Discerning patterns in public transport demand using access information from smart card data to a transportation system

Juan Benavente Ponce
Doctorando

José Luis Moura Berodia
Director

Borja Alonso Oreña
Director

17/05/2019

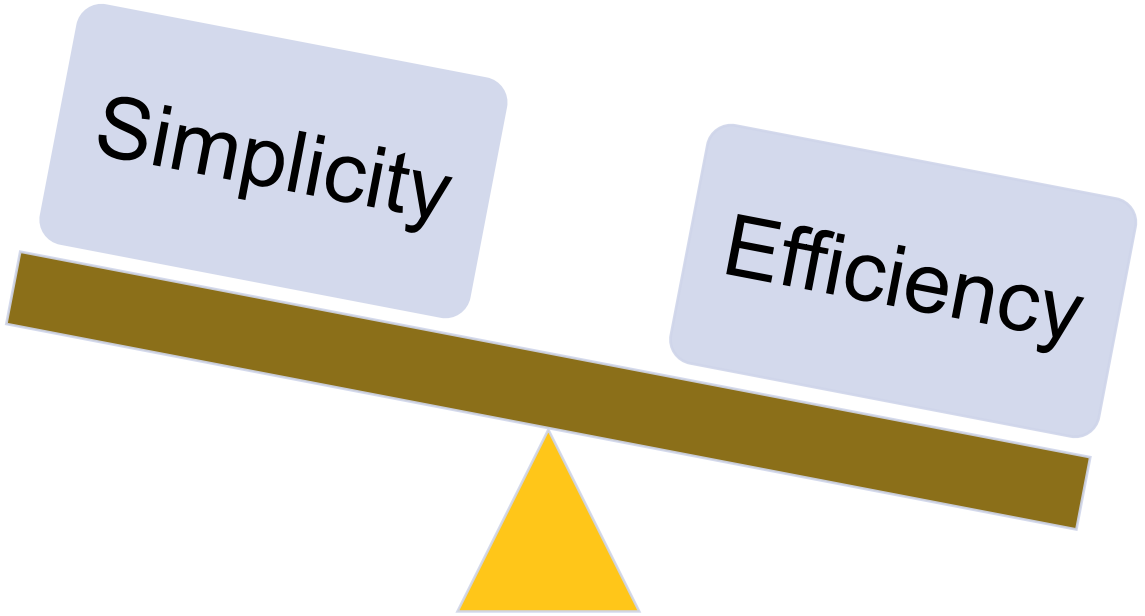
International Meeting of Civil Engineering Doctorates 2019

Motivations

Planning of a public transport system

Yearly fluctuations

Weekly fluctuations



Opposed objectives

Motivations

Proposing a format to exchange passenger validation information

The goal: to create an architecture with the same role for smart card data as the General Transit Feed Specification format is for transit data.

Index					number_of_ tap_ins
card_id	tap_in_time	line_id	veh_id	stop_id	
A unique id across the whole dataset for each card				Same as AVL data	Number of tickets paid during a single boarding

Outline of a Smart Card Data Exchange Specification

Motivations

Develop a series of useful tools to work with AVL and AFC data

Open Source

matplotlib

pandas

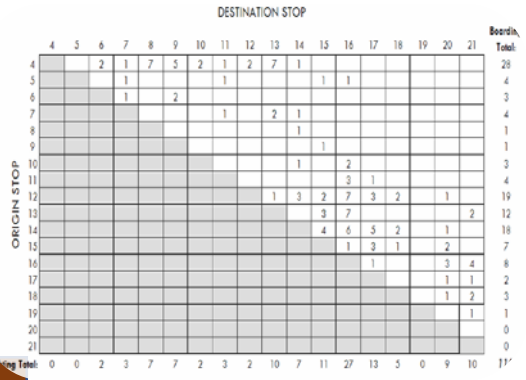
numpy

Methodology

Cleaning and analysis

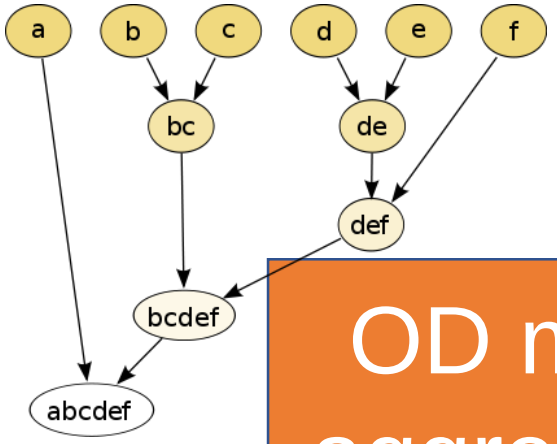
Pre-processing

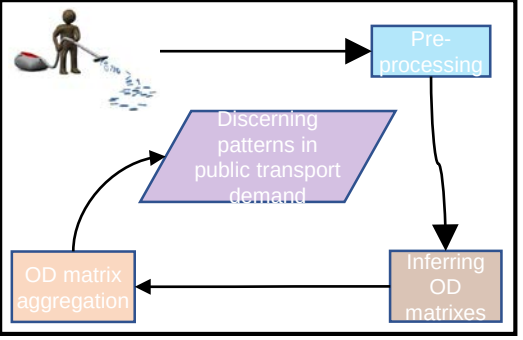
Discerning patterns in public transport demand



