Unravelling Mobility Patterns using Longitudinal Smart Card Data

Final report for Trafik och Region 2019

SLL-KTH research project

December 2021

Research team: Oded Cats*

Francesco Ferranti

Isak Jarlebring Rubensson

Matej Cebecauer

Lukas Kolkowski

Erik Jenelius

* Principal Investigator, cats@kth.se / o.cats@tudelft.nl

1 BACKGROUND

This project followed-up on a project called *FairAccess* which was granted in Trafik och Region 2018. In *FairAccess*, we processed Access card data and performed a sequence of inferences to derive time-dependent origin-destination matrices for the entire Region Stockholm system. Tap-in records were matched with corresponding inferred tap-out locations and time stamps for about 80% of all records. Moreover, we implemented an algorithm to generate a journey database based on our transfer inference method. We used the outputs of this process to evaluate the impacts of the fare scheme change (i.e. from zone-based to flat fare) on different user profiles. Access card products and zonal attributes were used for analysing policy impacts on different market segments.

The "*Unravelling Mobility Patterns using Longitudinal Smart Card Data*" project was granted on May 27, 2020 and the contract was signed on July 17, 2020. In this project, we capitalise on the capabilities of the inferences performed in previous work to conduct a series of market segmentation and advanced data analytics to empirically analysis demand patterns for public transport in the Stockholm County. The growing travel demand in Stockholm County is accompanied by an increased diversity of sub-centres within the region as well as in individual travel patterns. It is thus increasingly important to understand how demand patterns evolve over time, what the key market segments are and how different users are affected by changes in service provision. The latter is studied in the contact of the opening of the Citybanan project.

As stated in the SLL Research and Innovation Plan, the development of transport solutions for the Stockholm region requires new knowledge regarding travellers' needs and preferences, and the impacts for different types of travellers.

2 AIM

The overarching goal of this project is to advance the analysis capabilities developed in the previous *FairAccess* project further by discovering the prevailing demand patterns and identify distinctive user profiles from the data. To this end we perform a series of clustering studies. The results thereof offer a more nuanced understanding of different user segments. Moreover, we select as a case study the impacts of a particular infrastructural investment, the opening of the Citybanan commuter train railway, on how different user groups travel.

The output of this project consist of the following:

(i) Data-driven zonal definition which reflects observed demand patterns

(ii) Origin-destination matrices based on the zones obtained

(iii) Detailed description of user profiles based on the market segmentation analysis

(iv) Analysis of the consequences of the Citybanan opening on the travel patterns of different user groups and their relation to long-term planning goals.

3 APPROACH AND WORK PROCESS

To realize the objectives stated above, the project constitutes of a series of algorithmic, theoretical and empirical research steps. The overall workflow is structured into three work packages (WP) as depicted in Figure 1.

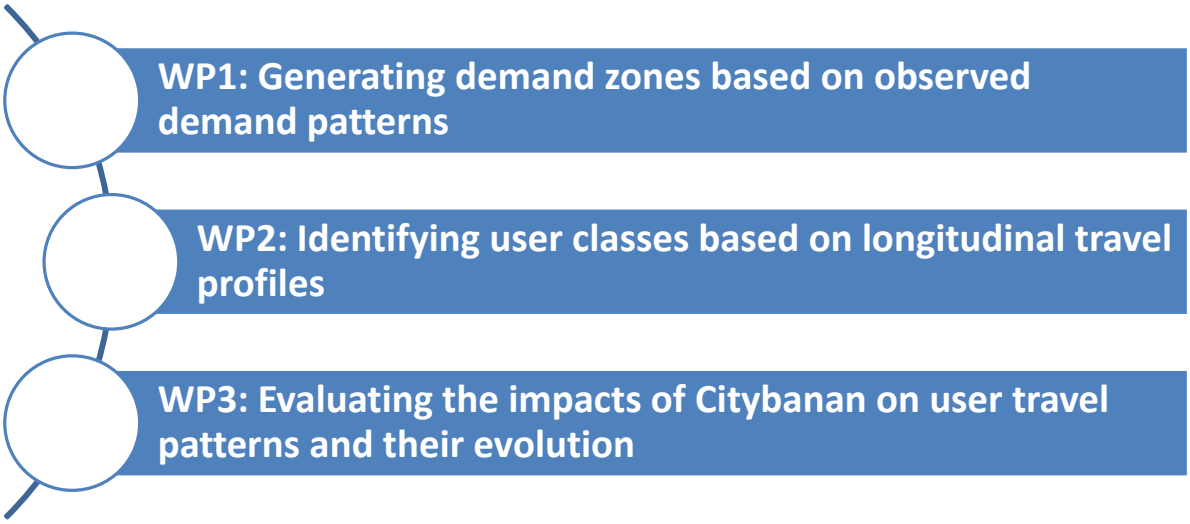

Figure 1: Project workflow

In WP1 we perform a spatial clustering for identifying a coherent and meaningful set of zones for constructing origin-destination matrices. A k-means-based stop aggregation method which can quantitatively determine the partitioning by considering both flow and spatial distance information has been adopted. This analysis allows determining zones based on the prevailing demand patterns rather than adhering to census zones which may have arbitrary boundaries in terms of consistent origin-destination travel relations. The method proposed determines for each stop which of the neighbouring zones manifests the most similar demand characteristics and assign it accordingly.

