Better Understanding Public Bus Passenger Behaviors in Jakarta

Preliminary Results from Mass Transportation Card Records, Waze Reports and Weather Data

ABSTRACT

Automatic fare collection (AFC) systems have been adopted by many public transportation authorities across the world, one of which includes Jakarta, Indonesia. The smart card AFC system administered by Transjakarta, the Jakarta Bus Rapid Transit (BRT), was deployed on all main transport corridors from February 2015. Despite the evolving structure of the BRT system, there has been limited research conducted on the influence of external factors on BRT passenger behavior.

In this study, we investigate the hidden travel patterns of regular TransJakarta passengers and how external factors, specifically, traffic and weather dynamics, affect to the behavior of regular commuters. In investigating these relationships we analyze 72 million transaction records from approximately 3.12 million smart transport cards, more than one million traffic reports from Waze, and the weather data from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS). The results show that external factors have a significant impact on the behavior of regular passengers. Moreover, we identify a case study which reveals a typical postponement of regular trips as a result of bad traffic. These results can be used as valuable inputs for the optimization of public transport services.

ACM Reference format:

. 2017. Better Understanding Public Bus Passenger Behaviors in Jakarta. In *Proceedings of CIKM, Singapore, Nov 2017 (CIKM'17),* 4 pages.

DOI: 10.1145

1 INTRODUCTION

Jakarta is the most densely populated province in Indonesia with a permanent population of more than 10.15 million, and a population density of 15,331 persons/km^2 according to the Indonesian Bureau Statistics. Based on a commuting survey conducted in 2014, 14.09% of the population in Jakarta are commuters and about 1.38 million people from Greater Jakarta commute to Jakarta.Despite the high population and commuting behaviour, Jakarta was named the worst city in the world for traffic congestion, according to new index published by Castrol[1].

Facing the high demand for public transportation and the challenge of traffic congestion in the city, the Government of Jakarta has made several attempts to improve transportation services, including implementing the Bus Rapid Transit (BRT) system, called the TransJakarta busway, since 15 January 2004 [6].

The implementation of the AFC system has brought opportunities to study the travel patterns and daily commuting behavior of Jakarta's commuters. Currently, insights on the daily travel behavior of commuters are still largely drawn from a commuting survey that was conducted by The Indonesian Central Bureau of Statistics (BPS) in 2014[2] which lacked granularity. The time that has passed since the production of this survey notwithstanding, challenges remain in understand the behaviors of public bus passengers, in particular the details of the spatial and temporal variety.

This study aims to develop a better understanding of the travel patterns of TransJakarta's regular passengers and the influence of external factors, such as traffic dynamics and weather conditions on the passengers' behavior. This information is relevant for the optimization of the public transportation system.

In particular we ask the following three questions by analyzing AFC data spanning four months, which records each passenger event, including the passenger id, boarding and alighting time, boarding and alighting station, and the trip corridor:

- Q1: Can we discover any hidden pattern of behaviour among regular passengers from AFC data?
- Q2: Do two obvious, external factors affecting the conditions of the traffic in Jakarta (*i.e.*, traffic dynamics and weather dynamics), lead the regular passengers to change their behavior?
- Q3: If so, what is the relative strength of the two factors in terms of their influence on the behavior of regular passengers?

In this paper we analyze commuting behaviour to determine the proportion of regular trips. We observe how the proportion of regular trips vary across different conditions. We use traffic report data from Waze and weather data from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS).

Our results show that more regular trips are presented during weekdays morning and afternoon peak times. Taking account the variation of these proportion of regular trips, we investigate that bad traffic and heavy rainfall have significant impact on the behavior of regular passengers. In addition, bad traffic leads to a greater variation in the ratio between regular to irregular trips than weather dynamics. This insight is statistically significant in five out of ten most popular sub-corridors.

The rest of the paper is organized as the following. Related work is briefly summarized in the Section 2 and the three types of data we use in this study is explained in Section 3. Section 4 and Section 5 explain our methodology and findings including passenger regularity and behavioral changes. We finalize this paper with discussion and a short summary in Section 6 and Section 7.

2 RELATED WORK

A set of studies have been conducted using the smart card AFC data in different aggregation level to infer urban mobility at a relatively fine granularity, facilitating transit planning and optimization.

