# Real-Time Riders: A First Look at User Interaction Data from the Backend of a Transit and Shared Mobility Smartphone App

Word Count: 5,219 (text) + 250 * 6 (figures) = 6,719

Submission Date: February 1, 2017

Candace Brakewood, PhD (Corresponding Author)
Assistant Professor of Civil Engineering
The City College of New York
160 Convent Avenue, New York, NY 10031, USA
Email: cbrakewood@ccny.cuny.edu
Phone: (212)-650-5217

Niloofar Ghahramani
Graduate Student
The City College of New York
160 Convent Avenue, New York, NY, 10031, USA
Email: nghahra000@citymail.cuny.edu

Jonathan Peters, PhD
Professor of Finance
The College of Staten Island
2800 Victory Blvd., Staten Island, NY 10314, USA
Email: jonathan.peters@csi.cuny.edu

Eunjin Kwak
Graduate Student
The College of Staten Island
2800 Victory Blvd., Staten Island, NY 10314, USA
Email: Eunjin.Kwak@cix.csi.cuny.edu

Jake Sion
Chief Operating Officer
Transit
5333 Avenue Casgrain, Suite 803, Montreal, Quebec H2T 1C2, Canada
Email: jake@transitapp.com

## ABSTRACT

A fundamental component of transit planning is understanding passenger travel patterns. However, traditional data sources used to study transit travel have some noteworthy drawbacks. For example, manual collection of travel surveys can be expensive, and datasets from automated fare collection (AFC) systems often include only one transit system and do not capture multimodal trips (e.g., access/egress mode). New data sources from smartphone applications offer the opportunity to study transit travel patterns across multiple metropolitan regions and transit operators at little to no cost. Moreover, some smartphone applications integrate other shared mobility services, such as bike-sharing, car-sharing, and ride-hailing, which can provide a multimodal perspective not easily captured in traditional datasets. Therefore, the objective of this research is to take a first look at an emerging data source, which is backend data from user interactions with a smartphone application. The specific dataset used in this paper is from a widely used smartphone application called “Transit” that provides real-time information about public transit and shared mobility services. Visualizations of individuals’ interactions with the Transit app are created to demonstrate three unique aspects of this dataset, which are the ability to capture [1] multi-city transit travel, [2] multi-agency transit travel, and [3] multimodal travel, such as taking bike-share to access transit. This dataset is then qualitatively compared to traditional transit data sources, including travel surveys and AFC data, and the findings suggest that this dataset has some potential advantages over traditional data sources that could help transit planners better understand how passengers travel.

## INTRODUCTION AND MOTIVATION

A fundamental component of transit planning is understanding passenger travel patterns. However, traditional data sources used to study transit travel patterns have some noteworthy drawbacks. For example, manual collection of travel surveys can be time consuming and expensive; similarly, data from automated fare collection (AFC) systems, such as smart cards, often include only one transit agency and usually do not capture multimodal trips (e.g., access and egress mode). Rapid adoption of mobile phones has led to the creation of new data sources that can be used to study travel patterns, particularly smartphones that are GPS-enabled or location aware *(1, 2, 3).* Smartphone applications (or "apps") that focus on specific modes, such as those providing real-time transit information, can capture data that may be harder to identify in larger mobile phone datasets (such as from cellular network providers) due to difficulty differentiating between modes, cell phone service issues in underground tunnels, and other challenges. The widespread usage of apps to access information about public transit and new shared mobility services (such as car-sharing, bike-sharing and ride-hailing) presents a unique opportunity to utilize the data generated from user interactions with these apps to study transit and shared mobility travel behavior. In light of this, the objective of this paper is to take a first look at an emerging data source, which is backend data from user interactions with a transit-related smartphone application. The specific dataset used in this paper is from a widely used app called "Transit" that provides real-time information about public transit and shared mobility services *(4).*

This paper proceeds as follows. First, a brief of review of prior research is provided. Next, the smartphone application that is the focus of this paper is discussed in detail, including a description of the user interface, the data captured in various backend tables from user interactions, and the specific data samples used in this paper. This is followed by a visualization exercise showing individuals' interactions with the Transit app to demonstrate three unique aspects of the dataset. Next, this new dataset is qualitatively compared to traditional transit travel behavior data sources, including travel surveys and automated fare collection systems. Finally, conclusions and areas for future research are presented.

## PRIOR RESEARCH

One of the most commonly used data sources to study transit travel behavior is surveys, which are often conducted as household travel surveys or in stations/onboard transit vehicles *(5, 6).* Survey data has been extensively written about in the prior literature *(e.g., 7)* and will not be reviewed in detail here.

