Data Driven Spatiotemporal Analysis of e-Cargo Bike Network in Lisbon and Its Expansion: The Yoob Case Study

Bruno Gil[1], Vitória Albuquerque[2], Miguel Sales Dias[1](✉), Rui Abranches[3], and Manuel Ogando[3]

[1] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisbon, Portugal
{bruno_alexandre_gil,miguel.dias}@iscte-iul.pt
[2] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal
[3] Yoob, Estr. de Telheiras 159 Z, 1600-769 Lisbon, Portugal

Abstract. The adoption of more environmentally friendly and sustainable fleets for last-mile parcel delivery within large urban centers, such as e-cargo bikes, has gained the interest of the community. The logistics infrastructure network, had to adapt to the requirements of this new type of fleet, and micro-hubs and nano-hubs emerged. In this paper we tackle spatiotemporal characterization of e-cargo bike fleet behavior, by conducting a data centered case study where we explore data from Yoob, a last mile delivery e-cargo bike logistics startup that operates in the Lisbon area and outskirts. We also address the identification of potential expansion locations to the establishment of new hubs. Our data was collected during a 4-month period (January to April 2022). By adopting state-of-the-art data science and machine learning techniques, and following the CRIPS-DM data mining method, our innovative approach discovered five clusters that are able to characterize the Yoob fleet, with variations in distances traveled, times, transported volumes and speeds. In the perspective of expanding Yoob's e-cargo bike network, three new locations in Lisbon were signaled for potential new hub installation. To the authors knowledge this is the first study of this kind carried in Portugal, bringing new insights in the field of last-mile logistics.

Keywords: e-cargo bikes · micro-hub · K-Means · last-mile logistics

1 Introduction

The impact of urban logistics and logistics networks in urban mobility of the large cities are increasingly discussed by policy makers and logistics operators [1]. These last ones, along with service providers are beginning to introduce more environmentally friendly vehicles into their fleets. E-cargo bikes are one of the most widely implemented electric powered vehicles for deliveries within urban centers [2].

This study of based on data generated by e-cargo bike urban logistic operator, allows us to understand and find patterns and dynamics in the functioning of E-cargo bikes in urban centers, taking the example of the Lisbon case study.

A. L. Martins et al. (Eds.): INTSYS 2022, LNICST 486, pp. 23–39, 2023.
https://doi.org/10.1007/978-3-031-30855-0_2

1.1 Motivation and Topic Relevance

Performing last mile delivery with less impact on urban mobility in a sustainable and ecological way is the main goal of Yoob. This startup is a delivery logistics company operating in Lisbon's urban center, and it is the first of its kind operating in the city, and in Portugal. Operations started in the fall of 2021. At the time of this article writing, Yoob has a fleet of ten e-cargo bikes and two e-vans supported by five logistic hubs spread (referred to as micro and/or nano-hubs) throughout the city, including the city center. With the growth of their operations in the city, the need arose to get more insights on the behavior patterns of Yoob's e-cargo bike fleet. This data centered study provides insights for better strategic decisions for Yoob's future logistic operations and expansion of its network.

1.2 Research Questions and Objectives

This study aims to analyze and visualize the behavior patterns of e-cargo bike fleet based on anonymized real time data of a logistics company, collected in Lisbon from January 2022 to April 2022. It also intends, based on collected data, to evaluate the optimal sites for the new hubs locations to expand the e-cargo bike delivery area in Lisbon. Therefore, the following research questions are addressed by our research:

RQ1: How can we characterize the spatiotemporal traffic of the last mile logistic distribution performed with the e-cargo bike fleet, taking into consideration open data of the city and data collected during the performed routes?
RQ2: Based on the fleet behavior and the patterns detected, what are the best possible locations for the micro-hubs or nano-hubs expansion?

1.3 Structure

This paper is organized into four sections. In Sect. 1, we introduce the topic context, motivation and relevance, and we raise our research questions and objectives. In Sect. 2, we present a literature review by using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology [3]. In Sect. 3, we apply the CRoss Industry Standard Process for Data Mining (CRISP-DM) [4] methodology to our case study, presenting the results of each phase. Finally, in Sect. 4, we present and discuss our conclusions, limitations, and future work.

2 Literature Review

2.1 Methodology

PRISMA [3] is a standard methodology for generating systematic and objective findings from literature reviews. It is an approach that assisted us in describing literature findings, as well as to contribute to our goals.

2.2 Results

To kick start PRISMA in our systematic literature review (SLR), we run the following logical query on academic data repositories: ("e-cargo bikes" OR "electric-assist cargo bicycles") OR ("micro-consolidation hubs" OR "hub location") OR ("Last mile logistic" OR "urban logistic") OR ("spatial patterns" AND "data mining"). 34 articles met the eligible requirements.

Analyzed literature methods applied strong emphasis on visualization, with focus on study and detection of transportation traffic patterns [10–19]. K-means [10, 11, 20] was used to perform clustering analysis regarding travel activity for taxis and bikes and to find the places that gave rise to shorter travel distances. DBSCAN was implemented to found travel paths made by users of public transportations [12] and to study private car trajectories in the city [15]. In the decision taking for hub location, the two principal algorithms implemented were Genetic Algorithm [2, 21] and PROMETHEE [20, 22].

Moreover, we found that the study of e-cargo bikes is still very limited and focused on the scope of environmental impact and benefit of cargo bike usage. Very few papers analyzed the behavior and performance patterns of last mile delivery of e-cargo bikes in urban centers. The importance of using spatiotemporal analysis in comparison to traditional data mining approaches that consider instances to be "distributed equally and independently", is due to the possibility to find existing links between the various instances of available data in space and time [23]. Ignoring these connections can lead to misinterpretation and results that are difficult to understand [18, 23].

