UC Irvine
UC Irvine Previously Published Works

Title
Heterogeneity in activity-travel patterns of public transit users

Permalink
https://escholarship.org/uc/item/1m67x54f

Authors
Rafiq, Rezwana
McNally, Michael G

Publication Date
2021-10-01

DOI
10.1016/j.tra.2021.07.011

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Heterogeneity in Activity-travel Patterns of Public Transit Users: An Application of Latent Class Analysis

Rezwana Rafiq[1] and Michael G. McNally[2]

Abstract

Public transit is considered a sustainable mode of transport that can reduce automobile dependency and can provide environmental, economic, and societal benefits. However, with the typical temporal and spatial constraints such as fixed routes and schedules, transfer requirements, waiting times, and access/egress issues, public transit offers lower accessibility and mobility services than private vehicles and thus it is considered a less attractive mode to many people. To improve the performance of transit and in turn to increase its usage, a better understanding of daily activity-travel patterns of transit users is required. This study analyzes transit-based activity-travel patterns by classifying users via Latent Class Analysis (LCA). Using data from the 2017 National Household Travel Survey, the LCA model suggests that the transit users can be divided into five distinct classes where each class has a representative activity-travel pattern. Class 1 constitutes Caucasians employed males who make transit-dominant simple work tours. Class 2 is composed of Caucasian females who make complex work tours. Caucasian employed millennials comprise Class 3 and make multimodal complex tours. Transit Class 4 are non-Caucasian younger or older adult groups who make transit-dominant simple non-work tours. Last, Class 5 members make complex non-work tours with recurrent transit use and comprise single older women. This study will help transit agencies to understand the activity-travel patterns of various transit user groups and to consider market strategies that can address their travel needs.

Keywords: Public transit users; activity-travel patterns; tours; latent class analysis; NHTS

This is a draft of a paper submitted for a review for "Transportation Research Part A: Policy and Practice"

[1] Corresponding Author, Assistant Project Scientist, Institute of Transportation Studies (ITS), University of California, Irvine, CA 92697-3600, USA, Email: rrafiq@uci.edu
[2] Professor, Department of Civil & Environmental Engineering, Institute of Transportation Studies (ITS), University of California, Irvine, CA 92697-3600, USA, Email: mmcnally@uci.edu

Introduction

Public transit is considered a sustainable mode of transport that can reduce automobile dependency and thus can mitigate some of the negative consequences of automobile use, including congestion, air pollution, and energy consumption (Federal Highway Administration, 2018). However, with operations based on fixed routes and fixed schedules, public transit offers lower flexibility and mobility services than automobiles, particularly in satisfying complex travel needs (Hensher and Reyes, 2000) and thus is considered a less attractive mode to many people. A better understanding of daily activity-travel patterns of transit users is needed to allow transit operators to evaluate their services and to implement strategies to attract more people to transit. This study investigates the complex activity-travel patterns and tours of transit users. Here, the term *pattern* refers to a complete sequence of activities (in-home and out-home) and trips made by an individual over a full day whereas *tour*, a basic unit of a full pattern, is defined as a sequence of trips that begins and ends at the same location (here, at home) and contains single or multiple activities. A detailed classification of tours is available in Rafiq and McNally (Rafiq and McNally, 2020a).

In recent years, a wealth of research has been completed that has focused on techniques to extract information on transit user's daily activity-travel patterns by mining transit smart card data (Ma *et al.*, 2013; Ma *et al.*, 2017; Bhaskar and Chung, 2014; Morency *et al.*, 2007; Chu and Chapleau, 2010; El Mahrsi *et al.*, 2014; He *et al.*, 2020). These studies mostly covered the data-mining procedure but did not capture the user's actual activity-travel patterns, with a few exceptions (e.g., Goulet-Langlois *et al*., 2016). Also, the insights on activity-travel patterns are derived either from Australian, Asian, Canada, or European contexts. Thus, our knowledge of activity-travel patterns of transit users in the US context has been limited. Our goal in this study is to address this research gap.

We posit that despite the complexity of an individual's activity-travel patterns, transit users might fall into a small number of heterogeneous sub-groups, each of which has a representative activity-travel patterns. The purpose of this study is to analyze the heterogeneity in activity-travel patterns of transit users by classifying them in such a way that demonstrates similar activity-travel patterns within a class but different between classes. The findings will help transit operators to identify transit user groups with representative activity-travel patterns and to develop policies to address user travel needs and to encourage higher transit usage.

2. Data and Sample

This study analyzes data from the 2017 National Household Travel Survey (NHTS), a source of information on travel by US residents in all 50 states and the District of Columbia. This survey sponsored by Federal Highway Administration includes data on trips made by all modes of travel (private vehicle, public transportation, pedestrian, biking, etc.) and for all purposes (travel to work, school, recreation, personal/family trips, etc.). The dataset contains the following four data tables:

- Households (socio-economic and location characteristics of surveyed households)
- Persons (demographic characteristics of all household members)
- Trips (over 24-hours by all household members 5 or older and trip-related attributes)
- Vehicles (vehicles used by the responding households)

The NHTS dataset contains 129,696 households consisting of 264,234 persons who took a total of 923,572 trips. For this study, we identified *public transit users* as those individuals who start their first trip from home and ends their last trip at home and used public transit for at least one trip segment[3]. A choice of travel mode is treated as public transit if it is any of the following: public or commute bus, city-to-city bus (greyhound, Mega bus, etc.), Amtrak/commuter rail, and subway/elevated/light rail/streetcar. This generates a sample of 4,994 individuals who made a total of 20,222 trips where almost half of the trips are made by transit (10,011).

2.1 Socio-demographic Characteristics

Table 1 summarizes the household, personal, and location characteristics of the selected transit users who used a transit mode in at least one trip segment.

[3] When a trip involves change of modes, each mode defines a trip segment.

Table 1: Descriptive statistics of NHTS transit users (N = 4,994)

Variables	**Percentage of users (%)**
Household characteristics	
Household size	
Household size = 1	29.4
Household size = 2	34.7
Household size > 2	35.9
Number of household vehicles	
Number of vehicles = 0	36.2
Number of vehicles = 1	29.7
Number of vehicles > 1	34.1
Monthly household income (USD)	
Low income (less than $35K)	37.3
Middle income ($35K to $100K)	29.2
High income ($100K or more)	31.2
Presence of child aged 0-17	19.0
At least one vehicle per licensed driver	48.1
Personal characteristics	
Age groups	
Younger group (below 18 years)	6.6
Millennials (18 – 38 years)	33.8
Generation X (38 – 58 years)	32.3
Older adults (more than 58 years)	26.1
Gender: Male	48.6
Employment status: Employed	62.2
Race: Caucasian	59.3
Type of transit use	
Commuter rail	42.7
Public bus	62.4
Location characteristics	
Population density (persons per sq. mile) in census block group	
Low density (0-2000)	17.1
Medium density (2000-10000)	42.5
High density (>10000)	40.4
MSA has a rail connection	50.7

In terms of household characteristics, a larger fraction of transit users have more than two persons per household (35.9 percent) and belong to a lower income group (annual income less than $35K USD) (37 percent). Few of these households have children aged 17 years or lower (19 percent) and 51.9 percent are car deficient households (less than one car per licensed driver). The age distribution of transit users is similar for millennials (18 - 38 years) and Generation X (38 - 58 years) people and there are a considerable fraction of older adults among users (26 percent). The majority of the transit users are Caucasians (59.3 percent), employed (62.2 percent), and live in medium to high-density areas. Although APTA (2017) reported that Caucasian riders represent the largest group of riders consisted of 40 percent of all riders, our finding (59.3

percentage) from the NHTS 2017 dataset is consistent with Grahn *et al.*, (2019) who used the same dataset in their study.

2.2 Trip Characteristics

This section discusses the characteristics of individual trips made by transit users. Figure 1 shows that transit is utilized for a considerable fraction of work (24 percent) and return home trips (38 percent). Shopping or running errands (14 percent) is also a common trip purpose of transit. Only 5 percent of trips are made by transit to go to school or religious centers. Note that we did not consider school bus as a public transit category. Transit is occasionally used for transporting someone (pick up/drop off) or going to a restaurant or medical center. Similar results are found in APTA (2017).

