USING BIG DATA OF AUTOMATED FARE COLLECTION SYSTEM FOR ANALYSIS AND IMPROVEMENT OF BRT- BUS RAPID TRANSIT LINE IN ISTANBUL

Ilgin Gokasar, Assistant Professor
Bogazici University
34342 Bebek/Istanbul Turkiye
Tel: +90(212) 359-7278 Fax: +90(212) 287-2457; Email: ilgin.gokasar@boun.edu.tr

Kevser Simsek, Corresponding Author
Istanbul Metropolitan Municipality Belbim Inc., Bogazici University
Hava Limani Karsisi Istanbul Dunya Ticaret Merkezi A3 Blok Kat:2-3 34149 Yesilkoy/ISTANBUL
Tel: +90(212) 468-0000 Fax: +90(212) 465-5193; Email: kevser.simsek@belbim.com.tr

Kaan Ozbay, Professor
New York University
1 MetroTech Center, 19th Floor, Brooklyn, NY 11201
Tel: +1(646) 997-0552; Email: kaan.ozbay@nyu.edu

Word count: 4,469 words text + 12 tables/figures x 250 words (each) = 7,468 words

Submission Date: 2014-07-31

ABSTRACT

Istanbul´s smart card fare collection system generates large amounts of operational data from the BRT-Bus Rapid Transit line. In addition to ridership, it captures system-wide transactions and provides comprehensive data records on usage. Processing and analysis of these data open new opportunities in transportation and travel behavior research. This paper presents a qualitative analysis of smart card (Istanbulkart) activity for the BRT-Bus Rapid Transit and investigates its potential for understanding complexities of the system and characterizing travel behavior. In this paper, an assessment of spatial and temporal travel behavior of commuters including mode choice, travel, and waiting times, is performed using Rstudio, a free and open source integrated development environment (IDE) for R. As a result of this qualitative analysis an evaluation of the automated fare collection system and pricing policies for public transportation along with some recommendations for improving the planning and management of the BRT-Bus Rapid Transit line are provided.

Keywords: Smart card (Istanbulkart); Automated fare collection; Transportation planning; Transit data; BRT-Bus Rapid Transit

I INTRODUCTION

Automated Fare Collection (AFC) - Smart card (Istanbulkart) Technology

With the improvement of smart card technology, automated fare collection systems have become the most common collection method used by public transit authorities. Since their invention in the 1969, new uses have been added to the original purpose of smart cards. They are portable and durable, features that make them useful for many purposes, such as authorization, payment, and identification. One important use of the smart card is the collection and processing of data.

Smart cards have become the fast, contactless, wireless technology of choice. Since 1994 Istanbul first used the Akbil for fare collection, and then, with the improvement of contactless transport systems, which respond to the requirements of operators and end-users alike, the smart card, appropriately named the Istanbulkart, was put into service in Istanbul’s public transportation system in 2004. Today, the Istanbulkartcard can be used for payment for all modes of public transport.

The operators of Istanbulkart technology reduced the system’s fraud and maintenance costs by replacing the inefficient paper-based fare collection system. Istanbulkart saves time by getting rid of queues for paper tickets. With use, the data collection and reporting capacity of the Istanbulkart has been improved.

Istanbulkart is similar in look and size to a credit card. It is used to pay for transportation in place of more traditional methods, such as tickets and cards with magnetic stripes. Each Istanbulkart has a unique serial number. A card can be assigned to a specific individual or it can be anonymous, and each unique card ID represents one single person. This enables analysis of individual itineraries and opens new ways for understanding people’s travel behavior on short as well as long term scales.

In Istanbul, the fare charge for each customer is based on travel distance, transport mode, and certain demographic attributes, such as prioritized rates for elderly people, children, students, senior citizens, government employees, etc. A range of payment and fare options can be created electronically. Currently, for example, 37 different media types and 69 different fare types are provided by the Istanbulkart (Table 1). Fares are collected by an automated reader (Validator) next to the driver or at a turnstile before one boards a transport vehicle. The Validator is a smart device that does validity checks, collects fares in accordance with specified tariffs, and records the result of all transactions. Data collected from the stations are transferred from the stations’ data transfer computers to the automated fare collection server located in the data center by means of an established external network.

TABLE 1 Istanbul Public Transportation System Ticket Types

Ticket Types			Definition
CR Media		SJ	Single Journey Ticket
		AT	Anonymous Ticket
		RFT	Reduced Fare Ticket
		FET	Free Entry Ticket
		ST	Seasonal Ticket
		FC	Function Card
From Istanbul AFC		Akbil	Ancient anonymous ticket
	Limited Usage Tickets	birGec	1 Journey Ticket
		ikiGec	2 Journeys Ticket
		beşGec	5 Journeys Ticket
		onGec	10 Journeys Ticket
	Stored Value Tickets Istanbulkart	Anonymous	Anonymous Ticket
		Discounted	Reduced Fare Tickets
		Free	Free Entry Tickets
		Function	Function Cards

Fare collection methods usage percentages for the BRT line are Istanbulkart 95%, Limited Usage Card 3%, and Akbil 2%. (1)There are refund machines at the exits from BRT- Bus stations. These machines recognize the cards of travelers who have used only a portion of the line and credits them with refunds up to 46% of the full fare. Hence, the trip time of each traveler can be calculated from information taken from the refund machines (Figure 1). The graphics are obtained for approximately 46% of all travelers because only some travelers receive refunds to their Istanbulkarts (2).Thus, besides of the information on boarding time and location, the data collected from Istanbulkart cards contain detailed records of alighting times and destination location for Bus Rapid Transit (BRT) stations.

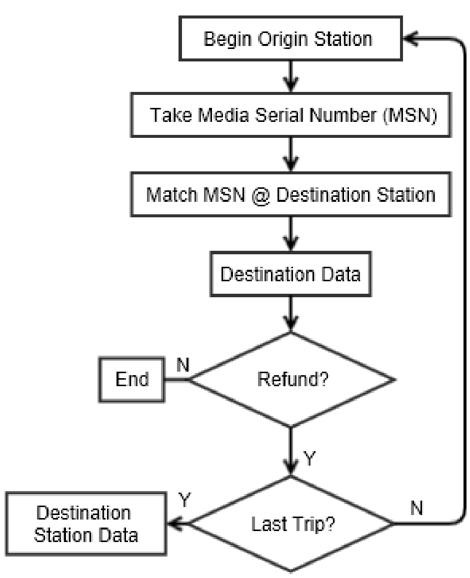

FIGURE 1 BRT-Destination station data process flow chart.