Inferring OD matrixes

OD matrix aggregation



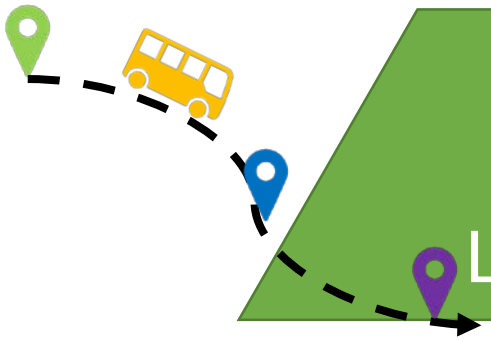


Methodology

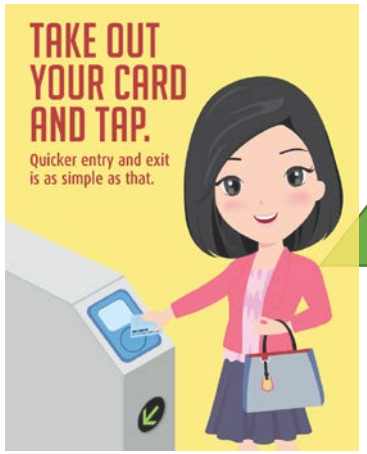
Cleaning and analysis

Geographic information

Automatic Vehicle Location logs



Automatic Fare Collection data



Cleaning and analysis

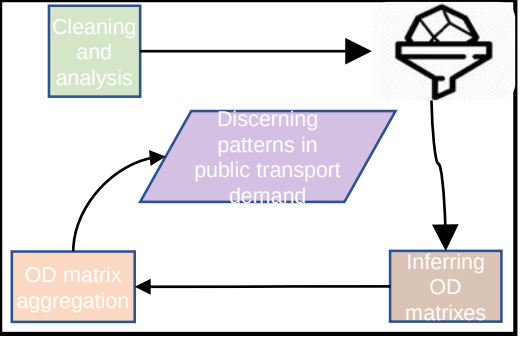
Input data

Pre-processing

Inferring OD matrixes

OD matrix aggregation

Discerning patterns in public transport demand