Aggregated data firstly analyzed to investigate the frequency and consistency of daily travel patterns [18], critical transfer points, and travel time map [8]. A rigorous framework was also proposed

[1] http://time.com/3695068/worst-cities-tra_c-jams/?iid=srlink1

[2] https://jakarta.bps.go.id/backend/brs_ind/brsInd-20150220094832.pdf

CIKM'17, Singapore

2017. 978-x-xxxx-xxxx-x/YY/MM.

DOI: 10.1145

to enrich the data to construct a passive travel demand survey [4], which could further complement the traditional household travel surveys [16].

Some studies has been dealing with these data quality enhancement and data accuracy issues, by using spatial-temporal logic [3][4][14]. Origin and destination is also estimated by trip chaining method [2], using Transportation Object-Oriented Modeling (TOOM) approach [17], Markov chain based Bayesian decision tree algorithm [11], and developing network-level OD matrix estimation [5][15]. Some research has focused on analyzing travel patterns at individual level. Travel pattern was studied by analyzing trip chain regularity and identifying passengers segmentation, using DBSCAN algorithm [10][12][13] and the Weighted Stop Density Based Scanning Algorithm with Noise (WS-DBSCAN) [9].

Although several studies to enrich data and interpreting passenger behavior by spatial temporal analysis have been conducted, only handful of studies examined the influence of external factors to passenger behavior. Impacts of weather elements on transit ridership were analyzed by using smart card data with multiple linear regression [1] and OLS methods [7].

One key difference between the study in this paper with previous studies is the developing environment for BRT system in Jakarta, where the bus lanes were not fully sterilized from other vehicles and sometimes disrupted by other infrastructure development. These caused the travel pattern vary since the traffic condition is quite unpredictable. Therefor, understanding the influence of such external factors will be determined at the foremost.

3 DATA

In this section, we briefly explain three types of data we analyze throughout this paper.

3.1 Passenger Tap Information

We analyze 'Passenger Tap' data collected between October 6, 2016 and January 31, 2017, from AFC systems installed at 12 main corridors and 219 stations[3]. The data contains 72,515,002 transactions produced by 3,124,174 smart cards and each record contains several information such as (a) corridor ID, (b) station ID, (c) type of transaction, (d) transportation card id, and (e) transaction timestamp.

Given the data, we group these 219 stations to 39 sub-corridors, based on a commonly used zone classification approach by the public transportation authorities. Each sub-zone contains approximately 3 - 5 stations.

From a basic analysis, the temporal frequency distribution of the transactions shows that there are two peak times on weekday, which are on 6 AM - 8 AM and 5 PM - 7 PM. Spatially, the evening peak time is more dense in certain areas compare to morning peak time which is more distributed, as people are coming from surrounding area at the morning while coming back from centralized area at the evening. Generally, 10 most popular sub-corridors serve more than 50% passengers while the 10 least popular sub-zones serve less than 10% of passengers. During the weekend, there is no significant peak time and the total number of transactions is reduced by 50%.

3.2 Traffic Condition Data

In order to understand the traffic dynamic in Jakarta, we use report data from Waze collected during at the same period, which contains 1,075,459 of reports. Each report records several entities including (a) location, (b) timestamp, and (c) category and subcategory. The reports are classified under four categories, named `Jam`, `WeatherHazard`, `Accident`, and `Road closed`.

3.3 Weather Data

We extract historical weather data in Jakarta from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)[4]. It has an 0.05° x 0.05° degree spatial resolution or about 5.67 km x 5.67 km square root per pixel resolution, so Jakarta are covered by 24 grids of CHIRPS data. Every grid has the information about its geo-location and rainfall data in mm.

4 PASSENGER REGULARITY

We investigate the temporal regularity of passenger by adopting DBSCAN algorithm. The application of DBSCAN has been used in the previous works to mine the habitual time of passenger, using parameters of maximum density reach ε denoting the variability of boarding time windows and *MinPts* as the core point minimum of activity considered to be "regular" [10][13]. After several of trial, we then select 30 minutes interval and 15 minimum sample (ε = 30 min, $MinPts$ = 15) as input for DBSCAN clustering. A transaction that belongs to a cluster, will be classified as a regular trips while a transaction that identified as noise by the algorithm will be classified as a irregular trips.