Typically, the date and time of the transaction, status of payment (transfer, acceptance, or refusal), card ID, fare type (student, adult, senior), route ID, and other related data are stored in the Validator and the central server (3). Because these data allow for more detailed assessment of travel behavior and mobility patterns they are invaluable for transit planning, long term planning, and daily management of the transportation system. Masses of data are collected and stored, complicating the process of analyzing the data.

In brief, by using a smart card system, transportation authorities have access to:

1. personal travel data of millions of people,
2. information about each card and/or traveler,
3. continuous trip data including refund information,
4. identity of user and frequency of use.

Despite the accumulation of so much information, it is difficult to improve the Istanbulkart's usability and accessibility for the following reasons:

1. Its prevalent purpose is not to monitor performance of the transportation system; hence, additional passenger trip information such as destinations and delays cannot be directly retrieved.
2. Each passive data collection method has its disadvantages, and processing usually requires additional knowledge. Interoperating and mining heterogeneous datasets would enhance both the depth and reliability of transportation studies.
3. The amount of data obtained is increasing tremendously (approximately 6 million distinct Istanbulkart data every day) and traditional data processing methods might not be equal to the task.

Such data barriers make the development of a large-scale transportation performance monitoring system cumbersome and slow (4).

Istanbul BRT –Bus Rapid Transit

The Bus Rapid Transit system (BRT) of Istanbul began service in 2007 with the intent of reducing traffic congestion on arterial roads while providing quick and comfortable transportation. The line connects the Asian and European sides of the city,so it is the only intercontinental BRT system in the world.Thanks to this connection, the duration of the 52 kilometer journey from Beylikduzu on the European side to Sogutlucesme on the Asian side has been reduced to 98 minutes (Figure 2).

The salient features of BRT-Bus Rapid Transit are

1. fast transportation (30 seconds service frequency);
2. appropriate for metropolitan area with high population;
3. environmentally friendly;
4. comfortable alternative transportation.

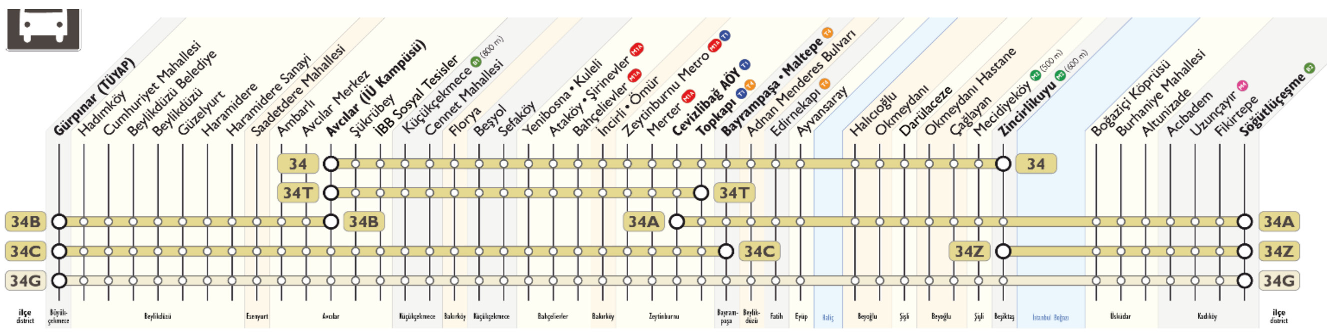

FIGURE 2 BRT-Bus Rapid Transit Line Map.

The yearly and monthly increase of ridership on the BRT-Bus Rapid Transit is illustrated in Figure 3. The increase is due mainly to improvement of the transfer points and the addition of new lines and stations.

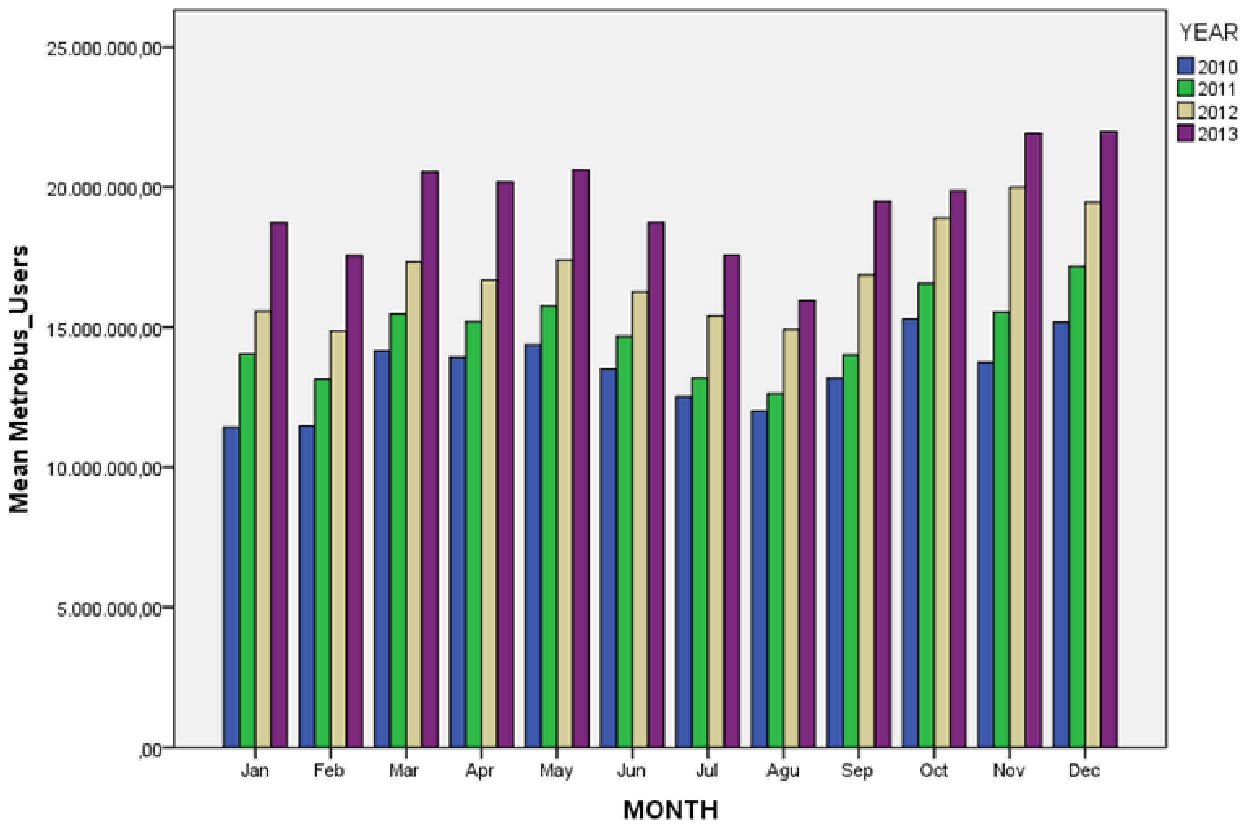

FIGURE 3 Number of Passengers per month.

