1. INTRODUCTION

Lake District is England’s largest national park which is also deemed as a World Heritage site by UNESCO. It is a rural area in the region of Cumbria with breathtaking landscapes, lakes, picturesque villages, and rich English heritage. There are national parks with wildlife and cultural heritage. The whole region covers about 2,362 sq kilometers with the beauty of its parks, lakes, and mountain regions. The region consists of ten of the highest mountains, sixteen lakes and natural parks.

Tourism being the main source of income for the lake district, there are many small businesses which benefit from this during the summers, which last for four months from May to August. The revenue grossing up to £1.35 billion in 2021 with 15.73 million tourists visit in 2021 (Cumbria Touism,2023).

The lake district is connected by railways and roadways from many parts of England. During the summertime, the destination attracts 20 million tourist each year (Cumbria Touism,2023) out of which around 83% of the tourist visit the destination by car. As a rural area this contributes to the congestion of the region, causing long waiting lines, pollution, and traffic. The A6 route runs from Kendal to Penrith, which contains most of the major settlements in the region.

The vast selection of public transport gives a wide range of selection where one can enjoy the places to the full extent. Using public transport helps the region reduce their carbon footprint during the peak visit times and reduce the congestion which occurs during this time, reduction of private vehicles also helps to preserve the region and its natural beauty to its fullest and brightest. The lake district council has started a new initiative called the go lakes that is funded by the Department of Transport, Cumbria Tourism, and the Lake District National Park Authorities. The project is estimated to take around 6.9 million to implement. The go lakes project is based on electric autonomous vehicle that would take tourists around to attractions around the cities, too few remote areas and other resources or responsibilities such as improving the public transport and traffic, bicycle hire fleets, developing safe networks for cycling, and walking for visitors (Lake district, 2023).

To overcome the increase in private vehicles used by the tourist for their visit to the lake district, to curb this problem the concept of shared vehicle is introduced. The increase in private vehicles has caused an increase in demand for parking spaces and congestion of the roads, the roads in the Lake district are not wide which intern forces the vehicles to be lines up in a single file leading to longer wait time. A shared used vehicle is a mode of transportation which encourages the public to use public transport and other services like bicycles and carpooling for their visits. As per the statistics survey (Cumbria, 2023) about 23% of tourists consist of individuals of the age group 25 – 34, as a majority and individuals of age group 45 – 55 coming with 21%. For the young group of individuals this mode of transportation seems fit to explore the various regions around the Lake district, making it adventurous for them. For the other age group shared electric cars can be considered to explore the area as a form of public transport

Tourists prefer to go about their day without any hassle or waiting required to board a bus or train when visiting a destination. Optimizing the shared used vehicle in the lake district would benefit for a better cost-effective mode of transportation and increase use of the public transport. The proposed model for the shared used vehicle is placing bicycle stands in the attraction points which would allow the tourist to cycle to the bus stop or any other destination without the time constraint of waiting for a bus or missing the bus while walking to the bus stop. Finding the best spots for locating the bicycle stands among the attraction points to be used as an efficient and effective way as mode of transport.

E- bikes, traditional bicycle, geared bicycle, and e-scooters as a mode of public transport are a part of mobility as a service (Oeschger et al; 2020), they are much cleaner and sustainable as transport. Optimizing the bicycle stands or docking stations (stations for charging) around the area will make the travel more efficient for the travelers. The ideal locations for these stations would be near hotels, attractions, lakes, and places to eat.

The shared vehicle method can be beneficial to the last mile and first mile transportation problem, which is the transport or the last journey between bus station to home or any workplace that usually is a short distance travelled by an individual during their daily journey. Shared vehicles can also be used as daily bases as a replacement for the private vehicle, at this day and age the generation is more cautious about the environment and would like to help in sustainable growth of the economy. The mobility can be useful as a replacement as they can be used to go around the city hassle free, considering the cities are not very big going from one end to the other should not be a problem. In the United Kingdom as per (Sustran.org, 2023) the National Cycle Network has dedicated bicycle lanes that runs through major cities and towns with 4.2 million people using this network which saves around 70.9 million car trips thus this would not cause a problem for other transport vehicles on the road and considering the safely of the cyclist at heavy traffic times (Lin, Yang; 2011 – bike sharing).

A person riding a bicycle on a graph

Description automatically generated

*Figure 1: Growth of bike sharing (source: statistica)*

The image above shows the growth of bicycle sharing worldwide, the statistics done by Statista. The growth up to the year 2018, the industry is growing constantly as many countries are implementing the bike sharing concept to gain an ecofriendly transportation method. Governments across the globe are adopting a new mode of transportation for the last mile problem which has increased the market value of bicycles and the manufacturing market. The increase in electric scooters and bicycles are on a rise compared to the conventional bicycle, these electric bicycles are mostly solar powered or rechargeable when they are docked at their stations.

The research question here is to develop a heuristic method for facility allocation to optimize the shared used vehicle in the Lake district. This paper consists of three technical aspects the initial analysis, model formation, and the model analysis.

1. LITERATURE REVIEW

“The mobility system of the future is likely to be very different from what exists in most of the world today. The individual traveler is the heart of evolution, so consumers will need to be open to adopting new technologies and services. However, both the public and private sectors will have roles to play in paving the way.” (Hannon et al., 2016)

Cycling is one of the cheapest modes of transportation besides walking and is considerably faster than walking, which means time is saved in reaching the destination.

Bicycle sharing can be defined as individuals can use the bicycle on an as needed basis but do not have to worry about the cost and responsibility of bicycle ownership. This can be termed as a flexible mode of transportation for short term usage on a daily basis which allows the public to attain the unused bicycle and use it on their bases (Lin, Yang; 2011- bike sharing).

To solve the facility location problems there are three approaches which can be taken, the first being the exact solution approach, the second is the heuristic approach and finally the simulation approach.