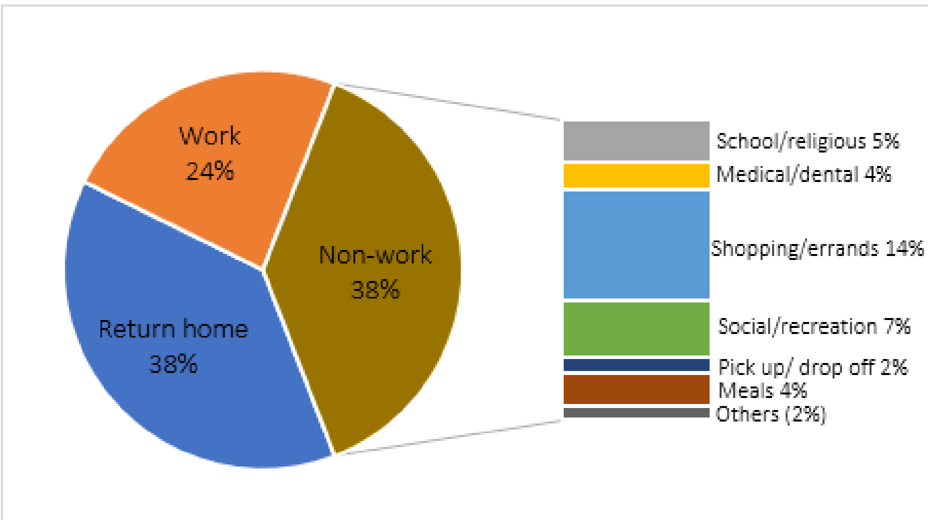


Figure 1: Distribution of transit trips by activity purposes

Next, we investigate how the demand for transit trips for three activity purposes -- work, non-work, and return home -- varies over time-of-day. Figure 2 shows that the overall demand for transit, represented by the fraction of trips made by transit, is similar (about 30 percent) for all conventionally defined periods of travel time during the daytime (i.e., AM peak, midday, and PM peak period). However, the purpose of trips varies among these three time periods. For example, during the AM peak period (6 am – 10 am), a majority of transit trips are made for work purposes (about 17 percent) whereas the higher fraction of midday (10 am – 3 pm) trips are made for non-work purposes (15 percent). On the other hand, a dominant share of PM peak (3 am – 7 pm) transit trips represent the return home trips (20 percent). Since transit services are typically unavailable or operate in less frequency during the late evening through early morning (7 am – 6 am), it is not surprising to observe less fraction of transit trips (11 percent) during this period.

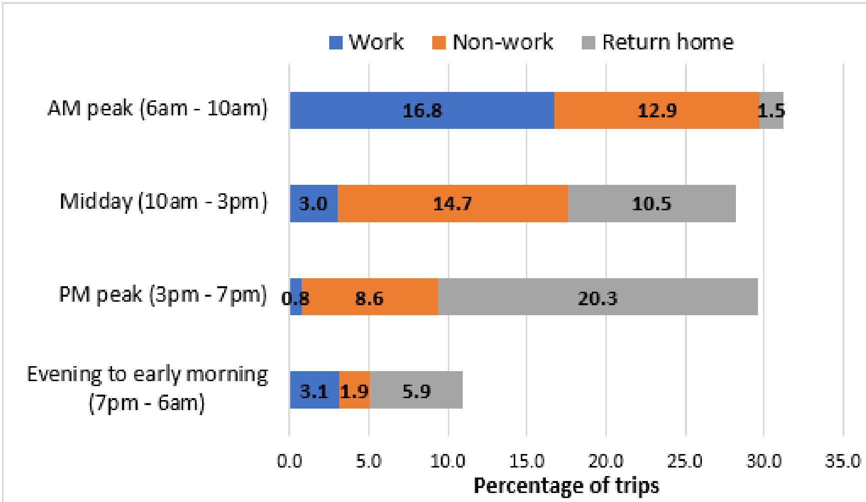
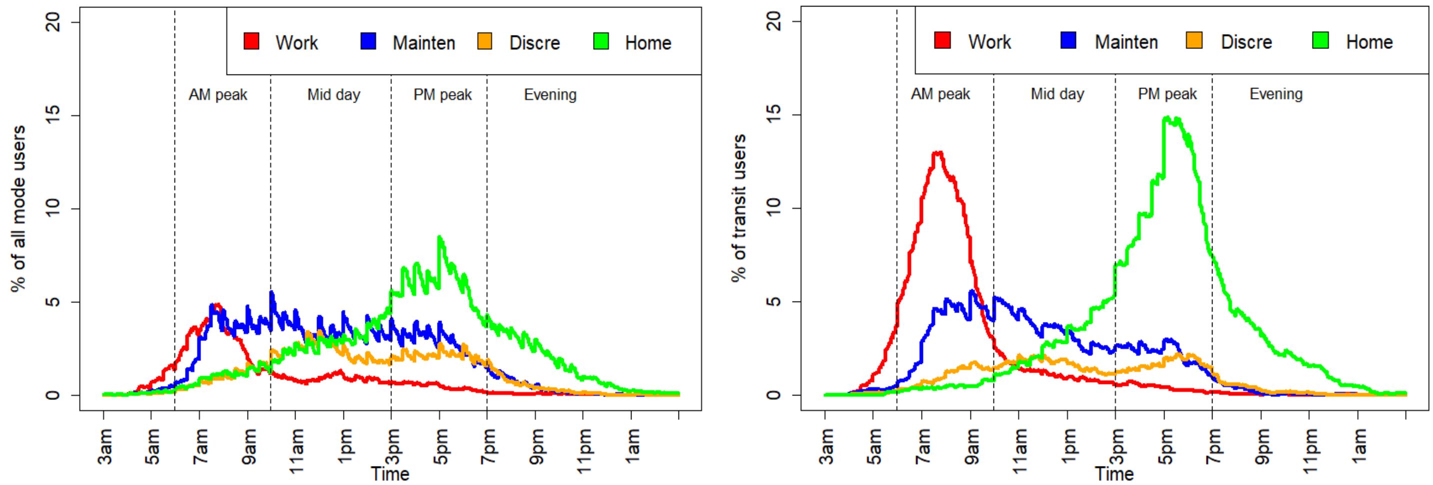


Figure 2: Distribution of trip purpose by time of day

The fraction of people traveling by activity purposes can be displayed in a *time in motion plot* as shown in Figure 3. The figure compares travelers making trips by (a) *all modes* versus (b) *public transit-only*. Note that we categorize trip purposes into four groups: (i) work: work- and work-related trips; (ii) maintenance: school/daycare/religious activity, medical/dental services, buying goods, buying services, other general errands, and drop off/pick up someone; (iii) discretionary: go out for a meal, snack, carry-out, recreational activities, and visiting friends or relatives; and (iv) return home.

Figure 3 shows that travelers typically commute to work during the AM peak period and return home from work during the PM peak period (Figure 3(a)). Transit riders demonstrate a similar trend but with higher peaks (Figure 3(b)). The higher peaks for work and return home trips indicate that among the transit riders, the majority of travelers are employed and use transit regularly, primarily for work and return home purposes (APTA, 2017). Maintenance trips are observed to occur at a constant rate throughout the day except in the evening period (Figure 3(a)). When travelers use transit for maintenance purposes, a similar trend is observed with a slight variation in the late midday and PM peak periods (Figure 3(b)). Regarding discretionary trips, no prominent difference appears between trips made by all modes and trips by transit-only.



(a) Traveler by *all modes*

(b) Traveler by *public transit*

Figure 3: Time in motion plot by trip purpose

The mode usage behavior of transit users for various trip purposes is shown in Figure 4. For any trip purpose, the majority of trips are observed to be made by public transit except for discretionary purposes. Besides, a similar fraction of trips (around 12-13 percent) is reported to be made by transit for both work and maintenance purposes. The second most frequent mode used by transit users to access any activity is walk followed by private vehicles.