Methodology

Pre-processing

Cleaning and analysis

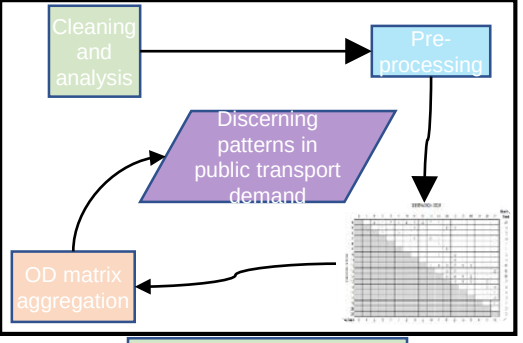
Pre-processing

Pre-processed data

Inferring OD matrixes

OD matrix aggregation

Discerning patterns in public transport demand



Methodology
Inferring OD matrixes

Cleaning and analysis

Pre-processing

Inferring OD matrixes

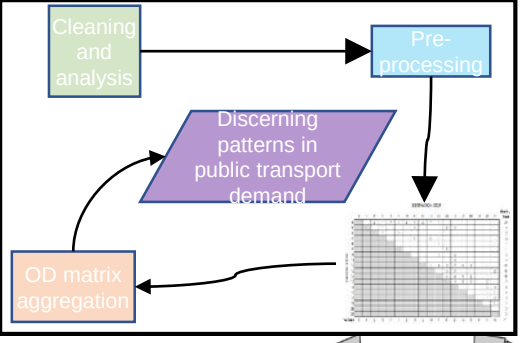
Trip chains

Assembly of OD matrixes by temporal grouping of trip chains

OD matrixes

