CP303

UNDER THE GUIDANCE OF DR. RAHUL T.M.

CAPSTONE PROJECT

MODE CHOICE MODELLING USING MULTINOMIAL LOGIT MODELS

PREPARED BY :

Rahul Matolia 2019CEB1024

Kanha Singhania 2019CEB1014

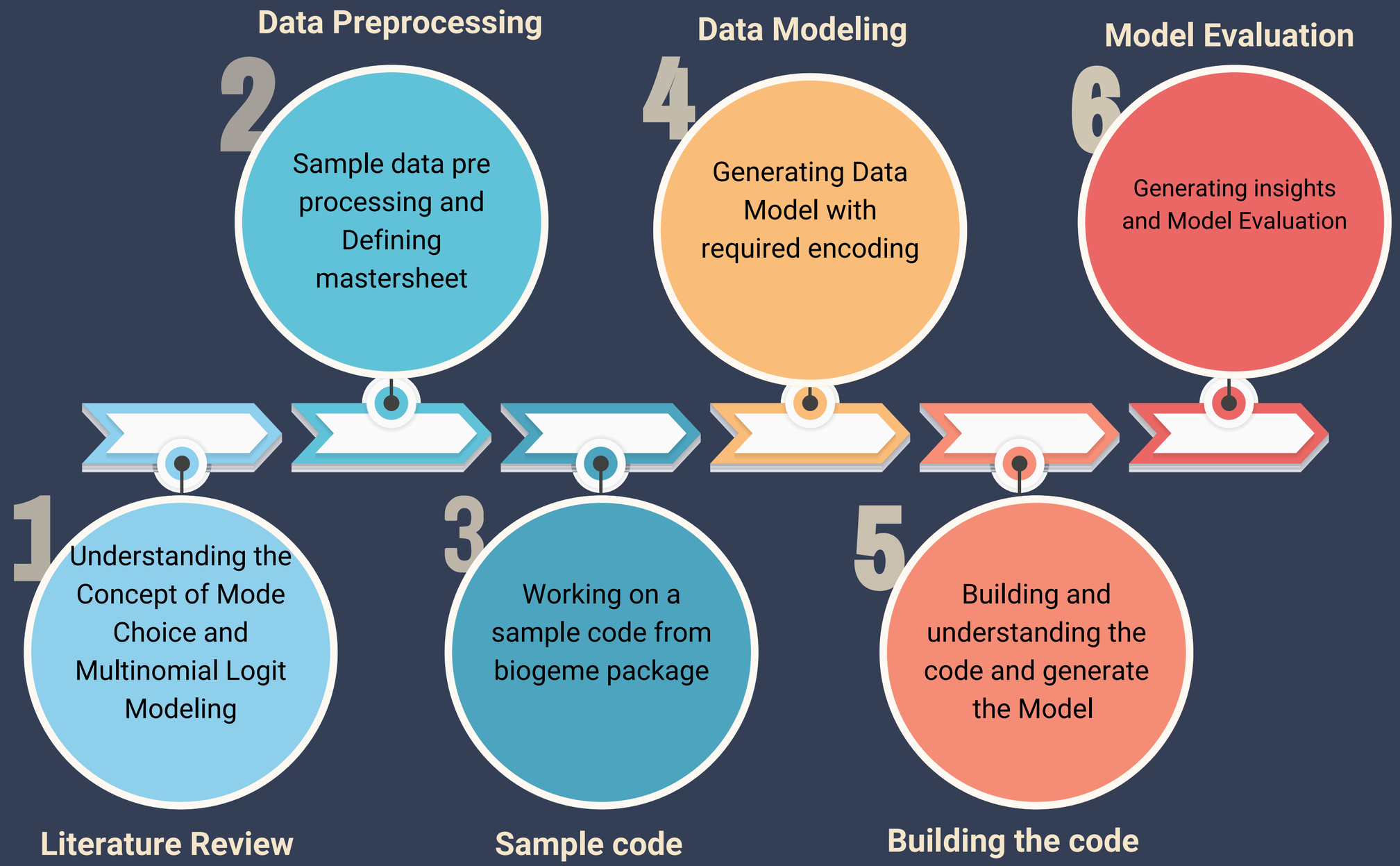
Manish Mahala 2019CEB1016

Ruheen Zeba 2019CEB1029

WHAT IS THE PROJECT ABOUT?