We observe homogeneity in the applied spatiotemporal methods. The clustering technique for pattern detection was the most present in our SLR. Zeng et al. [14] characterized the taxi travel patterns of Chongqing residents from two perspectives, hot spots and hot paths, by applying the GRIDBSCAN and ST-TCLUS (Spatial-temporal trajectory clustering) clustering algorithms. It allowed to conclude that depending on the time of day, these areas varied according to their land use. Y. Huang et al. [15] studied the travel patterns of private cars to identify the most frequented sites using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm and Markov chains, allowed them to identify that 59% of car trips exhibit regular spatiotemporal mobility and repeated travel patterns. By applying the ST-HDBSCAN clustering algorithm (combination of ST-DBSCAN and HDBSCAN clustering algorithms) Li et al. [18] made a spatiotemporal characterization of the hotspot characteristics, through the study of "Spatiotemporal Distribution", "Travel Distance Distribution" and "Travel Direction Distribution", concluding that the most frequented areas are the ones where there is a higher density of points of interest. Toro et al. [10] studied the mobility patterns of users of Milan's bike sharing systems and using the clustering technique with K-Means, allowed him to identify which stations have the same usage pattern. In the exploitation of the most frequent paths made in the Singapore Strait Ron, Wen et al. [16] applied the K-nearest neighbors' algorithm, to perform clustering on time series of waterways, which allowed them to identify the most congested areas spatially and temporally. Atluri et al. [23] state that, in exploring problems with spatiotemporal data, finding the similarities or dissimilarities between instances is the key to solving most challenges. In the collected studies, the evaluation of the performance of cargo bikes is highly focused on comparing with the performance of the cargo vans in the last mile delivery [17, 24–26].

Cargo bikes showed a greater flexibility and advantage in the routes they made. Most of the time the chosen bike route is shorter than the route made by vans [24]. This difference can be up to twice as large on shorter trips [17]. Also, it was found that cargo bike riders easily break traffic regulations by riding in the opposite direction during short trips [24]. Amaral et al. [17] identified that travel times were not as important for cargo bikes as for motor vehicles, because bicycles can easily "outrun" traffic jams. An interesting observation by Conway et al. [24], showed that the speed of cargo bikes on the bike paths is lower than when on the road for motor vehicles, with a speed figure lower than 20% on some of the routes. The impact of street topography was mentioned in Amaral et al. [17] who defined a scale between the elevation and the impact on cyclist performance. This scale sets as a reference, below 2%, with no effect, between 2% and less than 5%, already considered with impact and above 5%, representing a substantial impact. The speed considered in the studies was not homogeneous, varying between 11.6 km/h [24] and 24.0 km/h [25]. A literature review done by Büttgen et al. [7] finds an average speed of this type of vehicles between 8.0 km/h and 25.0 km/h.

Overall, all studies conclude that cargo bikes represent a more viable and advantageous alternative in last mile delivery, with greater gains in more congested areas [24], but with some constraints. Sheth et al. [25] concluded that the distance and the number of deliveries, are the most impacting factors on viability and cannot exceed 3.2 km and 20 orders per stop. In Amaral et al. [17], the capacity of the vehicle was not considered, but authors concluded that beyond 3.0 km, it was no longer efficient to deliver with this type of vehicle. The combination of cargo bikes and the implementation of micro hubs has helped the green alternatives for last mile delivery, to gain momentum [9]. Distribution networks with micro-hubs do promote a more organized last mile delivery [8] and benefit from economies of scale [27].

The definition of micro hub in the literature is vast, and for our paper we adopted the definition by Katsela et al. [8], which defined it as "logistics facilities where commercial transportation providers (or "carriers") consolidate goods near the final delivery point and serve a limited spatial delivery area in a dense urban environment". Finding and defining a location for micro-hubs is an important and complex task [2, 8]. The rising costs of urban land, lack of adequate infrastructure, changing demand, changing city characteristics [22] and regulatory requirements [8], do not ease the task of being able to find an optimal solution that minimizes operating costs and impact on communities. The most common characteristics addressed in the literature to study this problem were demand (e.g., residential, commercial, and/or employment density), infrastructure (e.g., pedestrian/bicycle infrastructure provision, road classifications, pedestrian zones, and measures to assess traffic), and land use constraints [6, 22, 28]. When the deliveries are made by cargo bikes, the location of the micro-hub should be the closest to the delivery point [22, 29]. Assman et al. [9] recommended locating them in areas of higher commercial density. This need for proximity comes from the capacity limitation of bikes compared to a delivery van, and multiple trips to the micro-hub may be required, so travel time and travel distances are minimized [8]. According to Assman et al. [9], the maximum distance between the micro-hub and the delivery point should not exceed 1000 m. In Rudolph et al. [22] a distance between 500 m and 1200 m is pointed out as the distance range that allows economic feasibility for deliveries made by cargo bikes.