One of the biggest hurdles for bicycle sharing systems is determining the best supply stations and the locations for bicycle hubs. The most widely used models are the p – median model and the maximum coverage location problem model (M.G. Werneck; 2004 -21). Both these models serve the same purpose of determining the locations, but the methodology is different. The p – median model seeks subsets of the hubs such that the serving of the customer cost is minimized whereas in the maximum coverage location problem the hubs the maximum distance to supply the customers is limited (C. Sohn; 2017 - 22). These are traditional methods with exact location solutions to the problem. These methods generally work for locations that are mostly distributed normally on the graphs. These models consider each node in a graph as a single point or facility which would not be usable when there is a collection of locations as pick up points. The p-median has a method of branching and bounding which considers the maximum coverage with maximum distance which would not be suitable for the facility location where distance should be minimum with maximum coverage of the facility (Oded; 1994 – p median).

There are many extensions to the traditional methods, these extensions consider few constrains to the models like stock of the bicycle and service level proposed by Lin et al. other such extensions like travel, set up cost of the bicycle lanes and set up cost of the bicycle stands finally the rate of pick up and usage of the bicycle proposed by Lin and Yang (Lin. Yang; 2011 - 25). All these extensions are a part of the objective function for the p – median and MCLP models (Wang et al; 2020 – placement optimization). The proposed model deal with a pre weighted graph with weighted nodes and edges, in cases where the graphs are not weighted these models might pose a problem, if the weights are assigned at random which might not be accurate only then the model will suffice. Most of the exact solution approach cannot be used in a particle setting with scattered and joined nodes as locations.

Solving the maximum coverage problem can have two variants, linear programming, and a heuristic method. The heuristic method follows a greedy algorithm, this algorithm starts with an empty set and adds nodes that have a maximum coverage of other points, but the only drawback is for solutions where p is greater than 1 the solutions obtained are not optimal. The solution contains all the points that have a maximum coverage which means some points are repeated or not feasible ( Church, Re Velle; 1974).

The quantitative tools to optimize the station allocation include linear programming, mixed integer, bi level, and queuing theory. These optimizing methods are well effective in bigger geographic areas. The constraint includes demand for bicycle sharing systems mainly for electric bikes.

The bi-level model framework proposed by (Wang et al; 2020 - placement optimization) shows two levels where the upper level determines the location of the bicycle stand the lower level determines the number of bicycles deployed per demand of the location. This model is based on the available distribution of the bicycle by solving the integer linear program and then optimizing the solution from the upper level. This model is effective for bigger cities where is a higher demand for bicycles on a daily basis.

An extension of the bi-level method is the vehicle sharing problem, which determines the placement of shared vehicles strategically placed in a network (Nair, Hooks;2014 -equilibrium network). All of the exact solution models have a defined demand and cost constraints which is not necessarily important for the heuristic method.

The heuristic model has a different approach to the same facility location problems. The main trade off for a heuristic model is between the computational time when there are multiple locations and the optimal solution with a cost factor. This approach has does not require a graph which is weighted, the heuristic method proposed by Leon Copper where the initial starting location is chosen at random.

An extension of this model is (Taylor online paper) is instead of finding the initial locations at random they propose to take into account the already found locations while generating the other remaining locations, this approach would be best suited if the locations are placed far away from each other in the graph, this approach would be useful only if the data of the whole Lake district is considered, in cities this would not be effective as the attractions and other locations are close by each other.

A different approach by (G R Sutanto et al; 2018 heuristic approach), where dealing with customers allocation to provide maximum coverage using capacitated facility location problem using k means clustering method. This approach looks at the capacity constraint for over capacitated situations and relocates the point. The author compares this approach with Liao’s approach, where the solution is evaluated and recorded. This author’s approach does not record and evaluate the solution which means that the solution provided might not be the best solution. (Liao, Guo; 2008) approach allocates the demand of each facility considering the capacity and minimizing the cost after which the optimization using the k means clustering method. The lack of robustness arises with use of k means clustering as the value is hard assigned the clustering algorithm becomes sensitive towards outliers or noise points, providing a fixed cluster size. It can also be sensitive to the initial centroids of the set of locations, since to clusters are assumed to be spherical clusters it would be difficult to use in a city location.

The final method for solving this problem is the simulation problem, as (Barth et al; 1999) there are multiple parameters which decide the final solution for a simulation model. The simulation model evaluates the usage of the multiple shared vehicle station. The extensive parameters for the model include the pick up and drop off rates of the vehicles, with the help of this analysis the cost analysis and evaluation could help in better improvisation of the system. This methodology could be helpful for the after implementation to analyze the demand and cost effectiveness of the whole system once deployed.

Several methodologies show the simulation models are run on established network routes, these network routes consist of the bus routes, car routes. The action based simulation model depicts the accurate demand of the distribution network, designing the transit network with integration of micro mobility in a controlled and safe environment with service supply and demand vectors. Simulation based optimization models are used by the transport planning department (Nnene et al; 2023). The extension of the simulation based optimization models can include minimization of the general cost of the multimodal transportation (Song et al; 2003). The most commonly used simulation models are the particle swarm optimization method, the area this particular optimization method is used mostly in locating the energy stations for electric vehicles in the city.

The traditional dB scan has a high computational cost as it calculates the distance between every two points and compares it with each other (Jing, Kang, Lie; 2018 taxi paper). In k means clustering method there must be a pre-defined number of clusters to be given to the algorithm, this would cause locations which might be included even to a cluster which does not satisfy the distance constraint, to avoid this misconception of random points being assigned to a cluster, db. scan is a better option.

1. METHODOLOGY

The public transport system in the Lake district is less frequent, as the demand for public transport is not great enough to fill public transport like the buses in the popular routes, many tourists prefer their private vehicles for the destination. The introduction of shared used bikes can help improve the public transport system. This paper explores the optimization and implementation of shared bike usage, by using a density-based clustering method and an allocation algorithm to locate the ideal spots for placement of the bicycle stands around the Lake district. The methodology depicts basic outline of the model, the finding locations section is for analyzing the locations of the attractions to the bus stops. This analysis helps in further cleaning of the data which is gathered.

