AGILE: GIScience Series, 4, 25, 2023. https://doi.org/10.5194/agile-giss-4-25-2023
Proceedings of the 26th AGILE Conference on Geographic Information Science, 2023.
Editors: P. van Oosterom, H. Ploeger, A. Mansourian, S. Scheider, R. Lemmens, and B. van Loenen.
This contribution underwent peer review based on a full paper submission.

Exploring Non-Routine Trips Through Smartcard Transaction Analysis

Nir Fulman[1,3], Maria Marinov[2], and Itzhak Benenson[3]

[1] GIScience Research Group, Institute of Geography, Heidelberg University, Heidelberg, Germany
[2] Department of Industrial Engineering and Management, Ort Braude College, Karmiel, Israel
[3] Department of Geography and Human Environment, Porter School of Environmental Studies, Tel Aviv University, Tel Aviv, Israel

Correspondence: Nir Fulman (Nir.Fulman@uni-heidelberg.de)

Abstract. Public transportation (PT) studies often overlook non-routine trips, focusing on commuting trips. However, recent research reveals that occasional trips comprise a significant portion of public transportation trips. Furthermore, traveler preferences for non-routine trips essentially differ from their preferences for regular commuting. We investigate non-routine trips based on a database of 63 million records of PT boardings made in Israel during June 2019. The behavioral patterns of PT users are revealed by clustering their boarding records based on the location of the boarding stops and time of day, applying an extended DBSCAN algorithm. Our major findings are that (1) conventional home-work-home commuters are a minority and constitute less than 15% of Israeli riders; (2) at least 30% of the PT trips do not belong to any cluster and can be classified occasional; (3) The vast majority of users make both recurrent and occasional trips. A linear regression model provides a good estimate (R^2 = 0.85) of the number of occasional boardings at a stop as a function of the total number of boardings, time of a day, and land use composition around the trip origin.

Keywords. Public transport, Smartcard, Spatio-temporal clustering, Demand-Responsive Transport

1 Introduction

Before the 2000s, motorized urban transport was largely stagnant. A traveler could choose between rigid public transport (PT) services - buses, light rail, metro, and trains with fixed routes and time schedules, and private cars or costly taxis offering complete schedule and route flexibility. In the early 2010s, new demand-responsive transportation (DRT) modes started to appear, exploiting mobile apps to match users to vehicles (Cohen and Shaheen, 2018). The new Ride-Hailing (RH) and vehicle-sharing services are fully or partially flexible in terms of routes, stops, and schedules. Typically, these services are operated by Transportation Network Companies (TNCs) that coordinate vehicle fleets to balance the operation costs and prices/level of service. The DRT services are more expensive than PT and, typically, the prices of RH may be close to the price of regular taxis, yet users describe them as advantageous in terms of convenience, comfort, and safety (Rayle et al., 2016). As a result, since being introduced, the use of DRT services steadily grows (Graehler et al., 2019) and the anticipated transition to autonomous vehicles is expected to strengthen this tendency (Schaller, 2021).

The introduction of DRT stimulated the hope that private car users would prefer this mode to their cars (Erhardt et al., 2021). This did not happen. Instead, recent studies show that the areas served by ride-hailing services experience a significant decline in PT ridership, that is, in the number of people who use these services (Graehler et al., 2019). Studies of the data accumulated by the TNCs demonstrate that their services are primarily used for leisure, errands, and other irregular and not work-related trips (Zhong et al., 2018).

The transition from traditional PT to flexible services for the occasional trips underscores the differences in traveler mode choice preferences for these two types of trips and emphasizes the necessity to examine the unique characteristics of occasional trips. This trip dichotomy is not typically drawn in public transportation studies, although recent works consistently demonstrate that non-routine trips are very common (Goulet-Langlois et al., 2018; El Mahrsi, 2014). In this paper, we investigate non-routine PT trips based on the dataset of 63 million smartcard validation records in Israel in June 2019. Our goal is to identify the characteristics of these trips and to estimate how frequent they are for different groups of travelers, in space and in time. Based on these estimates we consider possible changes in travelers' mode choice when flexible transportation modes will become available.

The following Section 2 presents the dataset, and Section 3 is devoted to the PT ridership analysis. We discuss the results of this analysis and the policy implications of our findings in concluding Section 4.

2 Data

The investigated dataset consists of the 63M records of smartcard (SC) ride validations in June 2019 over the entire Israeli public transport network. The Israeli SC system for buses is tap-on only and a ride is recorded when the traveler boards the bus. For a train ride, alighting is also recorded. When the SC is validated, the information recorded on the operator's database is as follows: User's unique ID (recoded for this study to protect privacy); Payment agreement that can be of two major types - Basic Fare Pass (BP) which is paid every boarding from an electronic purse, and Prepaid Pass (PP) which is paid in advance and allows unlimited rides within a region; Profile - General, Elderly, Student, etc.; Boarding stop ID; Line ID; Exact time of the onboard validation. To locate lines and stops we exploit the open GTFS dataset of the Israeli Ministry of Transport (Google 2021). According to this database, c.a. 3,000 bus lines and 19,000 stops operated in Israel in June 2019. Accounting for 90% of the boardings, buses are Israel's main form of PT. The remaining trips were made by train.

