Age-based Latent Choice Model using EM Algorithm

Behavior Modeling Summer School 2023

18th-20th September 2023

Group: t22_ut_bin_c (University of Tokyo)

Group Members:
Johannes Bernsteiner
Kasun Thalgaskotuwa
Niloofar Yavari
Pankaj Kumar
Jean Claude Ndarhuhutse

Problem & Objective

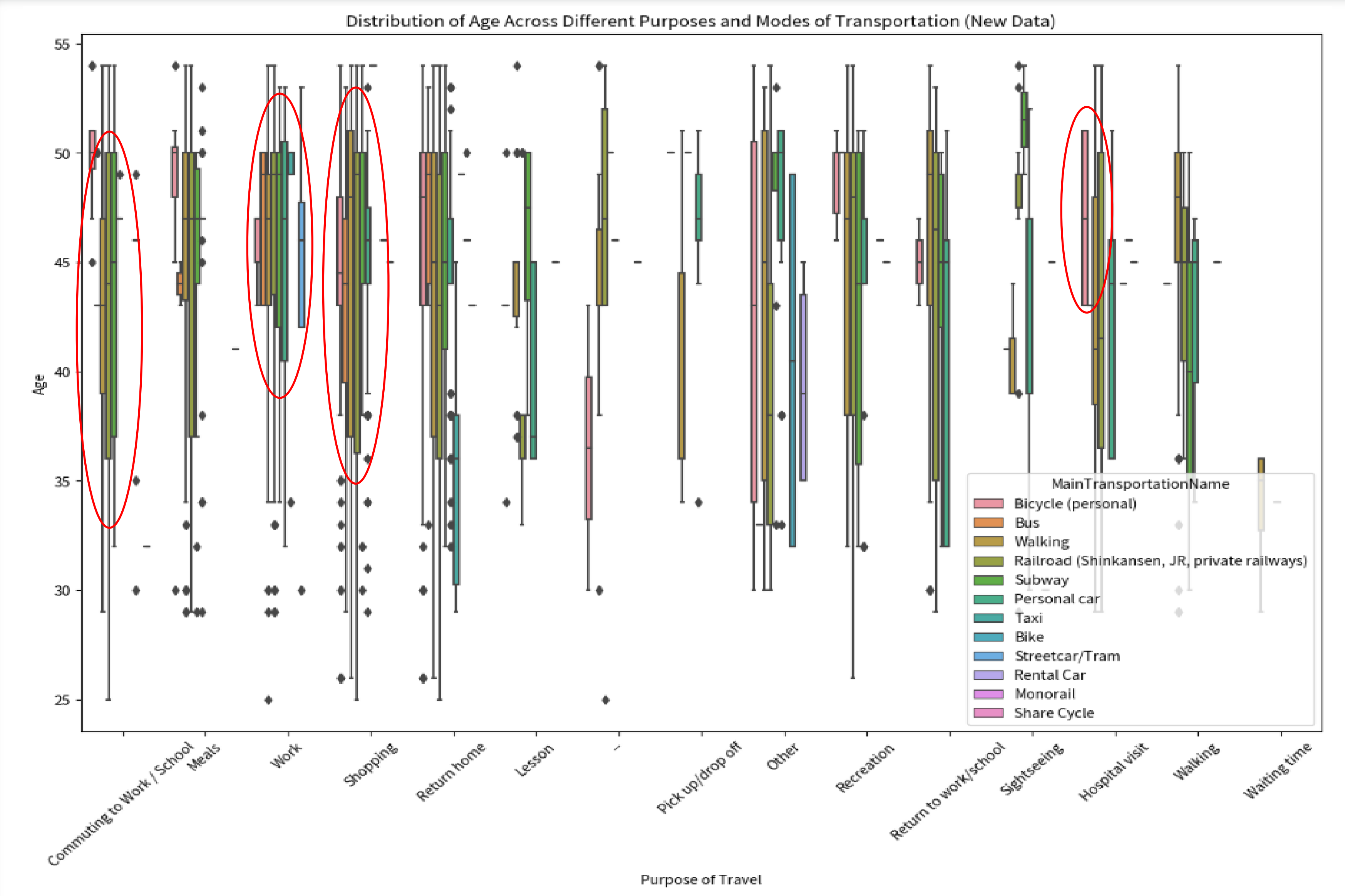
- Problem:
  - Age is an important parameter for policy-making. But, not counted adequately in various models
    - Important due to aging population in Japan
  - People's choices are heterogeneous which is not adequately represented in traditional mode choice models
  - Absence of Latent characteristics in decision-making process may adversely impact the estimation results