Automated fare collection (AFC) systems, particularly smart cards, have become commonplace in transit systems over the last two decades. Although smart card systems are installed for the purpose of revenue collection, they also provide a rich source of data about transit travel *(8, 9).* Passengers with contactless smart cards pay fares by "tapping" their cards at faregates or on fareboxes. With each tap, a record is created that includes the date and time, fare type, route or station ID, a unique card ID number, and possibly other things *(9).* Pelletier et al. provide a literature review of the many uses of transit smart card data, including "strategic-level" analyses that relate to travel behavior and demand forecasting *(9).*

More recently, data from GPS-enabled mobile phones has been utilized to study transit travel. Within this growing body of literature, one noteworthy segment pertains to smartphone-based household travel surveys that utilize prompted recall surveys on location aware smartphones to collect travel behavior data *(e.g., 10).* Other studies have used crowdsourced

GPS-traces from mobile phones to generate transit vehicle predictions *(11)* and map informal transit systems *(12).* A small number of studies have aimed to capture transit travel behavior using mobile phone location data for the purpose of understanding passenger movements; one noteworthy example was recently published by Carrel et al., who developed a methodology to merge transit automatic vehicle location (AVL) data with smartphone location data *(13).*

In summary, there is a growing body of literature that pertains to the use of mobile phone data to study transit travel. However, to the best of the authors' knowledge, previous studies have not utilized user interaction data from the backend of a smartphone application to study transit travel behavior. Therefore, this study aims to explore the potential uses of this emerging data source.

## TRANSIT APP

This section provides background information about Transit App, which is a company based in Montreal, Canada that has developed a freely available smartphone application known as "Transit" *(4).* In 2012, the company released the first version of its application for iPhone, and in the initial version, the app provided transit schedule information for Montreal, Toronto, and Quebec City. Since then, an Android version of the application was launched, and the app has expanded to over 125 cities in nine countries, including widespread coverage in the United States. Many additional features have also been incorporated into the app, including real-time transit information, transit trip planning, and multimodal support (including bike-sharing, car-sharing, and Uber). Transit app users can also store their favorite locations in the app, such as home or work, to facilitate quickly finding information that they commonly use. Figure 1 shows the Transit app Android interface displaying real-time transit information for nearby routes (left), trip planning (center left), storing a home location (center right), and bike-sharing (right), respectively.

Whenever a user opens the Transit app, data about their interactions with the app are created and stored in a backend database. Each user interaction is called a "session" and is identified by a unique identification (ID) number and timestamp. In order to provide relevant transportation information for nearby transit service, the application needs to identify the location of the device. Therefore, each time the application is opened, a session is generated, the device location is sent to the Transit app server, and this record is then stored in a backend database. More details about additional data fields that are stored are provided in the following section, but before continuing, it is important to note that names or demographic information are not requested nor stored, which protects the anonymity of users.

### *Tables*

The backend database generated by user interactions with the Transit app is divided into tables that capture data pertaining to the various functions within the app. Table 1 describes thirteen of these tables and the data that are captured in each. Notably, the structure of the backend can change when there are application updates to include new features, and the tables discussed in the following paragraphs represent a sample from the backend provided to the authors in 2016.

As can be seen in Table 1, the first file is the ***locations*** table, which is the primary file in the Transit app backend server that contains the users' interactions with the app. Every time a user opens the app, a session is created, and each session can be identified by a unique ID. The locations file records numerous items for each session, including the following: coordinates (latitude/longitude) of the user, timestamp, accuracy, speed (if the user is in a vehicle) and simulation (if the user has moved the GPS point in the Transit app map and searched for information in a location other than where they actually were).

The second and third files described in Table 1 are called ***session complete*** and ***placemarks***, respectively. The session complete table provides an event based view of a user. This includes starting and ending location in terms of user coordinates (latitude/longitude) for each session, starting and ending timestamps, number of records transmitted during the session, and whether the session is simulated or not. The placemarks table includes data about an optional feature to store frequently used places, such as home and work, which can be saved within the app to help users find relevant information quickly.

The next three files shown in Table 1 pertain to the shared mobility services that are integrated within the Transit app. The fourth table is called ***sharing system actions***, and it provides information on the booking of car-share, bike-share and other services. The file contains information on the location of the station for bike-share or the location of the vehicle for car-share systems, as well as the location of users when they are searching for sharing-system information. Further, it provides information on the booking and cancellation process for these systems, which can be done through the Transit app directly. The fifth table is ***sharing system purchase,*** which provides information on the successful purchase of shared vehicle passes, primarily bike-share. Variables include type of pass, time of request, number of passes, and the cost of the transaction. Next is the ***Uber request*** table, which lists requests for service from Uber. Once a user requests the service, the request is handed off to Uber through their smartphone application for fulfillment. This file contains the time of request, location of the user at time of request, type of service requested, and in some cases, the drop off location. For additional information about Transit app Uber request data, see *(14).*

The remaining seven tables are not used in the following analysis, but they are briefly described for completeness. As can be seen in Table 1, the seventh table is called ***trips***. This file contains information about usage of the trip planning feature in the Transit app, which provides A-to-B transit directions. This table includes start and end coordinates (latitude/longitude), date, and timestamps of trip planning requests. The eighth file is the ***nearby view*** table, which contains information about the routes presented to a user in each session upon opening the app. This includes the transit route and the corresponding transit agency, number of taps for each route, and if that route is designated as the user's favorite transit route. The ninth table is called ***user feed session***, and it provides information on the number of times the Transit app is opened by a user each day and the transit agencies for which the user requested information. The next file is entitled ***installed app***, and it reports on other apps installed on the user's device that can impact Transit app functionality. For example, a user that has Uber installed on their phone will be linked to the Uber app when requesting a ride, whereas a user that does not have the Uber app installed will be sent to the app store. The eleventh file is called ***feed download***, and this table provides an overall summary of Transit app activity by day. The next table is called ***favorite***, and it provides information on user designated favorites in terms of transit routes. Last is the ***device table***, which creates a Transit app specific identification (ID) number that is utilized in other tables to identify a unique device in the data archive. This table also includes the user's selected language, type of device, model of device, operating system used, version of the Transit app software installed, and last date of Transit app use.

