Infrequent Public Transport Use:
An Investigation with Smart Card Data

Mark Hickman
The University of Queensland
m.hickman1@uq.edu.au (corresponding author)

Tianwei Yin
The University of Queensland
t.yin@uq.net.au

Extended Abstract for TRANSITDATA 2019

It is generally noted in household travel surveys and in public transport on-board surveys that a majority of journeys taken by public transport represent commuters or students who are using the public transport system multiple times per week, in peak periods, with common weekday patterns of work and school trips (Neff and Pham, 2007). Naturally, sampling methods common in household surveys or even on-board surveys are less likely to capture infrequent travellers. While these trips are observed, far less is known about the frequency with which these passengers use public transport, or the commonality of origins, destinations, and the time of day and day of week (Neff and Pham, 2007).

Yet, smart card data allows segmentation of passenger types and passenger journeys using smart card data. As noted by many researchers, where a large density of smart card users can be identified, it is possible to use a variety of statistical methods and data clustering algorithms to understand transit use (Morency et al, 2007; Park et al, 2008; Chu and Chapleau, 2010; Wang et al, 2011; Kieu et al, 2014; Bhaskar and Chung, 2015).

In this study, we were motivated by two questions: (1) How likely are infrequent public transport users to increase their frequency of use? And, (2) In Mobility as a Service (MaaS), what kinds of journeys might be most appropriate for public transport? Regarding the latter question, various MaaS studies have proposed subscription services that may include some but not exclusive use of public transport. In this case, it may be useful to understand the travel patterns of infrequent public transit users.

Very little research has been done targeting infrequent users. Our own intuition suggested that there are some people who use public transport for a short period of time (e.g. a single week), representing either visitors to an area or others who have a trial period but then discontinue use. We also hypothesized a second group of users who are residents of an area who use public transport infrequently or irregularly for journeys that are well-served by transit.

Our exploration of the behaviour of infrequent transit users is based on analysis and data mining of smart card data from South East Queensland (SEQ), Australia, including the metropolitan areas of Brisbane, the Gold Coast, and the Sunshine Coast. In total, the area served has a population of 3.5 million, and the area records more than 151 million public transport journeys per year. Our data come from Go Card transactions from July 2015 to June 2016, with records of the time and location of tap-ons and tap-offs for about 86% of the journeys in the public transport network of bus, rail, and ferry. This data set allows examination of an individual card's use over the full period.

With our analysis of the Go Card data, the figures below illustrate the nature of our findings. Figure 1 shows the number of uses of each Go Card during the week of 12-18 October 2015, a period with no holidays and all schools and universities in session. One may see that the majority of cards are used infrequently, with a high number of cards used once or twice in the week, then a steep drop-off after 10 uses in the week. An even number of uses (2, 4, 6, 8, 10) are clearly more common than an odd number of uses, reflecting the preponderance of round trips.

Figure 2 illustrates that, for the cases of one or two card uses per week, the uses are spread somewhat uniformly during daylight hours (7:00 to 18:00), with the exception of a spike from 15:00-16:00 when primary and secondary schools are released for the day (green and orange colours in the figure). This finding identifies that infrequent travellers do not travel more in the peak periods than in other times of the day. Yet, we find that trip duration (in minutes) and the spatial location of the origins and destinations closely match those of more frequent trips; i.e. those persons making infrequent trips have similar spatial locations and durations as those who use transit on a daily basis. [This finding is not shown here, but is among our results.]