Figure 4 shows Fare types /Stations per year (2013). Regular fare tickets have the highest share among the total fare types. Figure 5 shows arrivals and departures (gray-dark gray); red sketches show direction to the Asian side, blue sketches show direction to the European side.

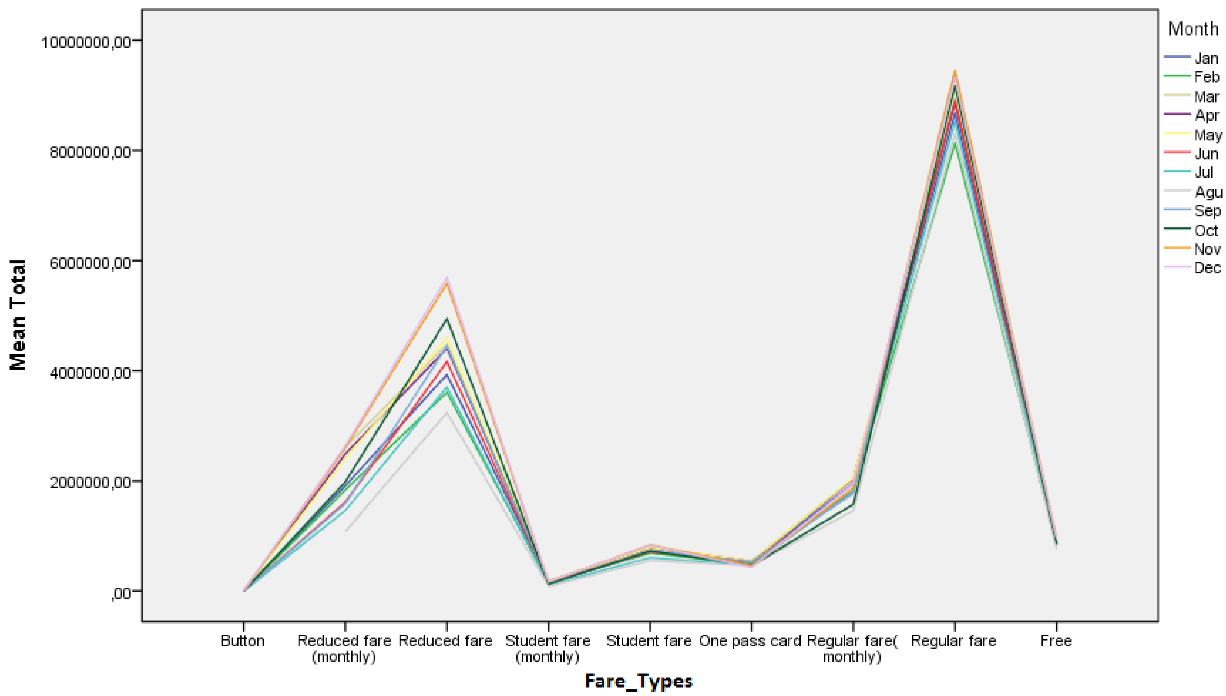

FIGURE 4 Fare types /Stations per year (2013).

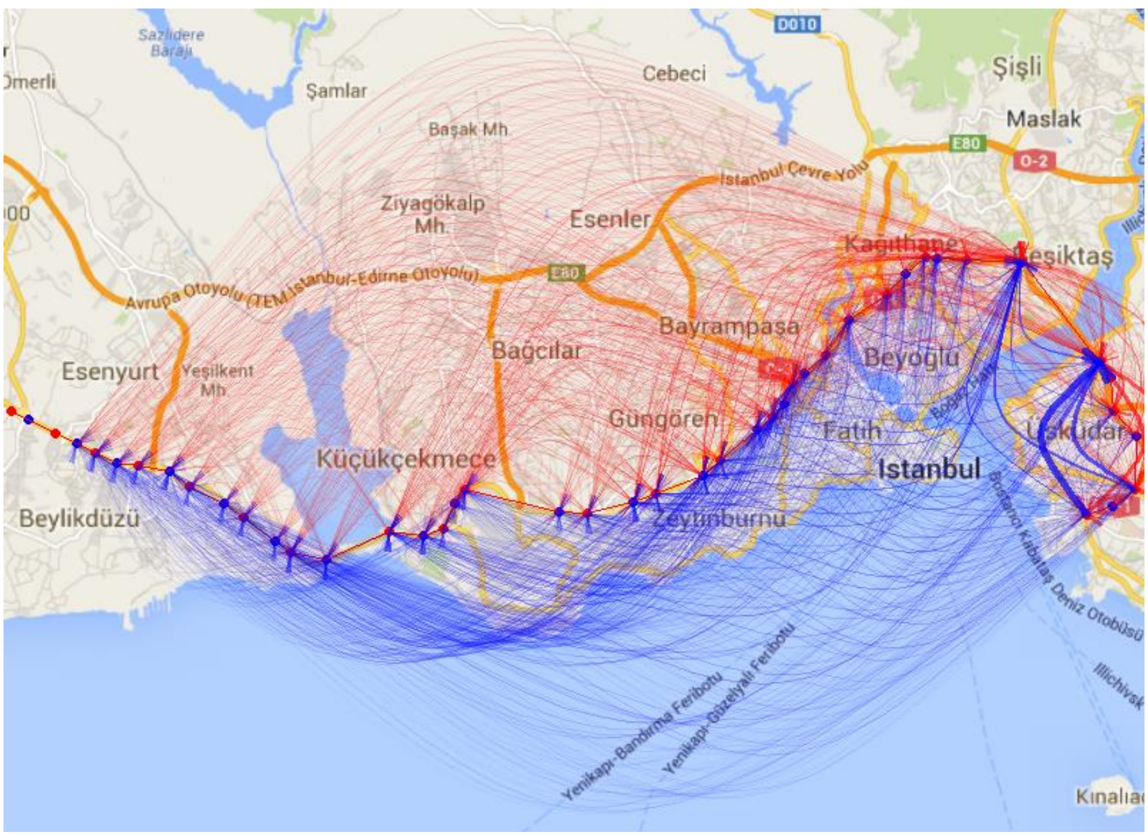

FIGURE 5 BRT journeys /Asian side (red) and European side (blue) (2013).

