Analyzing Last Mile Delivery Operations in Barcelona's Urban Freight Transport Network

Burcu Kolbay[1(✉)], Petar Mrazovic[2], and Josep Llus Larriba-Pey[1]

[1] DAMA-UPC Data Management, Universitat Politecnica de Catalunya,
C/Jordi Girona, 1 3 UPC Campus Nord, 08034 Barcelona, Spain
{burcu,larri}@ac.upc.edu

[2] Department of Software and Computer Systems,
Royal Institute of Technology, Stockholm, Sweden
mrazovic@kth.se
http://www.dama.upc.edu/en
http://www.kth.se

Abstract. Barcelona has recently started a new strategy to control and understand Last Mile Delivery, AreaDUM. The strategy is to provide freight delivery vehicle drivers with a mobile app that has to be used every time their vehicle is parked in one of the designated AreaDUM surface parking spaces in the streets of the city. This provides a significant amount of data about the activity of the freight delivery vehicles, their patterns, the occupancy of the spaces, etc.

In this paper, we provide a preliminary set of analytics preceded by the procedures employed for the cleansing of the dataset. During the analysis we show that some data blur the results and using a simple strategy to detect when a vehicle parks repeatedly in close-by parking slots, we are able to obtain different, yet more reliable results. In our paper, we show that this behavior is common among users with 80% prevalence. We conclude that we need to analyse and understand the user behaviors further with the purpose of providing predictive algorithms to find parking lots and smart routing algorithms to minimize traffic.

Keywords: Urban freight · Clustering · Partitioning Around Medoids · User behavior · Smart City · AreaDUM

1 Introduction

Barcelona is considered to be among the smartest cities in the planet. The IESE ranking [1] puts the city in position 33 with a significant amount of projects carried on. It is not necessarily the technology which makes Barcelona smart; the economy, environment, government, mobility, life and people are other indicators which help defining the city as smart.

Barcelona released an urban mobility plan for 2013/2018, where the need for a smart platform was pointed out in order to improve the efficiency, effectiveness and compatibility of freight delivery areas and the distribution of goods to reduce

possible incompatibilities/frictions with other urban uses [2]. Thus, in November 2015, the AreaDUM project was provided by public company *Barcelona Serveis Municipals (B:SM)* to serve the need [3,4].

AreaDUM (Area of Urban Distribution of Goods, Area de Distribucio Urbana de Mercaderies in catalan) intends to develop parking management in such a way that both freight delivery vehicle drivers and the city obtain a benefit. AreaDUM has several components and features:

– Uniquely identifies parking spaces indicated with zig-zag yellow lines in the streets of the city that can only be used by freight vehicles at certain times of the day (usually from 8:00 till 20:00).
– A maximum time to use the AreaDUM spaces (usually 30 min).
– A mobile app that every freight vehicle driver must install in their cellular.
– The enforcement for each vehicle driver to perform a check-in action with the mobile app every time their vehicle is parked in an AreaDUM space.
– It is forbidden to perform consecutive check-ins in the same Delivery Area.
– The analysis of the data collected.

In this paper, and based on the components of AreaDUM, we provide a set of analyses that we discuss in order to understand the user behaviors. The analyses show that the original data has a significant number of check-ins that behave in a special way, i.e. they are done in the same or close by locations to the original one, with different possible reasons. We detect those cases, analyse them and compute clusters of the parking actions, showing that the behaviour of the users is different from that one would expect with the complete dataset. Our conclusions show that there are significant differences among different quarters in the city, calling for further analytics that describe the actual use of the city and allow for a detailed understanding of each AreaDUM parking space and how they are re-dimensioned based on the data obtained.

The rest of the paper is organized as follows. In Sect. 2, we provide an account of the related work. In Sect. 3, we describe the data generated and used. In Sect. 4, we give an overview of the methods used. Then, in Sect. 5, the experiments and results are detailed. Finally, in Sect. 6, we conclude and make remarks about our future work.

2   Related Work

The demand for goods distribution increases proportional to the population, number of households, and development in tourism. There is a lot of research related to the management of urban freight in cities. Those include solutions for pollution, carbon creation, noise, safety, fuel consumption, etc. The main purposes are generally shaped around reducing travel distances (vehicle routing algorithms) and minimizing the number of delivery vehicles in the city [5,6].

