Presentation Number: TRBAM-21-03088

Scheduled time: Mon 1/25/2021, 2:30 PM- 4:00 PM ET

A Cluster Analysis of Uber Request Data via Transit app in New York City

Jing Guo, Antora Mohsena Haque, Cassidy Crossland, Candace Brakewood

jguo31@vols.utk.edu, ahaque1@vols.utk.edu, ccrossl3@vols.utk.edu, cbrakewo@utk.edu

Presenter: Jing Guo

THE UNIVERSITY OF TENNESSEE KNOXVILLE

1. Abstract

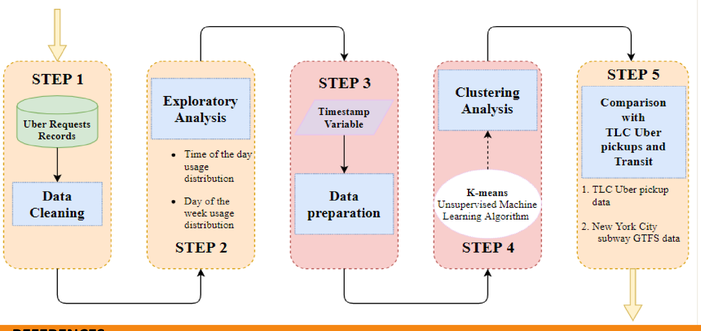
As ridehailing services like Uber become increasingly common in urban transportation systems, it is necessary to understand their usage patterns. Since private ridehailing companies do not publicly disclose their ridership data, usage patterns can be analyzed using other data sources, such as Transit app's Uber request data. The objectives of this research are three-fold: (*1*) explore the temporal characteristics of Uber requests through data visualization, (*2*) identify groups of users through cluster analysis, and (*3*) compare Transit app Uber request data with overall Uber usage data and transit data. The exploratory analysis results suggest that requests occurred most frequently during AM and PM peak periods. K-means clustering identified eight groups of Uber users: long duration and frequent users, off-peak users, PM peak users, AM peak users, party goers, long duration and infrequent users, holiday users, and weekend users. The main trip purposes determined by the clustering analysis were going to social events and to and from the workplace or home. Comparing Transit app data to the overall ridehailing usage data and transit data suggest that the time distribution pattern of Transit app Uber requests is a combination of transit and Uber usage while the time usage features are more similar to those of Uber users. These results will help transportation departments to better coordinate ridehailing services and public transportation to meet users' travel needs.

2. Objectives

1. **Explore the temporal Uber requests pattern** based on the time of day and day of week;
2. **Infer trip purpose and identify distinct groups of users** by **K-means** unsupervised machine learning algorithm; and
3. **Compare temporal characteristics** of **Uber requests** from Transit app with **Uber pickup data** released publicly through the TLC and General Transit Feed Specification (GTFS) service information for the **subway system** in **New York City**.

3. Methodology

Figure 1: Flowchart showing the methodology of this study



4. Data

- **Dataset:** Uber requests obtained from Transit app
- **Study period:** 10.30, 2016 to 10.29, 2017
- **Study area:** New York City