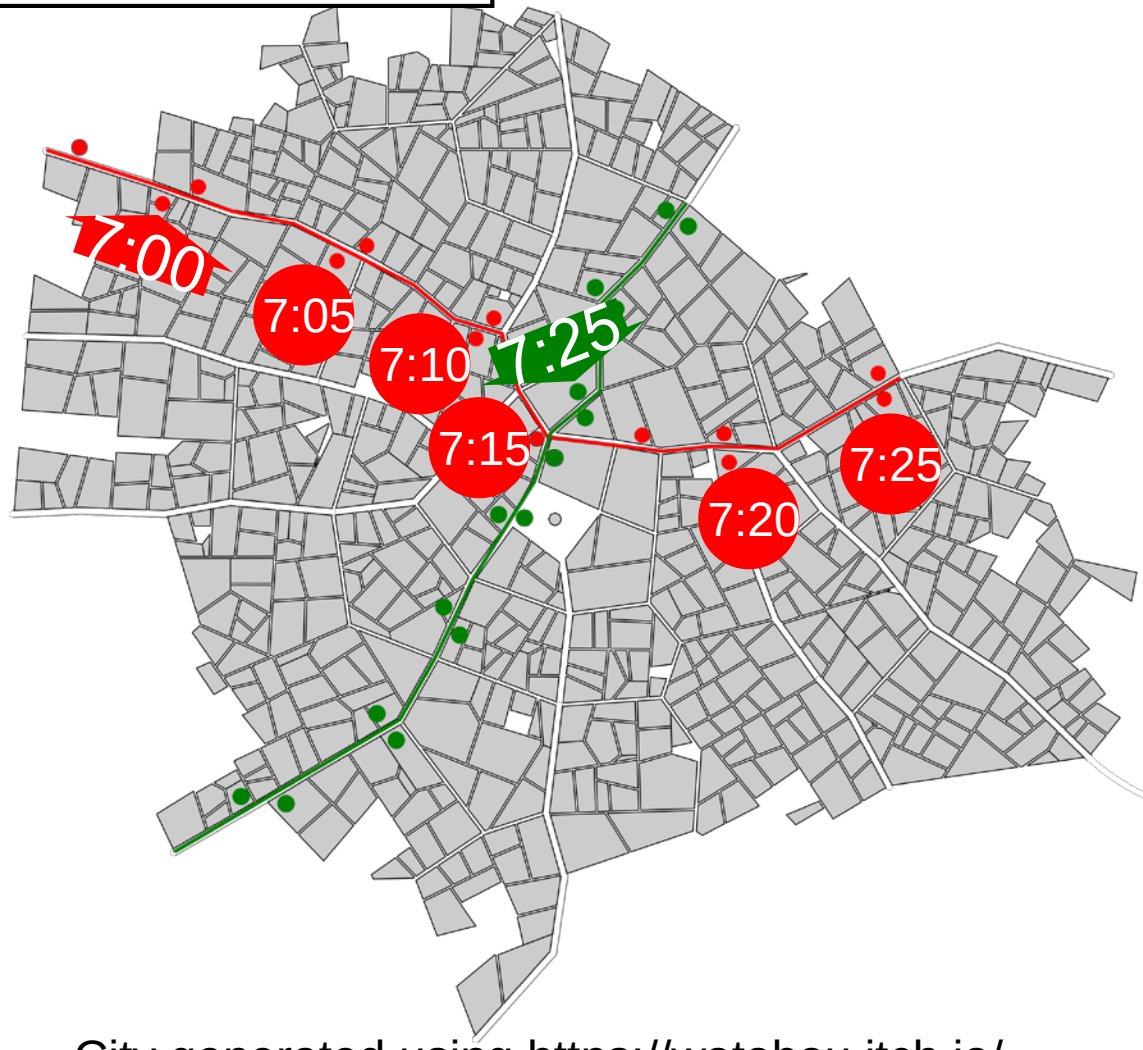
OD matrix aggregation

Discerning patterns in public transport demand



Methodology

Infer trip chains



City generated using https://watabou.itch.io/

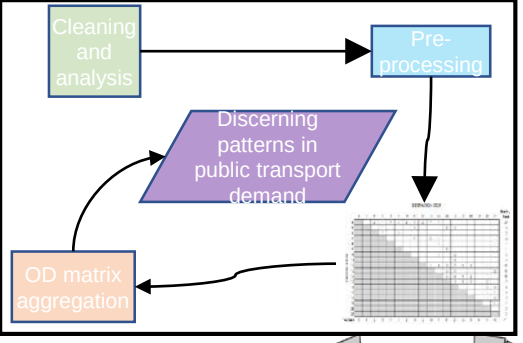
Path of a bus line

Bus stop

HH:M Passenger validation timestamp

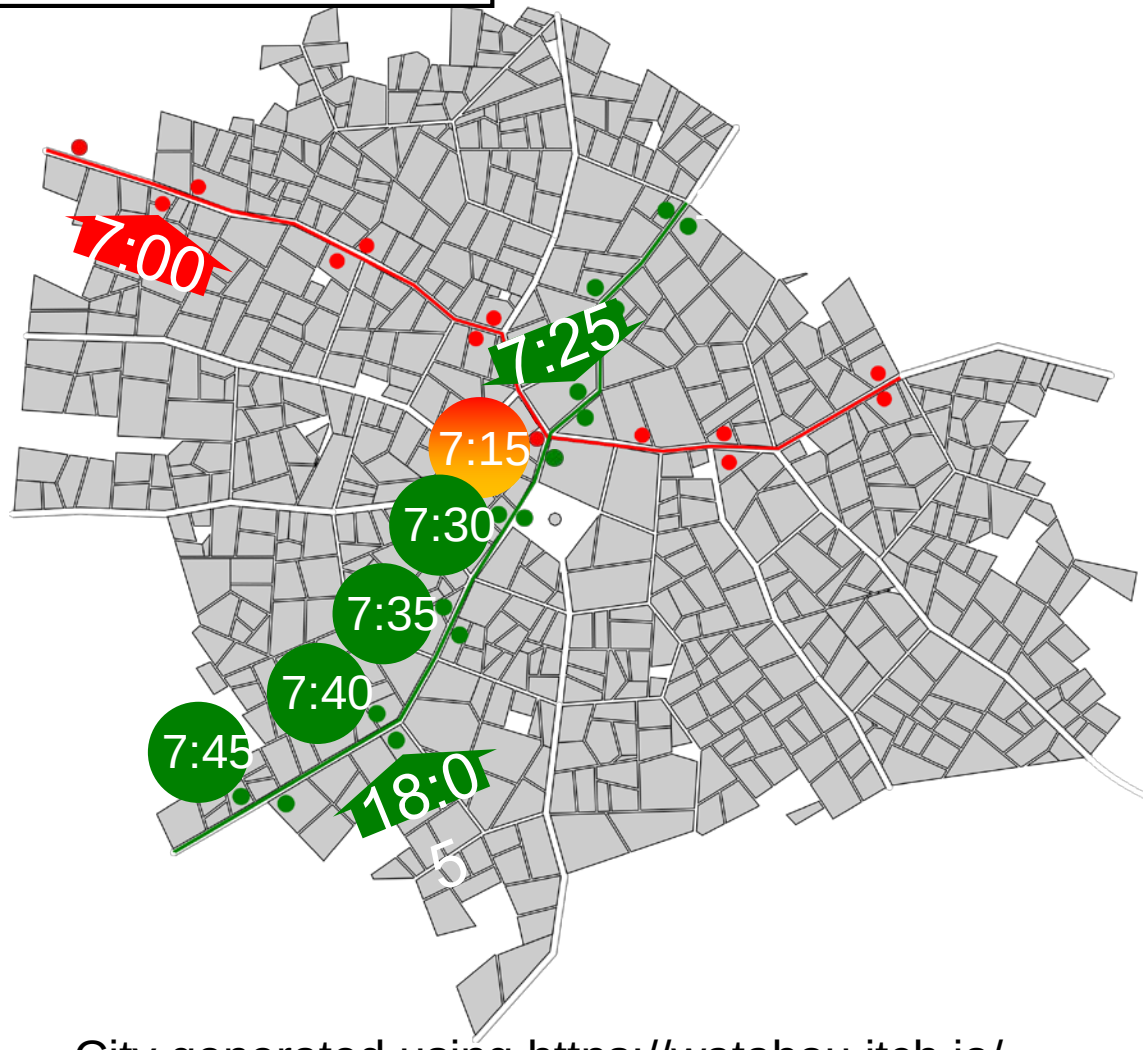
HH:MM Bus stop timestamp

HH:MM Inferred egress



Methodology

Infer trip chains



City generated using https://watabou.itch.io/