Other pieces of work are focused on what restrictions should be applied to vehicle moves in order to control the congestion and pollution level [7]. One of the most common restriction is the time access restrictions for loading/unloading

areas [8]. By finding optimal solutions for urban freight management, it is possible to reduce the pollution and traffic congestion, and minimizing fuel use and Carbon emissions. With this purpose, we believe that it is important to understand the vehicle drivers and manage their mobility for their satisfaction. We base our analysis in the observation of the user behaviors for loading/unloading trucks, rather than stablishing punishment policies for the drivers. Providing solutions comes after the problem detection and analysis. This is what we do in this paper, we observe the user behaviors, think about possible reasons of the behaviors and propose solutions in order to keep win-win strategies for the city.

3   Data

The data set used in this study was obtained through a web service which is used to export the data of the AreaDUM application (or SMS) developed by B:SM[1]. The time span of the available AreaDUM data sets ranges from January 1st, 2016 to July 15th, 2016. The sample data set consists of roughly 3.7 million observations described using 14 attributes. Some attributes are not relevant since they include information of the AreaDUM application itself. The most relevant attributes for each check-in, apart from the specific Delivery Area ID, are:

– Configuration ID, which tells us about the days when each Area can be used, the number of parking slots and their size, the amount of time a vehicle can be parked and the use times for similar Delivery Areas.
– Time, which tells us about the time, day of the week and date of the check-in.
– Plate number, which contains a unique encrypted ID for each vehicle.
– User ID, which links the vehicle with a company.
– Vehicle type, which describes the size and type of vehicle: truck, van, etc.
– Activity type, which describes whether the objective is to carry goods, or to perform street work, etc.
– District and Neighborhood ID, which tell us about the larger and smaller administrative geographical area of the AreaDUM parking slot.

After some data cleansing, we ended up with 14 attributes which include: Delivery Area ID, Plate Number, User ID, Vehicle Type, Activity Type, District ID, Neighborhood ID, Coordinate, Weekday, Date, Time.

4   Methods

One of the objectives of the paper is to understand the rough data provided in order to cleanse it if necessary. By exploring it, we noticed that there are a significant number of check-ins by the same vehicle ID, in the same or close by Delivery Areas during one day. This is an abnormal behavior because AreaDUM does not allow making consecutive check-ins in the same Delivery Area. However,

---

although consecutive check-ins in close by areas are not forbidden, it would be interesting to isolate them.

Thus, we first create a brand new attribute that we name *Circle ID*. The Circle ID will allow us to detect check-ins in close by areas. Thus, we will be able to isolate the abnormal check-ins from those of other vehicles, allowing the cleansing and the study of the those check-ins in an isolated way.

4.1   Creation of a Circle ID Attribute

The Circle ID attribute is needed since we want to group close loading/unloading areas by distance. Because of the square-shaped blocks in Barcelona, loading/unloading areas at the block corners are close to each other, and the maximum distance is 46 m among corners in "Eixample of Barcelona" by design in "Pla Cerda" from the XIX century.
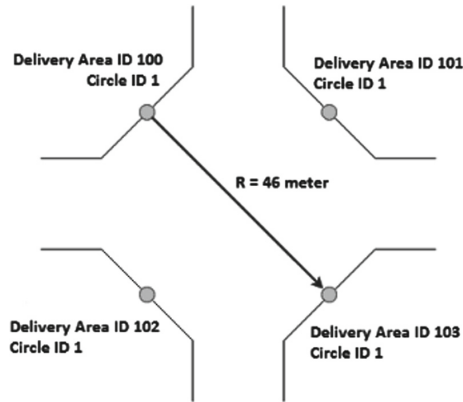
Fig. 1. Square-shaped Block and Location of Loading/Unloading Areas in the "Eixample" of "Pla Cerda"

We think that it does not make any sense for a user to iterate among the corners of a crossing of the "Pla Cerda" grid. It will very seldom happen that a user will go to the opposite corner of a crossing to make a new delivery since the distance is very short. In the case that they do iterate, we need to understand the underlying reason for this.

In Fig. 1, we can see that there 4 loading/unloading areas, and they have their corresponding *Delivery Area ID*s, whereas they have the same *Circle ID*. In order to achieve this, we calculated pair-wised *Haversine* distance among all loading/unloading areas in Barcelona. Haversine is the chosen method to approximate the earth as a sphere, since it works good both for really small and large distances [11].

The distance matrix is created using Haversine formula, where each row and column represents a *Delivery Area ID*. From this distance matrix, we extracted

the pairs of Delivery Area IDs with distance less than or equal to 50 m. If the extracted pairs have a common element, we combined these pairs and removed the common one in order to have only unique elements. After the combination process, we check the distance between the first and the last element in the list. If their corresponding distance value in the distance matrix is less than or equal to 100 m, we keep the last element, otherwise we remove it. As a last step, we assigned the same id for the delivery areas which are located in the same group.