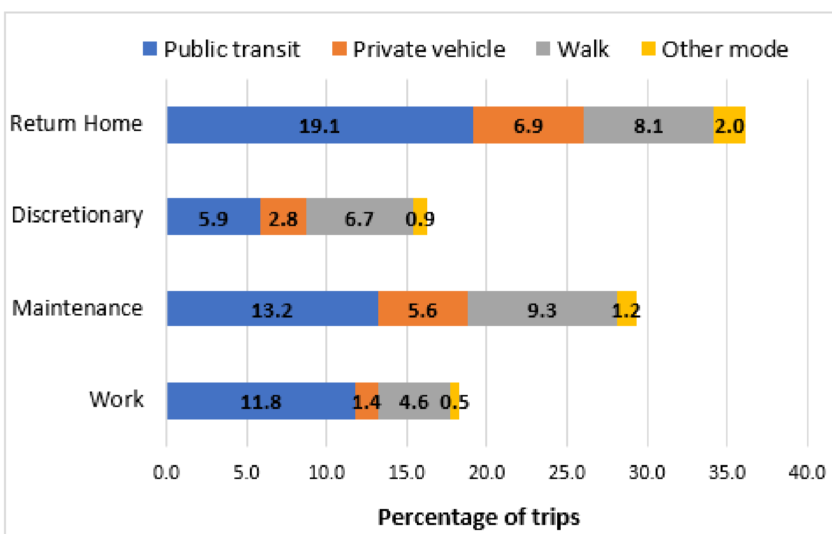


Figure 4: Distribution of travel mode by trip purpose

3. Heterogeneity in Activity-travel Patterns

The heterogeneity in activity-travel patterns of transit users can be captured by segmenting the transit users into a set of sub-groups (with representative activity-travel patterns) by using Latent Class Analysis (LCA). This technique offers several advantages over the traditional K-means cluster technique. For example, while LCA can simultaneously configure both the classification and prediction of classes with a single maximum likelihood estimation algorithm, K-means clustering requires a separate discriminant analysis to predict classes based on exogenous variables. Moreover, LCA provides various goodness-of-fit measures (e.g., AIC, BIC), which are useful in determining the number of classes whereas the K-means technique does not provide any such fit measures (Magidson and Vermunt, 2002).

LCA is commonly used in a range of travel behavior research, including to classify immigrants based on their travel behavior (Beckman and Goulias, 2008), individuals based on their residential location preferences (Liao *et al.*, 2015), millennials based on their mode usage (Ralph, 2017; Lee *et al.*, 2019), ride-hailing users depending on individual lifestyles (Alemi *et al.*, 2018), tours based on various non-work long-distance travel (Davis *et al.*, 2018), individuals based on mobility patterns (Schneider et al., 2020), and individuals with respect to their attitudes towards mobility as a service (Alonso-González *et al.*, 2020). While prior studies capture heterogeneity among target groups primarily in terms of individual demographics, lifestyles, attitudes or preferences, neighborhood characteristics, and travel behavior attributes, we identified heterogeneity among transit users based on their activity-travel patterns and tour behavior. A similar method is applied to analyze the activity-travel patterns of ride-hailing users by Rafiq and McNally (2020b). The mathematical formulation of LCA, required variables, and model estimation results are discussed next.

3.1 Latent Class Analysis (LCA) Model Definition

Latent class analysis is a mixture model that hypothesizes that there is an underlying *unobserved* categorical variable that divides a population into mutually exclusive and exhaustive latent classes (Lanza and Rhoades, 2013). Suppose each member of a population (indexed by i) contains J "indicator" variables (indexed by j), each of which can take a value from a set of K_j possible outcomes (all indicators variables are categorical). Let $Y_{ijk} = 1$ if respondent i takes k-th outcome for its j-th categorical variable, and $Y_{ijk} = 0$ otherwise (Y_i denotes the corresponding vector). For a given number of classes, say R, LCA attempts to simultaneously compute: (a) the probability that a respondent falls into a certain class, denoted by p_r, for $r = 1, 2, \ldots R$, and (b) the class-conditional probability, denoted by $\pi_{jrk,}$ that observation in class r produces the k-th outcome on the j-th variable. The likelihood of observing a certain respondent is therefore given by:

$$f(Y_i|\pi, p) = \sum_{r=1}^{R} p_r \prod_{j=1}^{J} \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}}$$

The parameters that the LCA model estimates are p_r and π_{jrk}, which are found via maximum log-likelihood estimation (MLE). In a more generalized LCA model, the class probabilities, p_r's, are regressed (by using a logit link function) from a set of observed variables, called "covariates". Hence, the estimation technique finds a set of per class coefficient vectors, β_r (instead of p_r), along with π_{jrk} (more details on this technique can be found in Linzer and Lewis (Linzer and Lewis, 2011).

3.2 LCA Model Indicator Variables and Covariates

LCA requires a set of indicator variables that defines the characteristics of each latent class and a set of covariates that help to predict the probability of an individual belonging to a latent class. Figure 5 shows the conceptual latent class model with a set of indicator variables and covariates used in this study. To capture the heterogeneity in activity-travel patterns, we used various *trip* and *tour attributes* of transit users as the indicator variables, such as day of travel (weekday or weekend), number of daily tours (one or more), a work tour is made or not, number of daily non-work trips, the timing of non-work trips, fraction of daily trips made by transit, and employment status of transit users. The covariates are to understand the class membership profiles that consist of various *socio-demographic* characteristics, such as gender, age, household income, household size, vehicle ownership, use of rail transit on the travel day, and population density (persons per square mile) in the census block group at the home location.

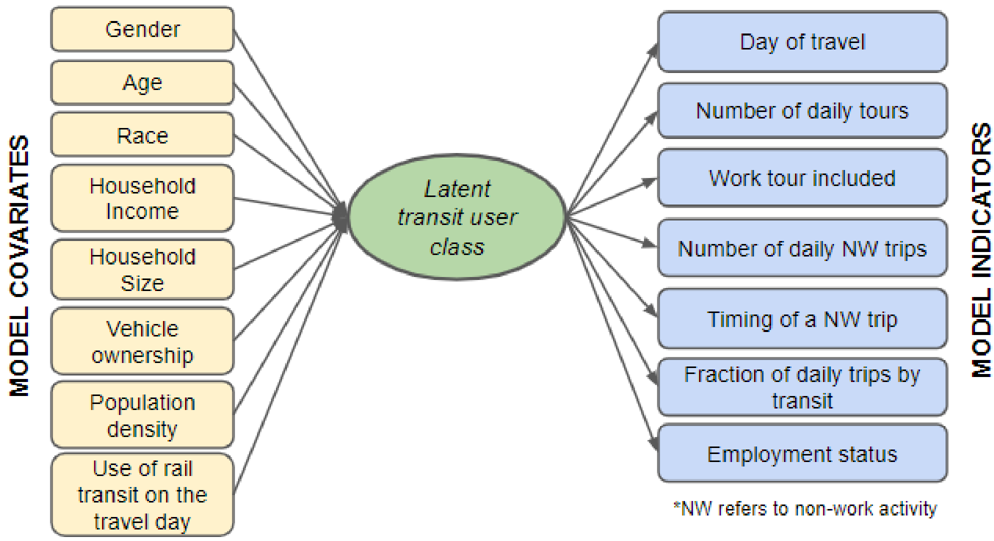
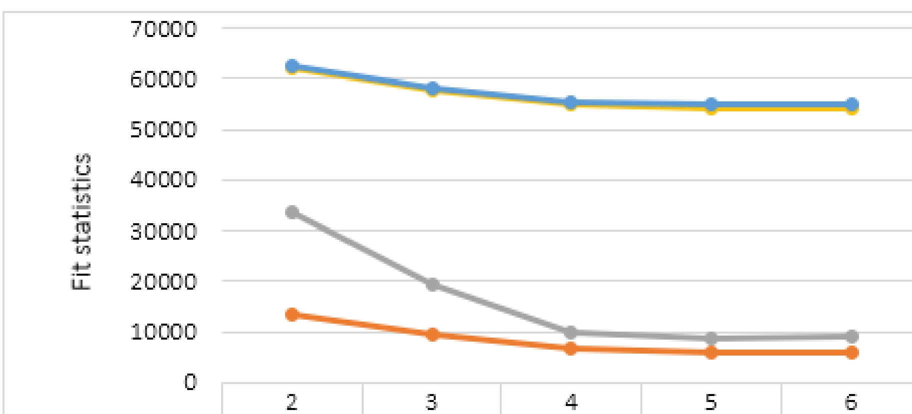


Figure 5: Latent class cluster model

3.3 LCA Model Estimation and Fit Statistics

We used poLCA (Polytomous variable Latent Class Analysis) in the statistical software package R to run LCA. R provides model parameters and goodness of fit measures, (chi-square with degrees of freedom and information criteria AIC or BIC). AIC or BIC can be used to compare the relative fit of models with different numbers of latent classes, where a lower value suggests a better model fit. In this study, we varied class sizes from 2 to 6, observed the corresponding fit measures, and empirically assessed the extent of the interpretability of the resulting classes.



	2	3	4	5	6
G^2	13566.3	9334.3	6569.5	6055.2	5813.5
Chi-square	33745.5	19246.9	9866.9	8494.5	9137.3
AIC	62325.0	57651.1	54937.0	54094.1	54161.5
BIC	62585.6	58087.7	55549.5	54882.6	55125.9

Figure 6: Model fit statistics for two to six-class models

Figure 6 shows the fit statistic values for two to six-class models. With the increase in the number of classes, the values of all fit measures decrease until the class size becomes six. The rate of decrease varies, with a sharp decline after class 2 and then flattening after class 5. Since the five-class model has the lowest AIC and BIC values and classes are easily identifiable and logically interpretable, we accepted the five-class model for our study.