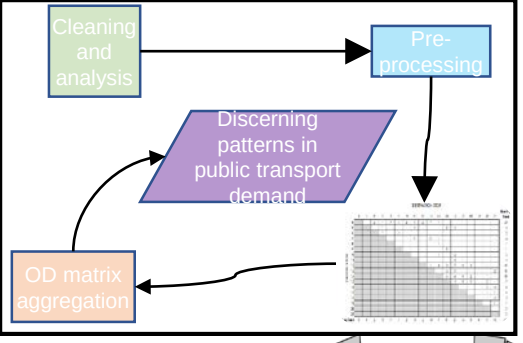
Path of a bus line

Bus stop

HH:M Passenger validation timestamp

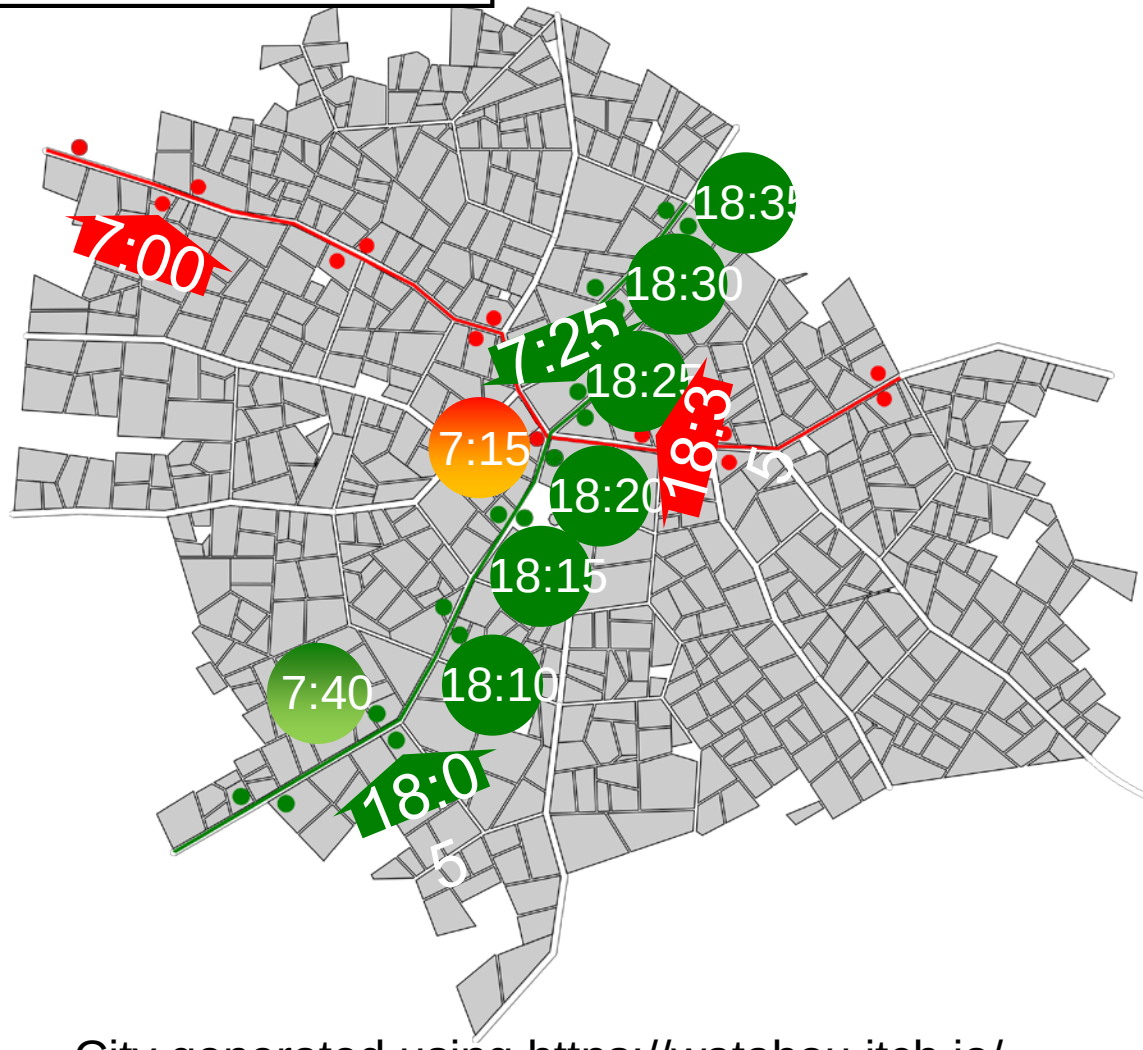
HH:MM Bus stop timestamp

HH:MM Inferred egress



Methodology

Infer trip chains



City generated using https://watabou.itch.io/

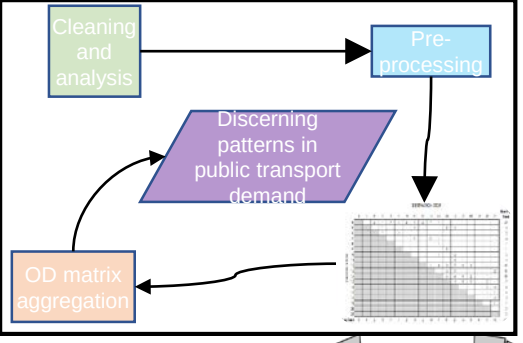
Path of a bus line

Bus stop

HH:M Passenger validation timestamp

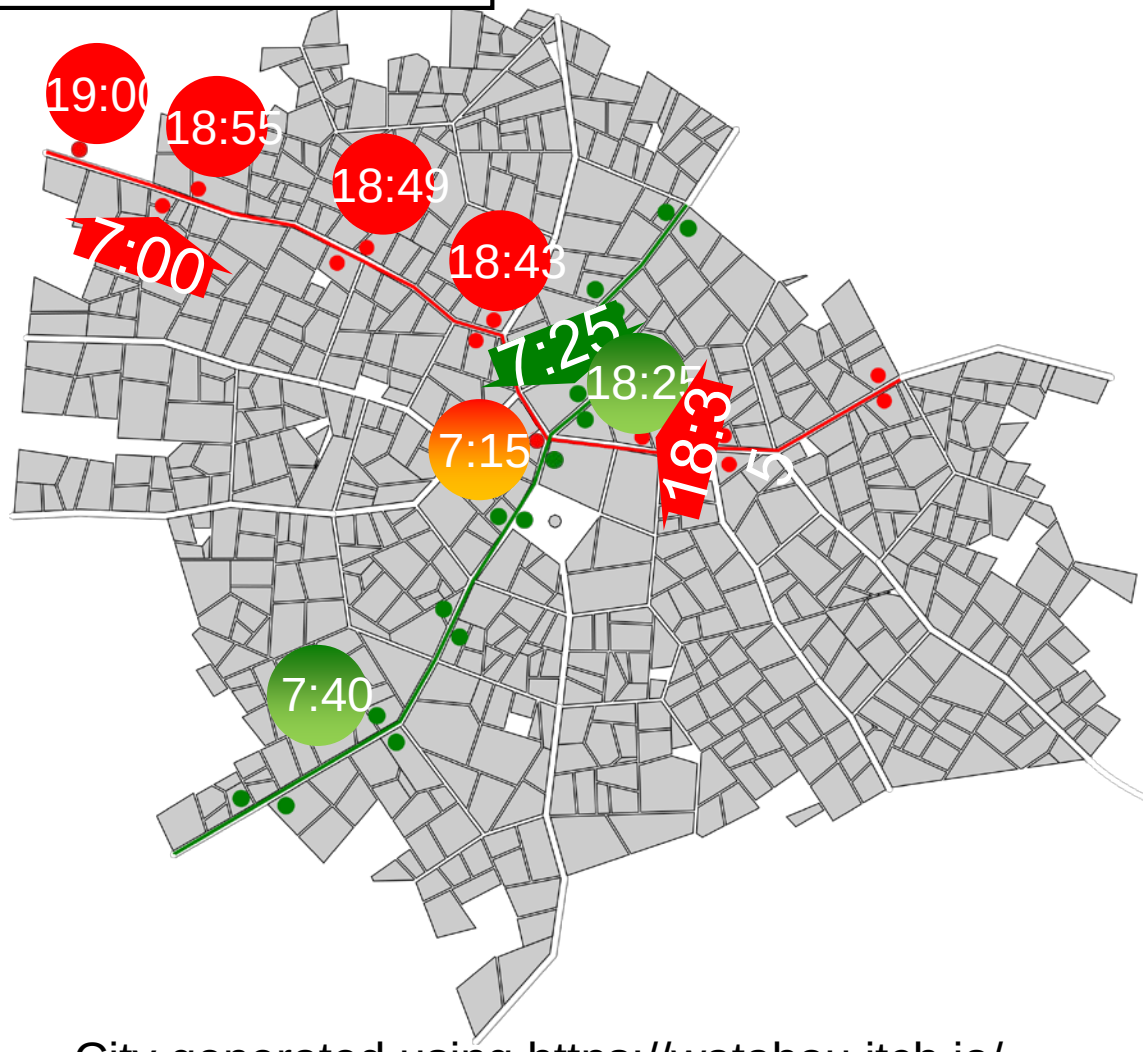