In Faugère et al. [5] and Srivatsa Srinivas et al. [30], the implementation of this type of infrastructure in mobile units was evaluated and, in both studies, they concluded that it can be a viable alternative but under very restricted conditions. Faugère et al. [5] indicated as a condition, the requirement to transport a high volume of orders and a very short maximum transit travel time. In Srivatsa Srinivas et al. [30] the need for a strong analytical engine that can accurately predict demand for a given geographic location and the dynamic optimization of the route and parking location of the mobile warehouse, was the only way to make this alternative viable. The study of stationary micro-hubs is the most widely covered in the literature, but the methods vary among literature papers. When the targets' location points are already known [2, 7, 31, 32], only an evaluation of the performance of each of the locations was done to find the one that best suited the purpose. Naumov et al. [2] developed a mathematical model representative of the network and its behavior and by applying Monte Carlos simulation, evaluated which of the five pre-defined locations allowed minimizing the transportation work. In Kedia et al. [32], the Location-Allocation model, was used to find the locations that minimized the distance that had to be traveled. Bütten et al. [7] uses the Two-Echelon Vehicle Routing Problem 2E-VRP model to find an optimal solution that minimizes costs. In Leyerer et al. [31], the Split Delivery Vehicle Routing Problem with Multiple Products Compartments and Time Windows (SPVRPMPCTW) model is solved, to minimize costs throughout the three stages (LRP, VRP with time window and VRP considering multiple products) that compose model. When there is no pre-knowledge of such locations, other approaches are needed, and possible solutions can be found based on the knowledge of the demand or the geographical characteristics of the cities. Rudolph et al. [22] uses a multi-criteria method to find the most suitable locations and employs the Analytical Hierarchical Process AHP and PROMETHEE algorithms, defining that the main criteria to use are demand, road type and land use. The optimal locations should minimize travel times and travel distances. Song et al. [19], use the LCRS (Longest Common Route Subsequence) algorithm, complemented with a voting system, to find the paths most traveled and where there is a higher concentration of deliveries. This approach allows them to calculate which locations can minimize the time and distance traveled. In the literature we found that, in the approach to this problem, the computational capacity and the time required to explore all possible options, limit the calculation of the optimal points [19, 21, 33, 34]. The implementation costs of micro-hubs and vehicle capacity are often not considered. We can argue that minimizing distances, travel times, and costs are among the most relevant objectives in hubs location.

3 Data Analysis and Modeling

The CRISP-DM methodology, applied in our research, attempts to reduce the cost and increase reliability, repeatability, manageability, and speed of big data mining operations. According to this methodology the life cycle of data mining projects is divided into six parts: business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

3.1 Business Understanding

The data explored was provided by the e-cargo bike urban logistics startup Yoob [35]. As mentioned, the purpose of the study is two-fold: the first, to provide a spatiotemporal characterization of the Yoob e-cargo bike fleet in the parcel collection and delivery processes in Lisbon as well in its outskirts; the second, to propose locations for the new hubs and adjustments to the existing logistics network, in order to strengthen and expand the fleet operations. The company has two types of hubs, the micro-hub, with an area of 36 m^2, a relatively smaller option compared to the values found in SLR, which range between 92 m^2 to 920 m^2 [36]. The functional definition is in line with that found at SLR, with various services being done at the micro hub, namely, consolidation of goods, storage, and recharging of e-cargo bikes. The nano-hubs, which is an innovative concept developed by Yoob, emerged from the adaptation of the pick-up/drop-off concept to last mile delivery logistics, characterized by having relatively small areas ranging between 3 m^2 and 120 m^2, exclusively dedicated as a temporary transition point where the goods remain no longer than 48 h. The type of associated physical infrastructure varies depending on where it is implemented, given it only requires temporary storage capacity for goods [37].

3.2 Data Understanding

The data was extracted from Yoob's database and covered the period of January 1st to April 30th 2022, encompassing 9,175 records and 34 variables. The data does not provide the routes (trajectories) done by the fleet. The geographic information on the route is characterized by latitude and longitude of origin and destination. There are some variables that generated based on mobile devices used by the employees during the entire logistics operation.

In our approach, each record in the data represents a "story", which is geographically composed of two points, one for pickup and the other for delivery. Within each story there are two "sub-stories", where each "sub-story" refers to a geographical location (pickup or delivery) and is always associated to a "route", where the "routes" can be composed of one or more "stories".

3.3 Data Preparation

The first data preparation step was the individual evaluation of all variables. Secondly, the unnecessary variables, outliers and incomplete stories were removed resulting in a dataset with 8,381 records (91,3% of the raw dataset) each one with 26 variables. The third step was to convert our dataset to have a sub-story granularity, by creating two datasets, one referring to the pick-up information and the other referring to drop-off information. These two datasets were merged.

To enrich our dataset, we added extra features:

- ['Elevation_point'] - Elevation of the sub-story geographic location, was obtained by consulting a DEM (Digital Elevation Map) [41].

- [‘order route’]: Number indicating the order in which the location is visited within the route sequence.
- [‘time_enRoute_sec]: Time period in seconds between the [‘history.enRoute’] and [‘history.arrived’].
- [‘time_points_sec’]: Time period in seconds between two consecutive points on the same route.

This dataset processing with sub-story granularity resulted in 15,828 records and 27 variables.

To perform the spatial analysis two geographic data frames were generated with the geopandas Python library [42]. In the first the granularity was the route level, and second the granularity was the sub-story level. To be considered valid, a route must have two or more associated sub-stories. Routes that do not meet this requirement were removed. With this procedure we were able to reconstruct 664 routes, representing 95% of the total routes in the original dataset (699 routes), at sub-story level. We have removed 20 records (<0.002%), ending up with a dataset with 15,808 records.

3.4 Modeling

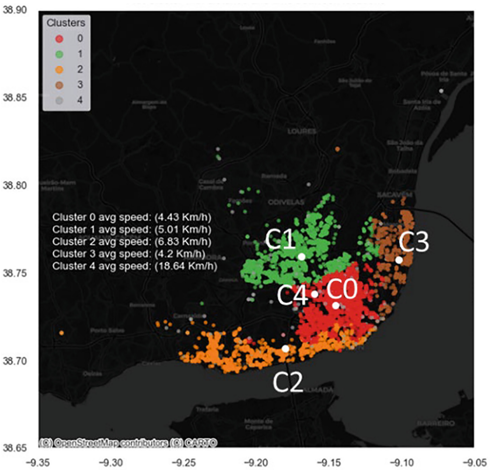

Fig. 1. Clustering the sub-stories with K-Means

In the modeling phase, we applied machine learning techniques, namely K-means, to developed three models to answer our research questions. In the first model we created clusters to identify the behavior of routes in certain geographical areas. In the second model we clustered the routes and evaluated their characteristics, providing answers to our first research question. In the last model we performed a gravity center analysis, with the goal to explore new locations for the implementation of new hubs, answering our second research question. To build the models we used the sklearn [43] library, for pre-processing we used MinMaxScaler [44] and LabelEnconder [45] and to perform cluster and the center of gravity analysis, we used K-Means algorithm [46]. To evaluate the optimal K value in the two first models, we adopted the Knee Elbow method with the knee library [47] and Davies-Bouldin index [48].