*Samples*
Two data samples of Transit app data were provided to the research team for this exploratory analysis. The first sample included records from one month in 2014 for any user that opened the Transit app at least once in the New York City region. This sample included approximately 10.8 million records and contained five tables, two of which were used for the following analyses (the ***locations*** table and the ***placemarks*** table). The second sample of data was much larger (approximately 12 terabytes) and included 418 days of data from 2015 and 2016 for all geographic regions available in the Transit app. This larger dataset included all 13 tables shown in Table 1; however, only the ***session complete*** and ***sharing systems*** table were used in the following visualization exercises.

## VISUALIZATIONS
Visualizations of individuals' interactions with the app were created to demonstrate the potential uses of this unique dataset. The three examples shown in this exercise are multi-city transit travel, multi-agency transit travel, and multimodal travel (e.g., from bike-sharing to transit). It should be noted that, for the purpose of this paper, all Transit app user location data (latitude/longitude) were offset by a random number. Anonymizing the location data in this way ensures that user privacy is protected, and throughout this document, whenever user locations are mentioned, they refer to the anonymized version of the data point.

### Analysis 1: Multi-city Transit Travel
Because the Transit app covers more than 125 metropolitan regions, this dataset can be used to identify intercity travelers and understand how they use transit systems in metropolitan areas other than their home city *(15)*. A simple example of this is shown in Figure 2. Data from the locations table of the first data sample (2014) is displayed to show how an individual traveler used the Transit app in Los Angeles, Houston and New York City over the course of five days. Each colored circle in Figure 2 represents that individual's interactions with the Transit app on a specific day. On the first day shown in Figure 2 (in light blue), the app user is observed in the vicinity of the Los Angeles Ontario International Airport. On the following two days, this individual uses the app in various areas of the Los Angeles metropolitan region. On the fourth day shown in Figure 2 (in pink), s/he has Transit app records in the vicinity of Houston Hobby Airport and New York LaGuardia Airport. This implies that s/he has traveled from Los Angeles to New York and has transferred planes in Houston, even though there is no observation of this user in a Los Angeles airport on that day. On the fifth day, this user has Transit app records in the Bronx, which is the location of his/her self-reported home location from the placemarks table, as well as records in other areas of Manhattan and New Jersey (not shown in Figure 2).

### Analysis 2: Multi-agency Transit Travel
Similar to the previous example, the Transit app dataset enables identification of users who search for transit information from multiple transit agencies operating within the same metropolitan region. Figure 3 shows an example of an individual's interactions with the Transit app using the locations table from the first data sample (2014). The figure shows two weekdays that are typical of this user, likely representing a commuting pattern in the New York region. During the two weekdays shown in Figure 3, the app user is observed in the Bronx in the early morning (shown in yellow) and late evening (shown in red), which is likely this users' home location (labeled as "Inferred Home Location"). S/he is also observed in New Jersey during the

daytime, implying that this location is likely his or her work location (labeled as "Inferred Work Location"). In addition to inferred home and work locations, this app user's commute route presumably includes multiple transit operators. S/he appears to take the New York City Transit (NYCT) 6 train route from the Bronx to Manhattan, and then transfer to the Port Authority of New York and New Jersey's PATH train from Manhattan to New Jersey. This user is observed on both weekdays in the vicinity of 23rd Street Station in midtown Manhattan, which is likely the transfer location between these two transit operators (labeled as "Inferred Transfer Location").

**Analysis 3: Multimodal Travel**

The third visualization demonstrates how the Transit app dataset can be used to identify multimodal travel. Bike-sharing is one of the shared mobility services available in the Transit app, and in some cities, such as Chicago, this app can be used to purchase bike-share passes and unlock bicycles. Figure 4 uses bike-share data from the sharing system actions table combined with the session complete table from the second (larger) sample of Transit app data to identify a multimodal traveler in Chicago on a single day in 2016. The individual unlocks a Divvy bike-share bicycle in the morning (labeled as "Divvy Request in the Morning"), which is likely near his/her home location (labeled as "Inferred Home Location"). Shortly thereafter, this individual is observed in the vicinity of the Chicago Transit Authority's Blue Line at Irving Park Station, where s/he has probably transferred from bike-share to rail (shown in yellow). Next, this user is observed during the daytime in the vicinity of Cumberland Station, which is also on the Blue Line and is likely near his/her work location (labeled as "Inferred Work Location"). In the early evening, this user is observed again unlocking a Divvy bike-share bicycle in a location close to Irving Park Station (labeled as "Divvy Request in the Afternoon"). This is nearly the same location s/he was observed transferring from bike-share to rail in the morning, implying that s/he is transferring from rail to bike-share and probably continuing his/her trip home via bicycle. Last, in the late evening, this individual is again observed near his/her inferred home location in Irving Park (shown in dark red).