- Objective:
  - Estimate people's choice for mode of transportation for different age groups using prove person data
    - When data for some users is missing
  - Examine the impact of inclusion of latent variables in estimation modeling on people's choice for mode of transportation
  - Develop scientific estimation models to evaluate available policy alternatives and compare the sensitivity based on elasticity of fares

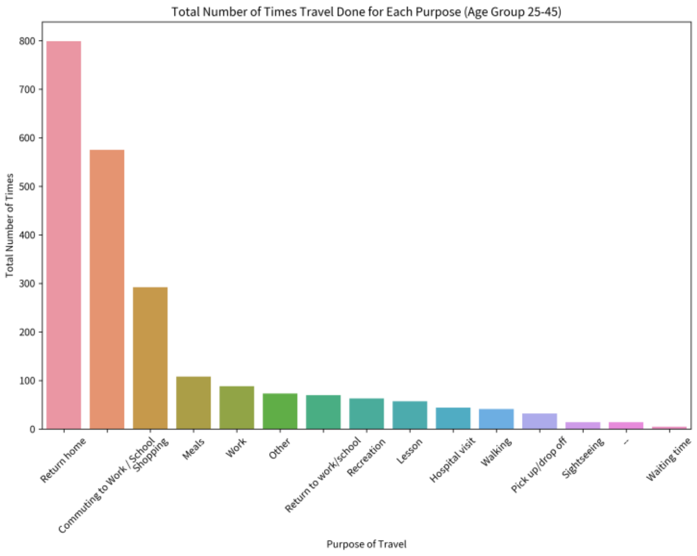
Data Summary

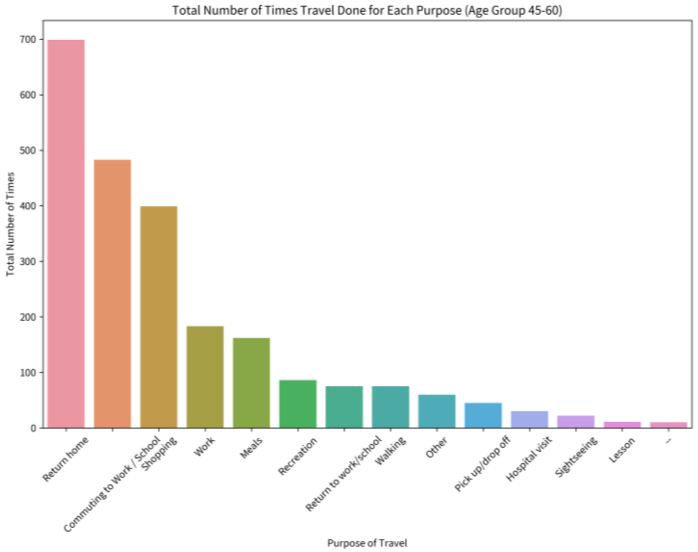
- Data collected for Shibuya city between September 22, 2021 to October 26, 2021
- Total Number of Trips: 4,619
- Total Number of Users: 136
- Different Purposes for Trips (14): Commuting to Work / School, Lesson, Work, Return home, Shopping, Meals, Hospital visit, Other, Return to work/school, Walking, Recreation, Pick up/drop off, Sightseeing, Waiting time
- Different Modes of Transportation (12): Bicycle (personal), Railroad (Shinkansen, JR, private railways), Personal car, Bus, Walking, Subway, Streetcar/Tram, Rental car, Taxi, Bike, Monorail, Share Cycle
- Number of users in different age groups: 74 (25-45 years), 62 (45-60 years)