The class-conditional membership probabilities for the indicator variables and covariates by each of the identified five classes are shown in Table 2a and Table 2b respectively. Also, the effects of covariates on class membership are presented in Table 3. Each of the identified five latent classes corresponds to an underlying group of individuals who are characterized by a particular activity-travel pattern and social-demographics features. Next, we provided a detailed description of (a) who belongs to which class among the five identified classes and their trip and tour characteristics, (b) class membership socio-demographic profiles (which factor influenced an individual belonging to a certain class), and (c) the activity-travel patterns of the five classes of transit users.

3.4 The Five Identified Transit User Classes

The first class corresponds to the *simple work tour transit commuters* (22 percent of total users, Table 2a), who, as the name suggests, make a single tour (96 percent) for work purposes (97.3 percent) on weekdays (92 percent). This group neither makes a nonwork stop in their work tour nor makes a separate nonwork tour in a day (100 percent). Besides, most of the members use transit for their work and return home trips (78.9 percent reported using transit for more than 50 percent of daily trips). This group constitutes Caucasian (63.8 percent), employed males who live with other household members (81 percent), have high vehicle ownership (79.5 percent), and typically use commuter rail (53.6 percent) for their work trips (c.f. Table 2b). The majority of this group (43.4 percent, Table 2b) resides in medium-density neighborhoods (2,000 to 10,000 people per square mile).

The second class is identified as the *complex work tour transit commuters* that constitute 22 percent of total users. Similar to class 1, this class also makes a single (67.2 percent, Table 2a) work tour (97.3 percent) on weekdays but it typically includes a non-work stop within the work tour that class 1 does not. Several users also make a separate non-work tour to perform a non-work activity (32.8 percent reported making multiple tours). Most of the users make one non-work trip (71.5 percent) per day, usually performed during the midday (10 am – 3 pm) or PM peak period (3 pm – 7 pm). The majority of the members (58.8 percent) depend on transit for making 25 to 50 percent of their daily trips. As per socio-demographic characteristics, these individuals are mostly Caucasian (62.9 percent) employed women with high income and high vehicle ownership (75.6 percent) who use commuter rail for work purposes (Table 2b).

Table 2a: Class-conditional membership probabilities for indicator variables by each class (N = 4,994)

	Class 1	Class 2	Class 3	Class 4	Class 5
	Simple work tour transit commuters (%)	Complex work tour transit commuters (%)	Multimodal complex tour transit users (%)	Simple non-work tour transit users (%)	Complex non-work tour transit users (%)
Class size[art]	1095	1127	733	977	1062
Class share	22%	22%	16%	19%	22%
Indicator variables					
Day of travel					
Weekday	**92.0**	**95.9**	**80.2**	**82.1**	**82.7**
Weekend	8.0	4.1	19.8	17.9	17.3
Daily tours					
Single tour	**96.0**	**67.2**	47.9	**100**	44.3
Multiple tours	4.0	32.8	**52.1**	0.0	**55.7**
Work tour included					
Yes	**97.3**	**97.3**	**59.3**	0.4	1.3
No	2.7	2.7	40.7	**99.6**	**98.7**
Number of daily non-work trips					
Non-work# 0	**100.0**	0.0	0.0	0.0	1.0
Non-work# 1	0.0	**71.5**	0.0	**76.8**	0.0
Non-work# 2	0.0	28.5	23.7	23.2	29.7
Non-work# >2	0.0	0.0	**76.3**	0.0	**69.3**
Timing of a non-work trip					
AM peak (6am – 10am)	0.0	18.8	**43.3**	**62.6**	**67.5**
Midday (10am – 3pm)	0.0	**33.5**	**70.6**	**41.6**	**81.3**
PM peak (3pm – 7pm)	0.0	**49.3**	**74.9**	6.1	**52.7**
Evening (7pm – 6am)	0.0	17.4	**42.8**	3.5	16.2
Fraction of daily trips by transit					
Less than 0.25	2.4	12.8	**42.1**	1.0	24.6
0.25 – 0.5	18.7	**58.8**	**49.9**	20.0	**48.1**
More than 0.5	**78.9**	28.4	8.0	**79.0**	27.3
Employment status					
Employed	**97.5**	**98.3**	**96.1**	16.8	9.6
Not employed	2.5	1.7	3.9	**83.2**	**90.4**

[a] Class of each sample is determined by modal assignment (so the percentage may not match).

Table 2b: Class-conditional membership probabilities for covariates by each class (N = 4,994)

	Class 1	Class 2	Class 3	Class 4	Class 5
	Simple work tour transit commuters (%)	Complex work tour transit commuters (%)	Multimodal complex tour transit users (%)	Simple non-work tour transit users (%)	Complex non-work tour transit users (%)
Class size[a]	1095	1127	733	977	1062
Class share	22%	22%	16%	19%	22%
Covariates					
Gender of the traveler					
Male	**54.1**	49.5	48.0	46.4	44.3
Female	45.9	**50.5**	**52.0**	**53.6**	**55.7**
Age of the traveler					
Younger group (< 18 years)	0.4	0.5	0.7	**16.7**	**14.2**
Millennials (18 – 38 years)	**40.4**	**42.6**	**47.9**	23.6	17.2
Generation X (38 – 58 years)	**41.1**	**37.8**	**37.2**	21.1	24.2
Older adults (> 58 years)	17.5	17.8	13.5	**36.2**	**43.4**
Race of the traveler					
Caucasian	**63.8**	**62.9**	**73.6**	45.7	52.9
Non-Caucasian	36.2	37.1	26.4	**54.3**	47.1
Household income					
Low income (less than $35K)	21.1	20.5	18.9	**60.9**	**62.8**
Middle income ($35K – $100K)	35.2	34.4	33.4	21.4	21.6
High income (more than $100K)	**41.4**	**43.4**	**46.8**	13.5	13.1
Household size					
One person	19.0	25.0	29.1	34.0	**40.3**
Two persons	**38.3**	**39.9**	**41.7**	25.0	29.5
more than two persons	**42.7**	**35.0**	29.2	**40.9**	30.2
Household vehicle ownership					
Own at least one vehicle	**79.5**	**75.6**	**71.5**	48.5	43.9
Does not own a vehicle	20.5	24.4	28.5	**51.5**	**56.1**
Used rail transit on the travel day					
Yes	**53.6**	**55.5**	**59.3**	24.3	23.2
No	46.4	44.5	40.7	**75.7**	**76.8**
Population density (persons per sq. mile) in census block group					
Low density (0 – 2,000)	21.4	15.4	11.7	18.6	17.0
Medium density (2,000 – 10,000)	**43.4**	39.1	38.9	**44.4**	**45.9**
High density (more than 10,000)	35.2	**45.6**	**49.3**	37.0	37.1

[a] Class of each sample is determined by modal assignment (so the percentage may not match).

The third identified class is deemed *multimodal complex tour transit users* (the smallest with 16 percent users) who are mostly employed (96.1 percent, Table 2a) and make work tours (59.3 percent) like the first and second class. The key difference is that class 3 typically makes multiple non-work trips (76.3 percent make more than two non-work trips) within a work or non-work tour while the other two classes do not. Despite most of the users being employed in this group, less than two-thirds of them made work tours on travel day (59.3 percent), which is contrary to class 1 and class 2 (more than 97 percent did so in these classes). Moreover, unlike the other two employed groups, this group makes a considerable fraction of multiple tours (52.1 percent compared to 4 percent (class 1) and 32.8 percent (class 2)). The users of this group are multimodal since most of them (more than 90 percent) use transit for at most 50 percent of their trips and depend on other modes for making the rest of the trips. Members of this class are mostly Caucasians (73.6 percent), millennials with high income (46.8 percent), and high vehicle ownership (71.5 percent) (c.f. Table 2b). Similar to class 1 and class 2, a higher fraction of this group uses commuter rail (59.3 percent). Unlike simple work tour transit users (class 1), a higher fraction of the two complex tour users (class 2 and class 3) live in high-density residential areas (more than 10,000 people live per sq. mile) (Table 2b).

