Masters Programmes

Assignment Cover Sheet

Submitted by:	2293758, 2285143, 2217946, 2282772, 2215107, 2288766, 1895583
Date Sent:	15th March 2023
Module Title:	Advanced Analytics Models and Applications
Module Code:	IB9190
Date/Year of Module:	2022/23
Submission Deadline:	16th March 2023
Word Count:	2000
Number of Pages:	4 pages
Question:	Investigate the customer allocation problem

Contents

Section 1

The optimization problem presented in the article is formulated as a mixed-integer linear programming problem. It is depicted in appendix 2.

The decision variables in the formulation represent the allocation of customers to facilities. Specifically, a binary variable X_{ij} is used to indicate whether customer i is allocated to facility j. Variable d_{ij} represents the driving time between customer i and facility location j. Variable h_i refers to customer *i's* demand in terms of number of required visits per year. a_{ij} represents the current allocation of customers to facilities. R_i is a binary variable which denotes whether a customer is reallocated to another facility or not.

The objective function of the problem is to minimize the total distance travelled by all customers, which is a measure of time, cost, and carbon dioxide emissions. The distance travelled by each customer is calculated based on the geographical coordinates of the customer and the facility, which are required parameters for the problem.

The problem is subject to a set of constraints, which ensure that the allocation of customers to facilities is feasible. Specifically, the model includes a capacity constraint for each facility, which limits the total number of customer visits that can be allocated to that facility. Additionally, the model includes a set of binary constraints that ensure that each customer is allocated to exactly one facility. Furthermore, there is an additional limitation in place to guarantee that a customer will not be reallocated to a facility that is more distant than their existing facility and is only reallocated when reallocation savings exceed reallocation cost T where R_i is the binary reallocation variable that indicates whether customer i is reallocated. If the customer is reallocated, $a_{ij} = X_{ij}$ and as such, $R_i = 1$.

The model first assumes that the number of facilities is fixed and doesn't take into consideration potential mergers, acquisitions, or closures. We also assume that the capacity for each facility is known and fixed. The model also does not take into consideration the nature of the relationship between a customer and a facility and so it is assumed that customers will always be willing to be reallocated if the facility is closer by. Furthermore, reallocation costs in this study are tangible and independent of the size of the customer.

Section 2

Construction heuristic

The construction heuristic assigns the nearest facility to each customer as an initial solution. If a facility is not available, the algorithm identifies the next closest one and assigns until all customers are assigned. Four versions of the construction heuristic were used: sequential, random, savings regret, and randomized adaptive greedy adding.

The sequential heuristic assigns customers to facilities sequentially and chooses the nearest facility for each customer and updates the facility's remaining capacity. While it is relatively quick, it is the worst-performing algorithm because it gets stuck in a local optimum. On the other hand, the random heuristic goes through the list of customers randomly and the savings regret heuristic sorts the customers in descending order based on their demand. The former one takes longest time and performs best while the other one is as quick as the sequential and has a better performance than the sequential.

The randomized adaptive heuristic finds a balance between the random and the savings regret versions by creating a shortlist of the largest demanding customers and randomly selecting from that list. The heuristic is helpful when assigning larger customers to specialized facilities, and its randomness prevents it from getting stuck in local optimum. For problems like Van Dorp's, it should theoretically be the best option, however, it ranking between than random and savings regret versions in terms of setup, runtime, and performance, according to the authors. However, it’s easily to fill up the nearest facility and result in many customers not being able to be allocated optimally. As such, random algorithms outperform random adaptive version in the real world as identified by author.

Improvement heuristic

The authors proposed three versions of improvement heuristics, each of which starts with the initial allocation given by Van Dorp.

The first version is the "first" improvement heuristic (FI). This version aims to improve the initial solution by reassigning customers to the first facilities that realised reduction in total cost, while also having randomness to avoid getting stuck in the local optimum. The algorithm begins by randomly selected customer and assess whether the current cost can be improved, given that cost saving exceed relocation cost (Appendix 3). This process is repeated until no further improvement is possible, and the algorithm will going through list multiple times. Although the process is time-consuming (~ 15 minutes for this problem), it finds the lowest value of the objective function.