5. Data preparation for clustering

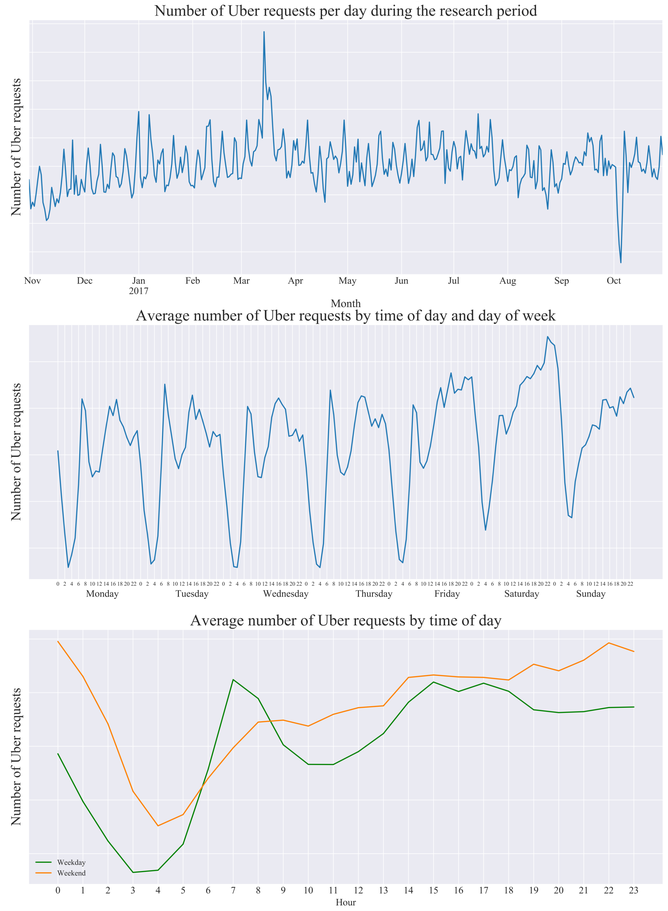
- **Goal for data preparation:**
  - Create new variables to explain the temporal usage patterns of Uber requests
  - Remove one-time users since it is hard to find patterns with only one request
- **New created variables:**

1. Count of Days
2. Duration
3. AM Peak Usage Rate
4. PM Peak Usage Rate
5. Party-time Usage Rate
6. Weekend Usage Rate
7. Holiday Usage Rate

6. Exploratory analysis

Transit app users' Uber requests were mostly concentrated in the AM and PM peak periods during weekdays and late night periods on Fridays and Saturdays.

Figure 2 Temporal Distribution of Uber Requests



7. Clustering analysis

Figure 3: Bubble Chart of Duration and Count of Days for Each Cluster

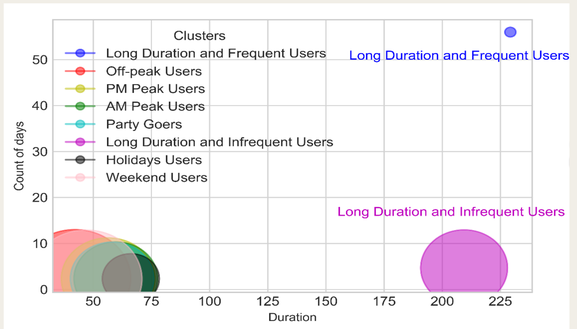


Figure 3 represents the count of days and duration for each cluster where the size of the bubble represents the number of users in the cluster.

Figure 4 describes how the clusters are related to the different time period usage rate variables.