The BRT-Bus Rapid Transit pricing system depends on the trip distance. When a traveler first uses his or her card, the full price is charged; then, at the destination station, he or she uses the card again, either at a refund machine or on another transit vehicle. First and last usage dates (by mining more than 6,000,000 different passenger data per day) were recorded to obtain mean travel time and origin-destination station data.

2.LITERATURE REVIEW

There are several BRT applications similar to Istanbul BRT line in other metropolitan cities of the world, such as Guangzhou, China, Bogotá, Colombia, Rio de Janeiro, Brazil, Lima, and Peru (6),and the number of published articles, and papers on smart card data analyse of BRT systems is increasing. A comprehensive survey on this topic is offered by CALSTART (2005) reflects what the transit properties are saying regarding the effect of BRT vehicles on ridership. Specific issues addressed include 1) whether the vehicles are branded and/or styled differently than the communities' regular buses, 2) whether the BRT vehicles themselves were responsible for changes in ridership levels and 3) the effect of the vehicles on community acceptance of the BRT system. The survey results indicate that ridership levels increased after BRT system implementation, and, in some cases, up to one third of the new riders came were new to transit and an additional third were riding more often.(7) Moreover previous work with focus on Case Studies in Bus Rapid Transit identifies the potential range of bus rapid transit (BRT) applications, and provides planning and implementation guidelines for BRT was released in January 2004 (8). Chu et al. (2010) presented a methodology for characterization of trips based on socio-demographic characteristics, multiday travel patterns and association of travel with specific locations (9).

Main motivation of this paper is the analysis of smart card (Istanbulkart) data generated by the BRT-Bus Rapid Transit in Istanbul and the investigation of its potential for understanding complexities of the system and characterizing travel behavior. An assessment of spatial and temporal travel behavior, including mode choice, travel, and waiting times based on smart card is another goal of this paper.

3. PRESENTATION AND ANALYSIS OF BRT-BUS RAPID TRANSIT DATA

In this paper, we adopt the definitions used by the Istanbul Land Transportation Authority (ILTA), and define a journey as one-way travel from one activity to another. Each journey consists of one or more consecutive journey stages or trips on the BRT line or a different transportation mode. To deal with the problems of analyzing big data, we enable visualization and analysis of transportation performance measures by using Rstudio and Microsoft Excel to highlight connections among heterogeneous transportation data sets, including Istanbulkart data (10).These programs provide a data-rich visualization platform from which to monitor transit network performance for purposes of planning and operations. This method will also prove to be useful for data-driven transportation research.

This data set was already pre-processed by ILTA and the single trips of each customer were aggregated to journeys according to fare rules and stations (2014). For further processing of the data, we used the MySQL open-source database in combination with MS Excel.csv file format and Rstudio software to extract passenger origin information from the Istanbulkart data. All BRT stations are geocoded using information provided by ILTA. The original aggregated one-day data set contains ca. 800,000 journey records and ca. 460,000 unique card IDs. In the process of preparation and reviewing of these data, about 4,000 journeys (1.1%) were removed from the data set because of errors, illogical journey routes, and missing values.

A variety of methods can be used to characterize and analyze public transport usage on BRT lines. Description in spatial and temporal dimensions are often used by transportation planning authorities as indicators of performance and quality of service. The objective of this section is not to provide comprehensive statistical description of the BRT line system but to show how Istanbulkart data can be specifically used as indicators of region- and culture-specific travel behavior and external factors that influence people's travel decisions and waiting times.

In the analysis of times, the time at which people decide to start their travel can provide valuable data. Figure 6 shows the temporal distribution of boarding times at the first stop or station of the journey for all BRT line journeys during 24 hours. The sharp peaks in the morning and evening hours are noteworthy. The evening peak shows a high number of journey starts concentrated within a 25-45 minute period. The fact that many people start their journeys almost synchronously during the morning as well as the evening peak hours suggests that there is not much flexibility in the travel time choice for commuting, probably as a result of commonly inflexible working hours in Istanbul. Additionally, it is worth noting that the small peak around 1:00 p.m. in the afternoon might be the result of lunch time travel, part-time workers going to or leaving work, and class hours in educational institutions.

The representation of trip distribution highlights the main advantages of analysis based on AFC data as opposed to the commonly used self-report survey. High temporal resolution and reliability of records allow aggregation on a minute basis and thus the detection of sharp peaks, which are clearly visible in Figure 6.

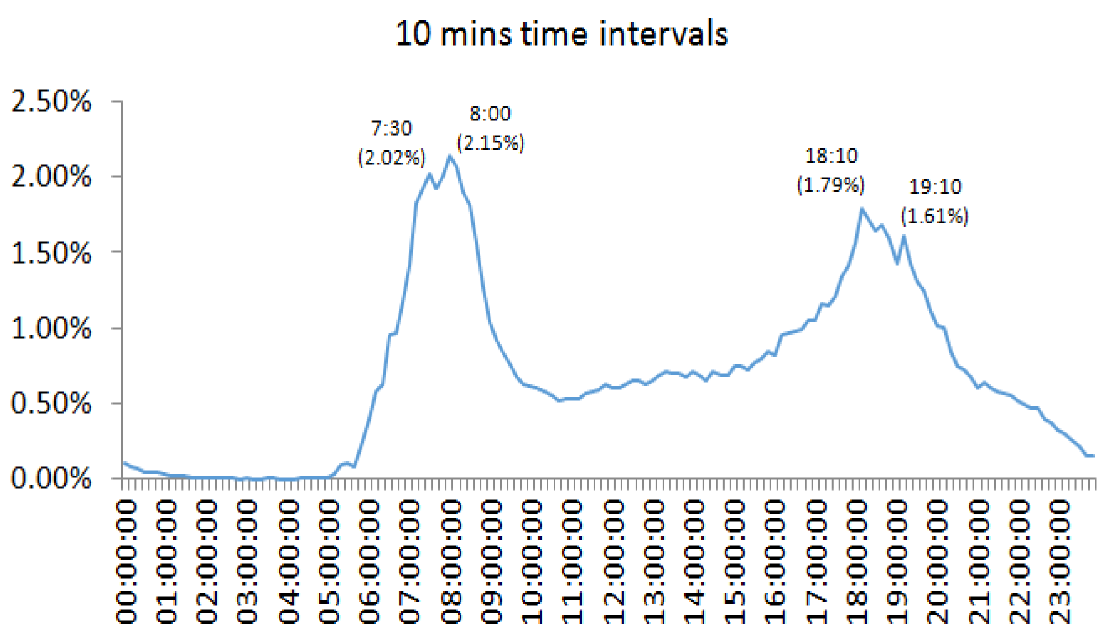

FIGURE 6 Trip start distribution.

