The study of hypermobility through the lenses of big data and machine learning

Yacouba Kone*[1] and Ouassim Manout[1]

[1]Transport, Urban Planning and Economics Laboratory (LAET), ENTPE-CNRS-University of Lyon, France.

Short summary

Cities are facing many mobility challenges. Between the increasing use of private vehicles and the challenges posed by climate change, many cities are wondering how to reconcile private vehicles and public transport. A great deal of research is being carried out, but in most cases, it focuses on how mobility is achieved but does not examine the reasons why this mobility is necessary. This paper explores the concept of hypermobility, defined as excessive mobility. A case study was carried out in the city of Lyon to identify and characterize hypermobility using a unique GPS data set.
Keywords: hypermobility, travel behavior, GPS Data, Big data, Floating Mobile Data, Artificial Intelligence.

1 Introduction

Cities around the globe are facing serious challenges in meeting their mobility needs while reducing their consequent negative externalities. Most of these are due to the widespread use of private cars. Public authorities, with the help of research, are consequently pursuing various paths to mitigate these externalities. Most of these solutions focus on how mobility is performed but do not question why this mobility is needed. Even the most ambitious decarbonization pathways limit their scope to making cars less pollutant and to shifting travel demand to less pollutant or more efficient mobility services like public transport. As such, these pathways are challenging to climb. However, these pathways rarely envisage the need to question mobility needs, especially when unsustainable travel modes are necessary. Yet, research can rightly question excessive mobility practices and specifically hypermobility. This research aims to understand hypermobility practices and its characteristics using a unique GPS dataset from 5% of the French population. Findings of this research can help understand the present and future of this practice.

There is no consensual definition of the concept of hypermobility in the transportation literature. Hypermobility is often defined as an *excessive* mobility in opposition to *normal* mobility. This concept is often discussed by transportation sociologists and anthropologists in relation to lifestyles, social and network capital (Cohen & Gössling, 2015; Khisty & Zeitler, 2001). However, to observe and investigate hypermobililty as a mobility behavior, there is a need for a clear definition of this concept. We define a person as a hypermobile if he/she performs more than twice the average number of trips per day and per person at least half of the days when he/she was observed. In the case of France, the average number of trips per person and day is around 4. In this case, a hypermobile person should perform at least 8 trips per day during at least half of the tracked days. The threshold of 8 trips is also adopted by the French standard of household travel surveys (HTS) (SYTRAL, 2015) to define hypermobile profiles. Nevertheless, this standard does not consider the time recurrence of this behavior, i.e. 50% of the days, since French HTS are a one-day snapshot of mobility routines. Contrary to HTS, this research relies on a unique dataset of GPS tracks from nearly 5% of the French population over one month to investigate hypermobility.

2 Methodology

Data

In this research, we exploit a unique dataset of GPS tracks collected all over France for a full month (January 2023) without interruption. The data is collected by hundreds of smartphone applications (weather condition apps, sports, cooking, news, etc.) targeting a large population. This massive database contains around 4 billion observations from 3.4 million different devices/persons. Each observation records the GPS trajectories of tracked devices over a month. These records are characterized by a set of variables, like device ID, latitude, longitude, precision, and timestamp.

Processing this data requires the development of a pipeline of different algorithms for formatting and enriching the raw data. After data cleaning, a method is developed for trajectory segmentation into trips, which is based on criteria like the stationarity time, the number of GPS records and the stationarity radius. These trips are then enriched by other variables such as trip distance, travel time, stationarity time, departure time, arrival time, mobility rate, day of the week, visited places, etc.

Despite its volume, collected GPS data is not representative of the French population. Even if 77% of the French people have a smartphone (INSEE, 2021), observations are not sampled randomly. To correct this bias, a weighting correction is needed. In this abstract, we use a simple reweighing procedure based on the population ratio. A more sophisticated method is under study.