2.1. Selection of data for analysis

In Israel, the PT system is largely suspended between Friday afternoon and Saturday evening. In what follows we consider transactions collected over 20 working days of June 2019. We filter out sequential boardings made by a user in two or fewer minutes; users who boarded 12 or more times on one or more of 20 working days; and transfers, which happen in close succession to previous boardings. For the analysis below, we have selected the users of the two major agreement types.

1. Periodic Pass (PP): Monthly, yearly, special student semester tickets - 12.2M records of 308K users.
2. Basic Fare Pass (BP): Prepaid multiple-entry, stored-value pass - 23.6M records of 2.2M users.

We analyze users of the four major profiles, disregarding minor ones such as Soldiers.

1. General Users: The default smart card profile. 19.5M records of 1.39M users.
2. Elderly/Seniors: 6.2M records of 396K users.
3. Youth/Teenagers: 6.8M records of 522K users.
4. Students: 2.0M records of 94K users.

The investigated dataset contains 34.7M records that describe the use of PT by 2.4M users during June 2019.

3 Analysis

3.1 General view of PT ridership in Israel

On a typical working day in June 2019, 800K users boarded public transport and made 1.7M trips. 100K (12.5%) of them used Periodic Pass (PP) cards and 700K (87.5%) used Basic Fare Pass (BP) cards. On average, Israeli travelers board 1.7M/0.8M = 2.1 times a day. The distribution of users by the number of boardings per day for the PP and BP agreement types is presented in Tab. 1, with 33% of users boarding once a day.

The intensity of monthly PT ridership is different for the PP and BP holders (Fig. 1). The majority of PP users board PT most workdays of the month, the average number of days of use is 15.6, and the share of travelers who use PT grows with the increase in the number of travel days and reaches 25% for those who use PT every working day. For BP holders, the average number of travel days is 5.5, and the share of those who used PT one working day is the highest, 25%. This share monotonously decreases with the increase in the number of travel days.

A similar tendency is characteristic of the number of monthly PT trips. The average number of boardings for PP card holders is 39.6, while for BP card holders it is 10.6. More than 60% of BP card owners board PT less than 10 times a month, and the median number of boardings for them is 7, while the median number of rides of PP owners is 38 (Fig. 2).

Table 2 specifies the statistics presented in Figs. 1-2 by agreement type and user profile. The statistics of the major General profile are close to the overall average. Teenagers and Student BP holders ride more often than other BP holders, yet PP Teenagers and Students ride less than average for their agreement type.

Table 1: PT Use by the number of boarding per day, for the Prepaid and Basic Pass users			
Rides/day	Total (800K, 100%)	PP holders (100K, 12.5%)	BP holders (700K, 87.5%)
1	33%	15%	40%
2	41%	46%	39%
3 – 4	22%	32%	18%
5 – 6	4%	6%	3%
7 – 12	1%	1%	<1%
Avg.	2.1	2.52	1.93

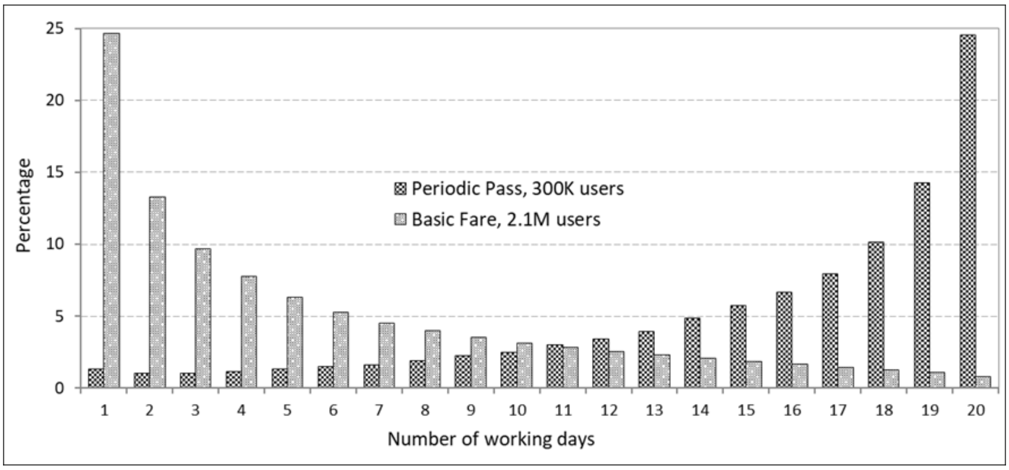

Figure 1. The distribution of the number of PT ridership days for PP and BP holders over the 20 working days of June 2019.