- The necessity for reliable transportation infrastructure to suit the needs of today’s ever-growing population is now more important than ever before.
- We propose conducting a mode choice experiment to learn about people's tradeoffs between private and public transportation system features. This would help in planning the transportation infrastructure to meet the needs of the people.
- By performing Mode choice modeling, we're looking for preferences and factors that influence transport mode selection among the population.
- Our goal is to employ choice experiments to elicit commuters' preferences for various modes of transportation.

TIMELINE


Data Preprocessing
2
Sample data pre processing and Defining mastersheet
Data Modeling
4
Generating Data Model with required encoding
Model Evaluation
6
Generating insights and Model Evaluation
1
Understanding the Concept of Mode Choice and Multinomial Logit Modeling
3
Working on a sample code from biogeme package
5
Building and understanding the code and generate the Model
Literature Review
Sample code
Building the code

WORK DONE PREVIOUSLY

LITERATURE READING

Studying various models which can be used to model the mode choice preferences. Further, extensively understanding the uses and limitations of the chosen model

DATA PRE-PROCESSING :

Level -1 : Cleaning of data and selection of non redundant values.
Level -2 : Defining Multi stage trips of individual's journey into separate trips and generating one hot encoded database.

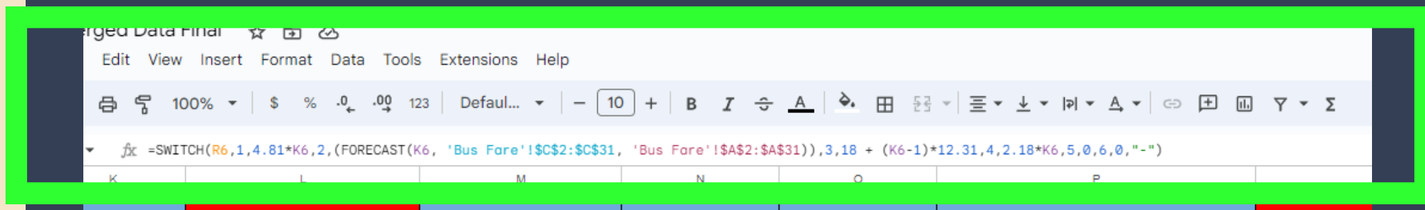
DATA ANALYSIS

After preparing the the data in the previous step, we have to analyze the data to gain some valuable insights to the distribution of various parameters.

RECAP

DATA PREPROCESSING

1. Removing redundancies
2. Filling missing data using excel tools and functions
3. Major functions learnt and used: Vlookup, Pivot tables, Nested If Else, Forecast using Linear regression etc
4. Merged the three different sheets for each ward into one so that information can be extracted for trips



Distance	STAGE DISTANCE(KM.& ROUNDED UPTO ONE DECIMAL)	STARTING TIME(24HRS.SCHEDULE)	FINISHING TIME	Time(in mins)	STAGE PURPOSE(CODES:HOME=1, WORK=2, EDUCATION=3, EMPLOYER BUSINESS=4, SHOPPING=5, SOCIAL & RECREATION=6, PERSONAL/SERVICE=7,OTHERS=8	MODE OF CYCLE= SCOOTER AUTO STU BUS=9,
1.71	1Km	8:00 AM	8.15Am	15	2	
1.71	1Km	5:00 PM	5015Pm	15	1	
7.97	1Km	7:00 AM	7.30Am	30	2	
7.95	1Km	7:00 PM	7.30Pm	30	1	
1.71	0.5Km	7:00 AM	7.15Am	15	3	
1.71	10Km	7.30Am	8.30Am	60	3	
14.25	0.5Km	8.30Am	8.45Am	15	3	
21.1	10Km	9:00 AM	10:00 AM	60	2	
9.6	500 M	11:00 AM	11.15Am	15	5	
9.55	500 M	12:00 PM	12.15Pm	15	1	
1.71	1Km	11:00 AM	11.45Am	45	3	
1.71	1Km	5.30Pm	6.30Pm	60	1	
13.49	10Km	9:00 AM	4.30Am	30	2	
13.46	10Km	6:00 PM	6.30Pm	30	1	
13.49	34m	9:00 AM	9.30Am	30	3	
13.46	34m	5:00 AM	5.30Pm	30	1	
13.49	15Km	11:00 AM	12:00 AM	60	3	
13.46	15Km	5:00 PM	6:00 PM	60	1	
13.49	15Km	9:00 AM	10:00 AM	60	2	
20.75	15Km	6:00 PM	7:00 PM	60	1	
21.56	10Km	9:00 AM	10:00 AM	60	2	
20.81	10Km	6:00 PM	7:00 AM	60	1	
21.56	15Km	11:00 AM	12:00 AM	60	3	
20.81	15Km	5:00 PM	6:00 PM	60	1	
19.42	10Km	9:00 AM	10:00 AM	60	2	
18.63	10Km	6:00 PM	7:00 PM	60	1	