First Model – Clustering the Sub-stories with K-Means

In the first model, we identified the behavior of routes in certain geographical areas with cluster analysis. The feature selection was made from the geodataframe data structure with sub-story granularity.

The selected features were [‘latitude’], [‘longitude’], [‘elevation_point’], [‘time_points_sec’] and [‘distance_to_prev’]. Before running the clustering model in our data, we had to scale the data, as it had different measurement units, with MinMaxScaler. When evaluating the Knee Elbow method and the Davies-Bouldin index through a range from 1 to 30 clusters, we found that the optimal value for K was 5 in knee elbow method, and 4 in the David-Bouldin technique. After testing the model with both values, the knee elbow value was selected as it gave us more information (later confirmed in YOOB briefings). Then we applied the K-Means algorithm with a K value of 5 to our data, and the output is depicted in Fig. 1. Four main clusters (C0 to C3) outstand in the visualization, and a fifth cluster (C4) with dissipated grey dots among the four other main clusters. In this model we can observe the e-cargo bikes’ performance according to the geographical area. In Fig. 1, we can see the four well defined clusters and a more disperse cluster (C4) where the e-cargo bikes have a higher average speed of 18.64 km/h, indicating that these are acceleration areas. In the other clusters the average speed is significantly lower. The zones with the second highest average speed were the ones in cluster C2 where e-cargo bikes achieved average speeds of 6.84 km/h, followed by the zones covered by cluster C1 with average speeds of 5.01 km/h. The areas covered by clusters C0 and C3 have a more homogeneous performance. However, in the areas covered by cluster C3 the e-cargo bikes tend to be slower, with average speeds of 4.20 km/h vs 4.43 km/h of the speeds practiced in the C0 areas.

Second Model – Clustering the Routes with K-Means

In the second model the selected features were based on the geodataframe with granularity of the route: [‘distancia_total’] and [‘distancia_maxima_do_ini’]; and were scaled with MinMaxScaler. Much like in the first model, we evaluated the Knee Elbow value and the Davies-Bouldin value in a range from 1 to 30 clusters and selected the optimal value for K (5) provided by the Knee Elbow method, since the optimal value in the David-Bouldin method was far bigger. Applying to our data K-Means with a K value of 5, the output results in five clusters (see Fig. 2, 3, 4, 5, 6, 7, 8 and 9).

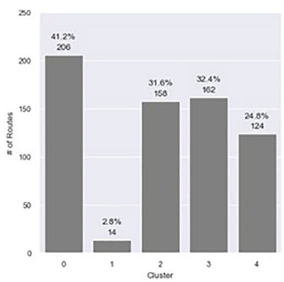

Fig. 2. Routes per cluster

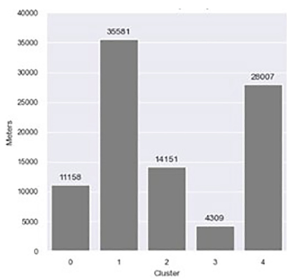

Fig. 3. Average total distance per cluster

In Fig. 9, the operation time metric was calculated by subtracting the average total en route time from the total time spent between two locations and dividing the result by twice the number of locations visited, representing the operation time spent at each location. In the presentation of results below, all figures are average numbers.