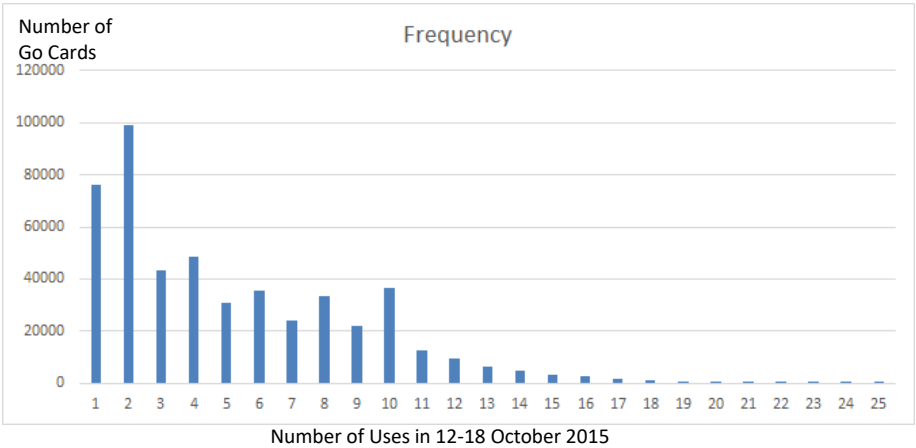
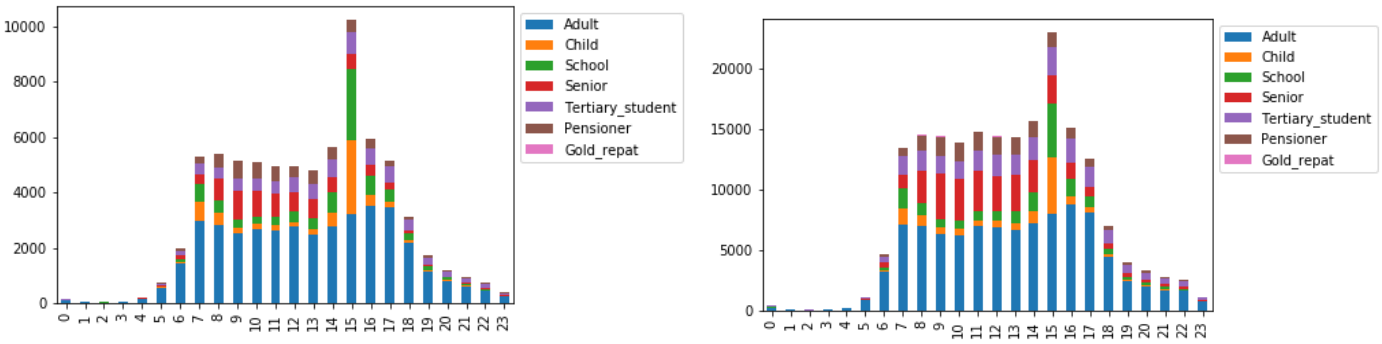


Figure 1: Go Cards by Number of Times Used in 12-18 October 2015



Travel times per week: 1 Travel times per week: 2

Figure 2: Time of Day of Go Card Tap-on, for 1 and 2 Card Uses per Week

Realizing that infrequent behaviours may be identified more clearly, we are also exploring methods of clustering smart card uses. Our investigation to date has used the measure of trip legs per week as the metric of interest. We expect infrequent users to have a low number of trip legs per week, regardless of the number of weeks when the card is used, or use only during specific weeks out of the year (e.g., for a visitor). We generated a sample of 10,347 Go Cards, taken from the approximately 2.5 million Go Cards recorded during the 2015-2016 year. These Go Cards represent about 128,000 journeys out of the nearly 151 million journeys. For each Go Card, we counted the number of trip legs in each of the 53 weeks in the 2015-2016 year (53 observations). Then, we generated a histogram of the uses per week, creating a probability density function on the 53 data points (sample in Figure 3).

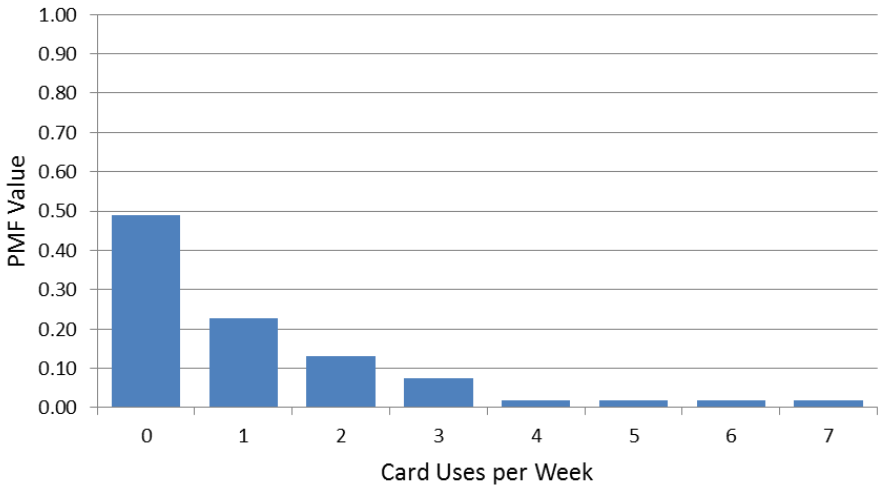


Figure 3: Histogram of Uses per Week for one Go Card User

We then proceeded to cluster these histograms, using the Wasserstein metric for density functions (Irpino et al, 2006). The results of this analysis are shown in Figure 4, using cumulative distribution functions (CDFs) for each

cluster. The number of clusters (10) was based on the Calinski and Harabasz index (CH), a common heuristic to identify a good number of *k* means in a given data set (Irpino et al, 2006). To identify infrequent users, we were more curious about the CDFs that are strongly right-skewed, with high densities toward 0-3 uses per week; these are clusters 2, 3, 7 and 9; cluster 6 is a single Go Card user – an obvious outlier. A further analysis of these infrequent smart card users is a work in progress, and these results will be reported in full at the workshop and in a later version of this extended abstract.