Table 1: Proportion of regular trips aggregated in time range, broken down by types of day

	All	Weekday	Weekend	Holidays, ...
Early	0.23	0.25	0.17	0.24
AM Peak	0.59	0.68	0.38	0.59
Inter	0.23	0.26	0.14	0.19
PM Peak	0.36	0.46	0.17	0.31
Evening	0.30	0.33	0.23	0.32

The volume of regular trips is then measured against to all trips, to calculate the proportion. The summary of proportion of regular trips broken down by time range from the four months data, is represented in the Table 1. It is shown that weekdays has more regular trips in all time frame, mostly during AM and PM peak times.

In [13], this proportion of regular trips is used to explore how the regularity varies across transportation modes, location, and time range . In this study, we use this proportion to explore how the regularity varies by the influence of external factors.

5 BEHAVIORAL CHANGES

In this section we test the behavioral changes of the loyal passengers of TransJakarta, across the traffic dynamics and weather dynamics. We analyze the variation of proportion of regular trips in aggregate

[3]By the time the data for this study was taken, the 13th corridor of Transjakarta was not operated yet.
[4]http://chg.geog.ucsb.edu/data/chirps/

level. Based on a set of descriptive analyses of two datasets, traffic reports from Waze and precipitation records from CHIRPS, we use two threshold values to determine bad traffic and heavy rainfalls, 15,000 reports and 25mm, respectively.

Once excluding a list of days which are not proper for a regularity study, such as weekends, national holidays and big events (e.g., city-level strikes), we define four types of periods of time for further studies,

- **All** consists all the data in 4-month duration without weekends, holidays and big events.
- **Normal Week** has relatively less traffic and rainfalls and we select the Week of November 14 - November 18.
- **Traffic Week** is one of the weeks with extensive traffic reports but little rain, and, based on our Waze data, we select the Week of December 5 - December 10.
- **Rainy Week** is one of the weeks with heavy rains, and, based on the rainfall data, we select the Week of November 7 - November 12.

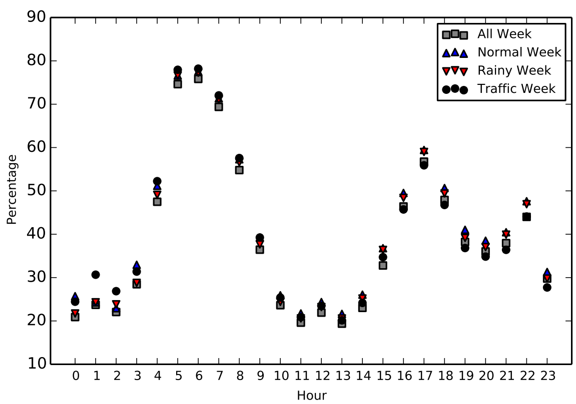


Figure 1: Average of proportion of Regular trips given multiple conditions

In Figure 1, we plot the average of proportion of regular trips on different condition and it shows that there is similar pattern of proportion of regular trip per hour across the periods with different condition. However, looking into detail, it is interesting to note the differences during peak time. The average proportion of regular trips remains high in the AM peak time regardless of external factors. In the PM peak time, proportion of regular trips in rainy week remains high, while the proportion of regular trips in traffic week is relatively drop. It is indicated that external factors influence more during PM peak time, and traffic dynamic are more likely has impact on passenger regularity.

To understand more detail on the variation of proportion of regular trips, we plot the box plot of each condition as shown in Figure 2. In addition, we also plot two different measurement on the same day, for instances, proportion of regular trips among the same normal Friday and one proportion of regular trips of normal day with traffic day.

Breaking down into each condition, there are different characteristics on variation of proportion of regular trips as illustrated in in Figure 2. Overall, big variation are presented slightly before the peak time and decreasing during the peak time. It is normal as during the peak time, more regular trips are made. In the normal week, the variation of proportion of regular trips in every hours are relatively small. In the rainy week, the variation is moderately higher, especially after 6 PM. Additionally, the variation is getting higher during the traffic week starting at 5 PM and remains high until late of night.