To show the potential of these data for the study of mobility behavior in general, and hypermobility in particular, we will focus on the region of Lyon, home to nearly 1.8 million inhabitants. Our dataset contains 2,007,450 trips from 80,000 unique devices that we deliberately assume as individuals. The comparison between the most recent HTS of Lyon of 2015 and the GPS data shows that the latter is capable of accurately replicating the number of trips per person and their distances as well as the relative OD matrices. The comparison of these two data sources is still ongoing. A comparison with more up-to-date HTS data from other cities is to come.

Methods

To investigate hypermobility, our research methodology is twofold: an exploratory analysis of hypermobility is first conducted, and then a discrimination analysis is run to identify the main aspects differentiating between hypermobile (HM) and non-hypermobile (NHM) individuals. The exploratory analysis compares the mobility profiles of HM and NHM using various indicators like the number of trips per day, distance, activity times, frequency of place visits, workday and weekend patterns. The discrimination analysis identifies the main differentiating characteristics of the two groups, i.e. HM and NHM, using a machine learning classification algorithm: Light Gradient Boosting Machine (LGBM). LGBM is used to select and rank the important variables that discriminate the most between HM and NHM individuals. LGBM is an assembling model such as Random Forest (RF). This model family is based on a set of decision trees. Unlike RF, LGBM forms a set of decision trees sequentially. The trees are updated by reducing the residual errors at each iteration. Individual decision trees are fitted using a differentiable loss function, a binary log loss and a gradient descent optimization algorithm. The choice of this algorithm was based on its low memory usage and faster learning speed. The hyperparameters are adjusted using a recursive process. First, the numbers of decision trees are adjusted using cross-validation and the early stopping quality present in the model. Then, the other parameters are adjusted using cross-validation. To rank features according to their importance, the algorithm is based on the frequency of feature selection.

3 Results

Exploratory results

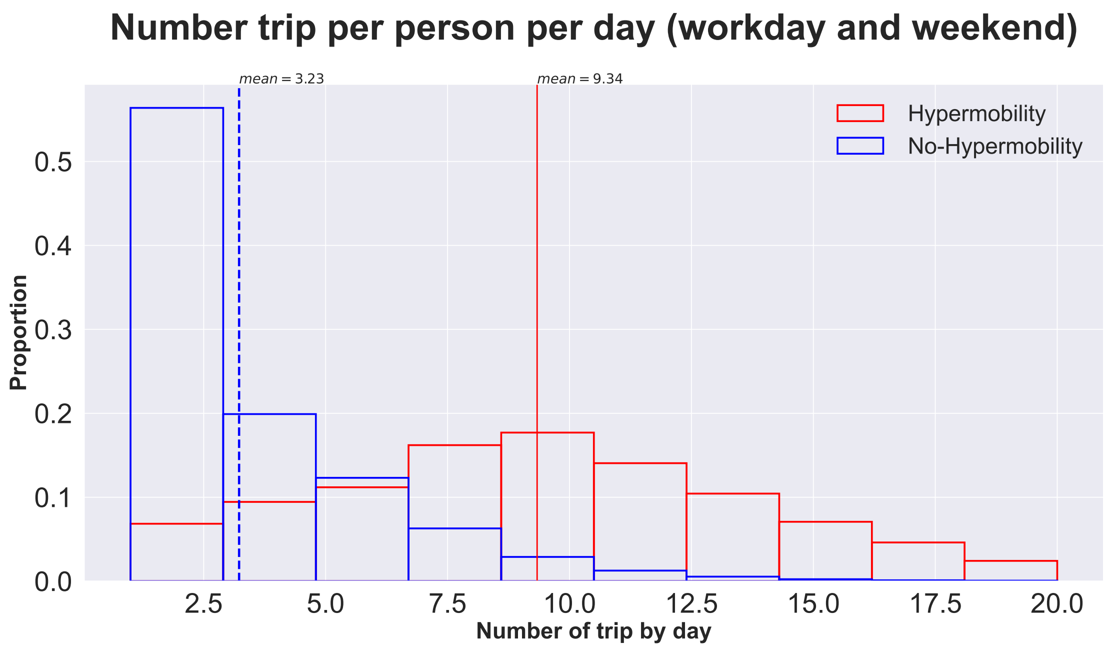
18% of the tracked individuals in Lyon are found to be hypermobile. Most of these live in the city center of Lyon. The comparison of HM and NHM shows significant differences between these groups. The HM group travels on average 9.3 trips per day vs. 3.2 trips for the NHM group (Fig