Figure 4: Radar Chart of Usage Time Rate for Each Cluster

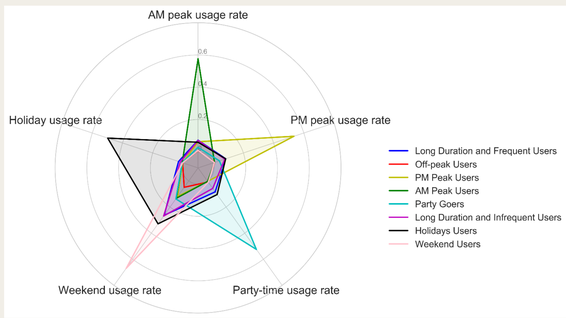


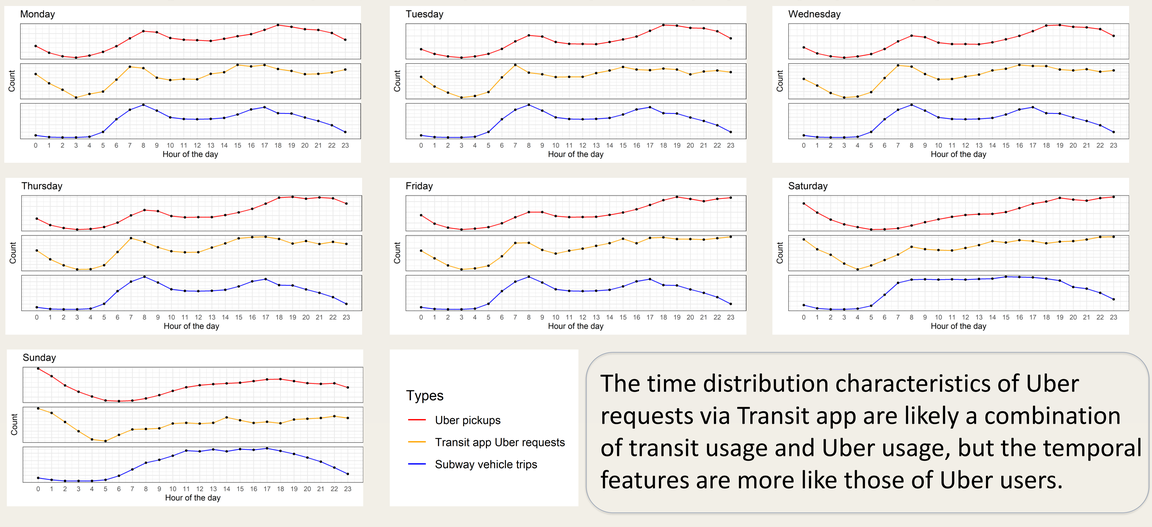
Table 1 Clusters of users and mean value of attributes for each cluster

	Long Duration and Frequent Users	Off-peak Users	PM Peak Users	AM Peak Users	Party Goers	Long Duration and Infrequent Users	Holidays Users	Weekend Users
AM peak usage rate	7.2%	0.7%	6.2%	**57.7%**	2.8%	7.1%	5.9%	0.9%
PM peak usage rate	8.3%	1.0%	**53.5%**	0.9%	4.1%	7.7%	8.3%	0.8%
Party-time usage rate	8.4%	0.8%	0.3%	0.3%	**52.5%**	5.7%	10.4%	1.0%
Weekend usage rate	26.3%	4.9%	11.9%	13.3%	13.8%	26.9%	32.9%	**67.0%**
Holiday usage rate	2.9%	0.2%	0.3%	0.5%	0.4%	1.8%	**49.8%**	0.2%
Percent of Users	0.2%	21.2%	15.0%	12.0%	11.9%	12.9%	5.6%	21.3%

Table 1 summarizes the results for each cluster. **K-means clustering** showed eight distinguishable clusters of users based on the temporal characteristic of Uber requests: long duration and frequent users, off-peak users, PM peak users, AM peak users, party goers, long duration and infrequent users, holidays users, and weekend users.

8. Comparison with TLC Uber pickup data and transit

Figure 5: Uber Requests, Uber Pickups, and Subway Vehicle Trips by Time of Day and Day of Week



The time distribution characteristics of Uber requests via Transit app are likely a combination of transit usage and Uber usage, but the temporal features are more like those of Uber users.

REFERENCES

[1] *Taxonomy and Definitions for Terms Related to Shared Mobility and Enabling Technologies.* SAE International. Accessed July 27, 2020.

[2] Clewlow, R. R., and G. S. Mishra. *Disruptive Transportation: The Adoption, Ytilization, and Impacts of Ride-hailing in the United States*. Research Reports UCDITS-RR-17-07, Institute of Transportation Studies, University of California, Davis CA, 2017.

[3] Brakewood, C., N. Ghahramani, J. Peters, E. Kwak, and J. Sion. Real-time Riders: A First Look at User Interaction Data from the Back End of A Transit and Shared Mobility Smartphone App. *Transportation Research Record,* Vol. 2658, No. 1, 2017, pp. 56-63.

ACKNOWLEDGMENTS:

The authors would like to thank Transit app for providing the data. We are particularly grateful to Jake Sion and David Block-Schachter for reviewing the paper.