Table 2: Statistical significance (p-values) of variation of regular trips on traffic day and rainy day

Sub-corridor	p-values	
	Traffic day	Rainy day
K-01 North	0.395	0.061
K-01 Inner North	0.054	**0.013**
K-01 South	**0.009**	0.086
K-03 Outer	**0.020**	0.083
K-03 Inner	**0.037**	0.926
K-05 Outer North	0.708	**0.016**
K-06 Inner	**0.003**	0.056
K-07 Middle	0.112	0.354
K-09 Inner East	**0.023**	0.187
K-10 North	0.395	0.103

In order to test whether there is significant differences between the proportion of regular trips in normal weeks and traffic week or rainy week, we perform a hypothesis test by running t-tests (comparison of mean tests). We select sample from 10 most popular sub-corridors. As the result shown in Table 2, the p-value for traffic day is significant (p-value < 0.05) on 5 out of 10 sub-corridors, while for rainy days only significant in 2 sub-corridors. It confirms that traffic dynamics has more significant impacts in larger area on variation of regular trips.

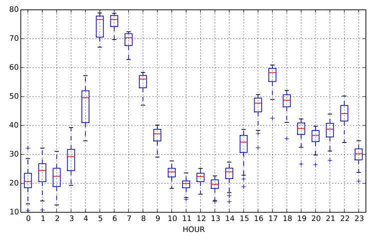
6 DISCUSSION

Our preliminary analysis results in a macroscopic level capture interesting insights including that the external factors affect passenger travel patterns but more detail analyses are needed in order to generate transportation policies.

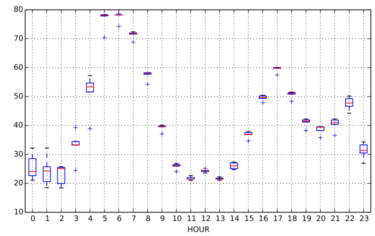
For instance, as an initial step towards further analyses, we investigate how traffic dynamics affect the regular trips, with a more localized, spatial and temporal dimension. When looking at the most crowded sub-corridor ('K-01 South') on a high-volume traffic day between 4PM and 9PM, we find that 56% of the trips during the selected time range are irregular trips, while two percent of the irregular trips are made by regular passengers who have a standard travel pattern in the time range. However, given the bad traffic conditions, they shift their trip beyond the habitual time, which we then classify as irregular trips and it leads the regular passengers to postpone their trips by 105 minutes in average. Likewise our future work will investigate in detail use cases, spatially and temporarily.

7 SUMMARY

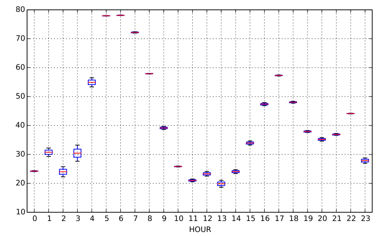
This preliminary study shows that there are regular patterns in the behaviour of Transjakarta passengers. During the weekday



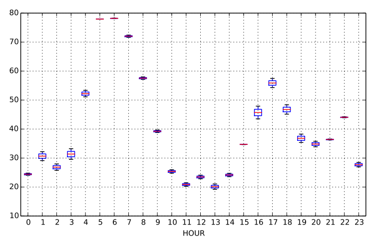
(a) All Days



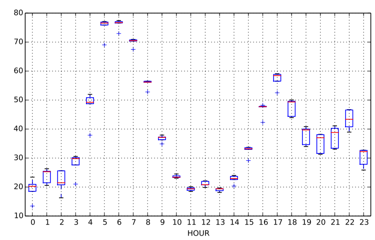
(b) Good Traffic & Weather (Mon-Thu)



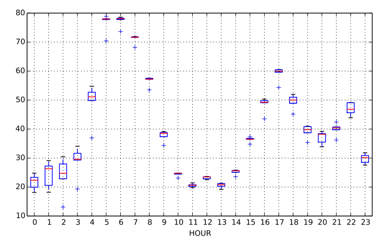
(c) Good Traffic & Weather (Fri)



(d) Good Weather but Bad Traffic



(e) Bad Traffic



(f) Heavy Rainfalls

Figure 2: Proportion of Regular Trips by Two External Factors

morning and afternoon peak times, a higher proportion of the trips were classified as regular trips. We also found out that the presence of external factors, traffic congestion and weather, lead to significant variations in the ratio between regular and irregular trips. These two external factors demonstrate a different impact on the proportion of regular and irregular trips, with traffic congestion being a stronger driver of behaviour change among regular passengers compared to heavy rainfall. In addition, through a case study, our research demonstrates the average amount of time that regular passengers on one of the transport corridors postpone their journey during periods of heavy traffic.