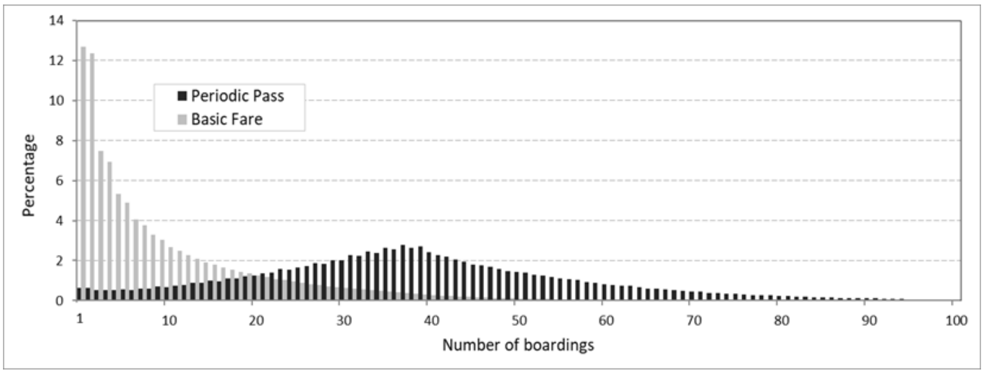

Figure 2. The number of PT rides on the working days in June 2019, for the PP and BP holders.

Table 2: Ridership statistics by four user profiles and two agreement types					
Profile	Agreement	Users per day	Boardings per day	Days of use per 20 workdays of the month	Boardings per 20 workdays of the month
Elder	Prepaid	22K	2.7	16	41
Elder	Basic	108K	2.0	5	10
General	Prepaid	65K	2.5	16	40
General	Basic	400K	1.9	5	10
Student	Prepaid	10K	2.6	13	35
Student	Basic	22K	2.0	7	14
Youth	Prepaid	3K	2.3	15	35
Youth	Basic	170K	1.9	7	13
Total	Prepaid	700K	1.9	6	11
Total	Basic	100K	2.5	16	40
Grand total	-------	800K	2.1	6.8	14.3

3.2. Clustering public transport trips

We use an extended version of the DBSCAN algorithm to cluster public transport rides by location and time, using spatial and temporal thresholds. The algorithm requires at least minPnt other boardings within certain spatial ϵ_s and temporal ϵ_t ranges to form a cluster. The values used in the study are minPnt = 2, ϵ_s = 400m, and ϵ_t = 60 minutes, with ϵ_s reflecting the maximum walking distance to a station and ϵ_t based on clusters of work commuting. The minimum demand for a trip to be considered regular is intentionally set low, at two similar trips per month. We classify a trip as *regular* if it belongs to one of the clusters, i.e., if at least one other trip started at a close PT stop at a close hour of the day. Otherwise, the trip is considered *irregular* or *occasional*.

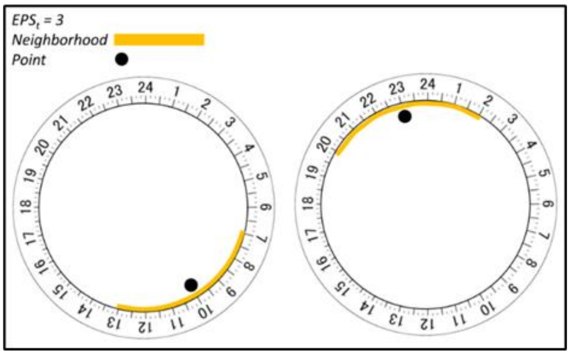

Figure 3. DBSCAN temporal neighborhoods around boardings made at 10:00 and 23:00 for ϵ_t = 3 hours

Typical spatio-temporal clusters of the June 2019 trips for a user with many - 58 boardings, and a user with a few – 10 boardings, are presented in Fig. 4.

3.3. Occasional trips by the user groups and hours of the day

Overall, 42% of all trips are occasional with 51% among the BP holders and 24% among PP holders. Fig. 5 shows that the share of occasional trips decreases drastically with the increase in the number of monthly trips made by the user. It declines from 60% for users with fewer than 10 monthly rides to 20% for users with 40 or more monthly boardings. BP users have only slightly higher shares of occasional trips than PP users with similar numbers of monthly trips. Students are the most sporadic while Teenagers are the most regular.

