has limitations. It may not include all existing charging points due to inconsistent data provision by operators, leading to potential omissions. Information on connectors, charging methods, and charge modes can also be inaccurate or incomplete, and the operational status of charging points might not be current. Lastly, the Substation Heatmap of Electricity North West, while useful for understanding electrical capacity, is based on the current network state and doesn't account for future upgrades or changes. The conversion from Easting and Northing coordinates to longitude and latitude could introduce errors, and there's no information about the physical space available at substations for new charging infrastructure. These limitations collectively present challenges to the accuracy and completeness of our analysis.

Data Compatibility and Integration: Merging data from various sources and in different formats was a complex task. We had to ensure that all the datasets were compatible with each other, which required significant preprocessing. For example, it was difficult to integrate Naptan dataset and NCR dataset as there were less matching column to join. Additionally, we had to convert all datasets to the geojson format to visualize them in the GIS system, which was a time-consuming process.

Geospatial Analysis Complexity: The complexity of geospatial analysis was another challenge. Conducting density overlay analysis, K-means analysis, and traffic weighted analysis required advanced skills and knowledge. The selection of the appropriate number of clusters in the K-means analysis was particularly challenging, as it directly impacted the results of our analysis.

Regulatory and Environmental Constraints: Our analysis had to account for various regulatory and environmental constraints, such as restrictions on building charging points in conservation areas, ancient woodlands, or greenbelts. This limited the number of potential sites for new charging points.

Assumptions and Predictions: Our project relied on several assumptions, such as the availability of charging capacity at bus depots and substations, which may not hold true in all cases. Similarly, our predictions about potential charging point locations were based on current data and may not accurately reflect future conditions, such as changes in bus routes or the construction of new charging points.

Resource and Time Constraints: Given the large scope of our project, we faced constraints in terms of time and resources. We had to balance the need for a detailed and comprehensive analysis with the need to deliver results within a specific timeframe. This required careful project management and prioritization of tasks.

6.2. Potential Blockers

As part of the project to develop a data file that interlinks with Optibus for viewing potential electric bus charging locations, it is important to consider potential blockers to site usage and planning systems that can help to resolve these issues. Some of the potential blockers and solutions are discussed below:

Availability of suitable locations: One of the biggest challenges to the roll out of electric bus charging infrastructure is finding suitable locations. This includes identifying locations that have the available capacity to charge EVs, the space to charge multiple buses simultaneously, and a centralised location where multiple routes pass through regularly to ensure maximum vehicle efficiency while maintaining charge. To address this issue, it is important to carry out a thorough site analysis to identify potential locations that meet these requirements. This can involve working with local authorities and bus operators to identify suitable sites and considering the feasibility of installing charging infrastructure in existing depots.

Cost of installation: Another potential blocker to site usage is the cost of installing charging infrastructure. The cost of installation can vary depending on the type of charging infrastructure required and the location of the site. To address this issue, it is important to consider the long-term benefits of installing charging infrastructure and the potential savings that can be made in fuel costs over the lifetime of the infrastructure. Additionally, working with local authorities and other stakeholders to secure funding for infrastructure installation can help to reduce the cost of installation.

Grid capacity: One of the key challenges to installing electric bus charging infrastructure is ensuring that the local grid has the capacity to support the additional demand from charging infrastructure. To address this issue, it is important to work with local utilities and other stakeholders to assess the capacity of the local grid and identify potential solutions to address any capacity constraints. This can include upgrading the local grid infrastructure or installing energy storage solutions to help balance the load on the grid.

Planning and permitting: Another potential blocker to site usage is the planning and permitting process. Installing charging infrastructure can require permits from a variety of different agencies, including local planning authorities, environmental agencies, and utilities.

To address this issue, it is important to engage with these agencies early in the process and work collaboratively to identify potential issues and develop solutions to address them. Additionally, working with local authorities to streamline the permitting process can help to reduce the time and cost of installation.

User adoption: Finally, user adoption can be a potential blocker to site usage. It is important to ensure that bus operators and drivers are comfortable using electric buses and that they understand the benefits of electric bus charging infrastructure. This can involve providing training and education to bus operators and drivers, as well as working with local authorities and other stakeholders to promote the benefits of electric buses and charging infrastructure

6.3. Potential Obstacles to Long-Term Data Access

In our project, we have identified several significant obstacles to long-term data access. These include:

Data Privacy and Protection: We must adhere to data privacy regulations such as the General Data Protection Regulation (GDPR). These regulations restrict how personal data can be collected, stored, and used. If our data includes any personally identifiable information (PII), we need to ensure we have the necessary permissions and protections in place. Failure to do so can result in severe penalties.