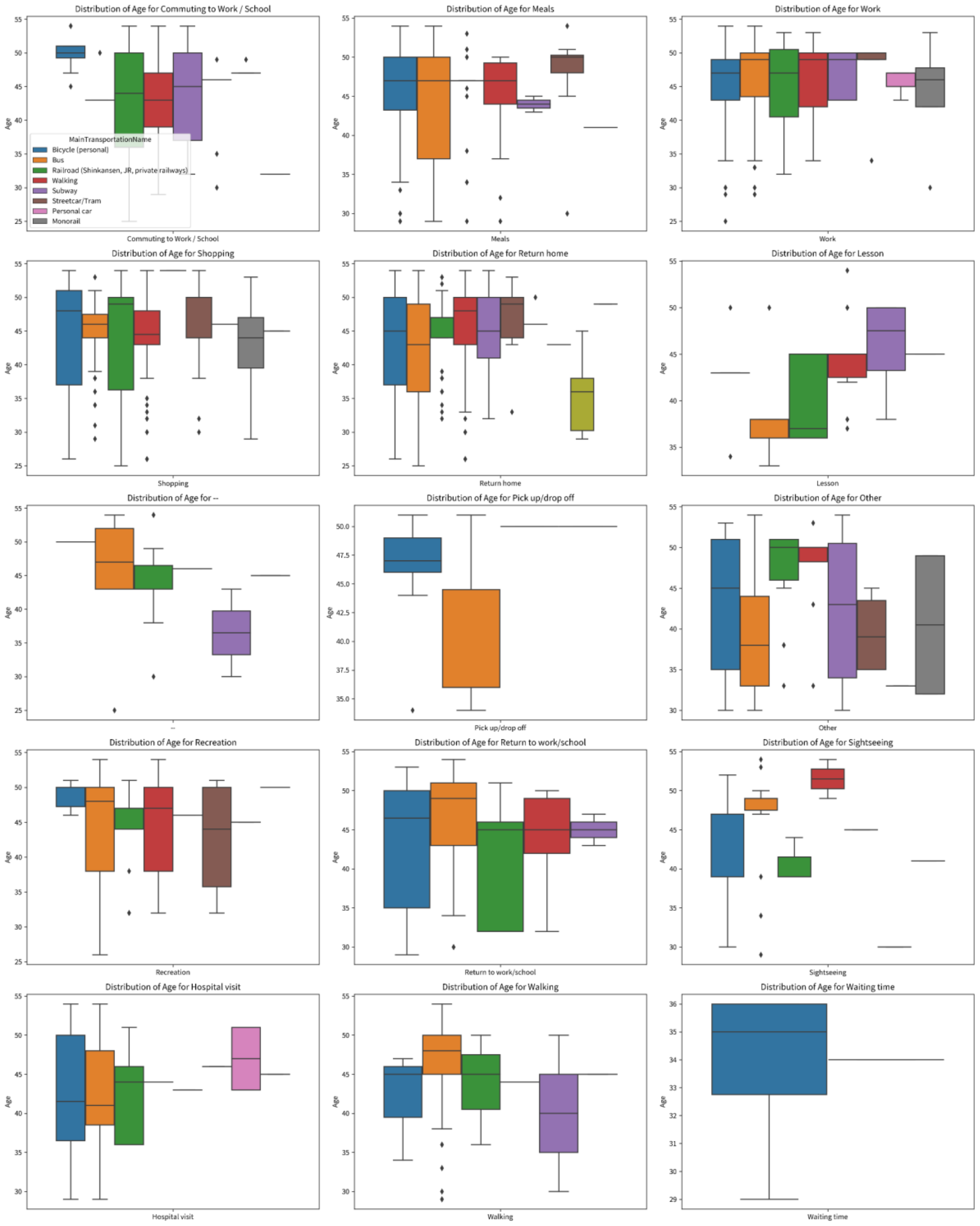
Cross-sectional Data Analysis



Data Summary







Model structure

Class 1

Car = β_1 Distance$_{car}$ + β_2 Travel Time$_{car}$ + $\boldsymbol{\beta_3}$ **Age**
Bus = β_4 Distance$_{Bus}$ + β_5 Travel Time$_{car}$ + β_6 Bus Fare + $\boldsymbol{\beta_7}$ **Age**
Train = β_8 Distance$_{Train}$ + β_9 Travel Time$_{car}$ + β_{10} Train Fare + $\boldsymbol{\beta_{11}}$ **Age**
Walk = β_{12} Distance$_{walk}$ + β_{13} Travel Time$_{car}$ + ε

Class 2

Car = β_{14} Distance$_{car}$ + β_{15} Travel Time$_{car}$
Bus = β_{16} Distance$_{Bus}$ + β_{17} Travel Time$_{car}$ + β_{18} Bus Fare
Train = β_{19} Distance$_{Train}$ + β_{20} Travel Time$_{car}$ + β_{21} Train Fare
Walk = β_{22} Distance$_{walk}$ + ε

Expectation – Maximization algorithm