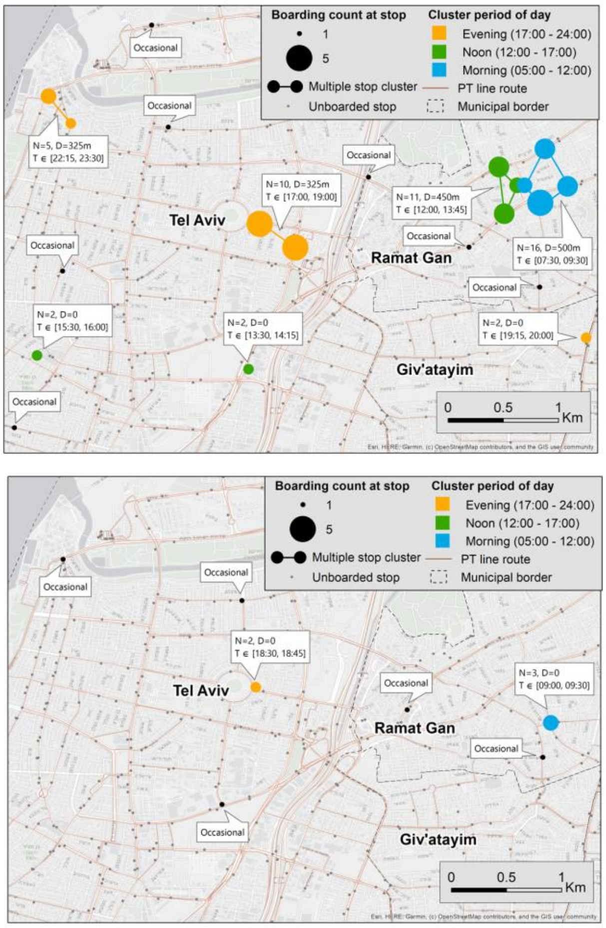

Figure 4. The clusters of June 2019 boardings for two travellers of the "General" profile: Top - 58 boardings of a PP holder are organized into 7 clusters and 10 occasional boardings, 4 of the which outside the map scope; Bottom - 10 boardings of a BP holder are organized into 2 clusters and 5 occasional boardings.

The share of occasional trips varies widely by the time of day (Fig. 6). It has a clear peak between 10:00 - 12:00 and declines later in the afternoon, with a typically higher percentage for BP cardholders compared to PP users. Teenagers' pattern is different from other groups' patterns, probably reflecting daily school schedules.

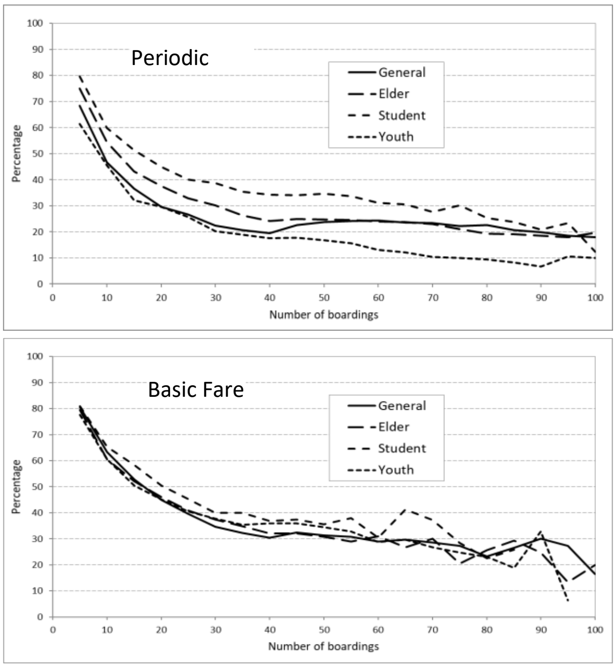

Figure 5. The share of occasional boardings by the users' agreement type, as dependent on the number of rides in working days of June 2019, for the holders of BP (top) and PP (bottom).

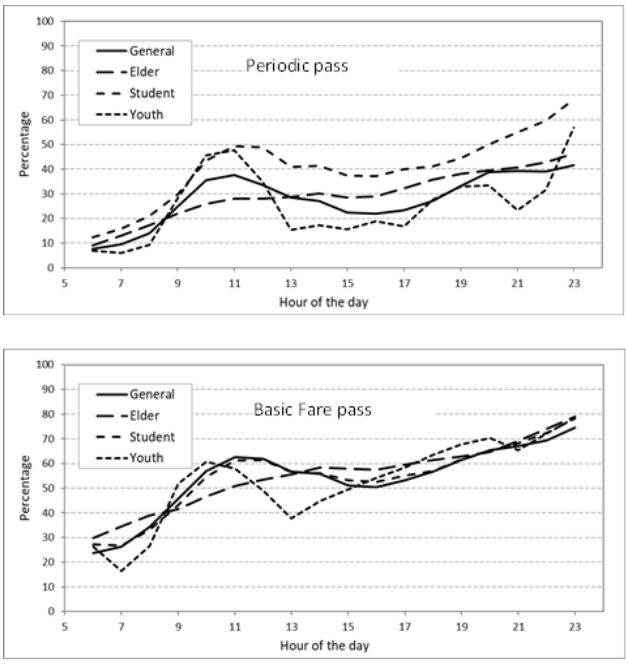

Figure 6. The share of occasional rides for the users of different profiles by the hour of the day: PP holders (top), BP holders (bottom).