In contrast to the previous three groups, the last two groups of transit users are not typically employed and consequently do not make work tours. Instead, they make single or multiple tours to perform one or more non-work activities. The fourth group, identified as the *simple non-work tour transit users* (19 percent of total users), primarily make a single tour (100 percent, Table 2a) to participate in only one non-work activity (76.8 percent). This group depends mostly on transit for making both of their trips (79 percent use transit for more than 50 percent of trips) (c.f. Table 2a).

Compared to class 4, the final class of transit users (class 5) mostly makes multiple tours (55.7 percent) to multiple non-work activities (69.3 percent make more than two, Table 2a). This class is, therefore, called *complex non-work tour transit users*, which comprises 22 percent of total transit users. These two non-work tour classes include a higher fraction of younger (age < 18 years) and older-adult (age > 58 years) groups and a larger proportion of low-income households with low vehicle ownership (nearly 45 percent in class 4 and 5 compared to about 75 percent in the other three classes) than the first three classes (c.f. Table 2b). Moreover, while a higher proportion of the three employed groups used commuter rail (more than 50 percent), the other two groups mostly used the public bus (more than 75 percent) on the travel day. Among all the classes, class 4 includes a larger share of non-Caucasian people whereas class 5 comprises a greater fraction of single-living people. Similar to class 1, the majority of users in both class 4 and class 5 resides in medium-density areas.

3.5 Prediction of Latent Class Membership

The socio-demographic factors (covariates) that influence an individual belonging to a certain class are shown in Table 3. The covariate coefficients for four classes are displayed relative to the first class (i.e. simple work tour transit commuters). Males are more likely to belong to the simple work tour class (class 1) compared to all the four classes. On the other hand, females are more likely to belong to all the complex tour classes. This is because females often have a greater range of activity responsibilities than their male counterparts (McGuckin and Murakami, 1999; Rafiq and McNally, 2020a). Both younger (< 18 years) and older adult groups (> 58 years) are more inclined to be the non-work tour transit users (class 4 and class 5) whereas millennials (18 – 38 years) are more likely to be the multimodal complex tour transit users (class 3).

Table 3: Prediction of latent class membership (N = 4,994)

Covariates	Complex work tour transit commuters vs. simple work tour commuters	Multimodal complex tour transit users vs. simple work tour commuters	Simple non-work tour transit users vs. simple work tour commuters	Complex non-work tour transit users vs. simple work tour commuters
Gender of traveler: Male	-0.168*	-0.255**	-0.226**	-0.313***
Age of traveler (baseline: Millennials, 18 – 38 yrs.)				
Younger group (less than 18 years)	0.324	0.646	4.542***	4.897***
Generation X (38 – 58 years)	-0.151	-0.300***	-0.249*	0.177
Older adults (more than 58 years)	-0.139	-0.666***	1.212***	1.581***
Household income (baseline: low income, < $35K)				
Middle income ($35K – $100K)	0.084	0.165	-1.210***	-1.097***
High income (>$100K)	0.295**	0.507***	-1.644***	-1.536***
Race of the traveler: Caucasian	-0.040	0.490***	-0.267**	0.006
Household size (baseline: single person)				
Two persons	-0.286**	-0.423***	-0.240	-0.167*
More than two persons	-0.525***	-0.880***	0.082	-0.269
Household vehicle: own at least one vehicle (baseline: does not own vehicle)	-0.033	-0.281**	-0.683***	-0.800***
Use of rail transit on the travel day: Yes	-0.033	0.047	-0.655***	-0.755***
Population density (persons per sq. mile) in census block group (baseline: low density, 0 – 2,000)				
Medium density (2,000 – 10,000)	0.198	0.463***	-0.162	-0.036
High density (more than 10,000)	0.507***	0.759***	-0.033	0.085

*, **, and *** indicate statistical significance respectively at 10%, 5%, and 1%.

Household income also affects class membership: transit users with low-income belong to class 4 and class 5, on the contrary, high-income users belong to class 2 and class 3. Likewise, the users who do not possess any household vehicle or do not use commuter rail are more likely to belong to class 4 and class 5. We found an association between household size and class membership: persons from single-living households tend to belong to class 5, whereas persons from larger households are more likely to belong to class 1. The effects of population density on class membership are limited, more specifically people living in high-density areas are more likely to make complex tours, hence more tend to belong to class 2 and class 3.

3.6 Activity-travel Patterns of Identified Classes

This section analyzes the activity-travel patterns of the identified five transit user classes. A graphical representation is utilized for each class that shows the sequence of *all* activities and travel reported in a travel diary day for a *randomly* selected 50 individuals from a given class. Ideally, we would depict the plots for all individuals in the class but space and clarity of display resulted a selection of 50 patterns yielding the clearest results. We generate the same plots for 10 different random samples, each time producing a similar set of plots. We report one of those ten results here. Figure 7 shows these results for each class (the x-axis denotes the time of day and the y-axis denotes sampled individuals with their activities and trips). The sequence of activities and travel is shown as segments based on activity and travel duration. The segments are color-coded based on activity purpose and mode use. In Figure 7, a summary of the major activity-travel characteristics of each class is also shown by a stacked bar chart beside the activity-travel pattern drawing.

3.6.1 Class 1. Simple Work Tour Transit Commuters

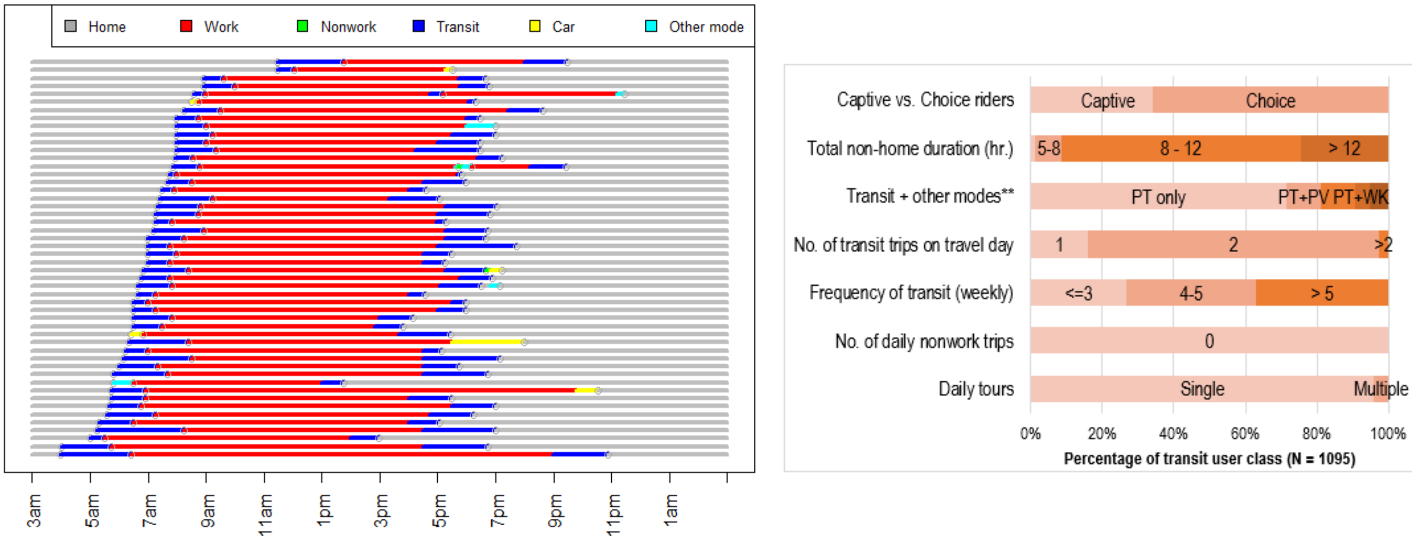
The dominance of red-colored segments in all the patterns in Figure 7(a) best illustrates the work focus in this class. The blue segments show transit use, predominantly preceding and following the red segments indicating transit as a commute mode to and from work. The departure time of transit trips during morning and evening hours and the length of red segments demonstrates that this is a 9-to-5 commuter group. In addition to the pattern diagram, the bar chart depicts that this group mostly make a single tour for work purpose and are primarily dependent on transit: 85 percent use only transit, 10 percent use transit in a combination of a private vehicle, and 9 percent combine walk trips with transit. The higher weekly frequency of transit use indicates that this class commutes regularly by transit. The majority of this class are not captive riders[4] but are rather choice riders (66 percent).