- Expectation step – Calculating expected value of latent variable

$$E[\text{missing value of } n \text{ when choice is } 1] = \frac{\sum_j p(j) * p(Choice\ 1|j, \boldsymbol{\theta}) * \ j}{\sum_{j'} p(j') * p(Choice\ 1|j', \boldsymbol{\theta})}$$

- Maximization step – Updating parameters for maximum likelihood

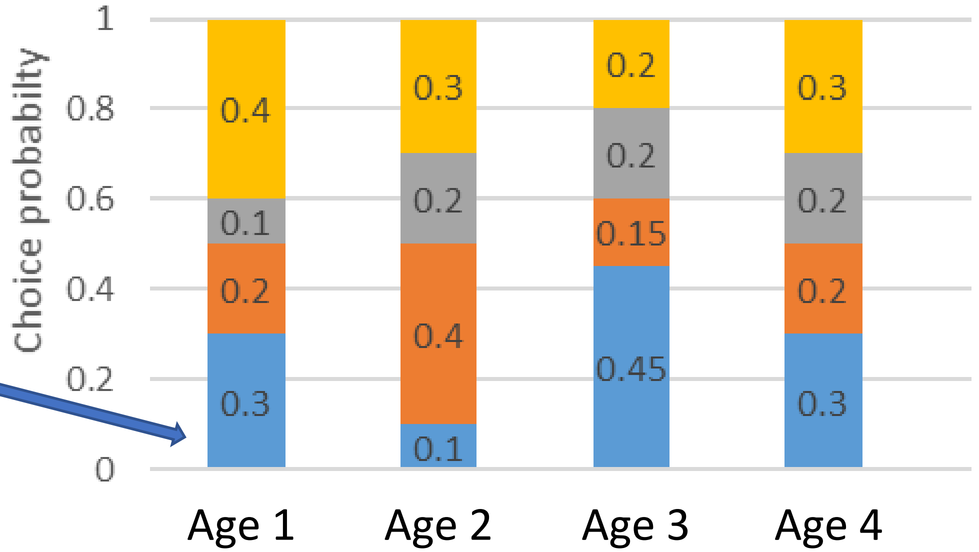
$$\boldsymbol{\theta} = argmax \sum_n \sum_i y_{in} \log\big(p_n(i|\boldsymbol{\theta})\big)$$

$$\left|LL^{new} - LL^{old}\right| < \varepsilon$$

Expectation step

1. Computation missing values

probability for individual n, select
choice 1,
when age level is 1



$$E[Agen] = \frac{\sum_j p(j) * p(Choice\ 1|j, \boldsymbol{\theta}) * Income\ j}{\sum_{j'} p(j') * p(Choice\ 1|j', \boldsymbol{\theta})}$$