The second version is the "best" improvement which only reallocates customers to the closest available facilities. This discourage reallocation when closest facilities are full, as a result, it’s the fastest improvement heuristic (~ 7 minutes) with fewest iterations. However, it prone to get stuck in a local optimum and may not allocate customers to better facilities even if they are available. As it does not outperform the first improvement heuristic and is more complex, it is not recommended for this problem. Similarly, the third version is "one-opt", which assigns customers to any better available facilities and compares the objective function values to determine whether to reallocate the customer. This causing sub-optimum allocation while also failed to improve over the “best” version.

Overall, the FI is recommended as it yield the best result with reasonable running time, in the case for Van Dorp. Alternatively, if the business objective is to avoid customer complain about not reallocating to the best facility, then best improvement heuristic should be more appropriate but at the risk of getting stuck in local optimum.

Section 3

To initiate the set-up, we first prepared the input by downloading the population data by Local Authorities (LA) for 2021. To obtain geographical data, we downloaded data containing area coordinates for 2022 and matched both datasets by using area code to filter out non-LA coordinates, which resulted in 374 LAs. As the service provider only operates in Great Britain, we removed 11 LAs located in Northern Ireland and resulted in final 363 LAs to be allocated. In order to find coordinates for the 13 facilities, we manually searched LAs based on their location and then added corresponding geographical information. As a result, the input is stored in two sheets of an excel file.

To choose the appropriate constructive heuristic, we first need to look at the demand vs. supply situation for the company. According to the distribution of customers in Appendix 4, we found that there was a concentration in the south-east of Great Britain. If random adaptive algorithm were chosen, there would be 61 high demand LAs that need to be allocated to further facility. This problem is similar to ‘filling up too quick’ situation in the paper where Random version is better than Random Adaptive. Hence, we decided to use Greedy Random heuristic.

To implement heuristics in Python, we read the data from the excel file and added 'remain' and 'demand' columns to the facilities and customer sheet respectively. The initial value of 'remain' is the same as the capacity and the 'demand' value is calculated by population multiplied by 0.1%. Then we prepared distance-cost matrix, on a one-way travel basis, by calculating a 363*13 matrix based on the latitude and longitude data. The data then converted from km to hours, assuming 60km/h vehicle speed. After performing a quasi-random distribution on 363 customers, we sorted the distance-cost matrix in the order of smallest to largest and assessed whether the customer demand can be fulfilled by the 'remain' of the closest facility. This iteration will stop when available facility is found, and the customer will be allocated accordingly. The value of 'remain' is then updated to the original value minus the demand. The loop will run until all demand has been satisfied. We ended up with an allocation matrix of 363*13, which is a zero-one matrix. To evaluate the result, we calculated the sum of the allocated distance multiplied by the customer demand and it resulted in total time of 64,196 hours.

Section 4

Improvement with OM and FI

As in the article, we tried to improve the allocation by applying the optimisation formulation (OM) as well as the First Improvement heuristic (FI) based on the previous result. The solution from Greedy Random (GR) was recorded as a 363*13 binary matrix, with consumers i in rows and facilities j in columns. Then we assumed that the GR result acted as the initial solution to develop both OM and FI. Apart from the input we prepared in Section 3, we proposed to add 2 additional parameters to the optimisation model: the capacity relaxation (representing flexible workload) and the reallocation cost (in terms of time), similar to the paper.

To compare the solution quality of two methods, we implemented the OM and FI, in a similar way as Section1 and Section2, under different facility capacity relaxation levels. The resulting table Appendix 5 shows the one-way costs when the reallocation cost is 0. It also shows the number of reallocations required to achieve the solution. We also did a sensitivity analysis on how reallocation costs could potentially affect the cost; however, the analysis only shows insignificant increased in cost even at 10 hours, hence we dropped reallocation cost from the analysis. It is worth noting that by removing the reallocation constraints ([5] and [6]) from the OM model, the solution can reach the global optimum by constructed the best possible allocation, which is 55,971 hours.