Data Quality and Upkeep: Maintaining the accuracy and reliability of our data over time is another significant challenge. Data can become out of date or the quality may degrade. For instance, charging station locations may change, or new stations may be added or removed. We need to commit to regular updates and checks to ensure that the data remains useful. However, this requires resources and a system in place to manage it, which could be a potential obstacle.

Data Sharing and Interoperability: Data is most useful when it can be easily shared and used by different systems. However, achieving this can be a challenge. Different systems may use different data formats or standards, making it difficult to share data between them. We need to ensure that our data can be easily converted or transformed to be compatible with these systems, which can be a complex and resource-intensive process.

Data Security: With the rise in cyber threats, keeping our data secure from breaches is another potential obstacle. As we plan to make our code and data available through an online repository like GitHub, we need to ensure that this does not expose our data to potential security threats. If we're using cloud-based storage or processing services, we need to ensure they provide robust security measures.

Data Ownership and Legal Issues: When obtaining data from external sources, we need to consider who owns the data and any restrictions on its use. There may be legal issues or licensing conditions that restrict how we can use the data or require us to pay fees or royalties. Over time, these conditions could change, potentially impacting our access to the data.

Data Storage and Management: Storing and managing large volumes of data over the long term can be a challenge. We need to ensure we have sufficient storage capacity, and that the data is backed up to prevent loss. As the volume of data grows, we may need more advanced data management tools or systems, which can be costly.

6.4. Comparison of the results with those of previous research

The findings of our research present a compelling alignment with previous literature, while contributing unique insights. Our Density Overlay Analysis resonates with the emphasis Lopez de Briñas Gorosabel et al. (2021) placed on land use, spatial constraints, and accessibility when selecting charging locations. By identifying high-value areas for the installation of charging points, we've underscored these factors, while also acknowledging the unsuitability of certain areas for installation, consistent with He et al. (2022)'s integrated approach to charging infrastructure planning.

In utilizing K-means Analysis, we've aligned with Dirks et al. (2023)'s emphasis on efficient route planning. Our method of overlaying bus routes with potential charging station sites contributes to route optimization, and by considering the proximity to substations, we minimize downtime, thereby addressing the limited driving range issue highlighted by Aldenius et al. (2023).

The Traffic Weighted Analysis, further, aims to distribute charging resources equitably, a holistic approach in line with Grigorieva and Nikulshin (2023)'s recommendations for charging infrastructure planning. By ensuring that each charging point can serve approximately the same bus flow, we're facilitating strategic decisions regarding charging infrastructure, which Guschinsky et al. (2021) found significantly impacts the overall performance of city electric bus services.

Overall, our study not only echoes the need for comprehensive strategies and collaborative efforts to overcome charging infrastructure integration challenges, as identified in the literature, but also provides data-backed methodologies for facilitating better route planning, scheduling, and overall system performance for electric bus networks, directly addressing the objectives of the Optibus project.

6.5. Recommendations for Future Research and Improvements to the Project

For the continuous improvement and evolution of this project, a number of strategic enhancements and future research directions can be considered.

One significant recommendation involves expanding the scope and diversity of data used in the project. This could encompass a more comprehensive set of variables such as real-time traffic data, detailed bus schedules, demographic information, and perhaps even aspects of commuter behaviour. Such an enriched dataset could provide a more nuanced understanding of demand patterns and usage of the charging stations, thus optimizing the infrastructure placement further.

The rapid evolution of electric vehicle (EV) technology necessitates that future studies adapt and evolve in parallel. As new charging technologies emerge, their unique requirements and specifications should be incorporated into the site selection and planning process. This could also include exploration of alternative power sources such as solar or wind, and understanding how these might integrate with the charging infrastructure.

Another avenue for future exploration is the intersection of charging infrastructure planning with socioeconomic factors. The impact of charging station location on local communities, the accessibility for different socioeconomic groups, and the potential for job creation in the sector are all pertinent considerations. Collaborative efforts with urban planners, local authorities, and community representatives could lead to more equitable and sustainable outcomes.

The real-world application of the research findings is an essential step for validating the project outcomes. Conducting pilot studies in the identified locations could offer valuable insights into the practical feasibility and efficacy of the recommended charging infrastructure. These studies could also provide real-world data on usage patterns and demand, which can then be used to refine the models further.