3.6.2 Class 2. Complex Work Tour Transit Commuters

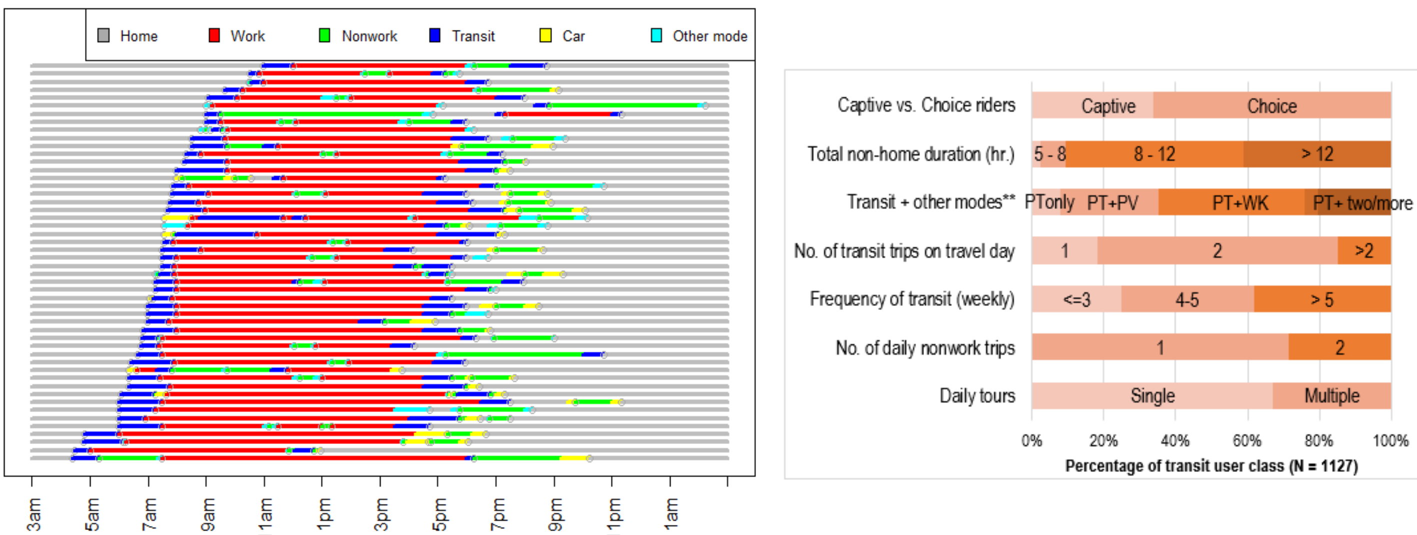
In Figure 7(b), class 2 demonstrates a similar pattern of red and blue colored segments like class 1, which means that class 2 also use transit (blue segment) as a commute mode to and from work (red segment). In contrast to class 1, this class depicts green colored segments in the middle of the red color and mostly on the right side of the diagram. The green segments depict non-work activities, usually performed either during work or after work hour (33.5 percent people make during midday and 49.3 percent during PM peak period, Table 2a). The after-work non-work activities are made either on the 'way to home' journey or via separate non-work tours. About two-thirds of people in this class make a single tour that typically mixed non-work with work whereas the other third made multiple tours, possibly one for work and another one for non-work. Data reveals that when this group makes non-work during work hours, they typically go out for lunch (spending 28 minutes on average) within walkable distance from their workplace. Again, when they stop on the way to home, the activity tends to be buying goods, groceries, or services spending about 40 minutes on average.

Figure 7(b) also shows that while transit (blue) is predominantly associated with work activity (red), private vehicles (yellow) and other modes (cyan) along with transit are associated with non-work (green) activities. More specifically, this class uses a variety of modes to access non-work activities, for example, in 32 percent of non-work trips transit is reported to be used whereas in 26 percent and 37 percent of trips private vehicles and walk are used to access non-work activities, respectively.

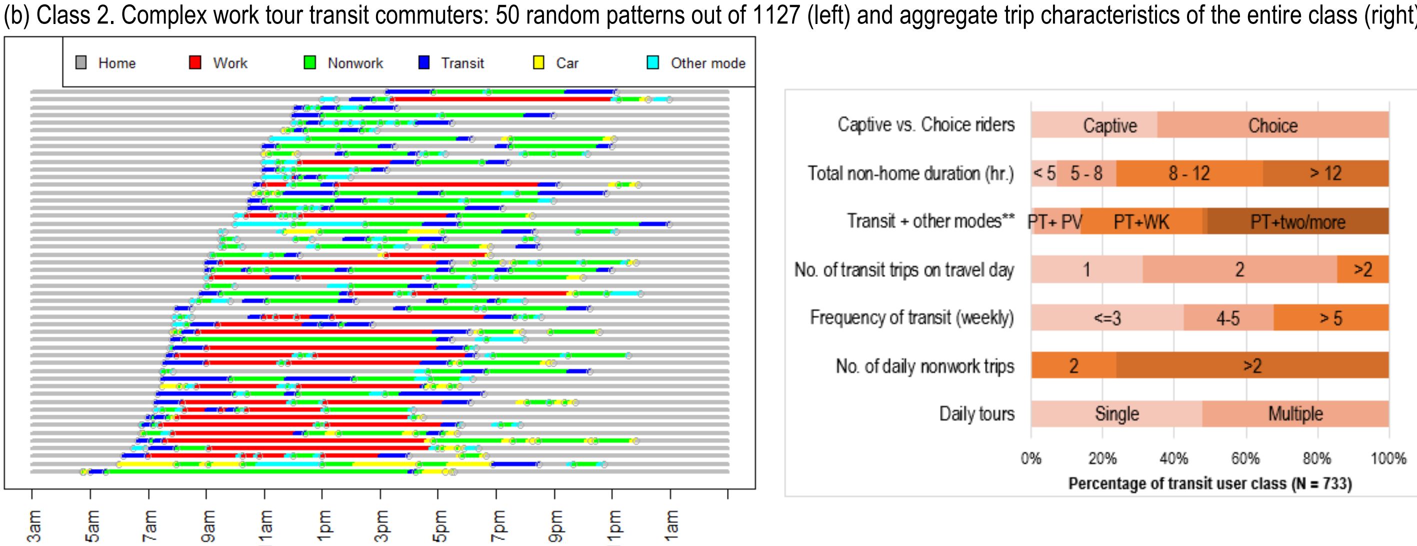
[4] Captive riders refers to those riders who either do not own a vehicle or do not have driving license or give up driving for a medical condtion.



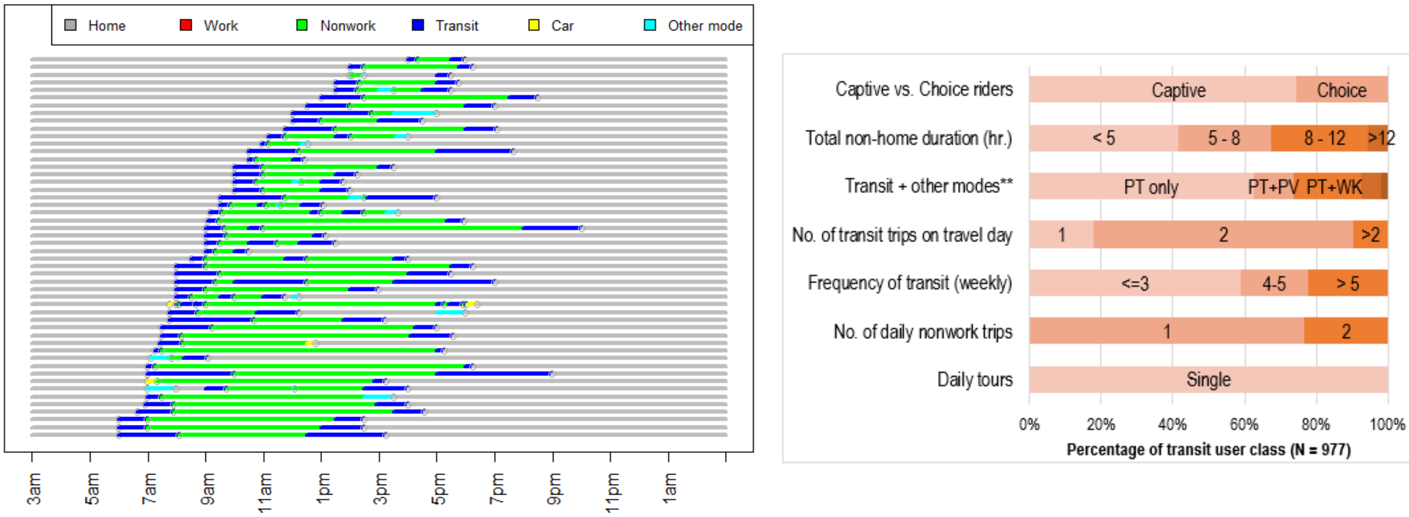
(a) Class 1. Simple work tour transit commuters: 50 random patterns out of 1095 (left) and aggregate trip characteristics of the entire class (right)