The results show that both methods failed to realise improvement at 0% relaxation. The reason why we cannot improve the cost with FI is the same as the constraints [5] and [6] in OM. This is because customers are unable to be reallocated to another facility, without breaking the constraints, since the facility already filled-up with sub-optimal customers. As a result, both methods are stuck in a local minimum and cannot find improvement. OM with constraints started to reach the optimal solution only at relaxation level of around 50%. Given that service type is similar to the paper and capacity relaxation in the paper is feasible up to 10%, OM with constraints able to reduce total cost by 8.5%, however, it's still 11.2% worse than the optimum level. Comparatively, FI performs 3.3% worse than OM at the same relaxation. Though the optimal solution might result in further cost savings, it required far more customer reallocations than two proposed methods, which also include reallocating customers to further facilities.

Conclusion

Similar to the paper, the result shows that FI failed to improve initial allocation at capacity relaxation 0% and only slight improvement when relaxed. However, it's surprising that OM with reallocation constraint could not improve the initial allocation at 0% as well. This can be explained by the demand vs. supply situation as described in Section 3, making it's impossible to improve allocation without moving some customer to a further facility. We also can be sure that this is not because RA allocation is at the global optimum, since the result from OM without constraint suggest that the best possible construction has lower total cost. Based on these results, we would recommend Prodnav Ltd. to opt for OM model to construct the best possible allocation. This is because the run time of OM and heuristics are indifferent for this problem size, and improving sub-optimal initial allocation would be impossible without moving customer to the further facility. Nonetheless, in the case where improving heuristic is the only option, we would recommend adjusting the algorithm by opting to start from the lowest capacity facility and assigning the closest customer instead to address demand vs. supply situation in Appendix 4. The result in Appendix 6 shows that this new proposed algorithm would outperform FI from the paper when capacity is relaxed, since facilities are no longer prone to filling-up quickly.

References

Haket, C., van der Rhee, B. and de Swart, J., 2020. Saving time and money and reducing carbon dioxide emissions by efficiently allocating customers. INFORMS journal on applied analytics, 50(3), pp.153-165.

ONS (2021) Estimates of the population for the UK, England, Wales, Scotland and Northern Ireland. Office for National Statistics. [Online], Accessed on March 1,2023 Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration /populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland

ONS (2023) Local Authority Districts (December 2022) names and codes in the United Kingdom, Open Geography Portal. [Online], Accessed on March 1,2023 Available at: https://geoportal.statistics.gov.uk/

Appendix

Appendix 1 – Tasks

1. (20%) Describe in your own words the optimization problem formulation presented in Section Method: Mathematical Model, including decision variables, constraints, the objective function, and all required parameters. State any necessary assumptions.

2. (20%) Explain in detail the construction heuristic (GA) and the improvement heuristic (FI) presented in Section Method: Heuristic Algorithms, including the motivations. State any necessary assumptions and you are encouraged to use figures, diagrams, and examples to explain the algorithms if needed.

3. (30%) You are asked to consult a new service provider Prodnav Ltd. operating in Great Britain. They have 13 regional facilities whose information, including their nominal capacities (number of customer visits), are presented in facilities.xlsx. The customers are grouped by local authorities and average customer demands are proportional to the population with the rate of 0.1%. The population data is available https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland. Finally, distances between customer locations and facilities can be calculated using their latitude and longitude values. Geography information of local authorities can be found here: https://geoportal.statistics.gov.uk/. Use a constructive heuristic to propose a customer allocation solution to the company. Explain in detail how you prepare data, implement the heuristic in Python, and briefly analyse the resulting customer allocation solution.

4. (30%) Use results from the previous parts to solve the problem using the optimization problem formulation as well as the proposed heuristics in Python. Explain in detail how you prepare additional required data/parameters for the problem. Analyse the solutions by varying relevant parameters and comment on how the proposed formulation, heuristics, and data preparation/settings can be improved or implemented differently?