3.4. Occasional trips by lines and stops.

The share of occasional trips varies greatly by lines (mean = 41%, STD = 24%) and stops (mean = 46%, STD = 24%), indicating that different lines and stops serve different purposes. Expectedly, the share of occasional trips for the BP holders is higher (line mean = 47%, STD = 15%, stop mean = 49%, STD = 17%,) and varies more than that for the PP holders (line mean = 20%, STD = 11%, stop mean = 24%, STD = 14%).

3.5. Spatial variation of the share of occasional trips

3.5.1. Correlation over stops of the same line.

Let us estimate the correlation of the daily shares of occasional boardings at the stops of the same line. The shares of occasional boardings at two sequential stops are strongly and positively correlated with r ~ 0.7 ($p < 0.01$) and with the increase in distance between stops the correlation decreases yet remains high, between $r = 0.2$ - 0.4 (Fig. 7). This trend is repeated for each user profile and period of the day.

3.5.2. Correlation between the share of occasional trips at nearby stops.

To estimate the correlation between the shares x_i of occasional trips at the nearby stops we applied the Moran's Index of spatial autocorrelation (Anselin et al., 2010):

$$I = \frac{N}{W} \frac{\Sigma_i \Sigma_j (x_i - \bar{x})(x_j - \bar{x})}{\Sigma_i (x_i - \bar{x})^2} \qquad (1)$$

where N is the number of stops "nearby" to a stop; x is the share of occasional trips, and $\bar{x}$ is the average of x. As a "nearby" to a stop, we consider stops at a distance less than 400 meters from a certain stop.

The value of Moran's I is positive and significant at $p < 0.01$ for all user profiles and periods of the day (Tab. 3), while, naturally, lower than the value of this index for the consecutive stops of the same line (Fig. 7).

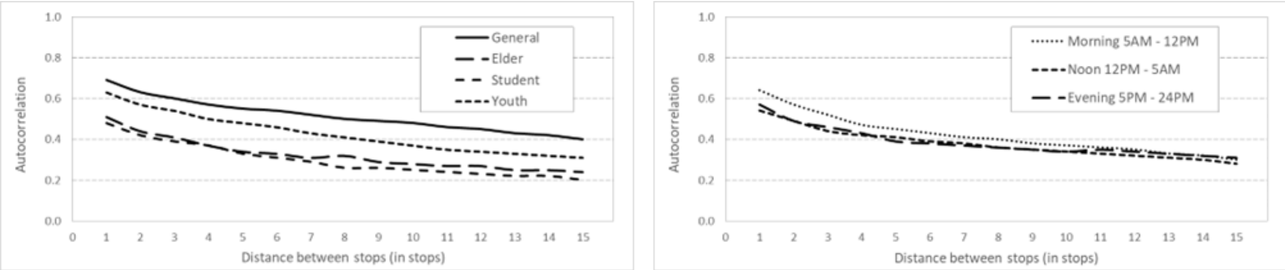

Figure 7. The correlation between the shares of occasional trips for sequential stops of the same line, by user profile (left), and by the period of the day for all profiles together (right).

Table 3: Moran's I spatial autocorrelation. All R values are significant at $p < 0.01$

User profile/period of day	Moran's I
Overall	0.45
User profile	
General	0.38
Youth	0.40
Elder	0.29
Student	0.20
Period of day	
Morning (5 AM – 12 PM)	0.42
Noon-afternoon (12 PM – 5 PM)	0.36
Evening (5 PM – 24 AM)	0.30

3.6. Spatiotemporal pattern of occasional boardings

We investigate the spatio-temporal pattern of occasional rides for the central part of the Tel Aviv metropolitan area (Tel-Aviv Metropolitan Center, TMC) that includes Tel Aviv and four neighboring cities. The general statistics of public transport ridership within the TMC and for the rest of the county are similar (Tab. 4). During the day, the total share of occasional boardings within the TMC varies from 34% in the morning (5 AM – 12 PM) to 44% at noon (12 PM – 5 PM) and 54% in the evening (5 PM – 24 AM), the same as in the rest of the country.

Table 4: Ridership statistics for the Tel Aviv Metropolitan Center (Figures 11-13) and the rest of the country

	TMC			The rest of the country		
Statistic	Periodic Pass holders	Basic Fare Pass holders	Total	Periodic Pass holders	Basic Fare Pass holders	Total
Boardings	3.6M	5.6M	9.2M	8.6M	16.9M	25.5M
Unique users	150K	900K	1.05 M	279K	1.8M	2.1M
Boarding days/month	16.0	6.0	7.4	15.7	5.7	7.0
Boardings/month	39.5	11.8	15.8	40.1	11.2	15.0
Percentage of occasional boardings	24.8%	52.6%	41.7 %	24.3%	50.1%	41.4%

Fig. 8 presents the shares and the volumes of occasional boardings at the TMC stops using the stop-based Voronoi coverage, for three periods of a day. The occasional volumes are relatively steady over the day (Fig. 8 right) compared to the shares of occasional trips (Fig. 8 left), which grow during the day.