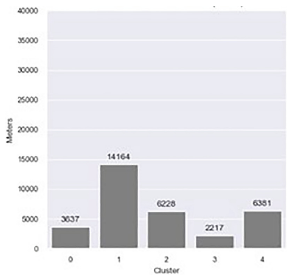

Fig. 4. Average maximum distance from initial location per cluster

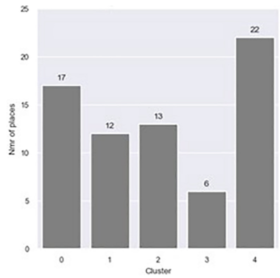

Fig. 5. Average visited locations per cluster

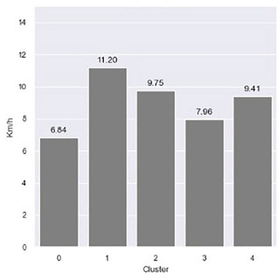

Fig. 6. Average speed per cluster

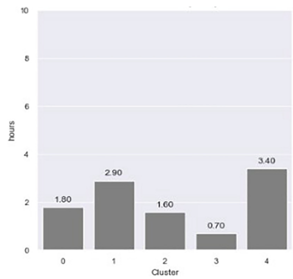

Fig. 7. Average total time in route per cluster

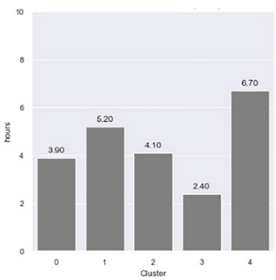

Fig. 8. Average total time between locations per cluster

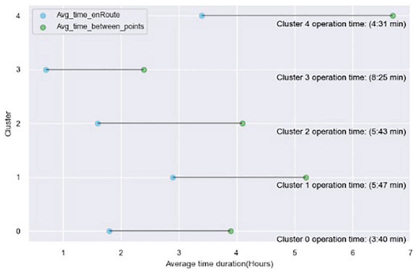

Fig. 9. Average operation time per cluster

The most common performance is the one observed in cluster C0, accounting for 41.2% of the total trips (see Fig. 2). This cluster features a speed of 6.84 km/h, which is the lowest speed of the five clusters, corresponding to a total traveled distance of 11.16 km. Yoob's e-cargo bikes travel at a maximum distance of 3.64 km, from their starting location. The total duration of cluster 0 trips is 3 h and 54 min, and the e-cargo bikes are only in motion for a period of 1h48m. Seventeen different locations are visited, and 3m40s is the shortest operating time per location visited, during trips of cluster C0. The second largest type of performance is observed in cluster C3, which includes 32.4% of the total trips (see Fig. 2). It is characterized by a total distance traveled of 4.31 km, at a speed of 7.96 km/h. In cluster C3, e-cargo bikes travel a maximum distance of 2.22 km from their starting location. These trips have the shortest and closest travel distances. They have a total duration of 2 h 42 m, and bikes are only in motion for 42 m. With six different locations, cluster C3 has the fewest number of locations visited from all five performances, but has the longest operation time per location visited, requiring 8 m 25 s. This may be associated with the high waiting time for customers according to Yoob partners feedback. The third most predominant type of performance is the one observed in cluster C2, with 31.6% of total trips (see Fig. 2). The total distance traveled is 14.15 km at a speed of 9.75 km/h. The e-cargo bikes travel at a maximum distance of 6.23 km from the starting location. The total travel time is 4 h 6 m, with the e-cargo bikes being in motion for 1 h 36 m. Thirteen different locations are visited, and bikers spend 5 m 43 s for each location. The fourth most observed performance type is the one of cluster 4, with 24.8% of the total trips (see Fig. 2). It is characterized by a total traveled distance of 28.01 km, at a speed of 9.41 km/h. The e-cargo bikes travel at a maximum distance of 6.38 km from their starting location, with a total duration of the route, of 6 h 42 m. Bikes are in motion for an average period of 3 h 24 m. These are the trips with the longest travel time and with the largest number of places visited, with a figure of twenty-two different places. At each location visited bikers spend 4m31s in operation time. The least observed type of performance is the one corresponding to cluster 1 (see Fig. 2), with only 2.8% of the total trips. These are the longest trips with the wider range, but also the fastest ones, with a total distance traveled of 35.58 km, at a speed of 11.20 km/h. In this cluster, the e-cargo bikes travel at a maximum distance of 14.16 km, from their starting location. The total travel time of a trip is 5 h 12 m, with the e-bikes being in motion for a period of 2 h 54 m. Twelve different locations are visited, and bikers spend 5 m 47 s in each location.

Third Model – Center of Gravity Analysis with K-Means

In the third model, we analyzed the centers of gravity of the sub-stories of our data. This model analysis was requested in one of the meetings held with Yoob. Although in our initial SLR there were no direct references to this specific topic, by doing some additional research, we found that Wen et al. [49] and Cai et al. [50], both approached this problem by applying K-means techniques with a weighted featured to find the best hub locations. In our approach, we adopted a similar method with a weighted K-Means algorithm.

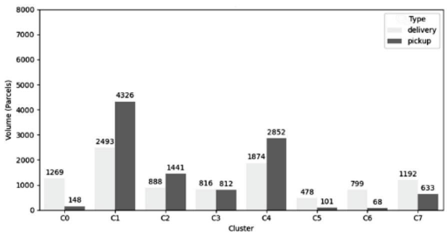

Fig. 10. Volume parcels per proposed new cluster centroid

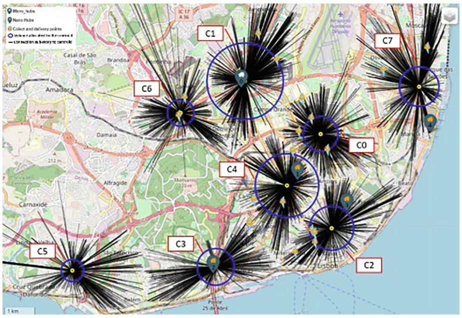

Fig. 11. Center of gravity analysis for eight hubs, using K-Means. Dark lines represent the distances from the hub center to the delivery points

Our model applied the number of locations intended to simulate, and a new variable was considered in the weighting of the cluster. In our model, the number of parcels was considered, as the effort needed to carry out the delivery. As most of the time the pickup parcels were in the hubs or at the collect/delivery locations, we added a penalty value in the delivery parcels, considering these last ones three times bigger in effort than the pickup ones. This forced the algorithm to locate the centroids of the cluster in places where distance and effort would be reduced. The data applied in this model was based on the variables [‘latitude’], [‘longitude’] and [‘parcels’] from the geodataframe with sub-story granularity. A new variable was created designated [‘calc_ajusto_de_custo_se_houver’], to include the penalty value. We simulated the center of gravity for 8 hubs, the result is shown in Fig. 10 and the volume associated for each location is depicted in Fig. 11.

3.5 Deployment

The models created were not applied in a real production environment. Software development was developed on a personal computer equipped with Windows 10 (64bits) operating system, Intel(R) Core (TM) i7-11370H 3.30 GHz, with 40 Gb of memory ram. We adopted the Python programming language (v3.10.4) [38], compiled with Visual Studio Code (v1.69.1) [39] on Jupyter Notebooks extension [40]. The developed software material and data sets are available for use by the Yoob company and for further academic research purposes.

4 End-User Evaluation

The end-user evaluation verifies that the findings are consistent with the proposed research objectives and the accuracy of the business requirements.