1a). It is noteworthy that our definition imposes 8 trips per day as a threshold. The difference between HM and NHM is more significant during working days (10 vs. 3) than on weekends (7.3 trips vs 2.9). The HM group performs more trips, but shorter in terms of distance than the NHM group (Fig. 1b). The average trip distance of the HM and NHM group is 5.1 and 7.1 km, respectively. This pattern is similar between work and weekend days, except that the NHM group travels 1km further, on average, during weekends than workdays. The HM group is found to travel 2 km less.

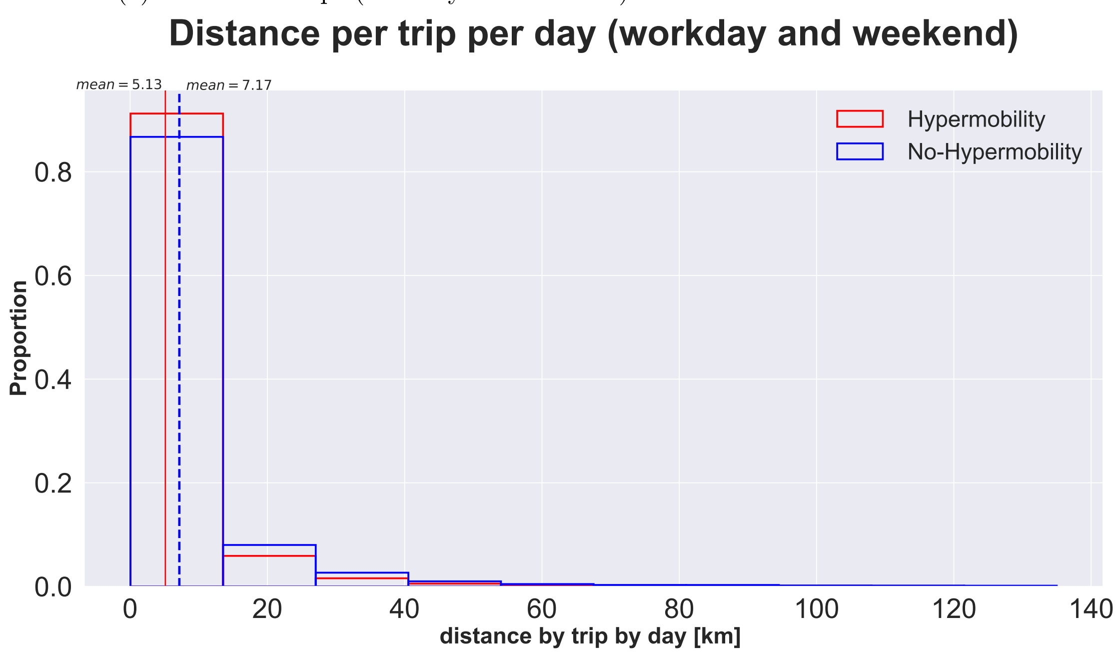
A higher number of trips induces shorter activity durations for the HM group (Fig. 2a). We found that for the NHM group, the average time spent at stops/destinations which is a proxy for activity duration, is nearly double (142 min) that of the HM group (76 min). This pattern is similar between working and weekend days but with an important difference within the same group (16 min, 30 min for NHM and HM respectively).

During 69% of the days, HM individuals are found to be hypermobile (Fig. 2b). Noteworthy is that our definition imposes only 50% of the days as a threshold for hypermobility. During 30% of the days, HM individuals are found to travel less than 8 trips per day. NHM individuals, on the other hand, travel more than 8 trips per day only 3% of the days.