4.2   Clustering

The next step is to cluster the behaviour of the vehicles by Neighbourhood. The *Hopkins Statistics* are applied here as a beginning step to see if the data is clusterable. The value of 0.1829171 from Hopkins statistics showed us that we can reject the null hypothesis and conclude that the data set is significantly clusterable [9]. Then, a clustering algorithm was needed in order to group similar neighborhoods by hourly check-ins frequencies in Barcelona. Most of all the clustering techniques (e.g. k-means, Partitioning Around Medoids, CLARA, hierarchical, AGNES, DIANA, fuzzy, model-based, density-based and hybrid clustering) were used for a comparison on the accuracy of results in order to choose the best for our data.

4.3   The Partitioning Around Medoids (PAM) Clustering Algorithm

PAM is a clustering algorithm like k-means in such a way that it breaks the data into smaller groups which are called clusters, and then it tries to minimize the error [10]. The difference is that k-means works with centroids whereas PAM works with medoids[2]. K-means uses centroids as representatives and minimize total squared error. On the other hand, PAM uses the objects in dataset themselves as representatives. We use PAM instead of K-means since K-means is highly sensitive to outliers and it is not suitable for discovering clusters of very different size.

After k representative objects are arbitrarily selected, a *swap* operation is performed for each medoid and for each non-medoid, and it continues until there is no improvement in the quality of clustering. The cost function is the difference in absolute value of error that appears on a swap operation, and it has to be the lowest to be chosen. In a nutshell, the main goal of PAM is minimizing the sum of dissimilarities of the observations to their corresponding representative objects.

5   Experiments and Results

In our data set, attribute Configuration ID holds the rules for parking (i.e. which day, which hour and how many minutes users can use the delivery areas, etc.).

---

[2] A medoid is a representative object of a dataset or a cluster with a data set whose average dissimilarity to all the objects in the cluster is minimal.

Using this attribute, we are able to check if a recorded delivery happened in the right day, right time etc. The *disallowed repeated check-ins* where detected through this attribute.

The rationale that we understand for those repeated check-ins is as follows:

– Some vehicles are used in household or street work. The time required for the work takes longer than the maximum allowance, and the workers keep doing abnormal check-ins.
– There can be some local store owners who have their own vehicles for their own transport of goods. It is possible that they face the problem for finding a parking slot. The reason of this situation is a necessity instead of an occupation purpose.
– The users just use the spaces as free parking for different purposes like having breakfast after a delivery, etc.

Type of Activities for Disallowed Repeated Check-Ins. In our data set, column Activity ID represents the activity type of the delivery. There are 6 different types: Public Work, Carpentry, Installation, Furniture, Transport, and Others. The results presented in this section confirm the types of reasons assumed above for repeated check-ins, as shown in Table 1.

Table 1. The percentage of activity types' disallowed repeated check-ins

Type of activity	Disallowed repeated check-ins
Public work	30.8%
Carpentry	7.0%
Installation	27.2%
Furniture	1.8%
Transport	19.9%
Others	13.3%

The results show that Public Work and Installation have higher percentages of disallowed check-ins than the others. This shows that professionals who spend time in specific locations, need some type of parking space that allows them managing their tempos in a better way. Transport also show quite a high number of disallowed check-ins, which may well be showing the case for local store owners who repeat their check-ins to preserve their parking space.

5.1   The Effect of Disallowed Repeated Check-Ins

In any case, being disallowed or not, the repeated check-ins in close-by Delivery Areas can be removed using the Circle ID explained above.

The new *Circle ID* that we computed created a total of 1484 circle areas, whereas we still have 2038 different delivery areas. The combination of both IDs allowed us for the analysis in the following paragraphs.

In this section, we present the effect of disallowed repeated check-ins removal using *Circle ID*. Table 2 shows the percentage of disallowed repeated check-ins per each district in Barcelona. In other words, these are the percentages of data we lose, in case that we remove the disallowed repeated check-ins occurred in the same circle.

**Table 2.** The percentage of disallowed repeated check-ins

District name	Percentage of data lost
Ciutat Vella	27.5%
Eixample	27.4%
Sants Montjuic	29.4%
Les Corts	27.6%
Sarria Sant Gervasi	29.2%
Gracia	28.9%
Horta Guinardo	30.4%
Sant Andreu	28.1%
Sant Marti	28.2%

After we see the percentages of disallowed repeated check-ins, it is easy to say that there must be some effects on possible preliminary analyses. We want to see how the clustering results will change in the case that we re-cluster neighborhoods by hourly check-ins frequency.