$p(j)$ = Prior probability of Age level j

$p(Choice\ 1|j, \boldsymbol{\theta})$ = Conditional probability on choice 1

Age j = Age value

$$= \frac{p(1) * 0.3 * Age\ 1 + p(2) * 0.1 * Age\ 2 + p(3) * 0.45 * Age\ 3 + p(4) * 0.3 * Age\ 4}{p(1) * 0.3 + p(2) * 0.1 + p(3) * 0.45 + p(4) * 0.3}$$

Expectation step

2. Estimate latent class probabilities

- Assume that samples can be classified as several latent class according to the heterogeneity.

Exp : 1. Class with Age

2. Class without Age

- Each class is evaluated by different utility functions.

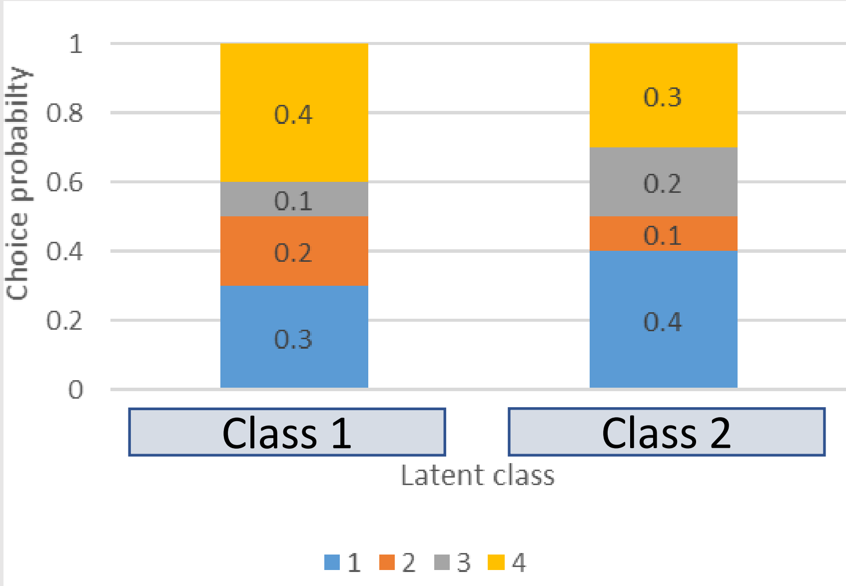
Probability to belong in class j for individual n: S_{jn}

$$S_{1n} = \frac{p_n(Choice\ i|Latent\ class\ 1, \boldsymbol{\theta})}{\sum_j p(Choice\ i|Latent\ class\ j, \boldsymbol{\theta})}$$

*i is the choice of n

When $P_{jn}(i|\boldsymbol{\theta_j})$ is the probability of selecting choice i when class is j,
The probability of selecting choice i is as follows:

$$P_n(i|\boldsymbol{\theta}) = S_{1n}P_{1n}(i|\boldsymbol{\theta_1}) + S_{2n}P_{2n}(i|\boldsymbol{\theta_2})$$



In this example, when n chose choice 1,
$S_{1n} = \frac{0.3}{0.3+0.4}$ (probability to be in class 1)

Maximization step

$$\boldsymbol{\theta} = argmax \sum_n \sum_i y_{in} \log\left(p_n(i|\boldsymbol{\theta})\right) \Rightarrow \boldsymbol{\theta} = argmax \sum_n \sum_i y_{in} \log\left(S_{1n}P_{1n}(i|\boldsymbol{\theta_1}) + S_{2n}P_{2n}(i|\boldsymbol{\theta_2})\right)$$

$$LL_{new} = \sum_n \sum_i y_{in} \log\left(p_n(i|\boldsymbol{\theta})\right)$$

Check convergence,
$\left|LL^{new} - LL^{old}\right| < \varepsilon$ → No → Back to expectation step

yes

Estimation end;