## COMPARISON OF TRANSIT APP DATA AND OTHER SOURCES

In light of the unique aspects of the Transit app dataset demonstrated in the previous paragraphs, this section provides a brief comparison of commonly used transit travel behavior data sources – specifically, surveys and automated fare collection (AFC) systems - with Transit app data. These three datasets are compared on five dimensions: [1] geographic coverage, [2] institutional coverage, [3] modes, [4] timescales, and [5] sample size. The results are shown in Table 2 and discussed in the following paragraphs.

The first dimension, ***geographic coverage***, refers to the geographic area included in each dataset. Surveys are typically conducted within a single metropolitan by the local transit agency or metropolitan planning organization, with a notable exception being the American Community Survey collected by the United States Census Bureau *(16)*. Similarly, AFC systems are typically not compatible across different metropolitan areas *(17)*, and subsequently, the data generated by these systems typically encompasses transit travel records for a single region. Because the Transit app includes over 125 metropolitan regions in nine countries, it collects user interaction data for many different metropolitan regions, and this data can then be used to understand multi-city travel patterns, as was demonstrated in Figure 2 in the previous section.

The second dimension, ***institutional coverage***, refers to the transit agencies within a single metropolitan region for which the dataset is collected. Surveys conducted by the local transit agency typically include questions pertaining to only one transit provider, whereas surveys conducted by the metropolitan planning organization are more likely to consider numerous operators in the region. AFC systems usually include only one transit agency *(15);* however, there are a small number of regions that allow for interoperability between multiple local operators. The Transit app, on the other hand, integrates many transit agencies that have open data (e.g., GTFS schedule information or real-time vehicle data), which has resulted in user interaction data for numerous transit agencies in the same region. This was shown in the visualization in Figure 3, where an individual likely transferred between New York City Transit and PATH.

Third, the term ***modes*** refers to the modes of transportation for which data are collected. Surveys conducted by the local transit agency typically focus on transit travel and occasionally include a limited number of questions about access and egress mode to transit stations/stops; surveys conducted by metropolitan planning organizations are more likely to capture multiple modes, including trips made by automobile, transit, and non-motorized modes. AFC systems typically include only data pertaining to transit travel, since data is captured when transit fares are paid. The Transit app, on the other hand, integrates information about numerous shared mobility modes, including bike-sharing, car-sharing, and ride-hailing (Uber) and allows users to purchase passes and utilize shared mobility vehicles in some cities, such as Divvy bicycles in Chicago. The multimodal nature of the Transit App dataset was shown in Figure 4 in the previous section.

The fourth dimension, ***timescales***, refers to the period when data are generated. Most travel surveys are conducted at a single point in time and therefore provide a cross-sectional snapshot of transit travel behavior. Panel surveys conducted at multiple points in time can also be conducted, but in practice, this is infrequently done primarily due to cost constraints. AFC data are collected whenever the transit system is in operation, which means this data source is (nearly) continuous in time. The Transit app functions twenty four hours a day, seven days a week and therefore continuously collects user interaction data. The continuous nature of this dataset could advantageously allow for future analyses examining travel behavior during events that are difficult to capture in cross-sectional datasets (e.g., extreme weather events).

Finally, the ***sample size*** refers to the quantity of data collected. Travel surveys typically sample only a small portion of the population of interest. Despite the relatively small sample size, it is worth highlighting that surveys are usually conducted using methods that aim to be representative of the entire population of interest. AFC systems generate vast quantities of data, and depending on the level of AFC adoption by riders in a region, they can represent all or nearly all transit riders. The Transit app also generates vast quantities of data because many riders use the app on a daily basis *(18).* However, it is important to note that the sample for which the Transit app data is generated depends on the app's adoption and utilization levels, which could be biased compared to the overall population (such as toward younger, more technology-friendly riders). Therefore, future research is needed to understand the potential biases of this new dataset.

## CONCLUSIONS AND FUTURE RESEARCH

This research took a first step at examining an emerging data source, which is backend data from user interactions with a smartphone application. The specific dataset used in this paper is from a widely used smartphone application called "Transit" that provides real-time information about public transit and shared mobility services. A visualization exercise was conducted to demonstrate three unique aspects of the Transit app dataset, which include the ability to study [1] multi-city transit travel, [2] multi-agency transit travel, and [3] multimodal travel, such as taking bike-share to access transit. These three aspects of the Transit app dataset have the potential to provide unique travel behavior information that is not easily captured in traditional transit datasets, such as travel surveys and AFC data.

Many areas for future research emerged from this study, and three specific areas are briefly described in the following paragraphs. First, this exploratory analysis highlighted some unique aspects of the Transit app dataset by creating simple visualizations. An important next step in this research is to develop algorithms to identify this type of travel behavior (multi-city, multi-agency, and multimodal transit travel) in the larger dataset or for desired subsets of the dataset (e.g., for a single metropolitan). While the visualizations are useful as a proof-of-concept, it is the development of algorithms that are critical to using the larger dataset for planning purposes. Data mining techniques could be used to implement this; however, challenges associated with manipulating this big dataset may arise.