Neighborhood Clustering: Before vs. After. In our analysis, there are 43 target neighborhoods which are associated with 9 different neighborhoods in Barcelona. We do PAM clustering two times. One for the original data set Before removing disallowed repeated check-ins and one After disallowing repeated check-ins.

For the clustering Before removing the disallowed check-ins, PAM clustering selected two Neighborhood medoids among the other observations in data. After that, PAM assigned each observation to the nearest medoid. These two neighborhoods are the representative objects which minimize the sum of dissimilarities of the observations to their closest representative objects.

Figure 2 shows the clusters obtained by PAM. The plot on the left of Fig. 2 is a 2 dimensional clustering plot which is done by *Principal Component Analysis*. It represents how much of the data variability is explained by a reduced dimension of principal components which are not correlated to each other. The plot on the right of Fig. 2 represents the silhouette widths which shows how the observations
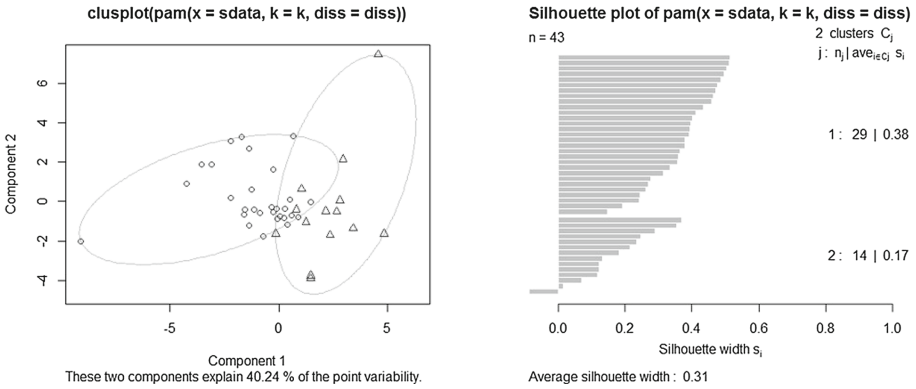
Fig. 2. PAM Clustering Results for the dataset before the disallowed check-ins were removed.

are well clustered. At the bottom of this plot, there is one horizontal line that located on the left side, which represents a misclustered object.

Two clusters are determined by PAM:

– Cluster 1 consists of 29 neighborhoods from 9 districts,
– Cluster 2 consists of 14 neighborhoods from 7 districts.

All neighborhoods from 2 districts (i.e. Gracia and Sant Andreu) are located into Cluster 1, whereas the other districts' neighborhoods are divided into two clusters.

Figure 3 shows the clusters after removing the disallowed repeated check-ins. In this case, the number of clusters increased to 9, and it shows us that variation is significant. On the left side of Fig. 3, there are some silhouette width values of 0, and these clusters have only 1 observation in their clusters. We can basically say that they are both representative objects for themselves. The ones who are
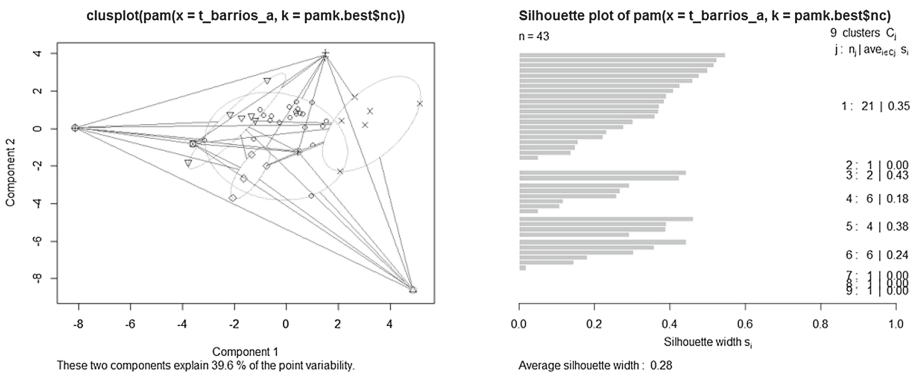
**Fig. 3.** PAM clustering results for data without disallowed repeated check-ins

located alone in the clusters tell us that these neighborhoods are quite different by hourly check-ins frequency than the others after the removal of disallowed repeated check-ins.

Nine clusters are determined by PAM:

– Cluster 1 consists of 21 neighborhoods from 7 districts,
– Cluster 2 consists of 1 neighborhood,
– Cluster 3 consists of 2 neighborhoods from 2 districts,
– Cluster 4 consists of 6 neighborhoods from 4 districts,
– Cluster 5 consists of 4 neighborhoods from 4 districts,
– Cluster 6 consists of 6 neighborhoods from 3 districts,
– Cluster 7 consists of 1 neighborhood,
– Cluster 8 consists of 1 neighborhood,
– Cluster 9 consists of 1 neighborhood.