Appendix 2 – Mathematical Model – Van Dorp

$$Min\, \Sigma\Sigma\, h_i \cdot d_{ij} \cdot X_{ij} \qquad (1)$$

$s.t.$

$$\Sigma X_{ij} = 1 \qquad \forall i \in I \qquad (2)$$

$$\Sigma h_i \cdot X_{ij} \leq C_{ij} \qquad \forall i \in J \qquad (3)$$

$$X_{ij} \in \{0,1\} \qquad \forall i \in I,\ \forall i \in J \qquad (4)$$

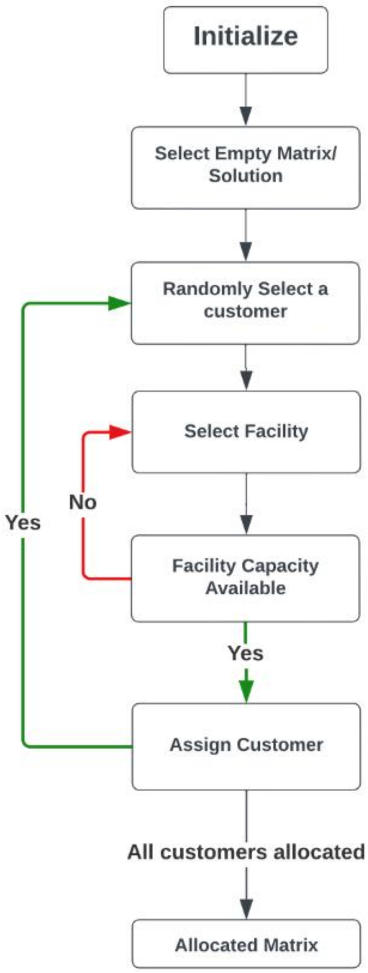
$$\Sigma h_i \cdot d_{ij} \cdot \left(a_{ij} - X_{ij}\right) \geq R_i \cdot T \qquad \forall i \in I \qquad (5)$$

$where$

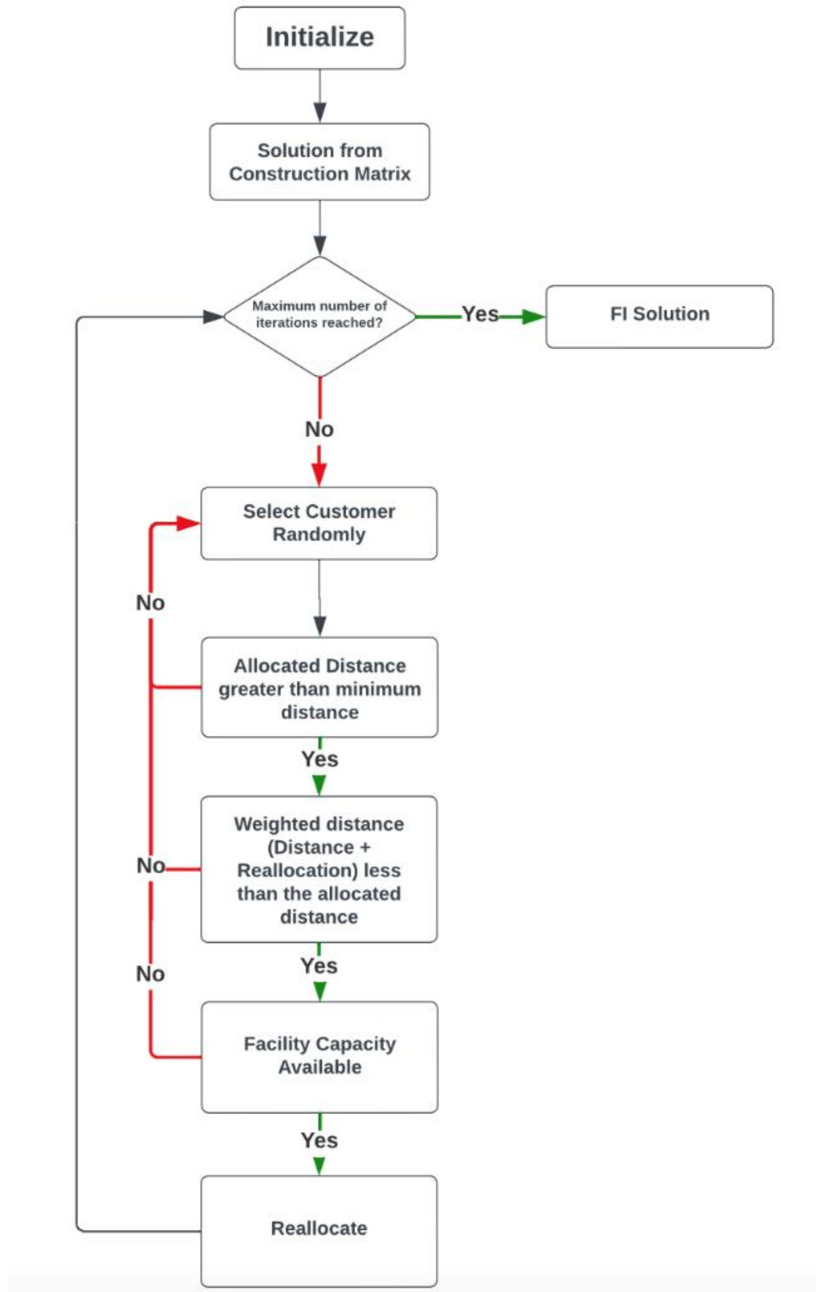
$$Ri = 1 - \Sigma a_{ij} X_{ij} \forall i \qquad \forall i \in I \qquad (6)$$