Second, once algorithms have been developed for the larger dataset, potential biases of this data source should be explored. This data could be biased in comparison to the overall transit riding population, such as toward younger, more technology-friendly riders. Comparing the Transit app data to traditional data sources (namely, travel surveys and AFC data) could identify systematic biases, and then, methods to correct for these biases can be developed.

Third, an important area for future consideration is the arrangements by which transit agencies and planners can access data from this and other smartphone apps. The company that provides the specific app used in this study has already begun sharing their data with a limited number of transit agencies. For example, they recently announced a partnership with the Massachusetts Bay Transportation Authority in which the agency promotes Transit as its preferred app, and as part of the agreement, the app developers provide data to the transit agency to be used for planning purposes *(19).* This example suggests that there are innovative and interesting data-sharing policies that can facilitate utilization of this type of data in the future.

In summary, after algorithms for expanding these analyses have been developed and corrections for systematic biases, the Transit app dataset is likely to be extremely valuable to transit planners, operators, and managers and has the potential to transform our understanding of public transit and shared mobility travel behavior.

**ACKNOWLEDGEMENTS**
The authors are very grateful to the Transit app for sharing data with the research team. They would also like to acknowledge CUNY Graduate Center student Adam Davidson for his contributions to an early draft of this paper. Additionally, thank you to Professor Liv Haselbach and Dr. Josephine Kressner for reviewing an early draft of this manuscript. This research was funded in part by a 2015 City University of New York (CUNY) Collaborative Incentive Research Grant (CIRG) grant and a University Transportation Research Center (UTRC) faculty-initiated research grant. Finally, thanks to the Transportation Research Board (TRB) for the opportunity to participate in the Chan Wui & Yunyin Rising Star Workshop for Early Career Professionals.

**REFERENCES**


1. Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q. and Bayen, A.M., Evaluation of Traffic Data Obtained via GPS-enabled Mobile Phones: The Mobile Century Field Experiment. *Transportation Research Part C: Emerging Technologies*, 18(4), 2010, pp.568-583.
2. Jiang, S., Fiore, G.A., Yang, Y., Ferreira Jr, J., Frazzoli, E. and González, M.C., A Review of Urban Computing for Mobile Phone Traces: Current Methods, Challenges and Opportunities. In *Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing*, ACM, 2013.
3. Wolf, J., Bachman, W., Oliveira, M., Auld, J., Mohammadian, A.K., Vovsha, P. and Zmud, J. NCHRP Report 775: Applying GPS Data to Understand Travel Behavior, Volume II: Guidelines, National Cooperative Highway Research Program of the Transportation Research Board, Washington DC, 2014.
4. Transit App, http://transitapp.com/, accessed October 11, 2016.
5. Schaller, B. TCRP Synthesis 63: On-Board and Intercept Transit Survey Techniques: A Synthesis of Transit Practice. Transit Cooperative Research Program of the Transportation Research Board, Washington DC, 2005.
6. Pratt, J., Lee-Gosselin M., and Burbridge, S. Chapter 3: Options for Travel Surveys. The Online Travel Survey Manual: A Dynamic Document for Transportation Professionals. http://www.travelsurveymanual.org/Chapter-3-1.html, Accessed October 13, 2016.
7. Transportation Research Board (TRB) Travel Survey Methods Committee (ABJ40). The Online Travel Survey Manual: A Dynamic Document for Transportation Professionals. http://www.travelsurveymanual.org/, Accessed October 13, 2016.
8. Bagchi, M., and White, P. R. The Potential of Public Transport Smart Card Data. *Transport Policy*, 12(5), 2005, pp. 464–474.
9. Pelletier, M.P., Trépanier, M., & Morency, C. Smart Card Data Use in Public Transit: A Literature Review. *Transportation Research Part C: Emerging Technologies*, 19(4), 2011, pp. 557–568.
10. Cottrill, C., et al. Future Mobility Survey: Experience in Developing a Smartphone-Based Travel Survey in Singapore. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2354, Washington DC, 2013, pp. 59–67.
11. Zimmerman, J., Tomasic, A., Garrod, C., Yoo, D., Hiruncharoenvate, C., Aziz, R. & Steinfeld, A. Field Trial of Tiramisu: Crowd-sourcing Bus Arrival Times to Spur Co-design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 1677-1686.
12. Williams, S., White, A., Waiganjo, P., Orwa, D., & Klopp, J. The Digital Matatu Project: Using Cell Phones to Create an Open Source Data for Nairobi's Semi-formal Bus System. *Journal of Transport Geography*, 49, 2015, pp. 39-51.
13. Carrel, A., Lau, P. S., Mishalani, R. G., Sengupta, R., & Walker, J. L. Quantifying Transit Travel Experiences from the Users' Perspective with High-resolution Smartphone and Vehicle Location Data: Methodologies, Validation, and Example Analyses. *Transportation Research Part C: Emerging Technologies*, 58, 2015, pp. 224-239.
14. Davidson, A., Peters J. and Brakewood, C. Interactive Travel Modes: Uber, Transit and Mobility in New York City. Proceedings of the 96th Annual Meeting of the Transportation Research Board, Washington DC, 2017.