Table 1. Method assessment questionnaire

Criteria	Objective statement	Eval #1	Eval #2
Utility	It can help business decisions regarding the behavior of the fleet and hub expansion	FA	FA
Understandability	Provides understandable results	FA	FA
Accessibility	Can be used without training	LA	LA
Level of detail	Provides knowledge regarding the mobility of the fleet and detailed location for expansion	FA	FA
Consistency	Gives consistent results	LA	LA
Robustness	Has enough detail to be used in other cases of e-cargo bikes and hub expansion	FA	FA

In the end of the study a questionnaire was sent to the two YOOB partners, with the questions and answers indicated in Table 1. The development of the questionnaire follows the standards defined by the ISO/IEC TS 33061[1] [51], primarily used to assess software development processes. Four levels of the NLPF were employed for evaluation: Not Achieved (NA) - [0–15%]; Partially Achieved (PA) - [15–50%]; Largely Achieved (LA) - [50–85%]; Fully Achieved (FA) - [85–100%]. In this evaluation, we obtained a rating of FA, in the criteria of usefulness, understanding, level of detail and robustness, and LA rating in the criteria of accessibility and consistency. Overall, this indicates that the work done represents an added value for the company, providing useful, detailed, and clear information, appropriate to support decision making, in the context of the e-cargo bike fleet as well as for the expansion of new hubs. The YOOB evaluators consider that this study can be replicated to other case studies with potential for improvement, and implementation readiness. Moreover, the outcomes are aligned with the objectives and requirements proposed for the research presented in this paper.

5 Discussion and Conclusions

We have presented an innovative data science-based study, the first regarding last mile delivery using e-cargo bikes operating in Lisbon, Portugal, as far as the authors are aware. To tackle our research questions, we developed and evaluated three intelligent computing models. Our second model (Clustering the routes with K-Means) in particular, allowed us to answer the first research question, and to characterize the behavior of the e-cargo bike fleet through the traveled distance, time, speed and number of visited locations. Overall, the average of total traveled distance ranges between 4.31 km and 35.50 km, distancing from their start location, between 2.20 km and 14.10 km. 63% of the routes show distance ranges very close or even lower than the values reported by Sheth et al. [25], which considered cargo bikes to have an efficient performance under

[1] "ISO - ISO/IEC TS 33061:2021 - Information technology—Process assessment model for software life cycle processes." https://www.iso.org/standard/80362.html.

3.20 km. The average number of different locations visited per route ranges between 6 and 22. The average observed speed varies between 6.84 km/h and 11.20 km/h, a value close to the study by Bütten et al. [7], where these authors looked at several cargo bike projects, and calculated average speeds between 8.00 km/h and 25.00 km/h. The temporal characteristics revealed a time in movement per route from 42 m minutes up to 3 h 24 m, and a total route duration time, ranging between 2 h 42 m and 6 h 42 m. Required transaction time within each route ranged from 3 m 40 s to 8 m 25 s. This higher time may be due to the particularities of certain customers requiring more waiting time. Excluding this last observation, the time metric ranges between 3 m 40 s and 5 m 43 s. This set of characteristics gave us an overview of the needs of each route and the respective performance of the e-cargo bikes in their operation conditions. As for the second research question, the third model (K-Means center gravity analysis), was used as our basis for analysis. The choice of new hubs locations, in the context of an expansion of the e-cargo bikes network, is a complex process due to the high number of constraints that are to be considered in the site search [2, 19, 21, 33, 34]. In the search for new locations the factors considered for the cost function of our model were the distance and the cost associated with each location visited. Then for evaluation of the hub type, the volume associated with each hub of this new structure was analyzed. When simulating an expansion of three more hubs beyond the five that are currently part of YOOB's network, our model suggests that the implementation of these new hubs should be located in the boroughs of Alvalade, Benfica and Algés (Fig. 10). When confronted with the results of this model, the YOOB partners considered that these three new proposed locations are valid options that required further analysis in terms of economic viability. Regarding the 3 remaining computed locations, in the case of C2 (Fig. 10), the choice of the current location of the hub (nr 1), which is within the radius of this cluster, was due to the geographical characteristics of the area, which is on top of a hill, causing the trips to have a downward direction, facilitating the effort required by the biker. In the case of C7 (Fig. 10), the divergence between the location of the hub (nr 4) and the location proposed by our model, raises additional challenges of further changes of location due to the high price of real estate in the area where the centroid calculated by our model is located. Considering the remaining proposed hub locations, the YOOB partners showed complete agreement. By analyzing the volume of parcels associated with each hub in Fig. 11, we can discuss what type of hub is the most adequate for micro-hub or nano-hub requirements. In our study all three new locations are more suitable for nano-hubs. In the already existing nano-hub located in the Saldanha, we observed that due to the high associated volume of parcels it could shift to a micro-hub, and this observation was positively validated by YOOB partners.

Research Limitations

The most significant limitations of our study are related to the dimension, granularity, and structure of the data. The information on the routes was limited to the visited geographical points, lacking information about the order of each visited location, and lacking complete information about the route trajectory (its 3D coordinates) taken from pickup to delivery as well city traffic. Having trajectory and traffic data would allow a deeper and more rigorous analysis of the e-cargo bike fleet route patterns, namely the real trajectories in which route was performed and the actual distances traveled. We collected data in the period from January to April of 2022, corresponding to the first four months of the company's registered activity (YOOB started operations in Lisbon in the fall of 2021). After data pre-processing, we came up with a dataset comprising 15 828 records and 27 variables, which was considered sufficient for our analysis, but that nevertheless can be limited for long-term trend analysis. The proposed hub locations can be considered the best possible locations with limitations, as many factors were not considered, such as street elevations and, specially, socio-economic factors that need to be taken into account, to tackle costs for the customer and the municipality.