Appendix 3 – Greedy Random Adding and First Improvement Heuristic

Greedy Random Adding



First Improvement



Appendix 4 – Demand vs. Supply situation

We focused on this region with a selection of 115 customers with a total demand of approximately 20,000, more than double the capacity of the three facilities in this area. We then filtered for LAs with over 200 demand and we found that 35 LAs were selected with a combined demand of 10,000. This already exceed the total capacity of this region. Hence, there were still 80 LAs, of which 61 over 100 demand, needed to be allocated to farther facilities, if random adaptive algorithm were chosen.

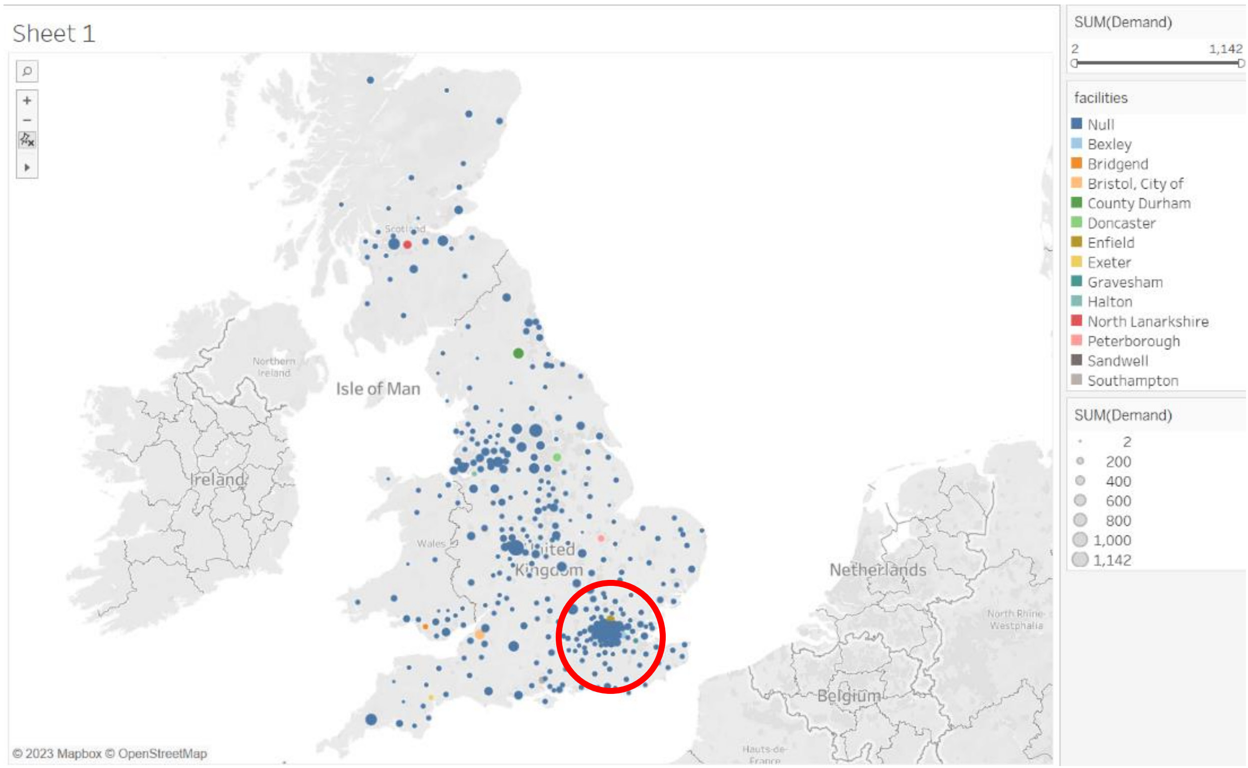
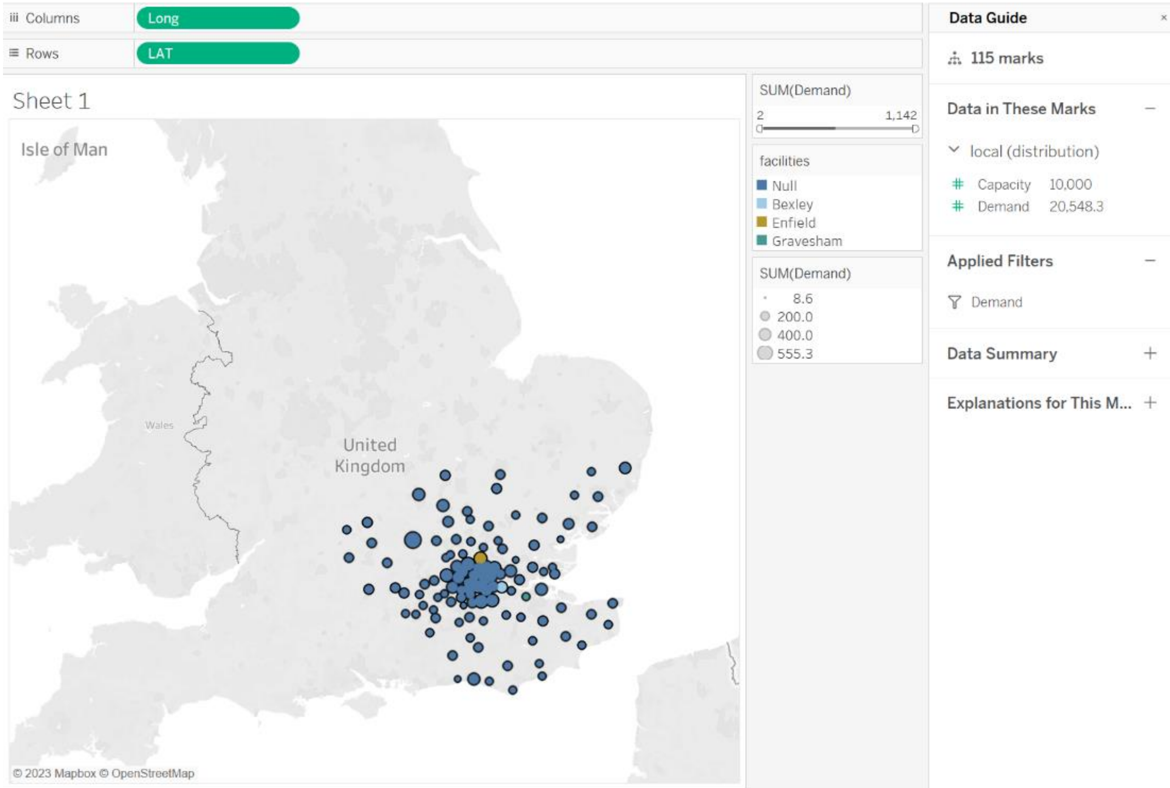
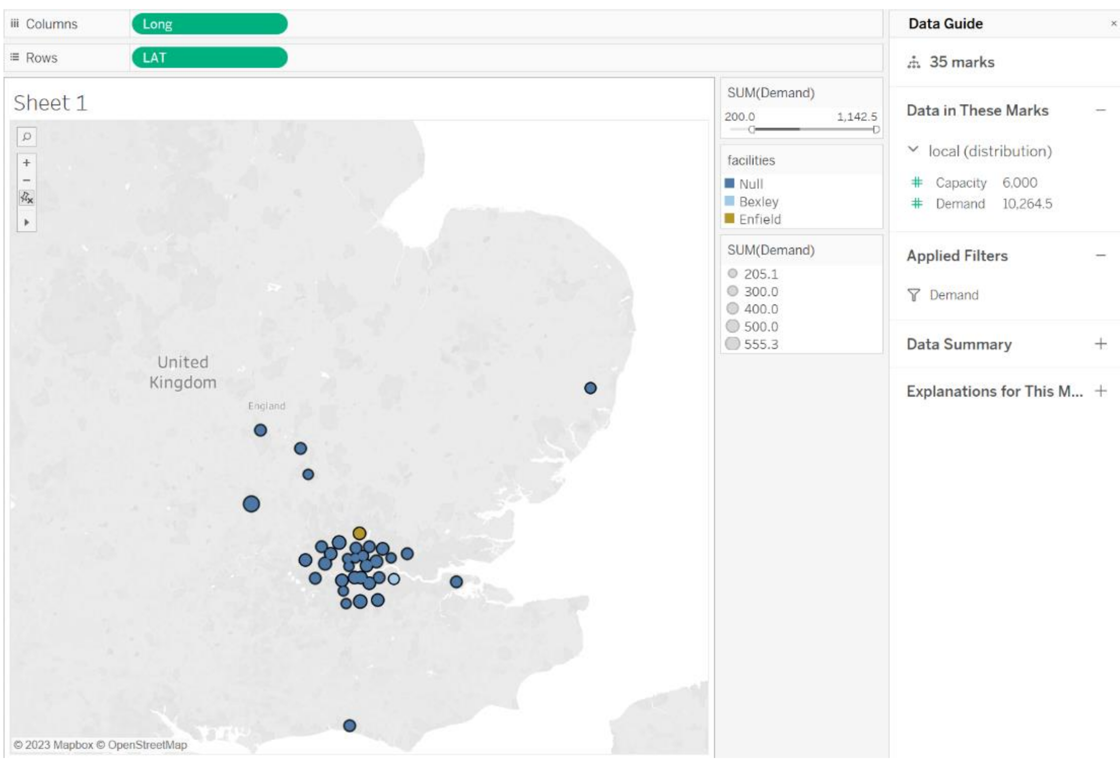