3.1 Gathering Data

The Lake district is a vast area of highlands, mountains, and lakes. For this model design the places are focused on four of the major cities namely Kendal, Windermere, Ullswater and Furness. These cities have high tourist populations during the summers which is beneficial to the model. Using Google Maps (Googlemaps.com, 2023) the main attractions, hotels, guest houses and restaurants were identified for these cities. The data collected contains the name of the places along with the latitude and longitude of the locations which is then put into separate excel file for each city.

Using this method there are an estimated 87 bus stations all around the Lake district and around 117 hotels, restaurants, cafés, and attractions in the four major tourist areas in the Lake district. The places outside the four cities have not been taken into consideration thus eliminated from the files.

3.2 Calculating Distance

The Earth is divided into many sections by longitude and latitude across the surface, there are different ways to calculate the distance between two locations on the surface of the earth. One such way is Cartesian coordinates or Nothing and Easting method, this method is geographical Cartesian Coordinates of a point (Encyclopedia.pub; 2023 Hand Wiki). Northing and Easting is a grid reference which describes the location on a topographical map. The grid square are lines which are called Easting and Northing. The easting line runs from west to east while the Northing line runs from south to north on the earth’s surface.

Easting is a vertical line top to bottom dividing the map from west to east and the Nothing line is horizontal line from left to right dividing the map from north to south. They are numbered on the map margining at 1km interval (Hunter-ed.com; 2023).

A red line on a grid

Description automatically generated

*Figure 2: Northing and Easting Grid*

The above image depicts a zone calculation of the nothing and easting coordinates.

The UTM or Universal Transverse Mercator is a cartesian coordinate system which divides the earth into 60 zones each of 6-degree wide longitude with each zone representing a pair of (x, y) coordinates. Here the x point is the easting and y point is the northing coordinate, this is generally measured in meters. Each zone is about 500 km wide; they are unique within zone the United Kingdome is in zone 30 which is the UTM.

The UTM system is used in a wide area for mapping and surveying by military units throughout the world. This method provides an almost accurate and consistent way of locating positions on the earth’s surface (GIS Lounge. Com; 2023).

A map of the world with a location on it

Description automatically generated

*Figure 3: UTM of the United Kingdome*

The disadvantage of using northing and easting locations it may require specific map projections for accurate large scale projections of the area and the distance may become inaccurate while calculating for a larger area as it does not consider the curvature of the Earth.

Another way of calculating distance in terms of real road distance can be done in another method called Haversine distance. As the Earth is not flat, we use a mathematical formula which includes the curvature of the earth to give us the actual walkable distance. This formula is called the Haversine Distance, this distance is defined as the angular distance between two points on the surface of a sphere (scikit-learn.org; 2023). Haversine Distance is calculated by taking the latitude as the first point and the longitude as the next point given in radius. This gives the approximate distance between two points of the Earth’s surface with 1% error on average (community.esri.com; 2023). This method is used in finding and grouping locations. The formula below represents the haversine formula, this formula is used in the finding location section for analysis purposes.

A picture containing text, line, font, white

Description automatically generated

*Figure 4: Haversine formula*

3.3 Finding Locations

As part of pre analysis, finding the ideal locations in the four cities using the distance formula, here we group the accommodations, attractions and restaurants which are within 1.5 km from a bus stop. For each of these bus stops are set as the center location, the surrounding hotels, cafes, and attractions which are within the range of 1.5 km from them are found. This algorithm checks the distance using haversine distance between two points.

From the gathered data, using the python package ‘haversine’ the locations are grouped together. The package takes latitude and longitude of the place and calculates the distance between them, the longitude and latitude are grouped together using the ‘zip’ function in python which returns the zip object, which is an iterator of tuples where the first item of one tuple is grouped together with the first item of the second tuple, using the ‘list’ function we return a list for the coordinates. The reason for using haversine distance is because the function takes into account the curvature of the earth and gives the real road distance.

An additional method, after calculating the Haversine distance the data must be converted to real road distance. This is done by using a factor which was proposed by Love & Morris (Love, Morris; 1972), it includes a complex formula and set of parameters based on lp norm, to estimate road distance from a planar surface. As both these distance calculations methods have exceedingly small errors this factor can be utilized to determine the road distance. The “deviation factor” of a road network defined by Cole & King (Cole, King; 2016) is a ratio between the road distance and Haversine distance (paper). After calculation, this factor was found to be between 1.2 to 1.4 in regions from mountains and rivers being the reason for the large value factors (Boyaci et al; 2021).

After the coordinates are put in a list with the corresponding accommodation names, attraction name, bus stops and the restaurant names, using a defined function for each city they are grouped together for a given radius and the green dots represent the locations which are withing the radius and the blue represent the ones out of the radius. The function distance will calculate the distance between the two location points using the distance formula or package and give a figure of two decimal places is specified in the function (Ashutosh Bhardwaj, 2020). Using the ‘Folium’ package in python we can develop a map figure in open street maps (openstreetmap.org).

As for Ullswater shown in the figure below, the clusters are better defined. The representation of the clusters can clearly be seen, and the number of bicycle hubs is not as much as those of other cities, giving an advantage of minimal usage with maximum capacity.

A map with blue circles and black circles

Description automatically generated

*Figure 5: Analysis of Ullswater*

For places like Windemere shown in the figure below the clusters are located close to each other, as all the places are clustered together. From the analysis of these places, it can be noticed that Windemere requires a greater number of hubs compared to other cities and major tourist attractions are placed around Windemere. From this we can approximate the number of hubs needed for each city.

A map with blue pins and circles

Description automatically generated

*Figure 6: Analysis of Ullswater*

Based on further analysis it can be concluded that places at Windemere are placed closer together thus can be accessed by both bike as well as on foot, while in the other places a combination of bike and public transport can make the city more accessible. This analysis is done to obtain the ideal and outlier locations within the city and help in the formulation of the model.