Several areas, mostly non-residential commercial, tourist, and business centers, have high shares of occasional trips, above 50% throughout the day. Some areas have both high rates and volumes.

To estimate the possible relationship between the land uses and the volume of occasional boarding at the stop, we exploited the layer of buildings of the National Geographic Data Base (BNTL) of the Survey of Israel (SOI 2018). This layer contains the foundation polygon of each building with the attributes of use – residential, commercial, industrial, public, transportation, agriculture, and the building's height. Based on this layer, we calculated the residential and non-residential floor area within a 400m radius of each stop. The correlation between the total volume of occasional trips (Fig. 9) that start at this stop and the non-residential floor area around is r = 0.3 ($p < 0.01$) and is only loosely dependent on the profiles and periods of the day (Tab. 5).

Table 5: The correlations between the number of occasional boardings at a stop and the size of non-residential/residential built-up area within a 400m radius around, all significant at $p < 0.01$

User profile/period of day	Non-residential	Residential
Overall	0.30	0.05
User profile		
General	0.30	0.04
Youth	0.28	0.05
Elder	0.28	0.05
Student	0.21	0.01
Period of day		
Morning	0.25	0.02
Noon	0.31	0.05
Evening	0.30	0.04

A simple linear regression provides a good estimate of the number of occasional boarding at a stop as dependent on the total number of boardings at this stop, the amount of non-residential built-up area around, and the period of the day. We construct it as:

$$y = X_1\beta_1 + X_2\beta_2 + D_1\gamma_1 + D_2\gamma_2 + e \quad (2)$$

Where y denotes the number of occasional boardings at a stop, X_1 represents the non-residential built-up area within the 400m neighborhood of a stop, X_2 is the overall number of boardings at the stop, D_1 and D_2 are dummy binary variables for noon and evening boardings, respectively, and $\beta1$, $\beta2$, γ_1, and γ_2 are regression coefficients.