For eg:

1. Calculated distance between zones using the formula:
=ARRAY_CONSTRAIN(ARRAYFORMULA(INDEX(Dist!$B$2:$IO$249, I3, J3)), 1, 1)

2. Estimated Bus fare using linear regression:
FORECAST(K9, 'Bus Fare'!$C$2:$C$31, 'Bus Fare'!$A$2:$A$31)

POST MID SEM TASKS

1. DATA MODELING

- One Hot encoding
- Scaling
- Division in slabs and categorization
- Building 3 models
- Finalizing variables to be considered

DATA MODELING

ONE HOT ENCODING

SCALING

CATEGORIZATION

BUILDING 3 MODELS

FINALIZING VARIABLES TO BE CONSIDERED

Trip Purpose					
Home	Work	Education	Employer Business	Shopping	Social & Reacreational
0	1	0	0	0	0
1	0	0	0	0	0
0	1	0	0	0	0
1	0	0	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	1	0	0	0	0
0	0	0	0	1	0
1	0	0	0	0	0
0	0	1	0	0	0
1	0	0	0	0	0
0	1	0	0	0	0

- We used One-Hot Encoding to determine the **Trip purpose.**
- One hot encoding involves the **Use of binaries** to define a value to be true or false rather than using indexing.
- This type of encoding is required to convert our **Nominal data into Discrete Numeric data**

DATA MODELING

ONE HOT ENCODING

SCALING

CATEGORIZATION

BUILDING 3 MODELS

FINALIZING VARIABLES TO BE CONSIDERED

Travel Distance (in KMs)			
DISTANCE	CAR_TC	BUS_TC	RICKSHAW_TC
1.71	8	8	27
1.71	8	8	27
7.97	38	14	104
7.95	38	14	104
1.71	8	8	27
1.71	8	8	27
14.25	69	20	181
21.1	101	26	265
9.6	46	15	124
9.55	46	15	123
1.71	8	8	27
1.71	8	8	27
13.49	65	19	172
13.46	65	19	171
13.49	65	19	172
13.46	65	19	171
13.49	65	19	172

Before scaling

- We scaled the data to build a **more uniform value** set of the variables. Rather than having haphazard distribution, the scaling helps **us and the code** to identify more clearly

- For. eg : We divided the **value of distance by 10**, the value for travel cost by 100 and the **value of income buy 10,000** for the entire data set

DISTANCE_SCALED	CAR_TC_SCALED	BUS_TC_SCALED	RICKSHAW_TC_SCALE
0.17	0.0823	0.0778	0.2674
0.17	0.0823	0.0778	0.2674
0.8	0.3834	0.1368	1.0380
0.8	0.3824	0.1366	1.0355
0.17	0.0823	0.0778	0.2674
2.11	1.0149	0.2607	2.6543
0.96	0.4618	0.1522	1.2387
0.96	0.4594	0.1517	1.2325
0.17	0.0823	0.0778	0.2674
0.17	0.0823	0.0778	0.2674
1.35	0.6489	0.1889	1.7175
1.35	0.6474	0.1886	1.7138
1.35	0.6489	0.1889	1.7175
1.35	0.6474	0.1886	1.7138
1.35	0.6489	0.1889	1.7175
1.35	0.6474	0.1886	1.7138
1.35	0.6489	0.1889	1.7175
2.08	0.9981	0.2574	2.6112
2.16	1.0370	0.2650	2.7109
2.08	1.0010	0.2579	2.6186
2.16	1.0370	0.2650	2.7109