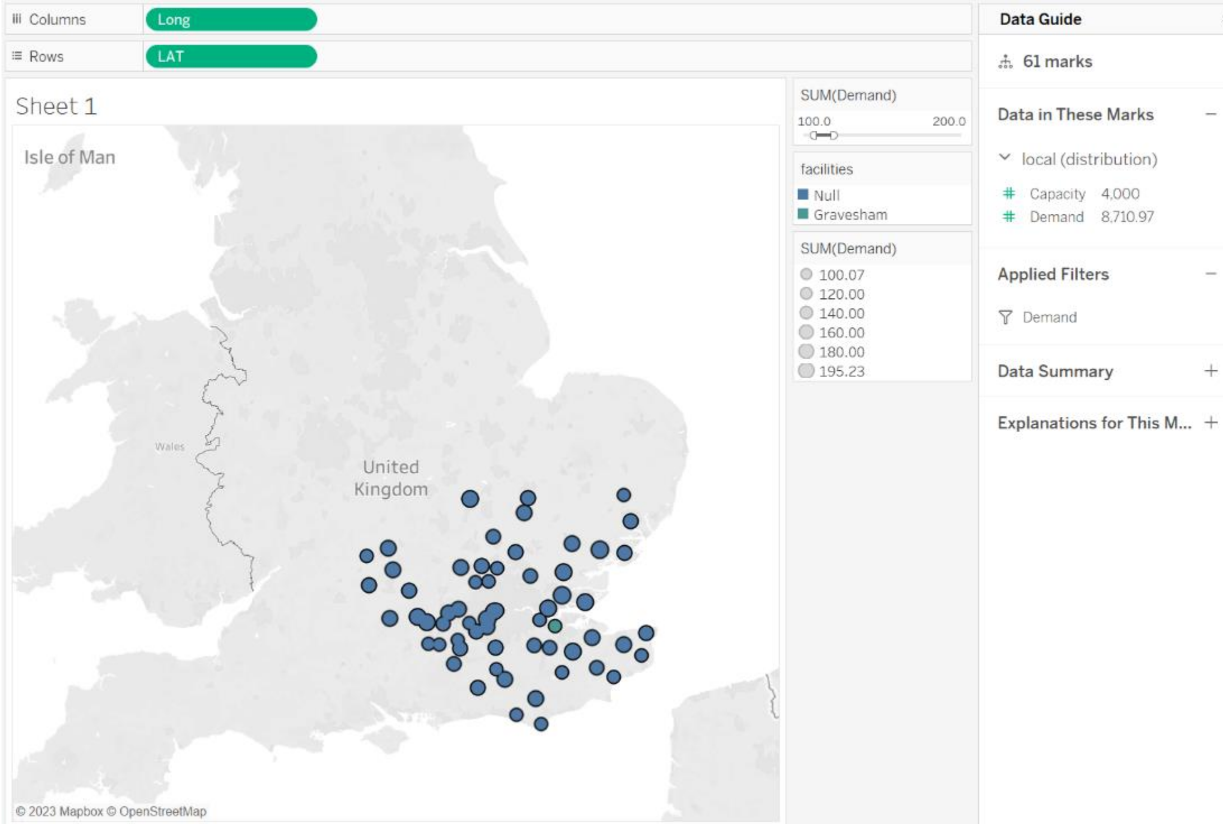
Illustration of demands (blue dot) and facilities (multi colour dot)