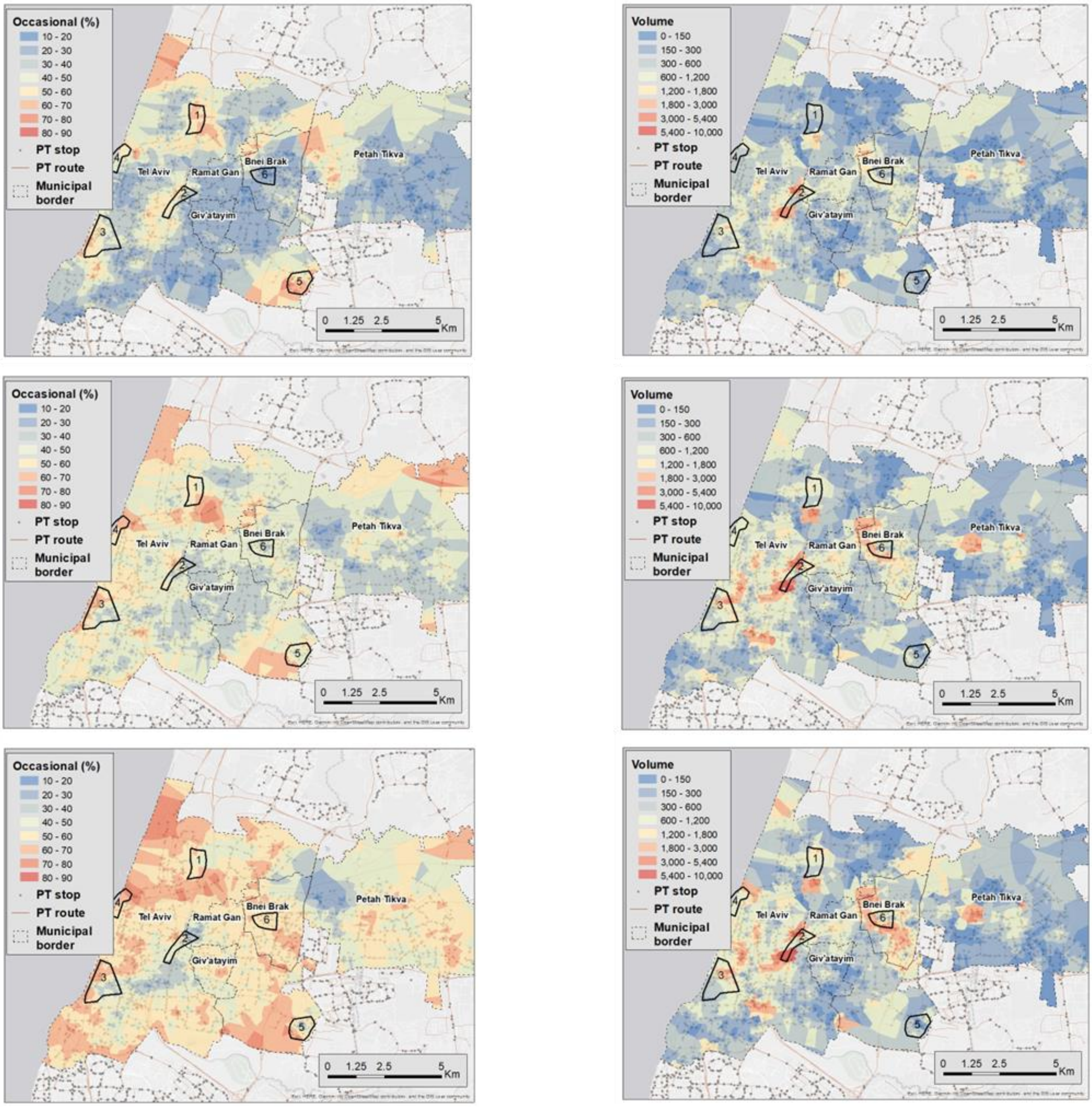

Figure 8. The share (left) and the number (right) of occasional trips at a stop (up) morning 5:00 – 12:00, (mid) noon 12:00 – 17:00, (down) evening 17:00 – 24:00. The areas marked: (1) - Tel Aviv University; (2) Ramat Gan – Tel Aviv Business Center; (3) Carmel Market, Neve Tzedek, and Nahalat Binyamin; (4) Old Tel Aviv Port Area; (5) Sheba Medical Center in Ramat Gan; (6) Bnei Brak city center.

The regression coefficients, all highly significant, are presented in Tab. 6, and $R^2 = 0.85$. As can be expected, the effect of the non-residential area size is positive.

Table 6: The parameters of the linear regression (2), all coefficients significant at $p < 0.01$.

Log-likelihood	-45,087
Intercept	-74.6
Total number of boardings	0.3238
Built-up area (measured in 1000s of sqm)	0.77
Morning (baseline)	----------
Noon	203.6
Evening	274.1
R^2	0.85
Moran's I on residuals	0.11

Fig. 9 presents the maps of the observed and predicted number of occasional boardings at stops and the scatterplot of the predicted versus the observed values.

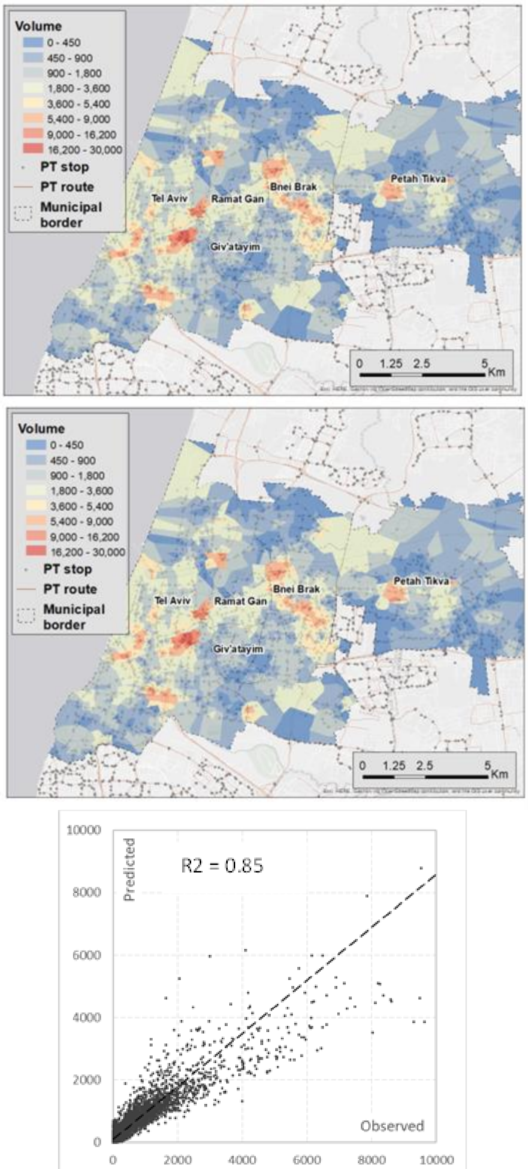

Figure 9. The observed (top) and predicted (mid) numbers of occasional boardings based on the Voronoi diagrams of stops, and the scatterplot of the predicted vs. observed values based on model 2 (bottom)

4. Discussion

This paper studies the characteristics of occasional public transportation ridership based on smartcard transactions. We analyzed a large database of public transport boardings in Israel for 20 working days in June 2019 differentiating between two types of riders (Basic Fare Pass and Periodic Pass) and several user profiles (Students, Youth, Elders, and General).

4.1. Major findings

Conventional commuting is not the primary part of public transport ridership in Israel: as many as 42% of all boardings should be considered occasional. The analysis shows that the share of occasional trips is highest among Basic Fare Pass users and varies by user profile, time of day, and PT line and stop. The number of occasional boardings at a stop is correlated with the number of occasional boardings within a 400m distance and increases with the overall number of boardings in the neighborhood and non-residential built-up area. A linear regression model based on smartcard and land-use data can accurately forecast the number of occasional boardings at a stop.