Proliferation of Disallowed Repeated Check-Ins. Up until this point, we have detected *disallowed repeated check-ins*, the effect of their removal, and the activity types which cause this situation. We know that 28% of check-ins corresponds to this behavior. The only thing we do not know is that how common it is among the deliverers. Figure 4 shows the results for this. The disallowed check-in practice has grown significantly as time passes, which means that, possibly, social networks or communication among drivers have worked very well.
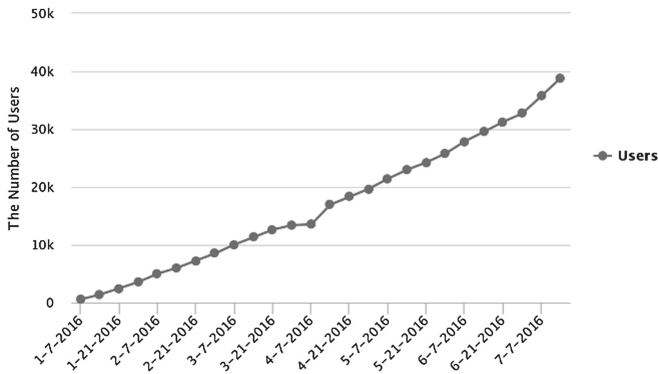
Fig. 4. The proliferation of disallowed repeated check-ins among deliverers

6   Conclusion

Among different components of urban mobility, freight transport is usually considered as the least sustainable. This naturally calls for new technologies, such as mobile data management and analytics, to achieve more efficient freight distribution systems, without the need for large investments or sophisticated sensor technologies. However, in smart and citizen-oriented cities, it is also important to

understand the citizens reactions to new technologies. Therefore, in this paper, we focused on deliverers as the main actors in the urban freight transport scheme recently deployed in the city of Barcelona.

We showed that people often look for different ways to bypass the intended use of new technology for their needs, and it can cause undesired effects. In our experimental study, we demonstrated the scope and significance of non-compliance of the parking regulations. For example, after filtering out the disallowed check-ins, we lost 28% of our data, and consequently increased the number of clusters from 2 to 9. Interestingly, we showed that *Public Work* and *Installation* activities which can be related to city governance, are usually associated with larger number of disallowed check-ins. Finally, one of the most important results of our study is the statistical proof that non-compliance of the introduced regulations is not an exception, but a common behavior. For the future work, the city governance needs to solve the issue using different *Configuration ID* for local store owners' vehicles, assigning new parking lots, or categorizing the parking lots into different purposes (e.g. short visit delivery, daily permission etc.).

References

1. New York Edges Out London as the World's "Smartest" City. http://ieseinsight.com/doc.aspx?id=1819&ar=6&idioma=2
2. Ajuntament de Barcelona. http://ajuntament.barcelona.cat/en/
3. Barcelona Serveis de Municipals (B:SM). https://www.bsmsa.cat/es/
4. AreaDUM Project. https://www.areaverda.cat/en/operation-with-mobile-phone/areadum/
5. Hwang, T., Ouyang, Y.: Urban freight truck routing under stochastic congestion and emission considerations. Sustainability **7**(6), 6610–6625 (2015)
6. Reisman, A., Chase, M.: Strategies for Reducing the Impacts of Last-Mile Freight in Urban Business Districts. UT Planning (2011)
7. Yannis, G., Golias, J., Antoniou, C.: Effects of urban delivery restrictions on traffic movements. Transp. Plan. Technol. **29**(4), 295–311 (2006)
8. Quak, H., de Koster, R.: The impacts of time access restrictions and vehicle weight restrictions on food retailers and the environment. Eur. J. Transp. Infrastruct. Res. (Print) 131–150 (2006)
9. Banerjee, A., Dave, R.N.: Validating clusters using the Hopkins statistic. In: Proceedings of the IEEE International Conference on Fuzzy Systems, vol. 1, pp. 149–153 (2004)
10. Kaufman, L., Rousseeuw, P.J.: Clustering by means of medoids. In: Dodge, Y. (ed.) Statistical Data Analysis Based on L1 Norm, pp. 405–416 (1987)
11. Shumaker, B.P., Sinnott, R.W.: Astronomical computing: 1. Computing under the open sky. 2. Virtues of the haversine. Sky Telesc. **68**, 158–159 (1984)