Creating a design plan for the locations of the bicycle hub is important, this could impact the usage of the bicycle more efficiently. The reason for choosing attraction points as destinations for the bicycle hubs is because tourist when taking their private vehicle always need to find a parking place and that can cause a disruption, for both the tourist and the area management as they need to allocate certain area as car park. Keeping a bicycle hub in the area would be beneficial as bicycles are a onetime transport facility one time here means they can use it a single journey to or from a destination, they can leave their bicycle at the hub and go to the attraction spot, these hubs also take in less space compared to big parking lots. The concept of deciding spots for the hubs is that the tourist can cycle from the bus stations to the attraction point or anywhere else in the city, these facilities a multimodal transport system which is the aim of this paper.

1. MODEL

There are many methods to determine locations for bicycle hubs. Two of the most effective methods are set covering problems and random destination algorithm. One is an exact solution while the other is a heuristic method. Both methods serve the purpose of locating potential spots for the hubs to be located.

This model is built on previous work by Leon Cooper (1964) with additional clustering algorithm for allocation of destinations of the bicycle stand. The paper provided a range of different location allocation problems like destination subset algorithm, random destination algorithm, successive approximation algorithm, and alternate location and allocation algorithm. All the heuristic method proposed is to solve the location allocation problem, the reason for choosing random destination algorithm was it provided near optimum solution with less computational time. The successive approximation algorithm may not usually provide a good result but is rapid in computing. The alternate location and allocation algorithm is different from the other three methods, as it is monotone with the solution and the final convergence of the solution are a set of solutions which are not considered by the other methods. And finally, destination subset algorithm takes up a larger computational time compared to the other algorithms, as both the cost and distance is calculated for any and all point in the set which in becomes extensive for memory. Furthermore, it is not clear if the correct allocations are included in the final destination source locations.

The heuristic method of solving the location allocation problem is well suited for smaller points as it would be easy to examine the solutions of all the possible allocations and the minimum points, for bigger problems it would not be feasible.

4.1 Set Covering Problem

Set covering problem is a combinatorial optimization problem under the NP-hard category. In the problem there is a given set of elements in this case the collective locations with the coordinates of all the locations [restaurants, hotels, and bus stations] and a subset which includes the bus stations with its coordinates. The set covering problem has many applications such as picking optimal locations for cell towers which gives the maximum coverage of the customers, other such applications are selections of portfolio (Liang et al; 2020), vehicle routing. This approach has several greedy approximation algorithms which can be used to understand computational algorithms. This model can be solved using Integer linear programming formula which minimizes the set covering problem. The disadvantage of using set covering for location allocation problem is that it depends on linear approximation which might not capture the real world cost accurately. It has a scalability issue, but this model is done on a small scale thus it would not cause a problem. The set covering problem focuses on its primary objective of cost minimization while ensuring maximum coverage, adding other objective functions would cause a conflict between the objective functions. The np hard complexity of the problem throws a challenge while finding the optimal solutions especially for large instances, the search space is growing exponentially. Despite having advantages, set covering problems is widely used in facility location, network design and scheduling problems.

4.2 Random Destination Algorithm

The heuristic method is an approach to find a quick and reasonable solution to a problem within a reasonable time frame. Heuristics make few or no assumptions about the problem to get the optimal or semi optimal solution with reasonable computational cost (Dantzig; 1963 ). The location problem starts with an initial solution after which the neighbors are explored of the current solution and then it will be replaced with the best of the two solutions, thus producing a semi optimal or optimal solution with regards to nodes of a given weighted graph (Singh; 2020).

The baseline for the heuristic location allocation method is that “for a set of n destinations and m sources, the location of the source is known the determination of the optimal solution is trivial with the cost to be minimal” (Cooper; 1964)

Random Destination Algorithm is a type of heuristic method proposed by Leon Cooper (1964) among the many other location optimization methods, the objective of the algorithm involves finding the optimal allocations of sources to the destinations with parameters of cost or distance. The method generates random allocations of sources to destinations which then terminates the generation when a solution is within a certain number of the standard deviation from the mean cost, considering the distribution of the left side tail to be a normal distribution to terminate the generation.

The algorithm allocates destinations as sources which are picked at random, checking the cost or distances from other destinations to either changed or not changed.

This algorithm is run multiple times to find the optimum spots of the destination. This produces a subset of the destination from the initial set. This method is also used in scheduling problems. Among the many heuristic methods developed by Leon Cooper (1964), the random destination algorithm gives a close approximation to the optimum solution with reasonable computation time.

This type of heuristic algorithm works by simultaneously locating and allocating by letting the destinations represent themselves as the sources as a set of sink locations which are generated randomly for m set of locations (Warsan, Jesse; 1970).

4.3 Assumptions

General assumptions,

The distance a person is willing to walk is approximately 1.5 km from any given place.

The candidate spots are within the cities namely Windemere, Kendal, Ullswater and Furness.

The model clusters the destination points.

The distance is calculated on a 2-dimensional plane.

There is no weight for the locations, but the algorithm considers the minimum distance between two points.

Cost of installing and maintaining the bicycle stand is not considered,

The number of bicycles per stand depends on the demand close to the attraction points.

The initial parameters can be changed based on preferences and usage of the model.

4.4 Model Formulation

Before going ahead with the heuristic model, the data collected in longitudinal and latitudinal coordinates should be converted into northing and easting coordinates. This can be done using a simple python package or using a web program like grid finder to change the coordinates (Grid Finder. com). The advantage of using northing and easting distance calculations is easier to use in a mathematical modeling using the Euclidean geometry, it is suitable for relatively small scale geographical areas. The only disadvantage of this method is that the curvature of the earth is not considered. This can be accounted for by using haversine distance calculation.

The model consists of two parts, namely the random destination algorithm and the DBSCAN algorithm.

The Copper’s algorithm of random destination allocation (Cooper; 1964) generates m subset of fixed points and solves each of those points using the same methodology of a single facility allocation problem. This random destination algorithm is used as an extension to the single facility weber problem of location allocation. The algorithm converges within the subset of destinations picked optimally, each of these fixed points is relocated to the nearest facility which can be the initial location assigned or a new location.

After going through the iterative process and all the locations are allocated, it is optimized to the best locations. Here the process of alternating allocation and location phase is created, and it terminates until no further improvements can be found or solved.