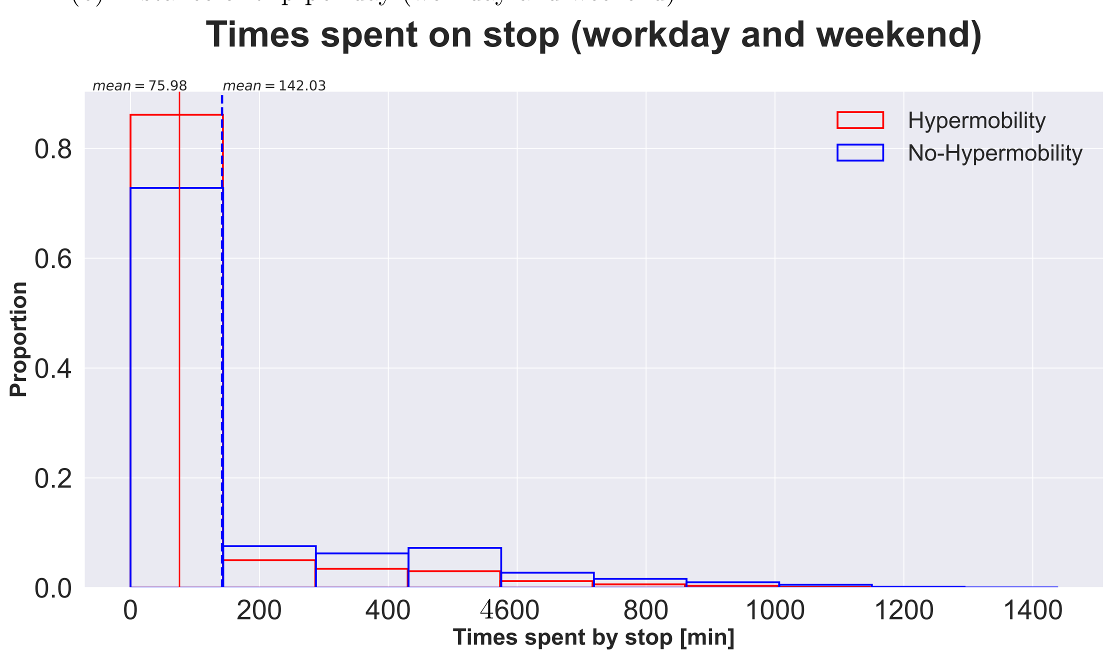
Regarding the hourly distribution of the trips of the two groups, both groups show the presence of 3 daily travel demand peaks (morning, midday and evening). The HM pattern is less marked by the morning and evening peaks than the NHM group. The HM group is also more likely to travel during off-peak periods (before 5 AM, between 9 AM and 12 PM and after 20PM) than the NHM group.

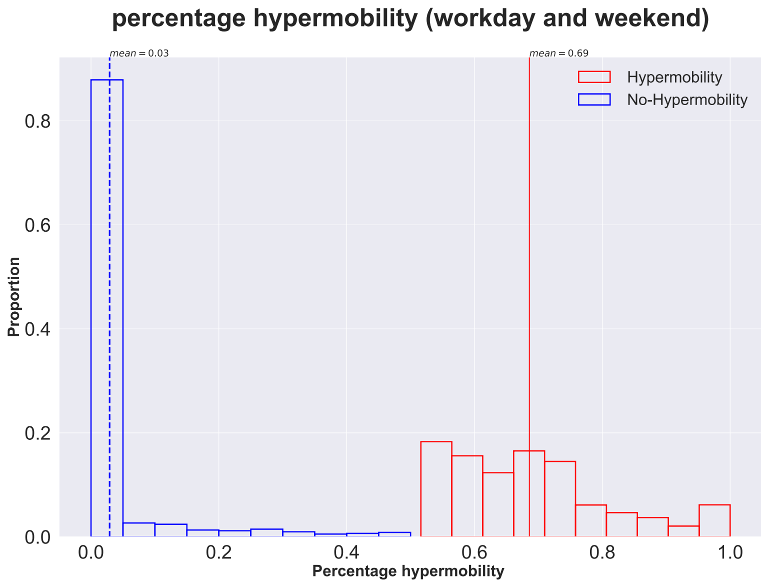


(a) Number of trips (workday and weekend)