In the context of the evolving urban landscape, static models may not suffice. Future research could focus on developing dynamic models that can adjust to changes in the urban structure, bus routes, and even broader changes in transportation trends. These dynamic models could ensure the continued relevance and applicability of the research outcomes.

Finally, the policy implications of charging infrastructure planning deserve further investigation. Understanding how existing policies impact the adoption of electric buses and the development of charging infrastructure can provide a foundation for advocacy work. Moreover, future research could propose policy recommendations that support the proliferation of EV charging infrastructure, promoting a smoother transition to green public transportation systems.

By taking these recommendations into account, the project can remain on the cutting edge of research in this area, contributing to the transition towards sustainable and efficient public transport systems.

7. Conclusion

In conclusion, this project presents a well-founded approach to tackling the urgent need for electric bus charging infrastructure in the Greater Manchester area. The team has exhibited a commendable effort in meticulously gathering, pre-processing, and visualizing data, and producing a comprehensive geojson data file. This file has the potential to significantly impact infrastructure planning and inform policymaking, ultimately promoting a more sustainable and environmentally friendly public transportation system.

Furthermore, the project's innovative use of GIS tools and techniques, such as density overlay analysis, K-means analysis, and traffic-weighted analysis, has enabled the identification of potential charging locations based on traffic flow, proximity to bus stops, and proximity to substations. By incorporating additional factors, such as Brownfield land and Ancient woodland, the analysis has achieved greater accuracy in determining suitable charging point locations.

Consequently, the study's findings provide valuable insights for decision-makers by offering a framework for the strategic placement of charging points within communities. These perceptions are essential for catering to the requirements of electric vehicle customers and supporting a more seamless transition to electric buses. The results also highlight the need for deploying electric vehicle charging infrastructure in metropolitan areas in an efficient and effective manner.

However, despite these successes, the study acknowledges the limitations and challenges it encountered, such as data quality, completeness, compatibility, and integration issues. The complexity of geospatial analysis, regulatory and environmental constraints, and resource and time constraints were also noted as obstacles. The project's results also highlight potential barriers to charging point installation, such as suitable location availability, installation costs, grid capacity, planning and permitting, and user adoption.

Building on this, this project expands on previous research findings related to land use, spatial constraints, and charging location accessibility while emphasizing the importance of route optimization and equitable distribution of charging resources. Recommendations for future research include broadening the scope and diversity of data, adapting to emerging electric vehicle technology, accounting for socioeconomic factors, conducting pilot studies, developing dynamic models, and investigating policy implications.

To address these limitations and challenges, future studies could delve deeper into incorporating factors such as population density, EV ownership rates, and accessibility analysis. The development of machine learning models to predict optimal charging point locations is another potential area for exploration. By taking these limitations, challenges, and recommendations into account, this project stands as a valuable resource for planners, decision-makers, and researchers working on electric bus charging infrastructure.

In summary, this project significantly advances the Greater Manchester region's demand for infrastructure for electric bus charging. Its exhaustive data gathering, analysis, and visualization, as well as its creative application of GIS tools and techniques, provide a strong platform for further study and advancement in this area. The initiative can be a useful tool in facilitating a smooth transition to sustainable and effective public transport systems in metropolitan areas by admitting and addressing the limitations and difficulties it has encountered.

8. References

Aldenius, M., Mullen, C., & Pettersson-Löfstedt, F., 2023. Electric buses in England and Sweden – Overcoming barriers to introduction. Transport Policy, 111, pp.43-53.
Available at: https://www.sciencedirect.com/science/article/pii/S0967070X22000730

Dirks, N., Schiffer, M., & Walther, G., 2023. On the integration of battery electric buses into urban bus networks. Transportation Research Part A: Policy and Practice, 159, pp.39-61.
Available at: https://www.sciencedirect.com/science/article/pii/S0965856422000347

Grigorieva, O., & Nikulshin, A., 2023. Electric buses on the streets of Moscow: experience, problems, prospects. Energy Policy, 160, 112759.
Available at: https://www.sciencedirect.com/science/article/pii/S0301421522000451

He, Y., Liu, Z., & Song, Z., 2022. Integrated charging infrastructure planning and charging scheduling for battery electric bus systems. Transportation Research Part C: Emerging Technologies, 138, 104548.
Available at: https://www.sciencedirect.com/science/article/pii/S0968090X22000730