Total demand in the south-east region vs. 3 facilities (Bexley, Enfield, Gravesham)



Allocated customer with over 200 demands



Unallocated customer with over 100 demands, these customers will need to be serve by facilities further away.

Appendix 5 – Solution Quality of OM and FI heuristic

Initial solution: 64,196 (GR)				
Capacity relaxation	Global optimum (OM w/o constraints)		OM with constraints	First Improvement
	Total cost	Number of reallocations	Total cost	Total cost
0%	55,971	96	64,196	64,196
1%	55,513	102	63,980	64,036
2%	55,183	95	63,040	63,366
4%	54,506	99	61,770	62,705
10%	52,793	97	58,710	60,876
15%	51,281	99	56,152	59,120
20%	49,816	93	53,991	57,665
50%	45,835	89	47,001	50,134
70%	44,769	90	45,190	47,628
Uncapacitated	43,677	89	43,677	43,677

Appendix 6 – Solution Quality of the FI and the new proposed algorithm

Capacity relaxation	First Improvement		New proposed algorithm		
	Total cost	Number of reallocations	Total cost	Number of reallocations	Gain over the original algorithm (%)
0%	64,196	0	64,196	0	0.00%
1%	64,036	1	64,036	1	0.00%
2%	63,366	4	63,304	3	-0.10%
4%	62,705	12	62,420	11	-0.45%
10%	60,876	28	60,075	18	-1.32%
15%	59,120	50	57,835	25	-2.17%
20%	57,665	77	56,153	33	-2.62%
50%	50,134	130	48,398	54	-3.46%
70%	47,628	198	46,386	62	-2.61%
Uncapacitated	43,677	89	43,677	89	0.00%