Parameter Estimation

Attribute/Parameters	With age (class 1)		Without age (class 2)	
	Estimated value	t statistic value	Estimated value	t statistic value
Travel Distance				
Car	74.753	4.329***	-	-
Bus	47.571	3.404***	-12.429	-0.858
Train	9.637	1.386	-10.591	-3.356***
Walk	-9.574	-1.391	-495.338	-0.060
Travel time				
Car	1.778	0.048	-27.0352	-3.783***
Bus	-64.689	-2.937***	-112.041	-3.317***
Train	11.764	0.443	-	-
Walk				
Fare				
Bus	7.395	2.481***	9.553	2.341**
Train	-6.224	-3.037***	4.845	4.335***
Age				
Car	-66.689	-5.386***	-	-
Bus	-23.225	-5.515***	-	-
Train	4.223	2.022**	-	-
ASC	-0.752	-0.880	5.815	0.029
Number of samples	1704			
Initial log likelihood	-2362.246			
final log likelihood	-550.084			
Likelihood ratio	0.767			
Adjusted Likelihood ratio	0.758			

Note –
$p < 0.1$: *
$p < 0.05$:**
$p < 0.01$:***

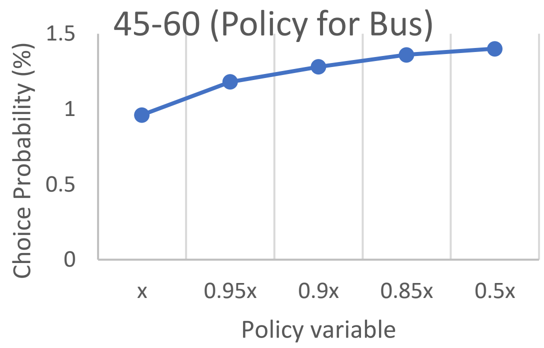
Policy Analysis

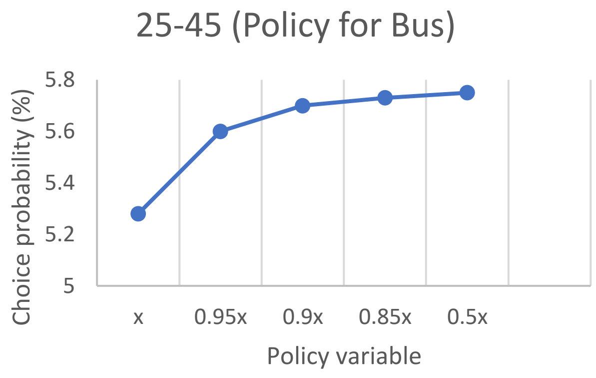
		Car		Bus		**Train**		Walk	
		45-60 years	25-45 years	45-60 years	25-45 years	**45-60 years**	**25-45 years**	45-60 years	25-45 years
No policy		6.64	11.28	0.96	5.28	89.58	79.48	**2.78**	**3.94**
Policy 1 (Bus)	Sub-policy 1: 5% discount	7.2	9.88	1.18	5.6	89.5	81.3	2.16	3.22
	Sub-policy 2: 10% discount	6.14	10.86	1.28	5.7	90.48	79.6	2.12	3.82
	Sub-policy 3: 15% discount	7.12	10.74	**1.36**	**5.73**	89.04	81.69	2.52	1.84
Policy 2 (Train)	Sub-policy 1: 5% discount	6.8	11.04	1.26	5.26	89.74	80.52	2.18	3.18
	Sub-policy 2: 10% discount	6.68	11.38	1.2	4.9	90.28	80.78	1.92	2.96
	Sub-policy 3: 15% discount	6.42	11.56	1.2	4.16	**91.08**	81.38	1.32	2.86
Policy 3 (Bus + Train)	Sub-policy 1: 5% discount	7.08	11.28	1.22	4.88	89.94	80.62	1.8	3.24
	Sub-policy 2: 10% discount	6.42	10.9	1.08	4.54	90.66	81.44	1.86	3.12
	Sub-policy 3: 15% discount	6.38	11.24	1.08	3.56	90.74	**82.2**	1.82	3.02