The other element which plays a role in building the model is a clustering method which is known as Density Based Clustering with noise. Clustering is a method of grouping data points with similarities. The aim of this clustering method is that a point belongs to a cluster if it is close to other points from the cluster. The two main parameters for the clustering method are ‘eps’ or epsilon which is the distance specified by the neighborhoods. If the two points are neighbors, the distance between them is the eps value. In this model the eps value is 1500. The next parameter is ‘minPts’ which is the minimum number of data points to define as a cluster. This can be modified as per usage of the model. DB Scan works well for scenarios where the clusters have an irregular shape and have persistent noise point. DB Scan has three-point clusters which are namely the core points, boarder point and the noise point. The core point is the point which has the minimum number of other data points with the specified distance from it, the boarder points are at the edge of the cluster which means that they are not close enough to any other cluster, and finally the noise points which are not part of any cluster groups, but they are a cluster as themselves. Density based clustering provides a flexibility in forming the cluster shapes, unlike k means clustering the number of clusters need not be specified, while the clustering method is robust to the variations of the clusters it makes minimal assumptions. The key point of this clustering method is that it can handle large data which is noisy and detect anomaly in the noisy data. While the clustering method is better than k means clustering for this case, density based clustering has its drawbacks. It might struggle with data with varying density in a higher dimension and struggle with the neighborhood size.

The basic principle of this model is that the attractions are clustered together within a certain distance from each other using DB Scan, the clusters along with the bus station points are given to the heuristic algorithm which locates the bicycle stands with reference to the bus stop. The algorithm optimizes the best or near best location within each cluster. The points that are considered as noise points are a cluster or can be eliminated as a defined point for the bicycle stand. This allocation algorithm treats each cluster as a set of locations and finds the optimum solution locations within each of these cluster locations.

The model is formed of 4 functions namely the distance calculation, cost calculation, heuristic random generator, and the main function.

The following represents a pseudocode of the heuristic algorithm,

For this method Nothing and Easting is considered as it is simple to solve and get the distance between two points. The number of sources, destination, iteration, and tolerance can be altered according to requirements based on the problem.

Function distance Calculation (point 1, point 2):

Northing\_difference = point 2 [0] – point 1 [0]

Easting\_difference = point 2 [1] – point 1 [1]

Returns square root (Northing\_difference ^2 + Easting\_difference ^2)

The first function distance calculation calculates the difference between two northing and easting values from the data and returns a list of the differenced value to the next function.

Function cost Calculation (location, destination):

totalCost = 0

For source in allocation:

Minimum\_distance = infinite

For destination\_point in destination:

Distance = distance\_calculation (source, destination\_point)

If distance < minimum\_distance:

minimum\_distance = distance

totalCost += min\_distance

returns totalCost.

The cost calculation function assigns cost to each link between two points which is the distance between the two points, after checking the shortest distance between any two points. Starting the minimum distance at infinity and if the value is lower it is assigned as the new cost between the link of two points and finally returns the cost to the function. As the locations are not weighted the distance between the points are considered. The minimum distance serves as the cost function.

Function Random Generator (a, b, iterations, tolerance, destinations):

Mean\_cost = 100

St\_deviation = 20

TotalLocation = combine (a, b)

BestLocation = []

For \_ in range (iterations):

locations = random. sample (destination, b)

Cost = cost\_calculation (location, destination)

If absolute (Cost – MeanCost) <= tolerance \* St\_deviation:

BestLocation. append ((location, Cost))

Returns BestLocation

The important function of random generator allocates the sources randomly to their destination creating a list of the best allocations which is returned to the main function. The standard deviation and tolerance are used as the stop point for generator of the subset.

The main function takes the input of the destination values as Northing and Easting coordinates, finally displaying the top 5 best solutions with their respective cost values. The solutions are based on heuristic optimization of the data with randomness of the source and destination allocation. The three parameters of the main function are the heuristic parameter, BDSCAN parameter and finally the clustering parameter. The function will return both the subset of the location in a northing and easting format and a cluster graph. With the cluster component the locations are cluster number zero to the number specified and the -1 cluster is the noise points.

Function Main ():

Inputs:

df = excel file with coordinates

number of Destination

number of Source

number of iterations

tolerance level

destination = list of coordinates

Assert number destination <= length (destination), “Error ! Number of sources should be smaller than number of destinations.”

# DB Parameters

dbScan = DBSCAN(eps=1500, minSamp = 5)

clusterLabs = dbScan. Fit predict(destinations)

# Clustering method

ClusterPoints= {}

For ClusterLabs, point in zip(ClusterLabs, destinations):

If ClusterLabs not in ClusterPoints:

ClusterPoints[ ClusterLabs ] = []

ClusterPoints [ ClusterLabs] . append (point)

# Visualization

For each ClusterLabs, ClusterDes in ClusterPoints. items () :

x\_value = [point [0] for point in ClusterDes]

y\_value = [point [1] for point in ClusterDes]

plot the graph ( x\_value, y\_value, label = f ‘ Cluster {ClusterLabs}’)

BestLocation = []

# Optimization

For ClusterLabs, ClusterDes in ClusterPoints . item ()

If length (ClusterDes) >= numberDestination:

ClusterLocations = Random generator ( length ( ClusterDes), NumberDestination, iterations, tolerance, ClusterDess)

BestLocation. extend (clusterLocation )

BestLocation. sort (key= lambda x: x [1])

For i, (allocation. cost) in enumerate (BestLocation [:3]):

Print ( f “Solution {i + 1}: Allocation = {allocation})

The clustering method is based on DBSCAN, which is a method of density-based clustering. The points are clustered as specified by the two parameters, the cluster includes core points and the border points. This clustering algorithm has a radius, if x is a point which is assumed to be a core point that is selected as an arbitrary data point from the dataset, after all the core points are identified by the algorithm it then goes to identify the boarder points by checking the points that is within epsilon distance from the core point but no other core points themselves, finally the left out points are considered as noise or outlier points. Each of these clusters formed contains at least one core point and the other points can be made of non-core point which also forms the edge of the cluster which justifies the other points which cannot be reached from that point making it a part of other cluster or an outlier/ noise point. The destinations are optimized in the random generator function and the best solutions or locations are returned and displayed. For the simplicity of the model here the package DBSCAN is used instead of the actual algorithm. The greater the available data the computational time for the clustering algorithm could consume time and memory of the device, this is the disadvantage of DB SCAN.