Identification of Waiting Times

Another important temporal measure, which has implications on travel behavior and route choice, is waiting time, i.e., the amount of time between a passenger's arrival at a station or stop and the boarding of a vehicle. No information on the waiting time of bus passengers can be obtained from Istanbulkart data, since the smart card readers, on which passengers tap their Istanbulkarts as they board a bus, are located inside the bus. However, Istanbulkarts are tapped when a passenger enters and leaves a BRT station, not on board the vehicle. This means that waiting time is a part of total recorded travel time and can be extracted from Istanbulkart data.

In this paper we use the term "waiting time" to describe the amount of time a passenger spends inside a station, which includes time spent walking in the station and actual waiting time. In order to extract this time for each station, all BRT trips from any one station to other stations on the same BRT line are considered. Hence only direct, uninterrupted trips without interchanges are used for the calculation of waiting times. Travel times of passengers traveling between each pair of stations in each direction are extracted. It is assumed that the fastest passenger arrived at the platform just in time and did not have to wait for the train at all. That passenger's waiting time is considered to be zero and is used as a benchmark. By subtracting this benchmark time from the travel times of all other customers travelling between the same two stations, waiting times of these passengers are obtained (Figure 7).

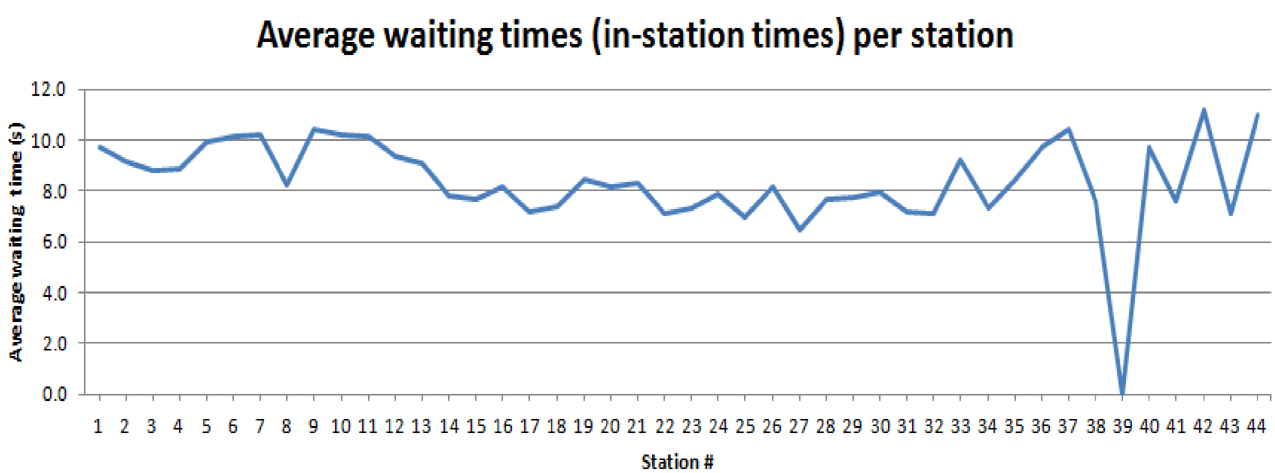

FIGURE 7 Average waiting times (in-station times) per station for West-East BRT Line during 24h

Dependent on the focus of the analysis of waiting times at BRT stations, information on significant service frequency changes during the day as well as potentially overcrowded trains can be gained. However, it is important to note that because of the method for defining and identifying waiting time, the absolute time values do not precisely indicate the service frequency at a particular station. Large hub stations with several access points tend to have larger variations and higher averages of waiting times, which reflects different lengths of access routes from different entry points. In Figure 7 the average waiting times per station for Istanbul's West-East BRT line are shown. As expected, large hub stations with long underground access routes such as Saadetdere, BRT Station 9, show longer waiting times. Notably the longest waiting time is found for Uzuncayir, BRT station 42. This exceptionally high value in comparison with other stations can be explained by high passenger volumes at the first station on the Asian side. Next to the common waiting time distribution around 7 minutes, a second peak of waiting times ranging from 10 to 13 minutes is clearly visible. This peak results from passengers who take BRT in the opposite direction towards the last stop at Zincirlikuyu and then stay on board to secure a seat during the morning peak hour. This observation leads to an interesting conclusion: the value of having a seat is exceptionally high and for some people it is worth about 6 minutes of additional travel time. This finding is also an indication of high passenger volumes and long BRT journeys. In the case of journeys from the Zincirlikuyu station, the average distance to the destination station for passengers taking the detour over the Edirnekapi Station is 7 stations longer compared to passengers taking the direct way.

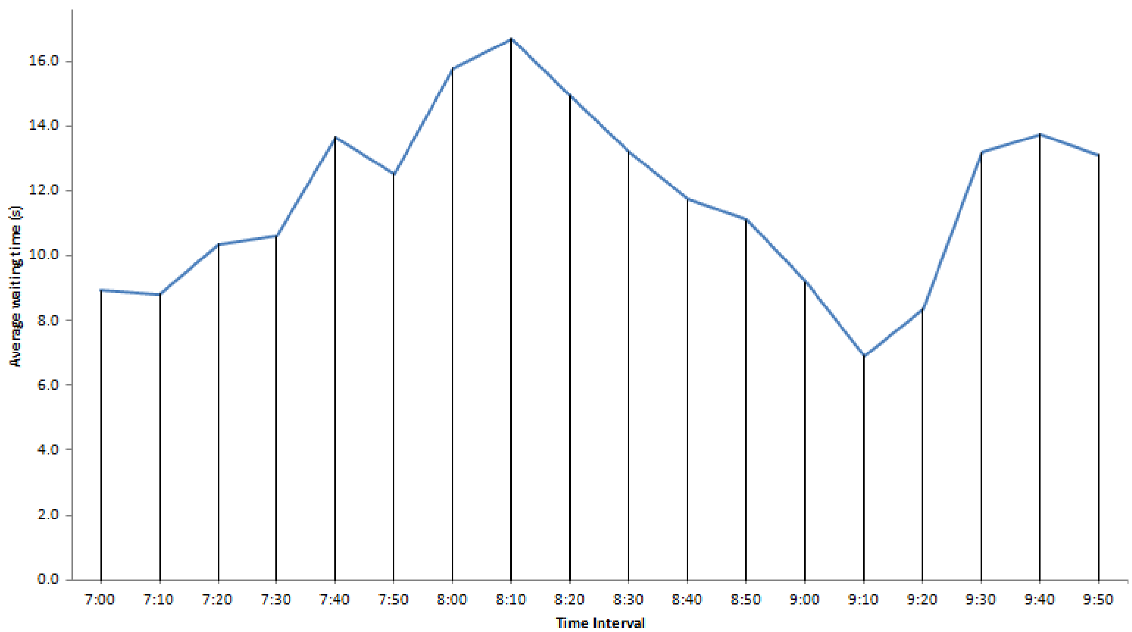

FIGURE 8 Waiting times at the Uzuncayir BRT Station for journeys starting between 7am-10am