HH:MM Bus stop timestamp

HH:MM Inferred egress



Methodology

Infer trip chains



City generated using https://watabou.itch.io/

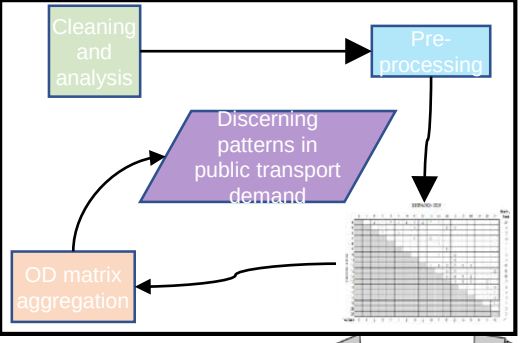
Path of a bus line

Bus stop

HH:M Passenger validation timestamp

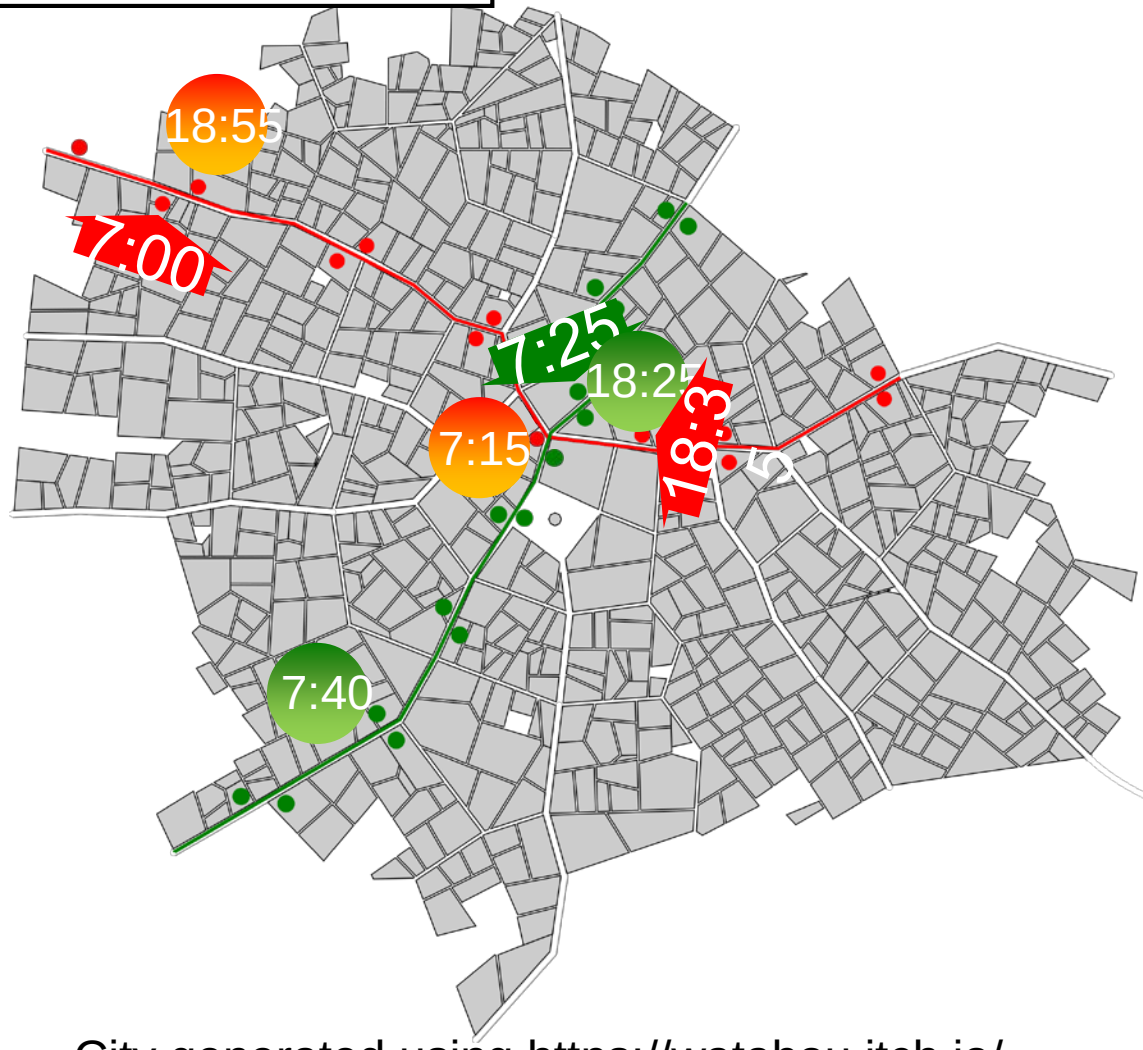
HH:MM Bus stop timestamp

HH:MM Inferred egress



Methodology

Infer trip chains



City generated using https://watabou.itch.io/

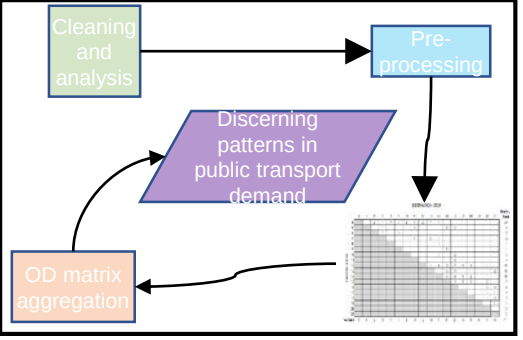