The code can be modified based on the number of solutions and sources as preferred. Since the result will be returned in northing and easting form it can be converted back to longitudinal and latitudinal value with respective.

4.5 Limitation

As for limitations, the geography of the area is not considered, the lake district consists of hill areas thus making cycling a bit hard for people who are not used to the elevation level. The weather can also play a role in the case of cycling as it might be dangerous for tourists who do not have a good experience cycling in the rain or wind. The elevation can be overcome by introducing electric bicycles or electric scooters which would make it much more efficient for all the tourists to use. Most tourists visit the lake district during the summer season, which might not be a problem when considering cycling in bad weather.

There are many places in the lake district which are in remote areas and accessibility by public transport will be more efficient and faster. For the model these places are not considered. A better estimate can be calculated by using the reuse rate of the bicycle and developing a redistribution strategy for the bicycle.

1. EXPERIMENTS

To obtain the ideal locations the parameters can be changed to have minimum noise points and maximum coverage with each cluster. The model is set to give different optimized solutions depending on the clustering parameters. The two main parameters are the number of points which are needed to be in a cluster and the other parameter is the distance between the points. For the optimization part of the algorithm, changing the tolerance level would show a difference in optimization of the solution. The tolerance level, the range allowable for a location to be considered a potential solution, this also acts as a trade between the computational time and the quality of the solution. A good range for the tolerance parameter is 10 as it gives a solution subset, any other number might not result in an optimum subset. The smaller the tolerance level the closer the solution is too optimal but there is a fine line between getting an optimal solution and a non feasible solution due to the clustering algorithm. The higher the tolerance level the solution may be semi optimal but faster computational time, to get an optimal solution using a higher tolerance level a greater number of iterations are required which could cost memory from the device.

For example, the parameter is set to 2000 as distance, 2 points per cluster which is the minimum number of points for a cluster and the tolerance of 10. The results for these conditions are,

A graph with many colored dots

Description automatically generated

*Figure 7: Cluster of experiment 1*

The image above shows the result of the clustering where the cluster -1 are noise points which are out of the boarder points, and the table below are the new set of locations for the bicycle hubs. All the places listed below are in the city of Kendal in the lake district. This is due to change is distance parameter from the initial analysis. It was noticed that the attractions in Kendal are further away from each other compared to other cities as Kendal is a bigger city.

|  |
| --- |
| Name of the locations |
| Blackhall Yard (no 67) |
| New Shambles |
| Premier Inn Kendal Central |
| Kendal Museum |
| Castle Green Hotel |
| Westmorland Shopping Centre |
| The Cottage Kitchen |
| Prince Charlie's House |
| Lound Road, Netherfield |
| Method, 37, Stramongate |
| *Table 1: Subset from experiment 1* |

The second example, decreasing the distance to 1000 which is the minimum distance for a person to cycle, number of points to 15 and finally the tolerance to 10. The results for this example are,

A graph with blue dots and white text

Description automatically generated

*Figure 8: Results of experiment 2*

The image above shows the cluster where the -1 cluster are the noise points, there are many noise points as most of the places are not within 1 km from each other. Cluster 0 is present mainly in the city of Windemere as seen since it is one of the most frequently visited cities by visitors. The table below gives the name of the hub locations. Most of these locations are in Windemere, this is because they are all placed in clusters. The algorithm finds locations within the cluster group as we have only one group and all the locations are placed within the same cluster which is the city of Windemere.

|  |
| --- |
| Name of the locations |
| Cedar Manor |
| Peter Rabbit And Friends |
| The Westmorland Inn |
| The Belsfield |
| The Old England |
| The Cottage |
| Stags Head Hotel |
| Jackson & Graham |
| Storrs Gate House |
| Holly Lodge |
|  |

Table 2: Subset of experiment 2

Example 3, where both the distance and number of points are set to max limit, which is 5000 and 15 respectively, this would give us a result where the cities are in cluster together and the points which are outside the city (remote attractions) are the noise points.

A graph with many colored dots

Description automatically generated

*Figure 9: result of experiment 3*

The table below shows the locations for the bicycle hubs for the above conditions. All the locations below are within the clusters in the cities close to a bus station.

|  |
| --- |
| Name of the locations |
| Hartley's, 51, Market Street |
| Lakeland Motor Museum |
| Abbey mill coffee shop |
| Museum Gift Shop |
| Thyme Bistro |
| Fell Foot |
| Furness General Hospital |
| Derwent Pencil Museum |
| Fins Thwaite Lane |
| Holly Lodge |

*Table 3: Subset of experiment 3*

After conduction the experiments with the parameters set to their maximum and minimum limits it can be found that, the parameters are set to 10 points per cluster and a distance of 1500 which is the distance between two points assuming 1500 m is the distance a person is willing to cycle to reach a destination. By changing the parameters, the model finds different solutions which can be optimal or close to optimal results.

This brings the topic of multi mobility which is the combination of different modes of transportation to complete a journey. This can be demonstrated by a small experiment where a simulation model is created which shows the journey line of the transport system.

The concept of multimodality is not new, the basic principle of multimodality in public transport is using different public transport vehicles to get to a destination. Multimodality can be faster compared to just pure public transport, here cycling is included as a part of the journey, and this reduces the waiting time for the bus or any other scheduled vehicle.

Simulation models are useful to assess what scenario is for a previously built model, they provide analysis on the optimality and risk of the model built. Simulation models provide continuous improvement of the model. From previous works simulation models are used on network model to access the efficiency of the proposed model, from (Jappinen et al; 2013) this simulation model is developed. This model shows the connectivity between the different modes of public transport. Simulation models are powerful tools for analysis. They provide a controlled environment to test out the scenarios in different settings. A simulation model can be further used in training and education in the area for network transit simulations.