Activities as reason for BRT travelling

People usually travel from one place to another because they want to conduct activities, such as work, education, leisure, or social intercourse, that cannot be performed in the desired way at their current location. Hence, in order to understand people's travel behavior in terms of mobility patterns and traffic volumes, it is necessary to look at their daily activities and their locations. In the absence of comprehensive data sources such as a whole population census, one of the main challenges in modelling travel behavior results from a lack of verified data of high spatial resolution on people's home and particularly work locations (11). However such information is important not only for transport planning and implementation of agent-based transport models, but is also invaluable in the areas of urban development and land-use planning. Travel patterns observed in smart card data from public transport can provide important information on people's primary activity locations and can help to verify and refine existing models and assumptions.

To describe the activities of a particular person, the recorded daily journey chain of that person must be consistent. Consistency in the context of AFC smart card data means that the person who arrived at an activity location by public transport has to leave at the end of the activity by public transport; otherwise the duration of the activity cannot be extracted. The assumption of consistency based only on AFC data is hard to verify since the use of any means of transport other than public transport, e.g., walking, cannot be detected. However, obvious cases of inconsistency can be detected by analyzing the distances between the alighting location of the last journey and the boarding location of the following journey (Figure 9).

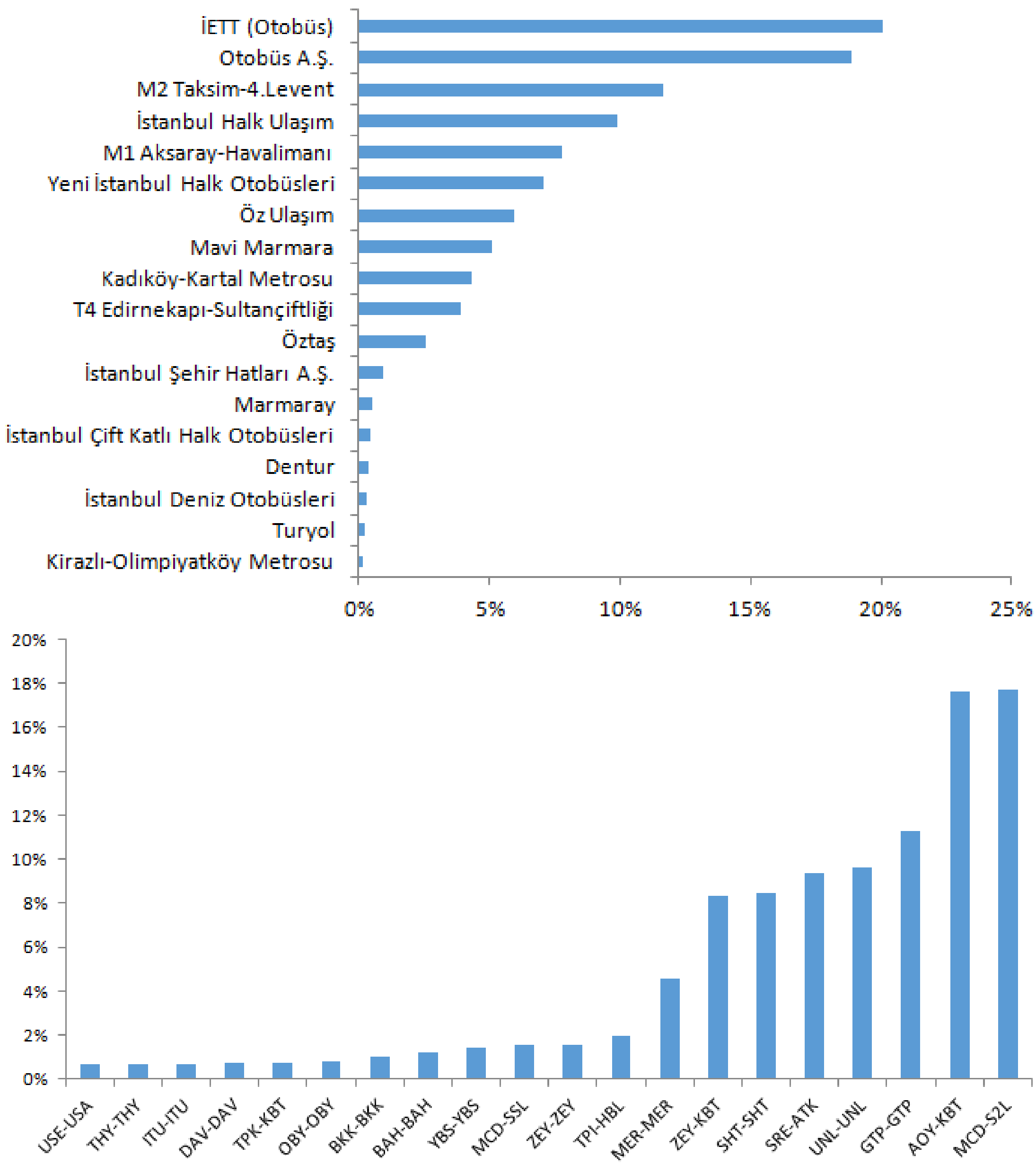

FIGURE 9 Metrobus transits and most common boarding stations.

The analysis of distances between Istanbulkart journeys is shown in the form of a cumulative relative frequency graph in Figure 10. Only persons with more than one journey recorded in the one-day Istanbulkart record were evaluated. The graph shows that 90.2% of journeys following a previous journey start less than 2 km away from the previous alighting location. This indicates that the majority of public transport users do not switch to other transport modes between public transport journeys.