Path of a bus line

Bus stop

HH:M Passenger validation timestamp

HH:MM Bus stop timestamp

HH:MM Inferred egress



Methodology
Trip chains - common data issues

Ambiguous ids for some elements

Missing timestamps for some events

Multiple timestamps for the same event

Trip ID inconsiste ncy

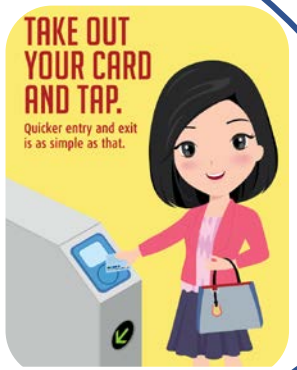
Unable to match rides to vehicle trips

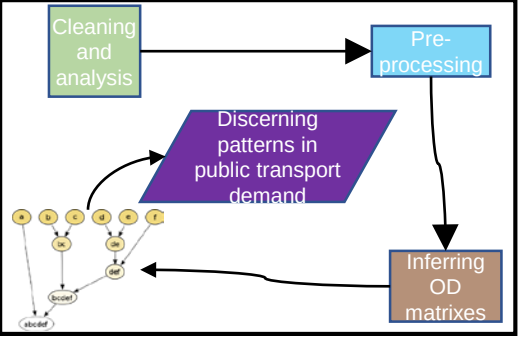
Uncertainty in trip cancellation/execution