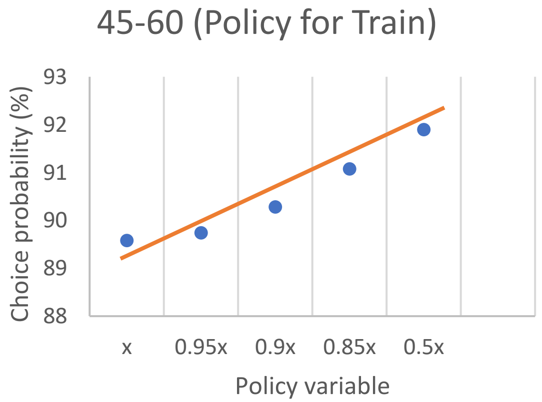
Policy Analysis

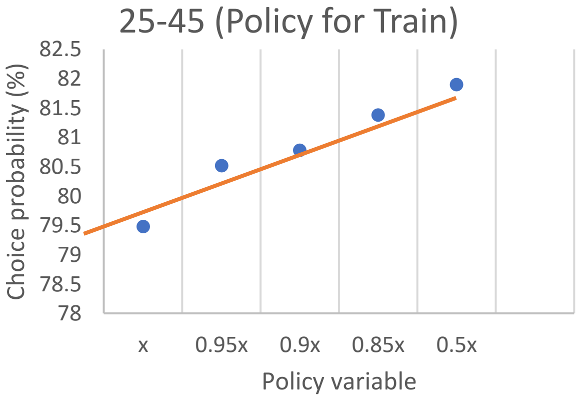
- The elasticity of age group (25-45) and age group (45-60) for both train and bus are relatively inelastic
- The choice probability for trains are significantly higher in all classes
- Probably, the results reflect the special demography and transport infrastructure of Shibuya because of lack of alternatives to trains
-
- The share of walk as mode choice decrease with discounts in public transport
- Not significant change in car as mode choice, possibly due to less impact of incentives such as discounts on shift towards public transport for

working population









FFPT policies in Other cities

Tallinn	Estonia	2013	Bus & Trolleybus	440000	8%
Chengdu	China	2018	Bus	16000000	12%
Dunkirk	France	2018	Bus	90000	85%
Hasselt	Belgium	1997	Bus	77000	132%
Seattle	USA (Waterfront Streetcar)	1982	Streetcar	750000	60%
Changning, Metro line 17	China	2018	Metro	7000000	15%
Adelaide	Australia (City Loop)	2013	Bus	1300000	25%
Sao Paulo	Brazil (Downtown Line)	2004	Bus	22000000	10%
Chambly	Canada	2012			
New Delhi	India	2019	Bus	26000000	20%
Tamil Nadu	India	2021	Bus	72000000	21%
Punjab	India	2021	Bus	27700000	40%
Luxembourg	Luxembourg	2020	Bus, Train, Trams	614000	
Washington	USA	2022	Bus & Trains	700000	
Scotland	Scotland	2022	Bus	5400000	
Romania	Romania	2022	Bus, Train, Trams	19000000	
Netherlands	Netherands	2023	Bus, Train, Trams	17300000	
Samokov	Bulgaria	2006	Bus, Train, Trams	27000	
Perth	Australia	2018	Bus	2100000	
Dewsbury	UK	2009	Bus	65000	
Avesta	Sweden	2013	Bus, Train, Trams	22000	39%
Mariehamn	Finland	2000	Bus	11000	

Challenges/ Recommendations

- The lack of data for age groups under 25 years, primarily comprising students, can have a substantial impact on the accuracy of the mode choice model.

- Using only broad age groups (25-45 and 45-60) may oversimplify diversity in transport mode choices within these groups.

- Lack of socioeconomic data such as income, employment status, and education level, among other factors, which are known to significantly influence transportation decisions, may lead to inaccurate or incomplete analyses and policy recommendations.

Thank you for Listening!!!