4.2. Policy implications

Our research reveals unique characteristics and usage patterns of non-routine public transportation (PT) ridership, including time of day, location, and user type. This approach can be replicated in other areas to uncover specific patterns, which can inform policymaking based on the travel preferences of locals, available transportation options, and the goals of decision-makers. While limited literature exists on mode choice preferences for occasional trips, evidence suggests that these trips differ significantly from commuting trips, and users may prioritize aspects such as comfort.

Policymakers are split regarding the effect of DRT services on urban transportation. Recent field and model studies reveal that the introduction of DRT services causes a decline in PT use, increases congestion, and decreases the overall effectiveness of the transportation system (Schaller, 2021; Ben-Dor et al, 2019). Our approach could be used to identify TP services for which improved service and reduced cost could potentially help to maintain ridership.

On the other hand, modeling studies have proposed shared DRT as a positive future of urban transportation, with potential benefits such as improved accessibility and travel comfort compared to conventional PT (Martinez, Viegas, 2017). Taking this perspective, reducing private car use should remain a major policy objective, with locally adjusted demand-responsive parking prices and/or congestion pricing as effective policy tools. Shared DRT may fulfill the demand for occasional trips in city centers, freeing up resources for conventional PT operators to focus on highly demanded commuting trips. This can be accompanied by reorganization and simplification of the bus network, leading to increased demand and higher service levels for all commuters.

4.3. Limitations and further research

Our study is performed at the very aggregate level of the smartcard transaction in the workdays of June 2019, over the entire of Israel. We do not consider variability in occasional ridership patterns between cities and towns, or different seasons, and ignore PT travel patterns on weekends. We thus plan, on one hand, to investigate the variability of the revealed phenomena by cities and towns and, on the other hand, to apply the proposed approach to the datasets for longer periods and in other countries. It would be also interesting to apply our methods for studying PT usage during periods of global and qualitative changes in travelers' behavior, like the Covid-19 outbreak**.**

References

Anselin, L., Syabri, I., & Kho, Y. (2010). GeoDa: an introduction to spatial data analysis. In *Handbook of applied spatial analysis* (pp. 73-89). Springer, Berlin, Heidelberg.

Ben-Dor, G., Ben-Elia, E., Benenson, I. 2019. Determining an optimal fleet size for a reliable shared automated vehicle ride-sharing service. Procedia Computer Science, 151, p 878 - 883

Canepa, B. (2007). Bursting the bubble: Determining the transit-oriented development's walkable limits. *Transportation Research Record*, *1992*(1), 28-34.

Cohen, A., & Shaheen, S. (2018). *Planning for shared mobility*. PAS Report 583, American Planning Association.

El Mahrsi, M. K., Côme, E., Baro, J., & 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*.

Erhardt, G. D., Mucci, R. A., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2021). Do transportation network companies increase or decrease transit ridership? Empirical evidence from San Francisco. *Transportation*, *49*(2), 313-342.

Fulman, N., & Benenson, I. (2019). Establishing heterogeneous parking prices for uniform parking availability for autonomous and human-driven vehicles. *IEEE Intelligent Transportation Systems Magazine*, *11*(1), 15-28.

Google (2021). GTFS Static Overview. https://developers.google.com/transit/gtfs. Accessed 20/11/2021.

Graehler, M., Mucci, R. A., & Erhardt, G. D. (2019, January). Understanding the recent transit ridership decline in major US cities: Service cuts or emerging modes. In *Transportation Research Board 98th Annual Meeting, Washington, DC, January*.

Goulet-Langlois, G., Koutsopoulos, H. N., Zhao, Z., & Zhao, J. (2018). Measuring regularity of individual travel patterns. *IEEE Transactions on Intelligent Transportation Systems*, *19*(5), 1583-1592.

Martinez, L. M., & Viegas, J. M. (2017). Assessing the impacts of deploying a shared self-driving urban mobility system: An agent-based model applied to the city of Lisbon, Portugal. *International Journal of Transportation Science and Technology*, *6*(1), 13-27.

Rayle, L., Dai, D., Chan, N., Cervero, R., & Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ride-sourcing services in San Francisco. *Transport Policy*, *45*, 168-178.

Schaller, B (2021). Can sharing a ride make for less traffic? Evidence from Uber and Lyft and implications for cities, Transport Policy, v102, 1-10

SOI - Survey of Israel (2018). BNTL – National Geographic Database.

Zhong, J., Ye, X., Wang, K., & Li, D. (2018). A modeling analysis of the impact of E-hailing service on non-work travel mode in Shanghai, China. *Transportation Research Record*, *2672*(47), 125-134.

Acknowledgments

The research was funded by the research grant "Estimating the demand for public transport based on the smartcard and mobile phone data" #0606915691 of the Israeli Ministry of Transport and Road Safety (MOT).