Lack of information to match rides to scheduled trips

Indirect availability of line information about stops

Erroneous check-in/out records





Methodology

OD matrix aggregation

Public transport demand patterns

Cleaning and analysis

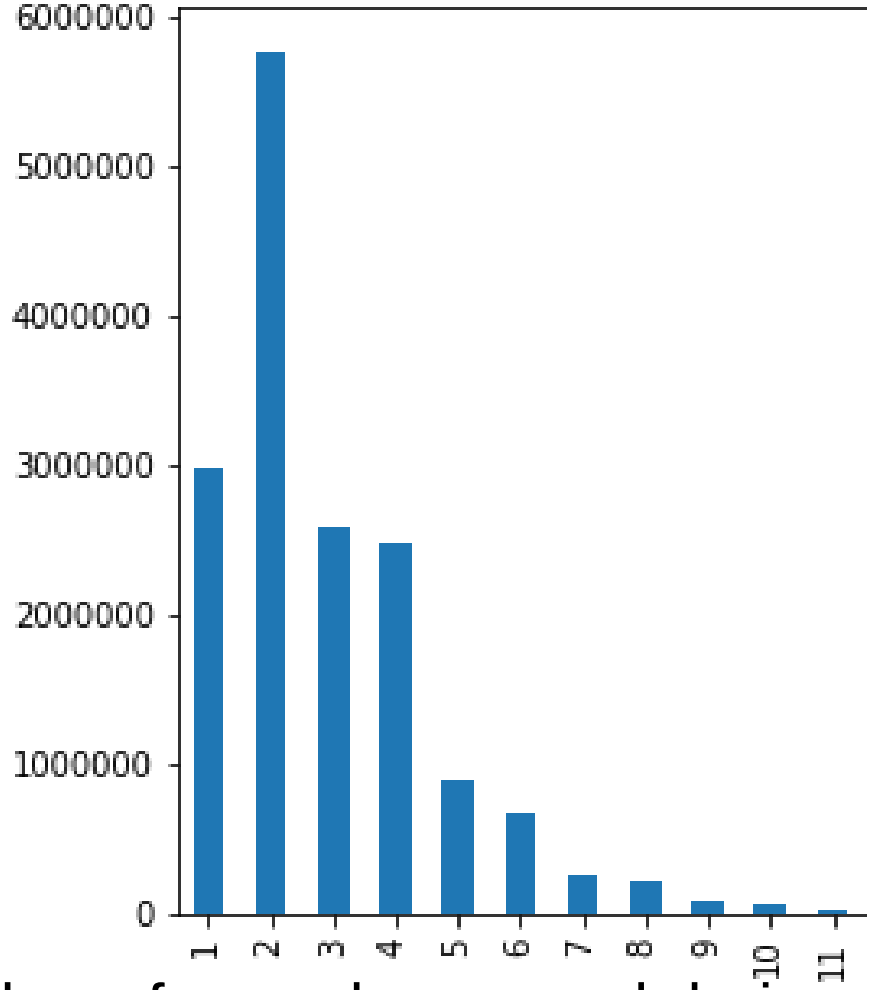
Pre-processing

Inferring OD matrixes

OD matrix aggregation

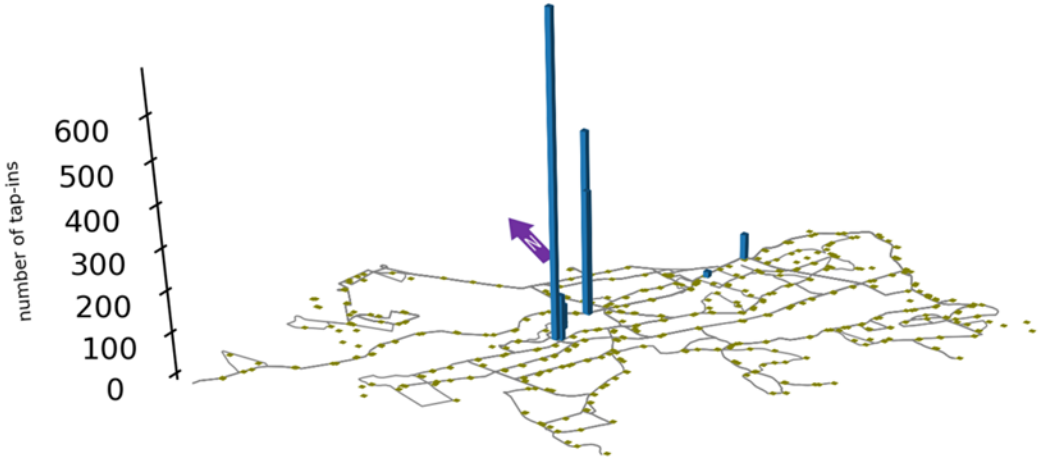
Discerning patterns in public transport demand

Some intermediate results



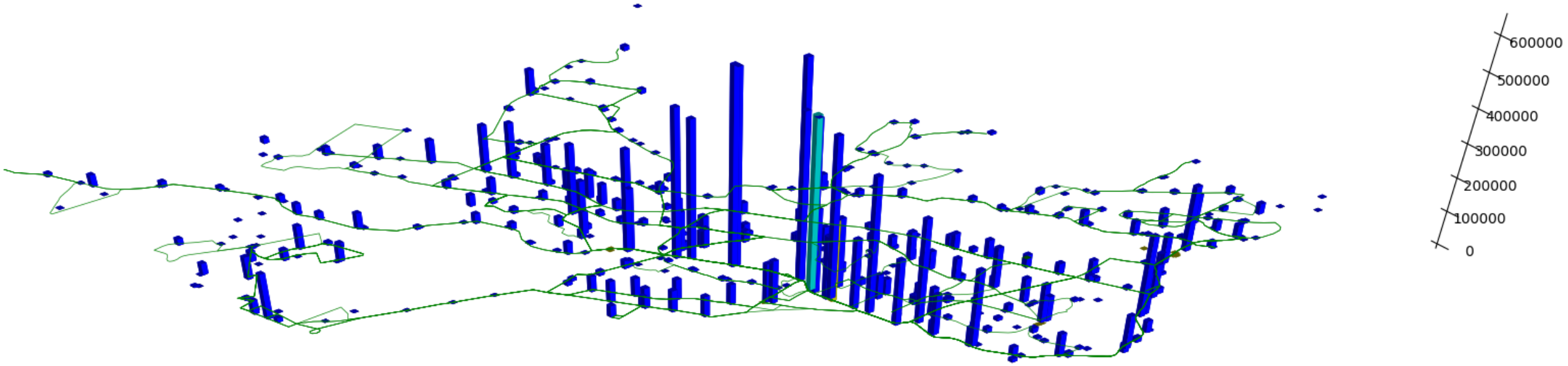
Number of records per card during a year

Card [redacted]: 1683 solitary tap-ins, 1731 total tap_ins.



Card used 40 times/day

Some intermediate results



Passenger validations at Santander bus stops during a year

Thanks