Future Work

The following suggestions are made for upcoming research work:

- Expand the number of observations analyzed to detect long-term trends and produce more insightful results, given that YOOB has the possibility to collect stories and route data on a regular basis.
- Study the shortest and flattest path.
- Perform more detailed cluster analysis, with an increased number of clusters when analyzing route typologies.

Acknowledgements. This work is partially funded by national funds through FCT - Fundação para a Ciência e Tecnologia, I.P., under the project FCT UIDB/04466/2020.

References

1. The future of the last-mile ecosystem transition roadmaps for public-and private-sector players (2020). www.weforum.org
2. Naumov, V.: Substantiation of loading hub location for electric cargo bikes servicing city areas with restricted traffic. Energies **14**(4), 839 (2021). https://doi.org/10.3390/en14040839
3. Page, M.J., et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. https://doi.org/10.1136/bmj.n71
4. CRISP-DM: a framework for data mining & analysis. https://thinkinsights.net/digital/crisp-dm/. Accessed 20 May 2022
5. Faugère, L., White, C., III., Montreuil, B.: Mobile access hub deployment for urban parcel logistics. Sustainability **12**(17), 7213 (2020). https://doi.org/10.3390/su12177213
6. Urzúa-Morales, J.G., Sepulveda-Rojas, J.P., Alfaro, M., Fuertes, G., Ternero, R., Vargas, M.: Logistic modeling of the last mile: case study Santiago, Chile. Sustainability **12**(2), 648 (2020). https://doi.org/10.3390/su12020648
7. Büttgen, A., Turan, B., Hemmelmayr, V.: Evaluating distribution costs and CO_2-emissions of a two-stage distribution system with cargo bikes: a case study in the city of Innsbruck. Sustainability **13**(24), 13974 (2021). https://doi.org/10.3390/su132413974
8. Katsela, K., Güneş, Ş., Fried, T., Goodchild, A., Browne, M.: Defining urban freight microhubs: a case study analysis. Sustainability **14**(1), 532 (2022). https://doi.org/10.3390/su14010532
9. Assmann, T., Lang, S., Müller, F., Schenk, M.: Impact assessment model for the implementation of cargo bike transshipment points in urban districts. Sustainability **12**(10), 4082 (2020). https://doi.org/10.3390/SU12104082

10. Toro, J.F., Carrion, D., Brovelli, M.A., Percoco, M.: Bikemi bike-sharing service exploratory analysis on mobility patterns. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **43**, 197–203 (2020). https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-197-2020
11. Guo, X., Xu, Z., Zhang, J., Lu, J., Zhang, H.: An OD flow clustering method based on vector constraints: a case study for Beijing taxi origin-destination data. ISPRS Int. J. Geoinf. **9**(2), 128 (2020). https://doi.org/10.3390/ijgi9020128
12. Ma, X., Liu, C., Wen, H., Wang, Y., Wu, Y.J.: Understanding commuting patterns using transit smart card data. J. Transp. Geogr. **58**, 135–145 (2017). https://doi.org/10.1016/j.jtrangeo.2016.12.001
13. Shen, Y., Zhang, X., Zhao, J.: Understanding the usage of dockless bike sharing in Singapore. Int. J. Sustain. Transp. **12**(9), 686–700 (2018). https://doi.org/10.1080/15568318.2018.1429696
14. Zheng, L., et al.: Spatial–temporal travel pattern mining using massive taxi trajectory data. Phys. A: Stat. Mech. Appl. **501**, 24–41 (2018). https://doi.org/10.1016/j.physa.2018.02.064
15. Huang, Y., Xiao, Z., Wang, D., Jiang, H., Wu, D.: Exploring individual travel patterns across private car trajectory data. IEEE Trans. Intell. Transp. Syst. **21**(12), 5036–5050 (2020). https://doi.org/10.1109/TITS.2019.2948188
16. Wen, R., Yan, W., Zhang, A.N., Chinh, N.Q., Akcan, O.: Spatio-temporal route mining and visualization for busy waterways. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 000849–000854). IEEE (2016). https://doi.org/10.1109/SMC.2016.7844346
17. Amaral, J.C., Cunha, C.B.: An exploratory evaluation of urban street networks for last mile distribution. Cities **107**, 102916 (2020). https://doi.org/10.1016/j.cities.2020.102916
18. Li, F., Shi, W., Zhang, H.: A two-phase clustering approach for urban hotspot detection with spatiotemporal and network constraints. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **14**, 3695–3705 (2021). https://doi.org/10.1109/JSTARS.2021.3068308
19. Song, H.Y., Han, I.: Finding the best location for logistics hub based on actual parcel delivery data. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11619, pp. 603–615. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24289-3_45
20. Barraza, R., Sepúlveda, J.M., Venegas, J., Monardes, V., Derpich, I.: A model for solving optimal location of hubs: a case study for recovery of tailings dams. In: Dzitac, I., Dzitac, S., Filip, F.G., Kacprzyk, J., Manolescu, M.-J., Oros, H. (eds.) ICCCC 2020. AISC, vol. 1243, pp. 304–312. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-53651-0_26
21. Hwang, J., Lee, J.S., Kho, S., Kim, D.: Hierarchical hub location problem for freight network design. IET Intell. Transp. Syst. **12**(9), 1062–1070 (2018). https://doi.org/10.1049/iet-its.2018.5289
22. Rudolph, C., Nsamzinshuti, A., Bonsu, S., Ndiaye, A.B., Rigo, N.: Localization of relevant urban micro-consolidation centers for last-mile cargo bike delivery based on real demand data and city characteristics. Transp. Res. Rec. **2676**(1), 365–375 (2022). https://doi.org/10.1177/03611981211036351
23. Atluri, G., Karpatne, A., Kumar, V.: Spatio-temporal data mining: a survey of problems and methods. ACM Comput. Surv. **51**(4), 1–41 (2018). https://doi.org/10.1145/3161602
24. Conway, A., Cheng, J., Kamga, C., Wan, D.: Cargo cycles for local delivery in New York city: performance and impacts. Res. Transp. Bus. Manag. **24**, 90–100 (2017). https://doi.org/10.1016/j.rtbm.2017.07.001
25. Sheth, M., Butrina, P., Goodchild, A., McCormack, E.: Measuring delivery route cost trade-offs between electric-assist cargo bicycles and delivery trucks in dense urban areas. Eur. Transp. Res. Rev. **11**(1), 1–12 (2019). https://doi.org/10.1186/s12544-019-0349-5
26. Caggiani, L., Colovic, A., Prencipe, L.P., Ottomanelli, M.: A green logistics solution for last-mile deliveries considering e-vans and e-cargo bikes. Transp. Res. Procedia **52**, 75–82 (2021). https://doi.org/10.1016/j.trpro.2021.01.010