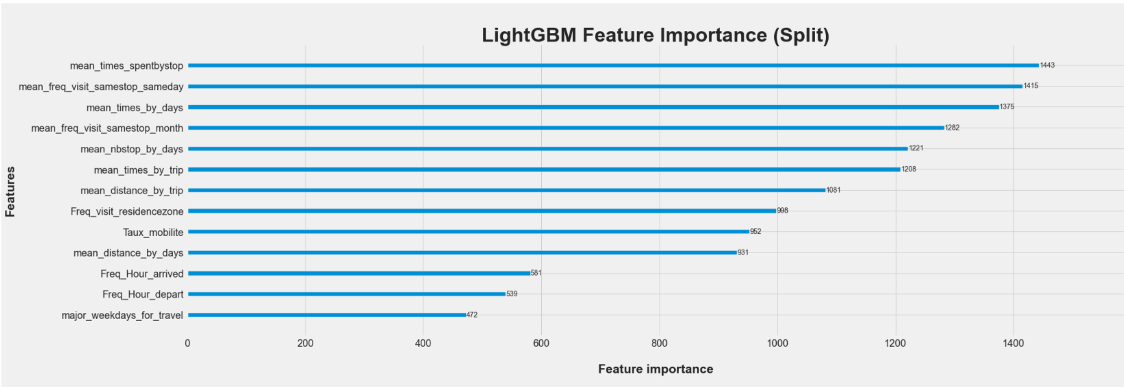


(b) Distance of trip per day (workday and weekend)





(a) Rate of hypermobility in Lyon (workday and weekend)



(b) Features importance

Results of the discrimination analysis

The discrimination analysis identifies various characteristics that are significantly different between HM and NHM. These characteristics are ranked according to their importance.

The algorithm achieved a classification score of 100% with 866 decision trees on the train set (85% of the sample) and 99% on the test set (15% of the sample). This score was used to classify the most important features for distinguishing between HM and NHM. Among all the features, the most significant is the time spent at stops. The more frequently people travel, the shorter the time they spend at their destinations. The frequency of visits to the same destination per day and over a long period is the second and fourth most important discriminator factor, respectively. Trip travel times and distances have a relatively lower power to distinguish between HM and NHM. The day of the week and the time of travel are not very important for distinguishing between MH and NHM.

4 Conclusion

The descriptive analysis shows a very significant difference between HM and NHM. HM individuals are found to perform more than double the number of trips of NHM individuals. These trips are often shorter and induce shorter activity durations. The HM group are also found to visit less frequently places than the NHM group. 30% of the days, HM individuals are less hypermobile, i.e. perform less than 8 trips per day. Inversely, only 3% of the time, NHM individuals are found to be hypermobile, i.e. perform more than 8 trips per day.

The discrimination analysis shows a very significant influence of stop/destination time and the frequency of visits to the same destination per day or over a long period on group membership. In 93% of cases, we were able to distinguish between HM and NHM using only these two variables.

Ongoing research focuses on indicators other than those mentioned in this abstract to compare HM and NHM. This research will be also extended to the whole France to understand the spatial influence on hypermobility. We will also explore the potential of new machine learning methods to better explain the differences between HM and NHM.

Acknowledgements

The authors acknowledge LAET and EGIS for their financial support.

References

Cohen, S. A., & Gössling, S. (2015, August). A darker side of hypermobility. *Environment and Planning A: Economy and Space*, *47*(8), 166–1679. Retrieved from http://dx.doi.org/10.1177/0308518X15597124 doi: 10.1177/0308518x15597124

INSEE. (2021). *94% des 15-29 ans ont un smartphone en 2021.* Retrieved 2024-02-02, from https://www.insee.fr/fr/statistiques/6036909

Khisty, C. J., & Zeitler, U. (2001). Is hypermobility a challenge for transport ethics and systemicity? *Systemic Practice and Action Research*, *14*(5), 597–613. Retrieved from http://dx.doi.org/10.1023/A:1011925203641 doi: 10.1023/a:1011925203641

SYTRAL. (2015). *Enquête ménages déplacements, de l'aire métropolitaine lyonnaise - 2015.* Retrieved from https://www.sytral.fr/474-sur-l-aire-metropolitaine-lyonnaise.htm