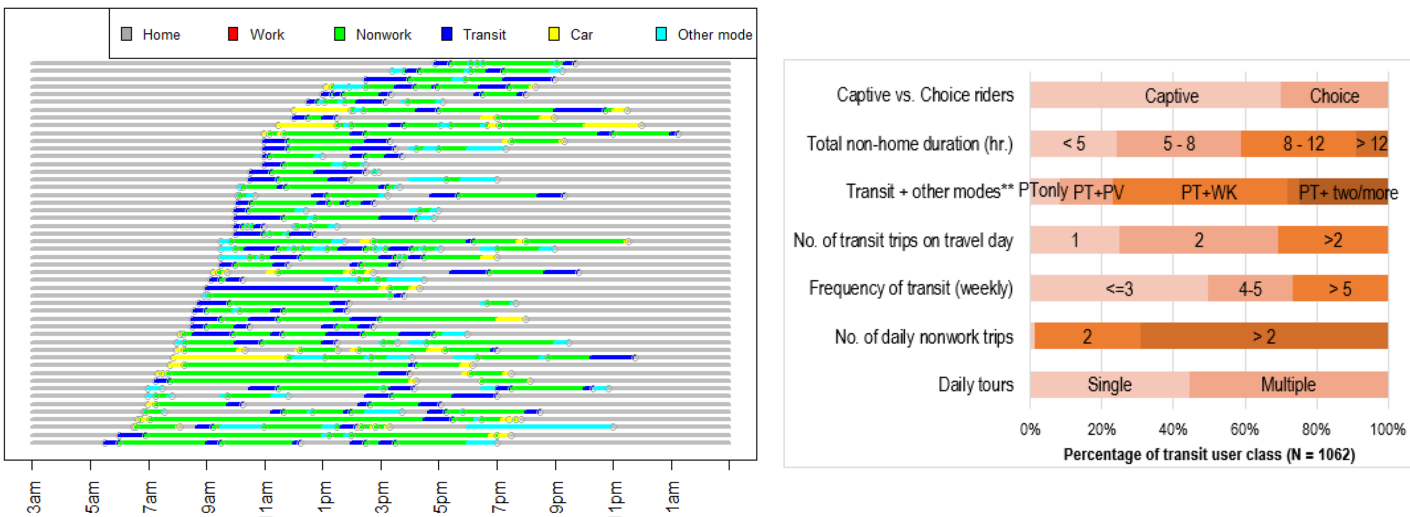
(b) Class 2. Complex work tour transit commuters: 50 random patterns out of 1127 (left) and aggregate trip characteristics of the entire class (right)



(c) Class 3. Multimodal complex tour transit users: 50 random patterns out of 733 (left) and aggregate trip characteristics of the entire class (right)



(d) Class 4. Simple non-work tour transit users: 50 random patterns out of 977 (left) and aggregate trip characteristics of the entire class (right)



(e) Class 5. Complex non-work tour transit users: 50 random patterns out of 1062 (left) and aggregate trip characteristics of the entire class (right)

** Modes denoted as PT, PV, WK, two/more refers to public transit, private vehicle, walk, and at least two modes respectively.

(These figures read better in color prints)

Figure 7: Sampled activity-travel patterns and aggregate trip characteristics by transit user classes

3.6.3 Class 3. Multimodal Complex Tour Transit Users

The transit users who belong to this class demonstrate different trip characteristics from the first two classes (class 1 and class 2), as evidenced in Figure 7(c). One difference is that not all people in this class make trips to work on the travel day (even though 96 percent of people in this class are employed). A possible reason may be that a higher fraction of class 3 reported weekend trips (20 percent compared to 8 and 4 percent for class 1 and class 2, respectively) or worked from home (12 percent compared to 3 and 4 percent) on the travel day. Another observation is that the non-home activities span from morning till late evening in this class, which is not visible in other classes (42.8 percent people make trips during evening compared to 17.4, 3.5, and 16.2 percent in class 2, 4, 5 respectively, Table 2a). Also, class 3 participates in more non-work activities by making multiple tours and depart late in their first trip made by transit within the first tour than the previous two employed classes.

The pattern also reveals that the transit users in the class mix private vehicles (yellow segments) and other modes (cyan segments) with their transit modes (blue segments). This class indeed has a higher fraction of "PT + two/more" group (the travelers who use two and more modes in addition to transit to complete their activities) than other classes (Figure 7(c) bar chart). This is why this class is called a multimodal transit user group.

3.6.4 Class 4. Simple Nonwork Tour Transit Users

The activity-travel pattern of class 4 is displayed in Figure 7(d), which shows a similar pattern of class 1, but instead of having red, class 4 illustrates green color. In particular, this class makes a single tour to perform one non-work activity and use transit to make the non-work and return home trips (blue segments juxtaposed with green segments). It is observed that the non-work trips mostly occur during the morning hours (blue segments that precede the green segments span between 8 am to 12 pm) usually to go to school (19 percent trips), to buy groceries or other goods (35 percent), to visit health care centers (14 percent), or to do discretionary activities (21 percent). As the pattern diagram shows, the total non-home durations for each individual varies quite a bit. For example, while the majority of them spend less than 5 hours (41 percent), a considerable fraction spends up to 8 hours (26 percent) or even up to 12 hours (27 percent) (Figure 7(d) bar chart).

This class is neither considered as choice riders nor as frequent transit riders as the other commuter classes. Most of them use transit constrained by their circumstances (74 percent are captive riders) and use it for at most 3 times a week (60 percent). They rarely use other modes to make non-work trips— only 11 percent and 19 percent of members combine private vehicles and walk with transit respectively.

3.6.5 Class 5. Complex Nonwork Tour Transit Users

Members of class 5 make multiple tours to make multiple non-work activities as illustrated by a high concentration of small green-colored segments in Figure 7(e). The green segments mostly span from morning through early evening period, which can be attributed to doing non-work activities during the daytime: 67.5, 81.3, and 52.7 percent users make trips in AM peak, midday, PM peak periods respectively (Table 2a). Non-work trips are usually made for school (8 percent), shopping (40 percent), discretionary activities (28 percent), and medical visits (10 percent). Similar to class 4, the duration of total non-work activities varies considerably among the class members (Figure 7(e) bar chart).

The scattered pattern of small blue-colored segments demonstrates a repetitive use of transit for making multiple tours or single tours with multiple trips by this class. Data reveals that compared to other transit user classes, this class makes more transit trips with shorter duration (average number of transit trips for class 5 is 2.3 compared to 1.9 for the other classes). The presence of cyan and yellow-colored segments denote corresponding walk and private vehicle trips integrated with transit to access multiple non-work activities.

4. Conclusions

This study analyzes the activity-travel patterns and tours of transit users by classifying them into a number of sub-groups via Latent Class Analysis (LCA). Here, the term *pattern* denotes a complete sequence of activities and trips made by a transit user over a full day whereas *tour*, a basic unit of a pattern, refers to a sequence of trips that begins and ends at home and contains single or multiple activities. Based on data from the 2017 NHTS, the LCA model results suggest that the transit users can be divided into five distinct classes where each class has a representative activity-travel pattern. Class 1 constitutes Caucasians employed males who make transit-dominant simple work tours. This is a regular 9-to-5 commuter group. Class 2 is composed of Caucasian females who commute by transit and typically make after-work non-work activities. Caucasian employed millennials comprise Class 3 and make a multimodal complex tour. Transit Class 4 consists of non-Caucasian younger or older adult people who make a transit-dominant simple non-work tour. Last, Class 5 members make complex non-work tours with recurrent transit use and are comprised of single-living older women. The summary of the characterizations of these five transit user classes is shown in Figure 8.