A slightly different picture is obtained by looking at the distances between the first boarding and the last alighting station of the day; if they are nearby they are likely to be in the vicinity of passengers' home locations. As shown in the cumulative relative frequency graph of these distances for all persons in the Istanbulkart data, including those with one journey (Figure 10), only 73% return to a station within a radius of 1 km from their first departure station of the day. This figure is largely accountable by the fact that 27% of BRT journeys are one-way journeys. To better understand mode choice and identify transportation modes used in combination with public transport in Istanbul, in the following section we refer to previous annual records of the Land Transportation Authority to obtain accurate data (2010-2014).

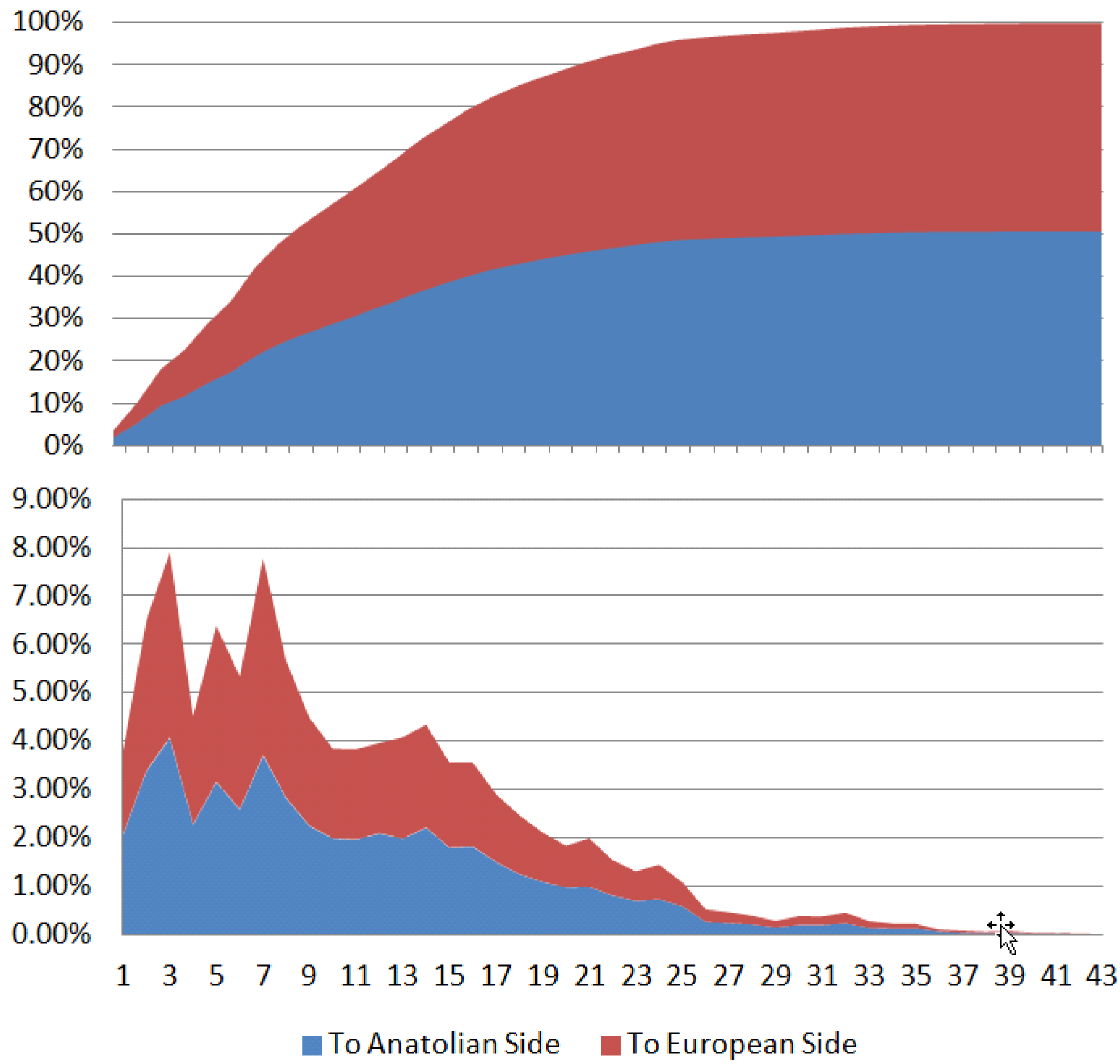

FIGURE 10 Cumulative and relative frequency graph of distances between journey stages

Looking at the distances between starting point and final destination of the day, we observe that more than 92% of all persons return to their starting location at the end of the day. Analysis of the alternative modes of transport used by BRT users reveals that a total of 64.7% of reported journeys are conducted completely on public buses, 7% are on ships, and 29.3% are on rail. We cannot detect other transit modes such as taxi, private car, cycle, foot, or some unknown mode in instances when a new journey starts more than 2 km away from a previous alighting location or more than 2 hours later. Almost all BRT journeys are made as the first or last journey of the day, which indicates the high usage of private buses for travel to and from work or academic locations (Figures 9 and 10).

Another important indicator of activity type is the regularity and frequency of trips to and from a particular activity area. In order to be able to conduct such analysis, a longer term Istanbulkart record of at least several consecutive days is obtained. This enables the analysis of multi-day travel behavior and identification of principal activity spaces. The regularity of activities performed there would give a strong indication of type. Additionally, analysis of activities can be based on demographic factors and differentiation of work activities for adults and educational activities for students (11). In this way, the accuracy and reliability of estimated home and work locations and their densities can be much improved.(Figure 11)