Lopez de Briñas Gorosabel, O., Xylia, M., & Silveira, S., 2021. A framework for the assessment of electric bus charging station construction: A case study for Stockholm's inner city. Energy Reports, 7, pp. 4093-4105.
Available at: https://www.sciencedirect.com/science/article/pii/S2210670721008751

Guschinsky, N., Kovalyov, M. Y., Rozin, B., & Brauner, N., 2021. Fleet and charging infrastructure decisions for fast-charging city electric bus service. Computers & Operations Research, 136, 105399.
Available at: https://www.sciencedirect.com/science/article/pii/S0305054821002021

National Geographic, 2022. GIS (Geographical Information System). National Geographic Society.
Available at: https://education.nationalgeographic.org/resource/geographic-information-system-gis/

ArcGIS, 2023. EV charging point project.
Available at:
https://www.arcgis.com/apps/instant/sidebar/index.html?appid=1432d1ba36ec4019b9c8cb69bac95b7a&locale=en

Woodland Trust, 2023. Ancient Woodland.
Available at: https://www.woodlandtrust.org.uk/trees-woods-and-wildlife/habitats/ancient-woodland/

Historic England[1], 2023. What is a Conservation Area?.
Available at: https://historicengland.org.uk/listing/what-is-designation/local/conservation-areas/

Oppenheimer, P., 2021. What is the green belt, and why does it matter?. The Countryside Charity.
Available at: https://www.cpre.org.uk/explainer/what-is-the-green-belt-and-why-does-it-matter/

Historic England[2], 2023. Registered Parks & Gardens.
Available at: https://historicengland.org.uk/listing/what-is-designation/registered-parks-and-gardens/

9. Appendices

- Function Code

```
import pyproj

# Read the substations csv file
substation = pd.read_excel('Heatmap Substations.xlsx')
# Define the projection systems
inProj = pyproj.Proj(init='epsg:27700')  # British National Grid
outProj = pyproj.Proj(init='epsg:4326')  # WGS84

# Define the input easting and northing
easting = 367434
northing = 426087

substation['Longitude'], substation['Latitude'] = pyproj.transform(inProj, outProj, df_coor['
Easting'].values, substation['Northing'].values)

# Show the converted coordinates to confirm completion
print(substation[['Easting', 'Northing', 'Longitude', 'Latitude']].head())
```

Code 1: Converting Easting and Northing data into Longitude and Latitude.

```
import json
from sklearn.cluster import KMeans
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.pyplot import figure
with open('GM_bus_stops.geojson') as f:
    data = json.load(f)
coordinates = [feature['geometry']['coordinates'] for feature in data['features']]
coordinates = np.array(coordinates)
#Train model
kmeans = KMeans(n_clusters=136,n_init= "auto")
kmeans.fit(coordinates)
#Plot clusters
figure(num=None, figsize=(8, 6), dpi=80, facecolor='w', edgecolor='k')
plt.scatter(coordinates[:, 0], coordinates[:, 1], c=kmeans.predict(coordinates), s=50,
cmap='viridis')
centers = kmeans.cluster_centers_
plt.xlim(min(coordinates[:,0]) - 10, -50)
plt.scatter(
centers[:, 0],
centers[:, 1],
c='black',
s=200,
alpha=0.5
```

```
);
import csv
from itertools import zip_longest
list1 = centers[:, 0]
list2 = centers[:, 1]
d = [list1, list2]
export_data = zip_longest(*d, fillvalue = '')
with open('centroid.csv', 'w', encoding="ISO-8859-1", newline='') as myfile:
      wr = csv.writer(myfile)
      wr.writerow(("longtitude", "latitude"))
      wr.writerows(export_data)
myfile.close()
```

Code 2: k-means clustering model

```
import os
import numpy as np
import pandas as pd
import time
import csv
with open("accumulate.csv","w") as csvfile:
    writer = csv.writer(csvfile)
    writer.writerow(["index","count"])
    tx=open('bus_stops_on_road.csv')
    df=pd.read_csv(tx)
    tx.close()
    Cname=df['properties.FID_GM_bus_stops_to_road_Features']
    arr=[]
    b=-1
    count=0
    for a in Cname:
        if a==b:
            count=count+1
        else:
            writer.writerows([[a,count]])
            count=1
            b=a
```

Code 3: accumulate the number of bus routes for bus stops as traffic flow value