In WP2 we perform market segmentation, clustering users rather than stops. Users are categorized into a limited and informative number of profiles which exercise distinctive travel patterns in terms of either their temporal profile - travel frequency with respect to time of the day and day of the week – or spatial properties of their travel pattern. We also examine the respective Access card products. The classification is based on a year-long travel diary to account also for seasonal variations. The outcomes of WP1 were used for analysing the spatial properties of user segments in WP2.

In WP3 we turn into analysing changes in user patterns with a special interest in changes related to the opening of the Citybanan and aspects related to what we refer to as mobility-segregation. By the latter it is meant that users might differ in the extent to which they are exposed to different-others (in this study considered in terms of income level) and that this might have also changed over time. To this end, we develop indicators that allow analysing variations in this regard and compare different areas across the county as well as how those evolve over time.

4 PROJECT MANAGEMENT

The project team consists of researchers with diverse training and skills, interests and backgrounds. Moreover, the research team includes researchers at different stages of their research career, affiliated with different groups and based in different countries. The pandemic crisis has on hand hampered physical meetings but on the other hand made it more acceptable for everyone to collaborate remotely. Slack was used as the main communication channel for discussing technical developments and sharing intermediate results along with supporting tools for software development and project management for the core developers. A research engineer collaborated closely with the PI and with a PhD student responsible for the data warehouse. In addition, a master student at TU Delft was co-supervised by a team member from KTH as well as received feedback at key milestones from the SLL contact person. The combination of these work patterns and communication channels proved very effective yet efficient in ensuring the successful completion of this project, including dependencies within and between WPs.

Two key meetings were organized with a large group of developers and planners at Region Stockholm on January 22, 2021 and November 17, 2021. During the first meeting, the results of the data-driven zoning and initial results from both temporal and spatial user clustering were presented. During the second meeting, the key findings of this project were presented, including the zonal clustering and mobility segregation studies, followed by a discussion of possible applications and follow-ups. In addition, the PI presented upon request this project on an internal SLL meeting on August 17, 2021.

5 PROJECT DISSEMINATION

Parts of the work performed in this project have been presented in the following international peer-reviewed scientific conferences:

- “Spatial-temporal clustering of travel patterns using smart card data” has been presented at TransitData 2021 on October 2021.
- “Spatial Visiting Profiles of Public Transport Users” has been accepted for presentation at the 101th Transportation Research Board Annual Meeting, Washington DC. January 2022.

In addition, three journal submissions based on the work performed in this project are either under preparation or under review: (i) spatial clustering of user visiting profiles; (ii) identifying zonal attractiveness using visiting patterns; (iii) analysing segregation in mobility patterns using smart card data.

6 OUTLINE

The research output appended to this report consist of the following:

- “Unravelling Individual Mobility Temporal Patterns using Longitudinal Smart Card Data”
- “Unravelling the Spatial Properties of Individual Mobility Patterns using Longitudinal Travel Data”
- “Voting with One's Feet: Unraveling Urban Centers Attraction using Visiting Frequency”
- “Measuring activity-based Social Segregation using Public Transport Smart Card Data”

The first three are working papers that are in different phases of preparation towards a scientific publication. The latter one is a master thesis report conducted in the context of this project, which includes a draft paper in Appendix A therein. Each of the abovementioned outputs contains a description of study context, objectives, method, case study, application results and conclusions. In the next section we discuss conclusions for the project as a whole.

7 CONCLUSIONS AND FUTURE WORK

The outcomes of this project have directly addresses the aforementioned project objectives. The outputs of this project are of direct relevance to the planning and operations of public transport services in the Stockholm County by acquiring knowledge on travellers' behaviour, differences in travel patterns among different travellers' groups and the impact of service change on different user groups. This is of special importance in supporting planning and policy goals, in particular facilitating travel for all by planning for a system that provides accessibility to all user groups.

The outcomes of this project demonstrate the powerful and rich insights that can be unlocked once a functioning pipeline for analysing smart card data is in place. In particular, we demonstrate how it can facilitate market segmentation analysis. Our analysis illustrates how a longitudinal analysis of user travel patterns over a sufficiently long period of time allows identifying recurrent and distinctive patterns in relation to travel frequency, time-of-day, destinations and products. Future research may also relate changes in those to service or contextual changes such as fare scheme, network changes, pandemic). A systematic analysis and automatic detection of such changes - identified as card holders switching from one user segment to another - will enable the development of targeted measures aimed at increasing customer retention and thus increasing public transport ridership.