After scaling

DATA MODELING

ONE HOT ENCODING

SCALING

CATEGORIZATION

BUILDING 3 MODELS

FINALIZING VARIABLES TO BE CONSIDERED

1. Working status :

7 categories

1. Full time (30+hrs per week)
2. Part time employee
3. Student
4. Self employed
5. Retired
6. Housewife
7. Unemployed

2. Trip purpose

5 categories

1. Home
2. Personal
3. Work
4. Education
5. Business

DATA MODELING

ONE HOT ENCODING

SCALING

CATEGORIZATION

BUILDING 3 MODELS

FINALIZING VARIABLES TO BE CONSIDERED

1. Working status :

7 categories

1. Full time (30+hrs per week)
2. Part time employee
3. Student
4. Self employed
5. Retired
6. Housewife
7. Unemployed

2 categories

1. Time bounded job

2. Non time Bounded job

2. Trip purpose

5 categories

1. Home
2. Personal
3. Work
4. Education
5. Business

2 categories

1. Work related

2. Personal

DATA MODELING

ONE HOT ENCODING

SCALING

CATEGORIZATION

BUILDING 3 MODELS

FINALIZING VARIABLES TO BE CONSIDERED

- **Gender**
- **Age**
- **Number of vehicles owned**
- **Personal income**
- **Distance Traveled**
- **Household income**
- **Trip Purpose** (Work Related and Non work related)
- **Working status** (Time bound and non Time bound)
- **Cost of travel** (For all mode choices, i.e Car, Bus, Rikshaw, Two wheeler, Cycle, Walk)

POST MID SEM TASKS

2. MODELLING USING BIOGEME PACKAGE IN PYTHON

Code :

Generated Output :

LOGIT MODEL

UTILITY FUNCTION

General utility function

$$V1 = (Bo) + (B1 * X1) + (B2 * X2) + (B2 * X3) + ...$$

- **V1 :** Denotes the utility of any particular mode of transport
- **B(beta) :** are the parameters affecting variables
- **X :** are the Variables which affect our mode choice and are considered for the modeling

Utility function of our model for any choice of mode

V1 = (ASC) + (B_COST*TC_SCALED) + (B1*TRIP_PURPOSE) + (B6*MALE) + (B7*AGE) + (B8*WORKING_STATUS) + (B9*NUM_TW) + (B10*NUM_CAR) + (B11*NUM_CYCLE) + (B12*PARKING) + (B13*DISTANCE) + (B14*PERSONAL_INCOME_SCALED) + (B15*HOUSEHOLD_INCOME_SCALED)

ASC	Alternative Specific Constant
B_COST	Parameter for travel cost
B1	Trip Purpose
B6	Gender
B7	Age
B8	Working Status
B9	Number of Two Wheeler Owned
B10	Number of Cars Owned
B11	Number of Cycle Owned
B12	Parking Availibility
B13	Distance
B14	Personal Income
B15	Household Income

SPECIFIC UTILITY FUNCTION

V_car = - 2.87 - (0.354 * CAR_TC_SCALED) - (1.12 * WORKING_STATUS) + (1.8 * PERSONAL_INCOME_SCALED) + (2.95 * NUM_CAR)

V_rickshaw = - 2.88 - (0.354 * **RICKSHAW_TC_SCALED)** - (0.479 * **MALE)** + (0.288 * **AGE)** + (0.792 * **WORKING_STATUS)** - (0.324 * **PERSONAL_INCOME_SCALED)** + (0.413 * **NUM_TW)** + (0.782 * **DISTANCE_SCALED)**