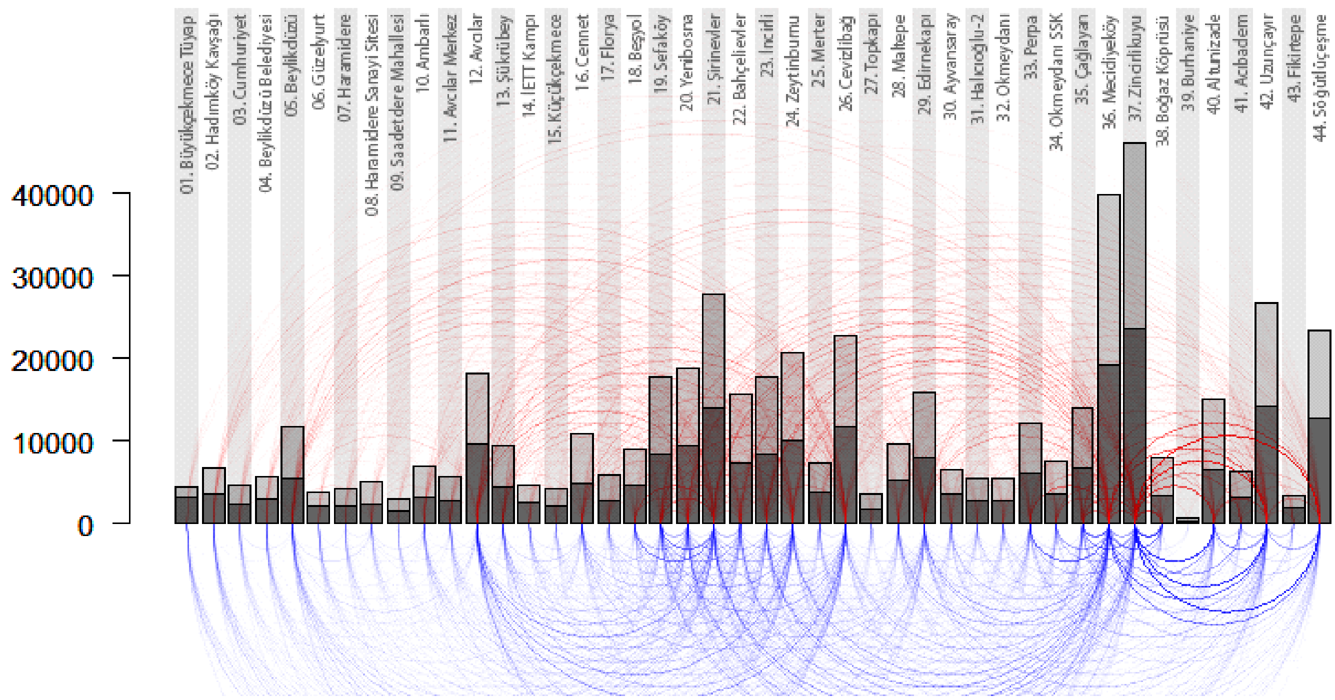

FIGURE 11 Arrivals/departures (gray-dark gray) /Asian (blue) and European (red) side (2013).

5.CONCLUSION

Improving the data-mining process while reducing the time needed for data processing has become an important problem for transportation decision makers who now have access to unprecedented amounts of operational data. In this paper, Istanbul´s automated fare collection system and pricing policies, the operational sources of big data, have been introduced, and approaches for the use of public transport smart card fare payment data for characterizing and analyzing travel behavior have been presented. Even though examination of this data source alone is not enough to measure the full potential of Istanbulkart data for planning purposes, the processed data used in this study reveal the potential for problems that might be caused by inappropriate handling of the big data that Istanbulkart makes available.In this study,we used the Rstudio simulation package which has an active user-base and has been applied to other large-scale scenarios.Using Rstudio,we detect demand management plans in such a way that utility is improved.Some of the immediate research questions that can lead to the improvement of BRT-Bus Rapid Transit line planning and operations that we were able to identify as a result of the current study can be summarized as follows:

1. Deployment of efficient data mining techniques to analyze large data sets for the purpose of discovering patterns of persistent problems;
2. Development of fast and easy to implement visualization tools for this type of big data;
3. Use of the Istanbulkart data to devise data-driven algorithms to generate products for operational and planning purposes including time- dependent OD tables and travel time estimates.

In this paper,the link between public transport usage and specific activities as a reason for travel was established, and methods for identifying primary activity locations were described. This is a first step towards an activity location model based on the public journeys record extracted from Istanbul BRT-line data. The disaggregated, multi-day data record, which is becoming accessible for research, will allow expansion and implementation of methods presented in this work.

Furthermore, obtained models of work locations based on Istanbulkart data can be compared with existing models used by transportation planners. Multi-day data open the door for further analysis of travel patterns and travel behaviors; for identification of demand profiles for bus routes, stations, and interchanges; and for development of a public transport route-choice model. Such future research possibilities add additional value to Istanbulkart data records and contribute to the development of methods for processing and analyzing smart card data records for the purpose of transportation planning.

In the second stage of this study, an integrated data fusion procedure that models the travel patterns and regularities of transit riders along the BRT-Bus Rapid Transit line will be developed. This procedure will incorporate transit riders' trip chains based on their temporal and spatial characteristics and effectively capture their historical travel patterns. We will examine big data in the rate of streaming data; and “veracity,” the relative certainty of data (12). Then, through examination of travel patterns and transfer data, rider-level destinations can be estimated from multi-day observations and “agent-based micro-simulations” (13). In these simulations, travelers and vehicles are modeled through agents that interact with the public transport system according to their individual goals.

In the future,we will use a numerical simulation models to test pricing strategies and get a better match between available capacity and the demand. We believe that an improved version of our SQL visualization methods and use of agent-based micro-simulations can improve the design of demand management systems, including origin-destination based schemes, in the realm of public transport.

6. ACKNOWLEDGEMENT

The authors thank Istanbul Metropolitan Municipality – Belbim Inc. for supporting this research and providing the data.

REFERENCES

1. IETT *www.iett.gov.tr*
2. IBB *www.ibb.gov.tr*
3. Blythe, P., 1998. *Integrated Ticketing-Smart Cards in Transport*. IEE Colloquium: *Using ITS in Public Transport and Emergency Services*, Paper No. 4, 20 p
4. Catherine Morency, Martin Trepanier, Bruno Agarda *Measuring transit use variability with smart-card data*, December 2010
5. Data Mining Solutions: *Methods and Tools for Solving Real-World Problems* (27 July 1998) by Christopher Westphal, Teresa Blaxton
6. Global BRT Data *http://www.brtdata.org/#/location-*
7. http://www.nbrti.org/docs/pdf/weststart_brt_ridership_analysis_final.pdf
8. *Recognizing Demand Patterns from Smart Card Data for Agent-Based Micro-simulation of Public Transport*, Paul Bouman, Milan Lovric, Ting Li, Evelien van der Hurk, Leo Kroon, Peter Vervest, ATT 2012
9. Chu, K.K.A. and Chapleau, R. (2010). *Augmenting Transit Trip Characterization and Travel Behavior Comprehension*. Journal of Transportation Research Board 2183, pp. 29-40.
10. Open source and enterprise-ready professional software-*http://www.rstudio.com/*
11. *Use Of Public Transport Smart Card Fare Payment Data For Travel Behaviour Analysis In Singapore*: A. Chakirov , A. Erath, 2012
12. Bagchi, M. and White, P.R. (2005) *The potential of public smart card data. Transport Policy* 12, pp.464-474.
13. Jang, W. (2010) *Travel Time and Transfer Analysis Using Transit Smart Card Data*. Journal of Transportation Research Board 2144, pp. 142-149.