27. Arrieta-Prieto, M., Ismael, A., Rivera-Gonzalez, C., Mitchell, J.E.: Location of urban micro-consolidation centers to reduce the social cost of last-mile deliveries of cargo: a heuristic approach. Networks **79**(3), 292–313 (2022). https://doi.org/10.1002/net.22076
28. Golini, R., Guerlain, C., Lagorio, A., Pinto, R.: An assessment framework to support collective decision making on urban freight transport. Transport **33**(4), 890–901 (2018). https://doi.org/10.3846/transport.2018.6591
29. Özbekler, T.M., Karaman Akgül, A.: Last mile logistics in the framework of smart cities: a typology of city logistics schemes. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **44**, 335–337 (2020). https://doi.org/10.5194/isprs-archives-XLIV-4-W3-2020-335-2020
30. Srinivas, S.S., Marathe, R.R.: Moving towards "mobile warehouse": last-mile logistics during COVID-19 and beyond. Transp. Res. Interdiscip. Perspect. **10**, 100339 (2021). https://doi.org/10.1016/j.trip.2021.100339
31. Leyerer, M., Sonneberg, M.-O., Heumann, M., Breitner, M.H.: Shortening the last mile in urban areas: optimizing a smart logistics concept for e-grocery operations. Smart Cities **3**(3), 585–603 (2020). https://doi.org/10.3390/smartcities3030031
32. Kedia, A., Kusumastuti, D., Nicholson, A.: Locating collection and delivery points for goods' last-mile travel: a case study in New Zealand. Transp. Res. Procedia **46**, 85–92 (2020). https://doi.org/10.1016/j.trpro.2020.03.167
33. Ghaffarinasab, N.: A tabu search heuristic for the bi-objective star hub location problem. Int. J. Manag. Sci. Eng. Manag. **15**(3), 213–225 (2020). https://doi.org/10.1080/17509653.2019.1709992
34. Huang, Z., Huang, W., Guo, F.: Integrated sustainable planning of micro-hub network with mixed routing strategy. Comput. Ind. Eng. **149**, 106872 (2020). https://doi.org/10.1016/j.cie.2020.106872
35. https://yoob.pt/
36. Clarke, S., Leonardi, J.: Agile Gnewt Cargo: parcels deliveries with electric vehicles in central London multi-carrier central London micro-consolidation and final delivery via low carbon vehicles (2017). www.london.gov.uk
37. Abranches, R., Ogando, M.: YOOB interview
38. www.Python.org. https://www.python.org/. Accessed 13 July 2022
39. Visual Studio Code. https://code.visualstudio.com/. Accessed 13 July 2022
40. Project Jupyter. https://jupyter.org/. Accessed 13 July 2022
41. EU-DEM v1.1 — Copernicus Land Monitoring Service. https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata. Accessed 14 July 2022
42. GeoPandas 0.11.0 — GeoPandas 0.11.0+0.g1977b50.dirty documentation. https://geopandas.org/en/stable/. Accessed 13 July 2022
43. Scikit-learn: machine learning in Python — scikit-learn 1.1.1 documentation. https://scikit-learn.org/stable/. Accessed 13 July 2022
44. Sklearn.preprocessing.MinMaxScaler — scikit-learn 1.1.1 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html. Accessed 18 July 2022
45. Sklearn.preprocessing.LabelEncoder — scikit-learn 1.1.1 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html. Accessed 18 July 2022
46. Sklearn.cluster.KMeans — scikit-learn 1.1.1 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html. Accessed 18 July 2022
47. Welcome to kneed's documentation! — kneed 0.6.0 documentation. https://kneed.readthedocs.io/en/stable/. Accessed 13 July 2022
48. Sklearn.metrics.davies_bouldin_score — scikit-learn 1.1.1 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.davies_bouldin_score.html. Accessed 18 July 2022

49. Wen, R., Yan, W., Zhang, A.N.: Weighted clustering of spatial pattern for optimal logistics hub deployment. In: 2016 IEEE International Conference on Big Data (Big Data), pp. 3792–3797. IEEE (2016). https://doi.org/10.1109/BigData.2016.7841050
50. Cai, C., Luo, Y., Cui, Y., Chen, F.: Solving multiple distribution center location allocation problem using K-means algorithm and center of gravity method take Jinjiang district of Chengdu as an example. In: IOP Conference Series: Earth and Environmental Science, vol. 587, No. 1, p. 012120. IOP Publishing (2020). https://doi.org/10.1088/1755-1315/587/1/012120