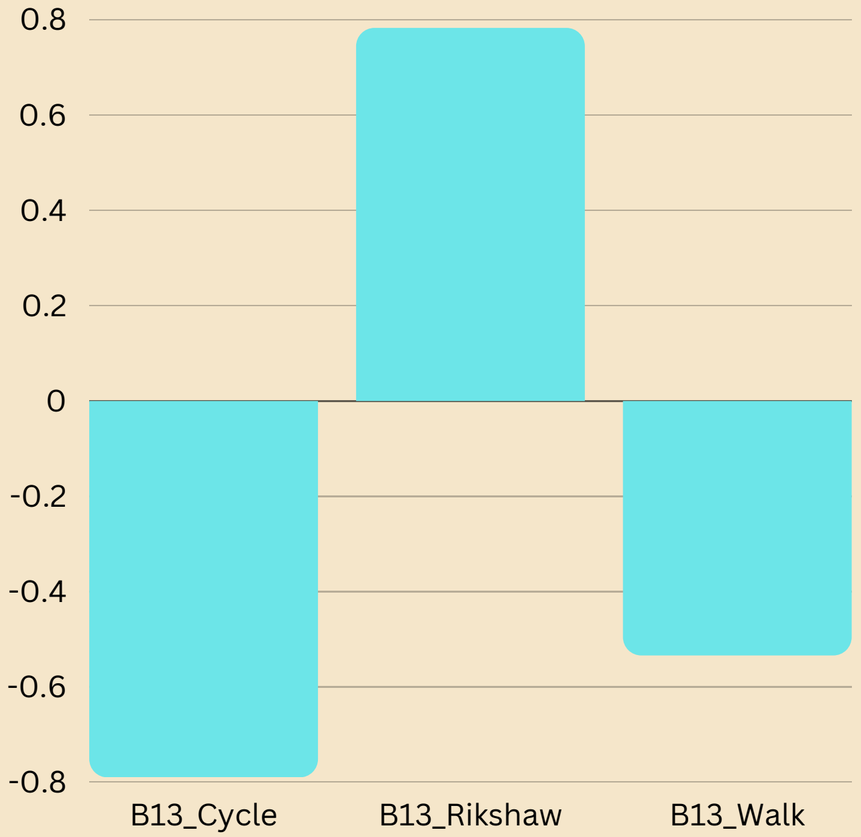
V_two_wheeler = - 2.67 - (0.354 * **TW_TC_SCALED)** + (0.851 * **MALE)** - 1.79 * **WORKING_STATUS** + 1.22 * **PERSONAL_INCOME_SCALED** + 1.72 * **NUM_TW** + 1.26 * **NUM_CAR** + 1.26 * **NUM_CYCLE**

V_cycle = - 0.851 - 0.588 * **WORKING_STATUS** - 0.665 * **PERSONAL_INCOME_SCALED** + 0.818 * **NUM_CAR** + 0.74 * **NUM_CYCLE** - 0.945 * **PARKING** - 0.789 * **DISTANCE_SCALED**

V_walk = 0.0497 * **AGE** - 0.107 * **HOUSEHOLD_INCOME_SCALED** - 1.28 * **WORKING_STATUS** + 1.03 * **NUM_CAR** + 1.69 * **NUM_CYCLE** - 0.533 * **DISTANCE_SCALED** - 0.695 * **TRIP_PURPOSE**