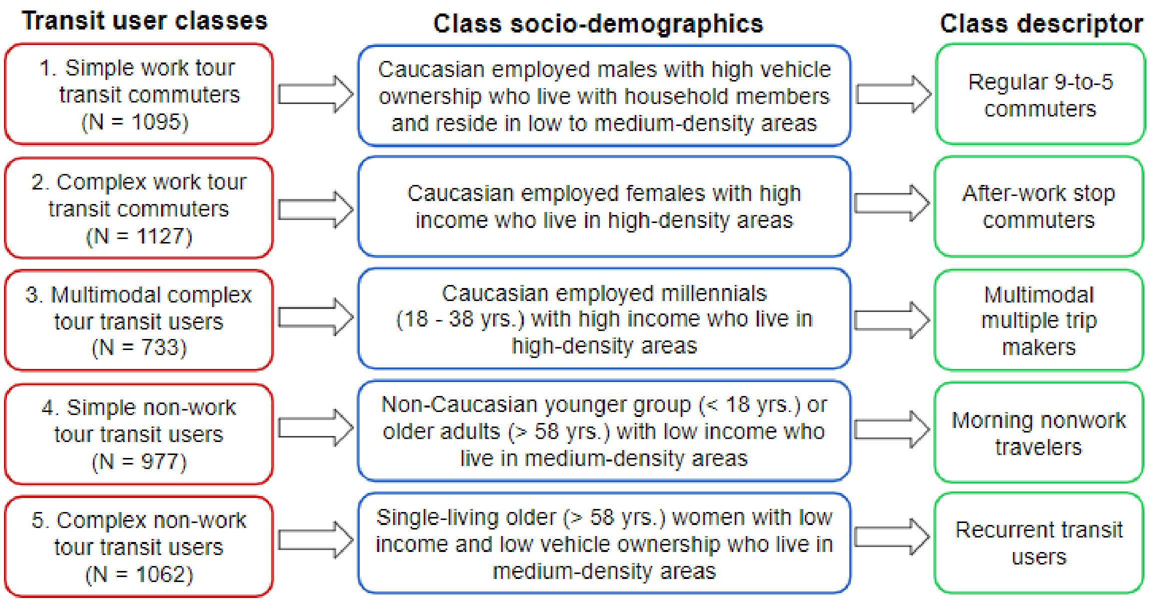


Figure 8: Five transit user classes and their socio-demographic properties

This study can help transit agencies identify potential market groups of transit users with particular socio-demographic characteristics and activity-travel patterns and to take necessary market strategies addressing different groups of users to meet their travel needs and to improve the quality of service provided. For example, frequent transit services and on-time strict schedules need to be ensured and monthly transit pass option can be offered particularly to those who regularly use it for commute purposes (Class 1 and Class 2). While making after-work non-work activities, a substantial portion of Class 2 members use private vehicles for non-work or return-home trips as transit use is not generally conducive to do so. To provide a convenient modal linkage for this class, transit stations should be designed to consider parking facilities and other activity services.

As individuals belonging to Class 3 make multiple trips to non-work activity locations and usually mix other modes in addition to transit, providing multiple activity centers (e.g. shopping/grocery, restaurants) at a single location might benefit them at length. This might reduce the number of transfers on their transit usage and might facilitate easier chaining of multiple activity purposes at a single location. On the contrary, since Class 4 and Class 5 comprise a large portion of older-adults, special attention needs to be given to design a better and convenient transit service for them addressing their mobility and accessibility needs. Finally, Class 5 transit users make use of transit quite often (multiple times in a single day) so discounted fare options (e.g., a transit day pass or free transfers) can be offered to them so that they can make multiple transit-stops in connecting non-work activities.

The findings of this study provide valuable information on the heterogeneity among transit users based on their activity-travel patterns. It, therefore, provides better insights on their pattern choice sets, which will help planning organizations in forecasting daily activity-travel schedules for transit users and subsequently predicting tour generation behavior in an activity-based travel demand forecasting model.

Author Contribution Statement

The authors confirm contribution to the paper as follows: study conception and design: R. Rafiq, M. G. McNally; data processing: R. Rafiq; data analysis and interpretation of results: R. Rafiq, M. G. McNally; draft manuscript preparation: R. Rafiq, M. G. McNally. All authors reviewed the results and approved the final version of the manuscript.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgment

This study was made possible with funding received by the University of California Institute of Transportation Studies from the State of California through the Road Repair and Accountability Act of 2017 (Senate Bill 1). The authors would like to thank the State of California for its support of university-based research, and especially for the funding received for this project.

References

Alemi, F., Circella, G., Mokhtarian, P., and Handy, S. (2018). Exploring the latent constructs behind the use of ridehailing in California. *Journal of choice modelling*, 29, 47-62.

Alonso-González, M. J., Hoogendoorn-Lanser, S., van Oort, N., Cats, O., and Hoogendoorn, S. (2020). Drivers and barriers in adopting Mobility as a Service (MaaS)–A latent class cluster analysis of attitudes. *Transportation Research Part A: Policy and Practice*, 132, 378-401.

APTA. (2017). *Who rides public transport: passenger demographics and travel*. American Public Transit Association.

Beckman, J. D., and Goulias, K. G. (2018). Immigration, residential location, car ownership, and commuting behavior: a multivariate latent class analysis from California. *Transportation*, 35(5), 655-671.

Bhaskar, A., and Chung, E. (2014). Passenger segmentation using smart card data. *IEEE Transactions on intelligent transportation systems*, 16(3), 1537-1548.

Chu, K. K. A., and Chapleau, R. (2010). Augmenting transit trip characterization and travel behavior comprehension: Multiday location-stamped smart card transactions.*Transportation research record*, 2183(1), 9-40.

Davis, A. W., McBride, E. C., and Goulias, K. G. (2018). A latent class pattern recognition and data quality assessment of non-commute long-distance travel in California. *Transportation research record*, 2672(42), 71-80.

El Mahrsi, M. K., Côme, E., Baro, J., and Oukhellou, L. (2014). Understanding passenger patterns in public transit through smart card and socioeconomic data: A case study in Rennes, France. In *ACM SIGKDD Workshop on Urban Computing*.

Federal Highway Administration. (2002). *Status of the Nation's Highways, Bridges, and Transit: Conditions & Performance*, http://www.fhwa.dot.gov/policy/ 2002cpr/pdf/execsummary_book.pdf. Accessed Dec 24, 2018.

Grahn, R., Hendrickson, C., Qian, Z. S., & Matthews, H. S. (2019). Socioeconomic and Usage Characteristics of Public Transit Riders in the United States. *Presented at 98th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Goulet-Langlois, G., Koutsopoulos, H. N., and Zhao, J. (2016). Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C: Emerging Technologies*, 64, 1-16.

He, L., Agard, B., and Trépanier, M. (2020). A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method. *Transportmetrica A: Transport Science*, 16(1), 56-75.

Hensher, D. A. and Reyes, A. J. (2000). Trip chaining as a barrier to the propensity to use public transport. *Transportation*, 27(4), 341-361.

Lanza, S. T., and Rhoades, B. L. (2013). Latent class analysis: an alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science*, 14(2), 157-168.

Lee, Y., Circella, G., Mokhtarian, P. L., and Guhathakurta, S. (2019). Are millennials more multimodal? A latent class cluster analysis with attitudes and preferences among millennial and generation X commuters in California. *Transportation*, https://doi.org/10.1007/s11116-019-10026-6

Liao, F. H., Farber, S., and Ewing, R. (2015). Compact development and preference heterogeneity in residential location choice behaviour: A latent class analysis. *Urban Studies*, 52(2), 314-337.

Linzer, D. A., and Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of statistical software*, 42(10), 1-29.

Ma, X., Wu, Y. J., Wang, Y., Chen, F., and Liu, J. (2013). Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, 36, 1-12.

Ma, X., Liu, C., Wen, H., Wang, Y., and Wu, Y. J. (2017). Understanding commuting patterns using transit smart card data. *Journal of Transport Geography*, 58, 135-145.

Magidson, J. and Vermunt, J. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, 20(1), 36-43.

McGuckin, N., & Murakami, E. (1999). Examining trip-chaining behavior: Comparison of travel by men and women. *Transportation Research Record: Journal of the Transportation Research Board*, (1693), 79-85.

Morency, C., Trepanier, M., and Agard, B. (2007). Measuring transit use variability with smart-card data. *Transport Policy*, 14(3), 193-203.

Rafiq R. and McNally, M. G. (2020a). A study of tour formation: pre-, during, and post-recession analysis. *Transportation*, https://doi.org/10.1007/s11116-020-10126-8

Rafiq R. and McNally, M. G. (2020b). How Do People Use Ride-Hailing? An Exploratory Analysis of Associated Travel Behavior. *Presented at 99th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Ralph, K. M. (2017). Multimodal millennials? The four traveler types of young people in the United States in 2009. *Journal of Planning Education and Research*, 37(2), 150-163.

Schneider, F., Ton, D., Zomer, L. B., Daamen, W., Duives, D., Hoogendoorn-Lanser, S., & Hoogendoorn, S. (2020). Trip chain complexity: a comparison among latent classes of daily mobility patterns. *Transportation*, 1-23.