Let us imagine a scenario where an individual has to travel to another city by train, to go to the train station the individual has to travel by bus first which is a long walk from the house, in order to not miss the bus the person decides to take a bicycle from the nearest bicycle hub which is relatively faster compared to walking, cycling to the bus stop and returning the bicycle to the hub near the stop makes the individual’s journey smoother and quicker. This scenario demonstrates multi mobility where there is a combination of different transport methods to reach a destination. This is common for people who use public transport as a main method of travel to daily work. To conquer the last mile problem bicycles can be implemented. Here we see a combination of public transportation the last or first mile is where many individuals choose to walk or choose an alternative, especially when using this public transport. This can be seen demonstrated in (Jappinen et al; 2013).

A bus and bicycle with arrows

Description automatically generated

*Figure 10: Simulation Scenario representation*

The image above shows the scenario of travelling using different modes of transport to reach a certain destination.

By understanding the above scenario, this paper discusses a similar situation in the Lake district were using bicycle for the first or last mile of a journey to or from the bus station or bus stops from or the attraction points in the cities. The table below shows such a scenario where cycling reduces the waiting time and total travel time from destination to destination. Cycling is preferred over walking only if there is a bicycle hub close by.

A simple simulation is constructed to demonstrate this example where the two inputs are the locations of the bus stations/ attractions and the location of the bicycle hubs. With predetermined speeds for walking, bus and cycling. The first function generates a simulation environment by choosing the origin, destination, and arrival rate of the tourist at random. The second function calculates the distance in kilometers between the bus stop or attraction to another bus stop/attraction. The third function is written to check if there is a bicycle hub close to the location of the bus stop or attraction if available the bicycle option is chosen else walking is chosen. For the purpose of demonstration there are three available bus stops or attractions and two bicycle stands close to the other place. The simulation runs 100 times which can be changed as per need. The next function calculates the time taken to reach the destination by the two modes of transport which are bus along with bicycle and bus along with walking. After the time is calculated the total sum of the times is calculated and displayed in a table form.

The simulation runs in these steps, first defining the parameters then calculating the time travelled, simulating the bus journey, simulating the walking journey, simulating the cycling journey, comparing the results finally analyzing and repeating.

The arrival rate of the tourist is set as an exponential rate and the choice of the initial attraction station is chosen at random. The simulation calculates the stright line distance, but other modes of distances can be used like the haversine distance mentioned above. The cycling and walking time is set to uniform distribution, where the average cycling time is assumed to be 3 minutes and the average walking time for the same distance is assumed to be 10 minutes. The distribution time can be changed according to the usage of the simulation. The simulation can be done by using geocoding and simulation pages in python and this returns the passenger id and time travelled in each journey. The conclusions can be made by looking at the table below,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Passenger ID | Bus time (min) | Cycle time (min) | Walk time (min) | Total time (min) |
| 1 | 14.70 | 0 | 23.13 | 37.83 |
| 2 | 12.85 | 0 | 24.79 | 37.65 |
| 3 | 5.87 | 8.66 | 0 | 14.54 |
| 4 | 9.07599 | 5.46 | 0 | 14.55 |
| 5 | 7.60 | 0 | 22.42 | 30.02 |
| 6 | 12.76 | 5.72 | 0 | 18.48 |
| 7 | 6.82 | 0 | 21.31 | 28.13 |
| 8 | 5.46 | 0 | 19.85 | 25. 31 |
| 9 | 11.37 | 0 | 21.41 | 32.78 |
| 10 | 7.41 | 3.22 | 0 | 10.63 |

*Table 4: Results of simulation experiment*

From the table it is proven that the combination of bus and bicycle is faster than walking and saves a lot of time for travelling. This example shows the importance of multimodality in a public transport network, the waiting time and the journey time is reduced which gives the user a better and efficient way to travel in daily life or to a new destination. Bicycle hubs placed at the right location can increase the maximum coverage in the group of attraction points with minimum distance to a bus stop or other attraction points. The journey can be enjoyed by the tourist to the fullest with the breath-taking view in the Lake district.

The model in this paper can be improved by gathering data from public transport data which would include the arrival of the bus to a particular bus stop, this would prove a better understanding of the journey from the origin to a destination and the useability of the bicycle during the journey.

1. RESULT AND DISSCUSIONS

After conducting experiments with different parameters ranging from the minimum number of points per cluster to the minimum distancer an individual is willing to cycle, the results of the proposed model with 117+87 destination points and bus stops from the four cities in the Lake district and allocation of 10 sources. Each of these locations represent at least one location in the clusters which are formed by the clustering method, that would require a bicycle hub and the top hub locations are as follows:

|  |  |  |
| --- | --- | --- |
| **Name** | **Latitude** | **Longitude** |
|  |  |  |
| The Lamplighter | 54. 380065 | -2. 9060954 |
| Windermere Jetty | 54.370807 | -2. 9219165 |
| Windermere Guest House and Hostel | 54.380687 | -2.9061101 |
| Beech Hill Hotel & Spa | 54.321226 | -2.9402782 |
| Stott Park Bobbin Mill | 54.285446 | -2.9656426 |
| Masons Arms | 54.297637 | -2.9031735 |
| Applegarth Villa | 54.379253 | -2.9074039 |
| The World of Beatrix Potter Attraction | 54.365404 | -2.9190698 |
| Stage Head Hotel | 54.3641894 | -2.92153461 |
| Storrs Hall Hotel | 54.3391272 | -2.93569584 |
|  |  |  |

Table 5: Subset of the optimal locations

The above table shows all the ten locations in the four cities for potential hub location. This result was obtained after converting the results from the model converting back to longitude and latitude value using a simple program and package in python and getting their respective names. The conversion can also be done on any web program such as Grid Reference (Grid Finer.com) which allows the entry of the X and Y coordinates and points to the location on the map.