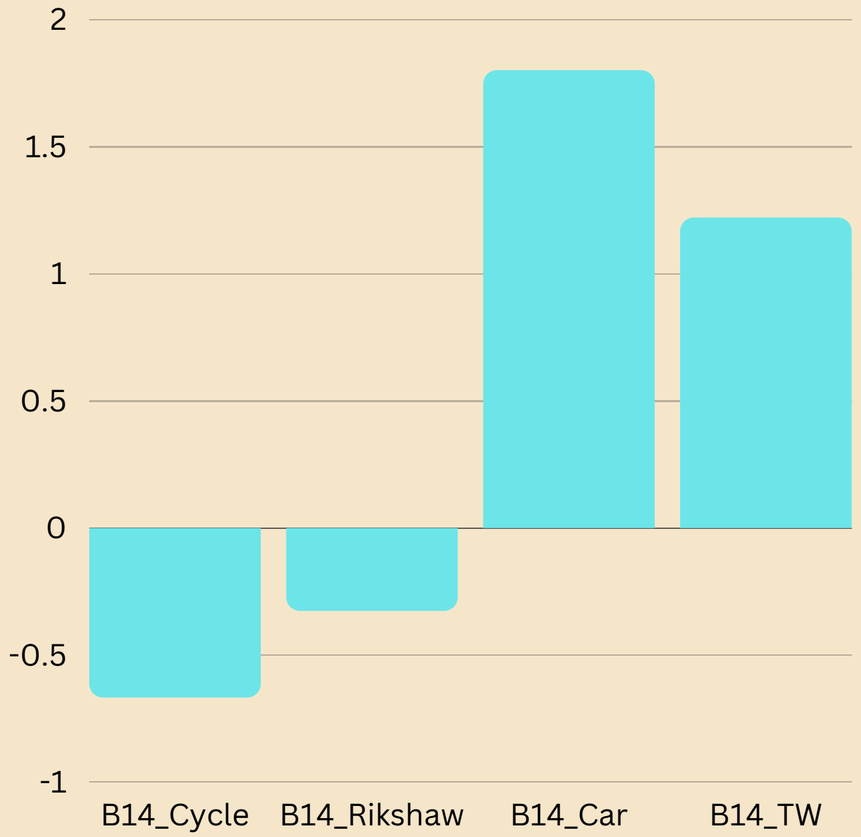
FOR B13 (DISTANCE PARAMETER)

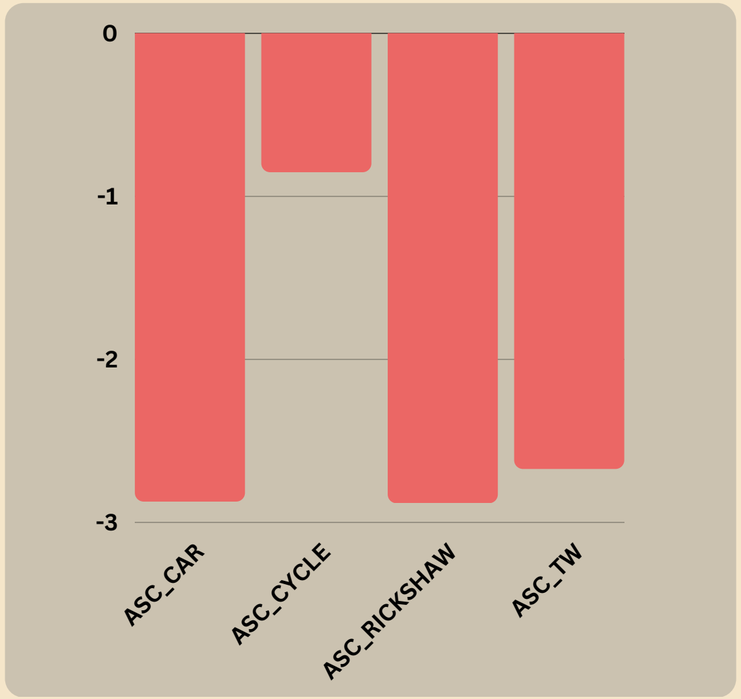
B13_CYCLE	-0.789
B13_RICKSHAW	0.782
B13_WALK	-0.533