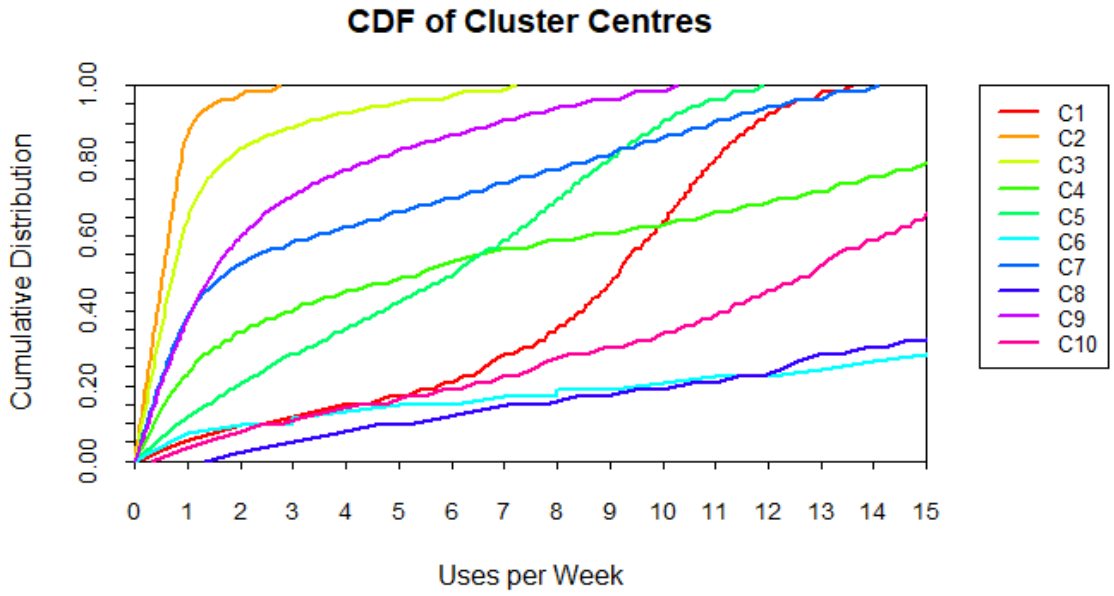


Figure 4: Cumulative Distributions for Centres of *k* Clusters by Number of Uses per Week (*k*=10 clusters, n=10,347 Go Cards)

References

Bhaskar, A, and Chung, E (2015). Passenger Segmentation Using Smart Card Data. *IEEE Transactions on Intelligent Transportation Systems*, 16(3), 1537-1548.

Chu, K K, and Chapleau, R (2010). Augmenting Transit Trip Characterization and Travel Behavior Comprehension: Multiday Location-Stamped Smart Card Transactions. *Transportation Research Record: Journal of the Transportation Research Board,* 2183, 29-40.

Irpino, A, Verde, R, and Lechevallier, Y (2006). Dynamic Clustering of Histograms Using Wasserstein Metric. In: Rizzi, A, and Vichi, M, *COMPSTAT 2006 – Advances in Computational Statistics*, 869-876, Heidelberg: Physica-Verlag.

Kieu, L M, Bhaskar, A, and Chung, E (2014). Transit Passenger Segmentation Using Travel Regularity Mined from Smart Card Transactions Data. Presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, DC, January.

Lee, S G, and Hickman, M (2014). Trip Purpose Inference Using Automated Fare Collection Data. *Public Transport*, 6(1-2), 1-20.

Morency, C, Trépanier, M, and Agard, B (2007). Measuring Transit Use Variability with Smart-Card Data. *Transport Policy*, 14(3), 193-203.

Neff, J, and Pham, L (2007). A Profile of Public Transportation Passenger Demographics and Travel Characteristics Reported in On-board Surveys. American Public Transportation Association.

Park, J, Kim, D J, and Lim, Y (2008). Use of Smart Card Data to Define Public Transit Use in Seoul, South Korea. *Transportation Research Record: Journal of the Transportation Research Board,* 2063, 3-9.

Wang, W. Attanucci, J P, and Wilson, N H M (2011). Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems. *Journal of Public Transportation*, 14(4), 7-22.