We plan to extend this work not only by further analyzing the public transport data but also by including other data sources such as mobile data and social media. With these new data sets and additional analyses, we hope to develop a deeper understanding of commuting behaviours in Greater Jakarta.

REFERENCES

[1] Pilar Arana, Sara Cabezudo, and Miriam Peñalba. 2014. Influence of weather conditions on transit ridership: A statistical study using data from Smartcards. *Transportation Research Part A: Policy and Practice* 59, C (2014), 1–12.

[2] James Barry, Robert Newhouser, Adam Rahbee, and Shermeen Sayeda. 2002. Origin and Destination Estimation in New York City with Automated Fare System Data. *Transportation Research Record: Journal of the Transportation Research Board* 1817 (January 2002), 183–187.

[3] Ka Kee Alfred Chu and Robert Chapleau. 2007. Imputation Techniques for Missing Fields and Implausible Values in Public Transit Smart Card Data. In *11th World Conference on Transport Research*. Berkeley, CA.

[4] Ka Kee Alfred Chu, Robert Chapleau, and Martin Trpanier. 2009. Driver-Assisted Bus Interview: Passive Transit Travel Survey with Smart Card Automatic Fare Collection System and Applications. *Transportation Research Record: Journal of the Transportation Research Board* 2105 (2009), 1–10.

[5] Alex Cui. 2006. *Bus passenger Origin-Destination Matrix estimation using Automated Data Collection systems*. Master's thesis. Massachusetts Institute of Technology, Cambridge.

[6] John Ernst. 2005. Initiating Bus Rapid Transit in Jakarta, Indonesia. *Transportation Research Record: Journal of the Transportation Research Board* 1903 (2005), 20–26.

[7] Zhan Guo, Nigel Wilson, and Adam Rahbee. 2007. Impact of Weather on Transit Ridership in Chicago, Illinois. *Transportation Research Record: Journal of the Transportation Research Board* 2034 (2007), 3–10.

[8] Wonjae Jang. 2010. Travel Time and Transfer Analysis Using Transit Smart Card Data. *Transportation Research Record: Journal of the Transportation Research Board* 2144 (2010), 142–149.

[9] Le Minh Kieu, Ashish Bhaskar, and Edward Chung. 2015. A modified Density-Based Scanning Algorithm with Noise for spatial travel pattern analysis from Smart Card AFC data. *Transportation Research Part C: Emerging Technologies* 58, Part B (2015), 193–207.

[10] Le Minh Kieu, Ashish Bhaskar, and Edward Chung. 2015. Passenger Segmentation Using Smart Card Data. *IEEE Transactions on Intelligent Transportation Systems* 16, 3 (June 2015), 1537–1548.

[11] Xiaolei Ma, Yinhai Wang, Feng Chen, and Jianfeng Liu. 2012. Transit smart card data mining for passenger origin information extraction. *Journal of Zhejiang University SCIENCE C* 13, 10 (2012), 750–760.

[12] Xiaolei Ma, Yao-Jan Wu, Yinhai Wang, Feng Chen, and Jianfeng Liu. 2013. Mining smart card data for transit ridersfi travel patterns. *Transportation Research Part C: Emerging Technologies* 36 (2013), 1–12.

[13] Ed Manley, Chen Zhong, and Michael Batty. 2016. Spatiotemporal variation in travel regularity through transit user profiling. *Transportation* (2016), 1–30.

[14] Catherine Morency, Martin Trpanier, and Bruno Agard. 2007. Measuring transit use variability with smart-card data. *Transport Policy* 14, 3 (2007), 193–203.

[15] Marcela Munizaga, Carolina Palma, and Pamela Mora. 2010. Public transport OD matrix estimation from smart card payment system data. In *12th World Conference on Transport Research*. Lisbon, Portugal.

[16] Martin Trepanier, Catherine Morency, and Carl Blanchette. 2009. Enhancing Household Travel Surveys Using Smart Card Data. In *Transportation Research Board 88th Annual Meeting*. Washington, DC, 15.

[17] Martin Trpanier, Nicolas Tranchant, and Robert Chapleau. 2007. Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System. *Journal of Intelligent Transportation Systems* 11, 1 (2007), 1–14.

[18] Mariko Utsunomiya, John Attanucci, and Nigel Wilson. 2006. Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning. *Transportation Research Record: Journal of the Transportation Research Board* 1971 (2006), 119–126.