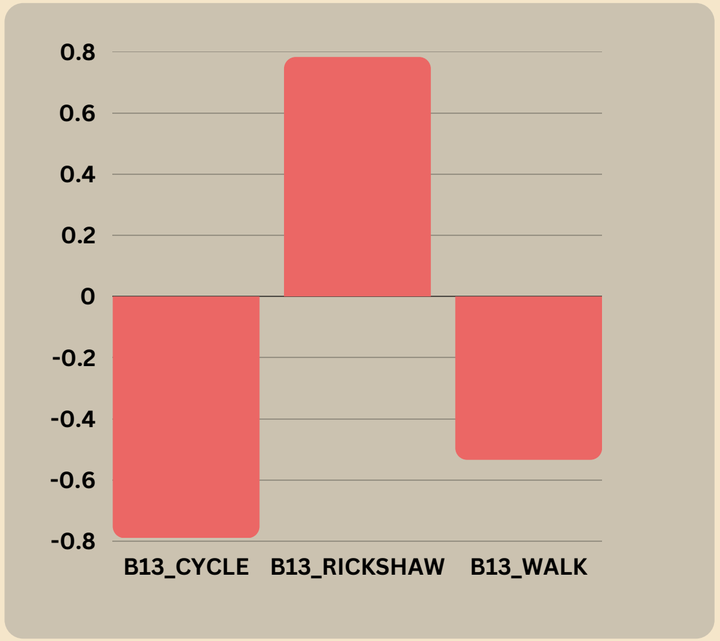


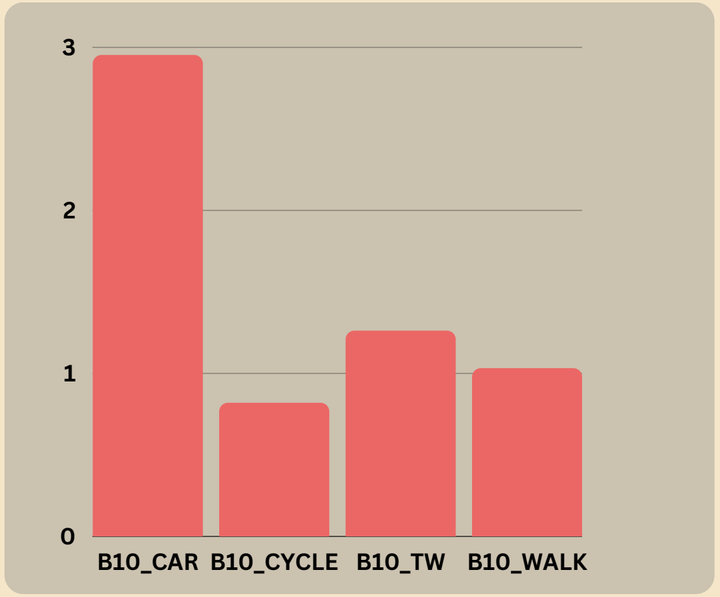
FOR B14 (PERSONAL INCOME)

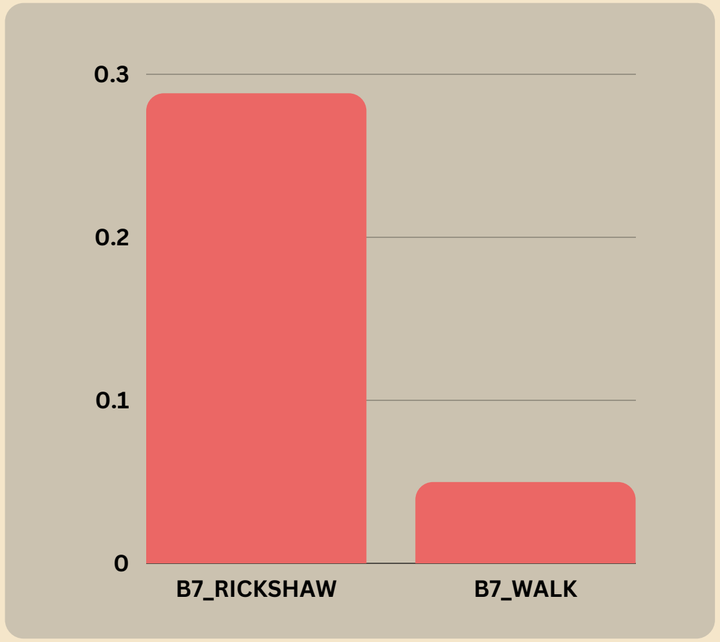
B14_CAR	**1.8**
B14_CYCLE	**-0.665**
B14_RICKSHAW	**-0.324**
B14_TW	**1.22**