15. Ghahramani, N., Brakewood C., and Peters J. (2017). An Exploratory Analysis of Intercity Travel Patterns Using Backend Data from a Transit Smartphone Application. Proceedings of the 96th Annual Meeting of the Transportation Research Board, Washington DC, 2017.
16. US Census Bureau. American Community Survey. http://www.census.gov/programs-surveys/acs/, accessed October 11, 2016.
17. Acumen Building Enterprise, Inc. TCRP Report 115: Smartcard Interoperability Issues for the Transit Industry. Transit Cooperative Research Program of the Transportation Research Board, Washington DC, 2006.
18. Ghahramani, N. and Brakewood, C. Trends in Mobile Transit Information Utilization: An Exploratory Analysis of Transit App in New York City. *Journal of Public Transportation,* Vol. 19, No. 3, 2016.
19. MBTA Selects 'Best Transit App' Winner. Metro Magazine. September 8, 2016. http://www.metro-magazine.com/management-operations/news/715324/mbta-selects-best-transit-app-winner, accessed January 31, 2017.

**List of Figures and Tables**

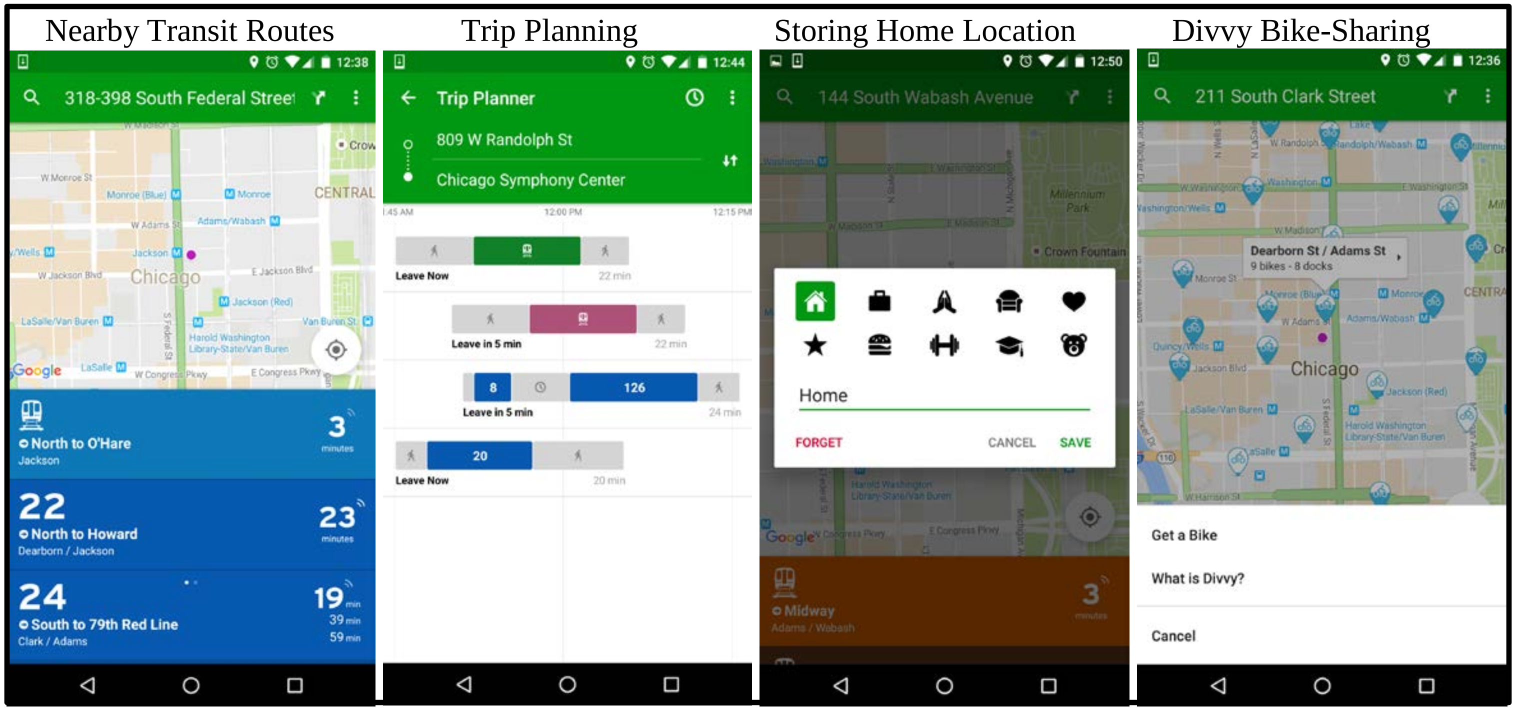


**FIGURE 1 Transit App Screenshots (Android Version 3.11.2)**


**FIGURE 2 Example of an Individual's Multi-city Transit Travel**

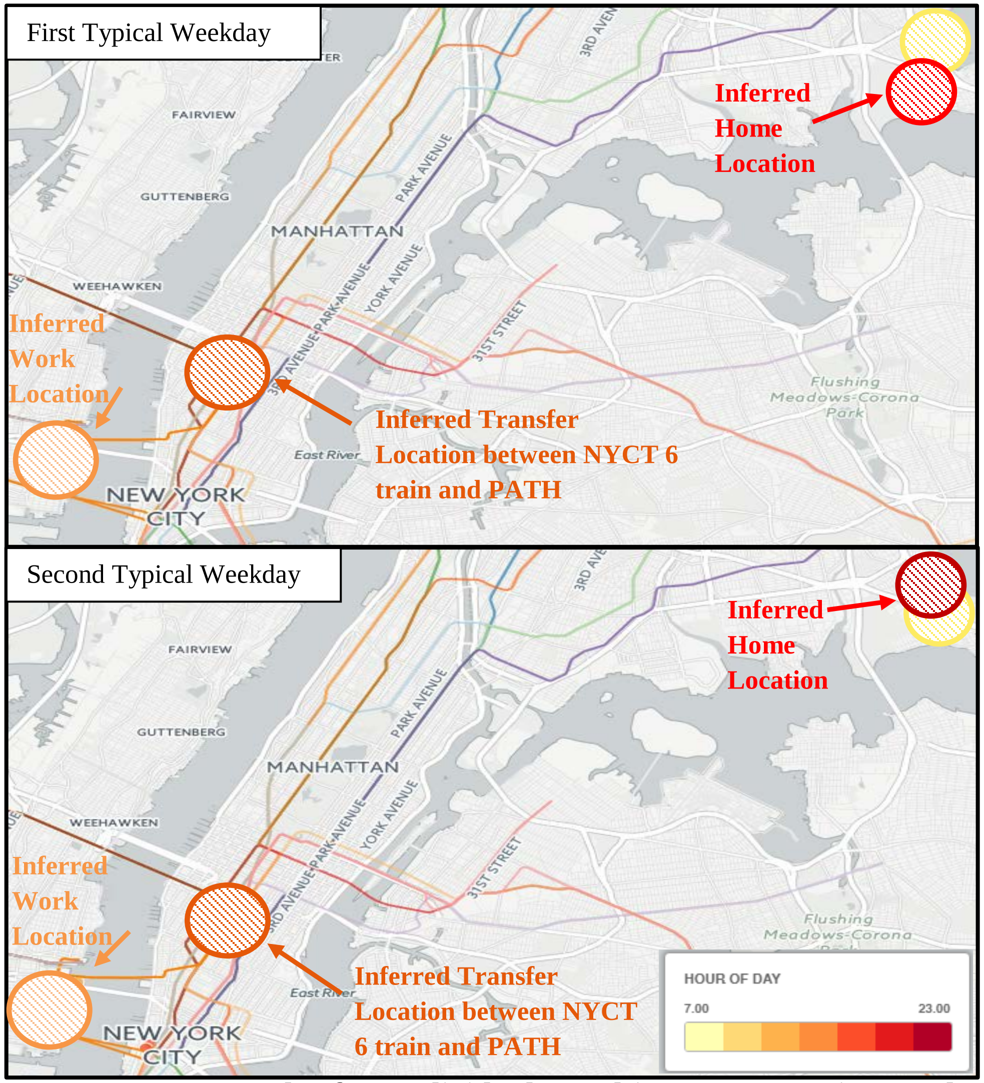


**FIGURE 3 Example of an Individual's Multi-agency Transit Travel**

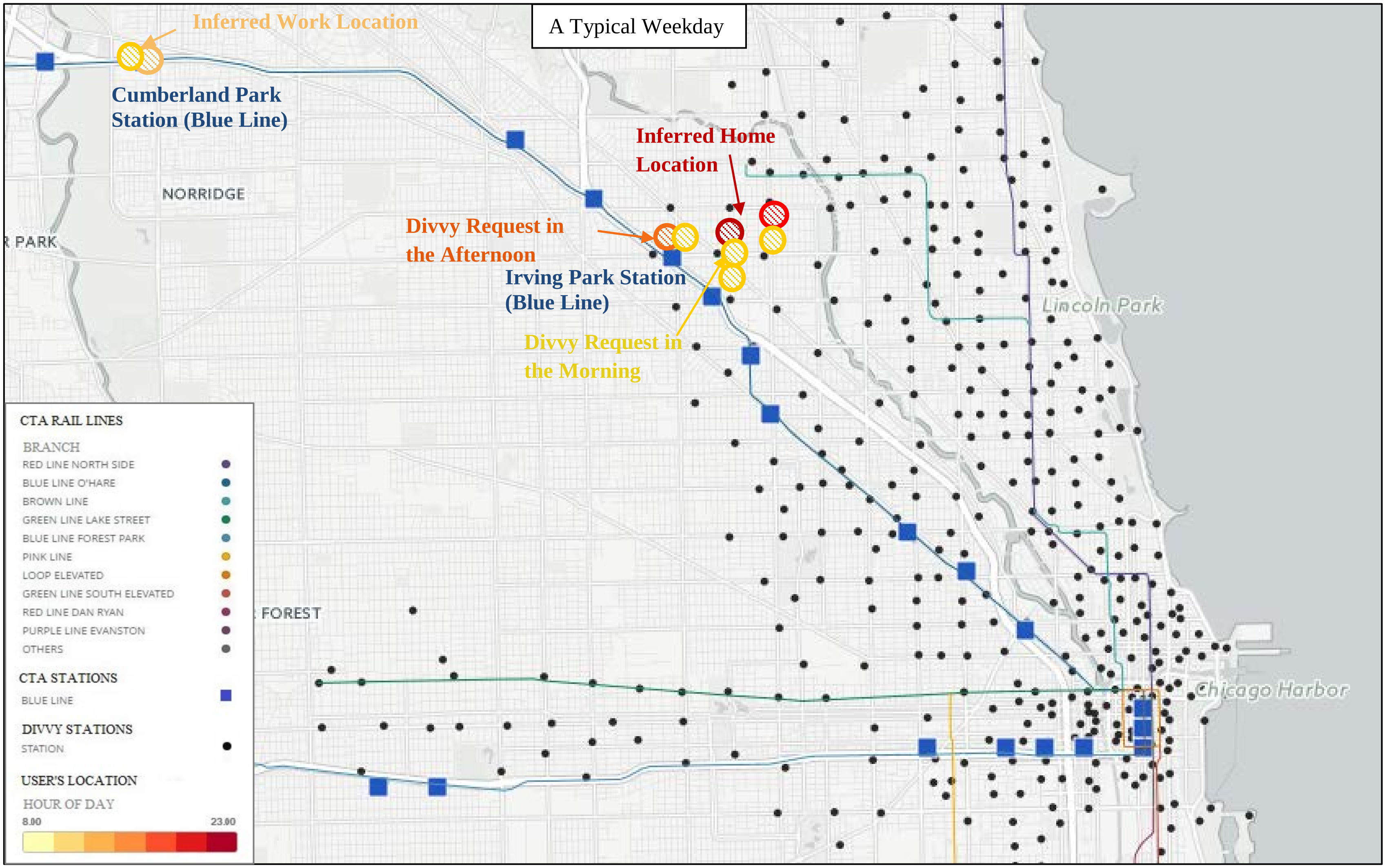


**FIGURE 4 Example of an Individual's Multimodal Travel**

**TABLE 1 Description of Tables from the Transit App Backend**

| No. | Table Name | Description of Contents |
|---|---|---|
| 1 | Locations | Includes location (latitude/longitude) for each time a user opens the app. Date, time, accuracy of their location, and speed (e.g., if they are in a vehicle) are recorded, and a unique session ID is created each time a user opens the app. |