The final cluster result can be visualized using a few methods either on a map using packages on python or as a graph with clusters with nothing and easting as the x and y coordinates. The image below shows the clustering of points in a form a graph. Cluster -1 represents the noise points which cannot be groups as the minimum points to form a cluster is ten thus these points are not in the cluster formation as there are no other points close by to make a core point. The noise points can be reduced by adjusting the parameters of the clustering methodology, which is demonstrated in the experiment section of the discussion. The noise points do not have a bicycle stand and there are no bus stops close by the points. They are mostly the remote locations in the lake district.

A graph with many dots

Description automatically generated with medium confidence

*Figure 11: Cluster of the optimal solution*

The combination of the heuristic approach with the clustering method satisfies the constrains for a smaller geographical area, the location of the bus stops and the attractions together are scattered around the city and few resorts are located further away from the main bus route or bus station. The last or first mile journey from these resorts could be completed by using a bicycle.

The main contribution of this model is to provide an optimized solution of subsets of the attraction points, each location in the subset is from one cluster with is formed using the density based clustering method. This heuristic approach with the clustering method provides locations which have maximum accessibility with minimum cost, here since there is no weight to each location or cluster the cost is determined by the distance between each point.

As per previous works the demand and station capacity are unknown. This results in the assumption of the bicycle fleet. The public can pick up a bicycle from any location and drop off at any other location, this is the basic principle for the bicycle sharing system in the proposed model. The demand depends on the attraction points. For example, at the bus stops or train stations the number of bicycles can be increased and for other attraction points there could be fewer bicycles on the basis that the individual would use the same bicycle if available again for the day. As the lake district attracts more tourists during the summer time (in May – August) the demand can be adjusted to the needs of the season. The demand is not fixed for each of these cities, Windemere being the most visited city and a bigger city in size comparison has a greater demand compared to Furness or Ullswater. Further development to the model can be made by adding a capacity and demand variable, these variables result in a real time solution to the private vehicle issue. The concept of shared used vehicles can be expanded throughout the Lake district, improving the public transport system, and implementing the shared used vehicle concept the Lake district could find itself reducing the number of private cars used by visitors and locals alike to explore the destination along with reduction of carbon emission and preserving the Lake district with all of its beauty.

Planners of the bicycle sharing system should decide on the allocation of the available resources and demands which would prompt the sharing system to be utilized to its maximum and provide the users with the right mobility for transportation. Shared bicycle provides a connection to improve the public transportation system by reducing the overall waiting time and the travel time (Nair, Miller-Hooks;2016).

1. CONCLUSION

Tourism and travel are interconnected, with leisure travel comes traffic which has been on a constant rise, the Lake district being no exception. Reduction of carbon emission due to seasonal tourism is a problem for rural protected areas like the Lake district. To curb the use of private vehicles in the area the concept of shared used vehicle is introduced. Shared used vehicles, mainly bicycles, in the Lake district can be beneficial in terms of cost effectiveness, control of traffic congestion, reduction in carbon emission and sustainability.

The lake district has public transport which runs in lesser frequency on the main lines which increases the waiting time for the tourist, which is not efficient, by using shared used bicycles the waiting time can be decreased or make the public transport more efficient. Bicycle tourism is gaining popularity and implemented in many European counties.

Gathering and converting the data into northing and easting coordinates to find the distance between two locations which is defined as the cost for this model.

The goal of the proposed model is to identify the locations for the bicycle stand which has a minimum distance from the attractions and has maximum coverage of the attractions making it accessible to the public. The proposed model has two parts, the clustering algorithm, and the heuristic method for optimizing the ideal or close to ideal locations within the clusters for the location of the bicycle stand. The heuristic algorithm gives the closest to optimal solution to the allocation problem, instead of exhaustively examining all possible allocations this method is used. The Density based clustering or DBSCAN algorithm clusters the attraction points together based on parameters given by the user. The objective of the heuristic method is to find the optimal allocations of sources to destination with minimum cost factor. The final solution is a subset of the attraction points where the bicycle hubs can be placed to make transport easy and accessible to the public. The attraction points are a set of all the restaurants, cafes, and other attraction points from the four most visited cities in the Lake district.

The results of the model show a subset of locations which are allocated as the sources to the destination around it. There is at least one location from each of the clusters formed with the clustering method. The subset of locations are the optimal spots for placing a cycle stand within in the area which has minimum distance from a bus stop as well as other locations with maximal coverage of the area.

Through a simple simulation model with an example, it can be concluded that the integration of a shared vehicle with the public transport journey providing a multimodal system saves travel time in a journey along with promoting a healthy lifestyle. As more individuals start using public transport for their daily use it reduces their carbon footprint creating a sustainable environment.

Further analysis can be done by considering the elevation levels in the geographical area of the lake district. Electric bicycles or electric scooters are the best alternative to normal geared bicycles, they serve the same purpose of sustaining the shared used vehicles. There can be charging stations as they have a battery life limit in the place of the attraction points which would serve the same purpose as the normal bicycle hub.

Public transport intertwined with shared used vehicles like the bicycle can solve the last and first mile problem, they are flexible, reliable, sustainable, and less expensive. The ongoing demand for this mode of transportation can transform the region into a much greener and safer destination for the public to visit. With integration of both the public and private sectors of the transportation department could make the public transport a most cost-effective change which would ultimately reduce the use of private vehicles not only in rural areas but cities as well.

Cycle tourism is a growing business, cities in the European region have well developed laws and infrastructure to incorporate bicycles into their daily lives. This model could help rural areas such as the Lake district to include bicycles into their growing tourism demands, given the investment in infrastructure of the cycling lanes and public transport this model could serve its purpose of allocating the optimum locations to place bicycle hubs to maximize the efficiency of tourism and protect the environment.

With more study and importance to electric bicycle and scooter the increase of demand of this mode of transport could be beneficial to both the government and the environment. By increasing the safety standards and making a few other modifications to the electric bicycles, a good fleet of them can be distributed around the city to help tourists to have a better holiday experience. These docking stations can be placed around the cities near offices, bus stations, train stations or attraction points for easy access along with rechargeable features. The same methodology with constraints for electronic features can be used to locate the various docking stations for the lake district region.

REFERENCES