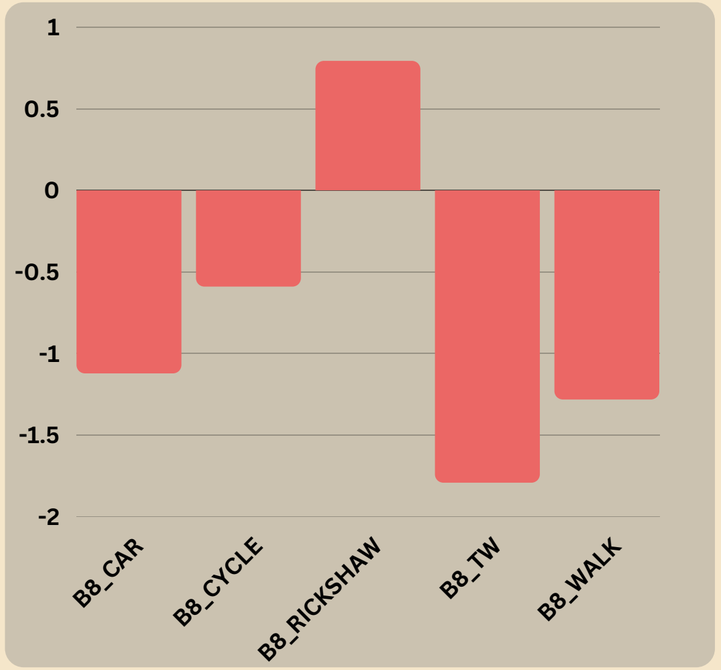


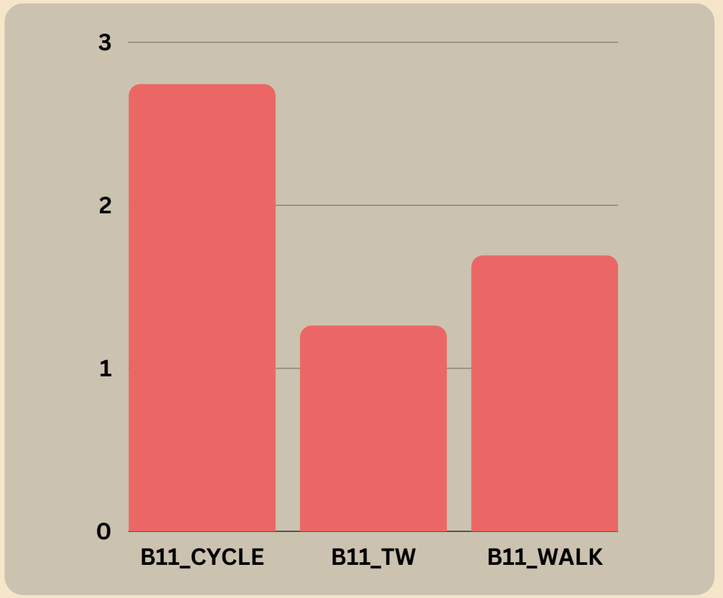

0
-1
-2
-3
ASC_CAR
ASC_CYCLE
ASC_RICKSHAW
ASC_TW


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
B13_CYCLE
B13_RICKSHAW
B13_WALK


3
2
1
0
B10_CAR
B10_CYCLE
B10_TW
B10_WALK


0.3
0.2
0.1
0
B7_RICKSHAW
B7_WALK


1
0.5
0
-0.5
-1
-1.5
-2
B8_CAR
B8_CYCLE
B8_RICKSHAW
B8_TW
B8_WALK


3
2
1
0
B11_CYCLE
B11_TW
B11_WALK

FURTHER SCOPE OF STUDY

- From the utility functions estimated in this study, we can further assess what are the parameters which affect the use of public transportation systems and to what extent. Following which we can also propose action plans to promote Public transport

- Similar modeling can be done for access trips and post main trips for assessing the choice of transport for these types of trips. The Study can be expanded to larger areas as well.

- Other than the type of trips, similar study can be done to create separate models for public transports like buses, metro, taxi, train and private transport like car, bike, etc.

Thank You

Heartfelt gratitude to Dr. Rahul T.M for giving us the opportunity
And Mr. Aaditya Bhamidipati for providing assitance throught the project