| 2 | Session Complete | Provides an event based view of a user. The Session Complete file provides a beginning location and ending location for each session of the Transit App. A session is loosely defined as when the app is opened until it is closed - with some variation in actual timing and frequency caused by phone "widgets" on Android devices. |
| 3 | Placemarks | Includes coordinates of home and work locations that users have stored in Transit App. This represents data from an optional function in Transit App where users can store places that they often go (e.g., home or work) to easily access relevant transit information for that specific location. |
| 4 | Sharing System Actions | Provides information on the booking of car-share, bike-share and other services. File contains information on the location of the station for bike-share or the location of the vehicle for car-share systems. Further, it provides information on the booking and cancellation process for these systems. |
| 5 | Sharing System Purchase | Provides information on the purchase of shared vehicle passes, which are primarily bike-share passes. Variables include type of pass, time of request, number of passes selected, and the cost of the transaction. |
| 6 | Uber Request | Lists requests for service from Uber, which are then handed off to Uber’s app for fulfillment. This file contains the time of request, location of user at time of request, type of service requested and in some cases the drop off location. |
| 7 | Trips | Contains information about usage of the trip planning feature in the Transit app, including start and end coordinates (latitude/longitude), date, and timestamps of trip planning requests. |
| 8 | Nearby View | Contains information about the routes presented to a user in each session upon opening the app. The information includes transit route, the corresponding agency, number of taps for each route and if that route is designated as the user’s favorite transit route. |
| 9 | User Feed Session | Provides information on the number of times the Transit app is opened by a user and the number of different transit agency's data that was accessed by the user. |
| 10 | Installed App | Reports on other installed apps on the user's device that can impact Transit app functionality. For example, a user that has Uber installed will be linked to the Uber app when ordering a ride, whereas a user that does not will be sent to the app store. |
| 11 | Feed Download | Provides an overall summary of activity on the Transit app by day. |
| 12 | Favorite | Provides information on user designated favorites in terms of transit routes. |
| 13 | Device | Contains a Transit app specific identification number that is utilized in other tables to identify a unique device. This file also includes information on user selected language, type of device, model of device, operating system used, version of Transit app software installed, and last date of app use. |

**TABLE 2 Comparison of Transit App Data with Traditional Transit Data Sources**

| | **Travel Survey Data** | **Automated Fare Collection (AFC) Data** | **Transit App Data** |
|---|---|---|---|
| Geographic Coverage | Single region | Single region | Multi-region |
| Institutional / Agency Coverage | Single or multi-agency | Single agency* | Multi-agency |
| Modes | Transit and/or others | Transit | Transit, Ride-hailing, Bike-sharing, Car-sharing |
| Timescales | Cross-sectional | Continuous (when transit system open) | Continuous |
| Sample Size | Small | Large | Large |

**A small number of AFC systems allow for use on multiple transit operators in the same region.*