CP-302

Capstone Project

Under the guidance of Dr. Rahul T.M.

Department of Civil Engineering
IIT Ropar

Mode Choice Modelling Using Multinomial Logit Models

By:
Kanha Singhania (2019CEB1014)
Manish Mahala (2019CEB1016)
Rahul Matolia (2019CEB1024)
Ruheen Zeba (2019CEB1029)

Introduction

- 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 hypothetical choice experiments to elicit commuters' preferences for various modes of transportation.

Objectives

1. **Literature Reading** : Studying various models which can be used to model the mode choice preferences and selecting the best one which satisfies the requirements for our project.

2. **Data Pre-processing** : Preprocess the collected data and make it suitable as an input to our mathematical model.

3. **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.

4. **Mode Choice Modeling** : The final objective of the project is to model the mode choice preferences of individuals based on the defined parameters.

What is Choice Modeling

Choice Modeling is a scientific method that attempts to model the process that an individual goes through while deciding a choice among many available alternatives. Mathematical models are used to represent these processes.

Using Choice Modeling, we can identify even those factors which influence the final decision of the person, but are outside the conscious awareness of the respondent.

Elements of choice modeling

There are various elements the affects the choices of the users and thus the choice modeling

CHOICE MODELING - ELEMENTS

An individual, group, or institution which has the responsibility to make the decision at hand.

Decision maker

The set of alternatives available to the decision-maker to make a unique choice.
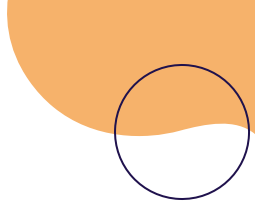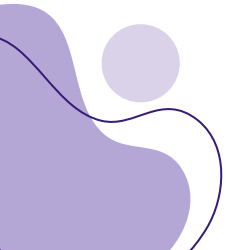
Alternatives

Attributes of Alternative

The values of an alternative based on which the decision-maker makes the decision are called the attributes of that alternative.

A mechanism by which the decision-maker processes all the inputs, and outputs the final choice out of available alternatives.

Decision Rule (Model)

Data Collection

Data can be collected in two ways:

- Data collected by the researchers themselves is called primary data. It is mainly collected from questionnaire surveys.

- Data collected previously by government organizations or others, from sources like household travel surveys etc, is called secondary data.

For our project, we are using secondary data. We are using a sample survey data from the Wadgaon region of Pune.

Site selection - Wadgaon, Pune

1. Wadgaon Budruk also known as Vadgaon in the local area is one of the many upcoming suburbs of Pune, Maharashtra, India. It is located on Sinhgad Road.

2. The locality Wadgaon Budruk falls in Pune district situated in Maharashtra state, with a population 56323. The male and female populations are 29192 and 27131 respectively. The size of the area is about 3.38 square kilometer.

3. The available mode of public transport in the city majorly include buses and rikshaw. The utilization of cars and bikes (as cabs) is not much as compared to them. Still there is a certain chunk of population which utilizes cabs as public transport.
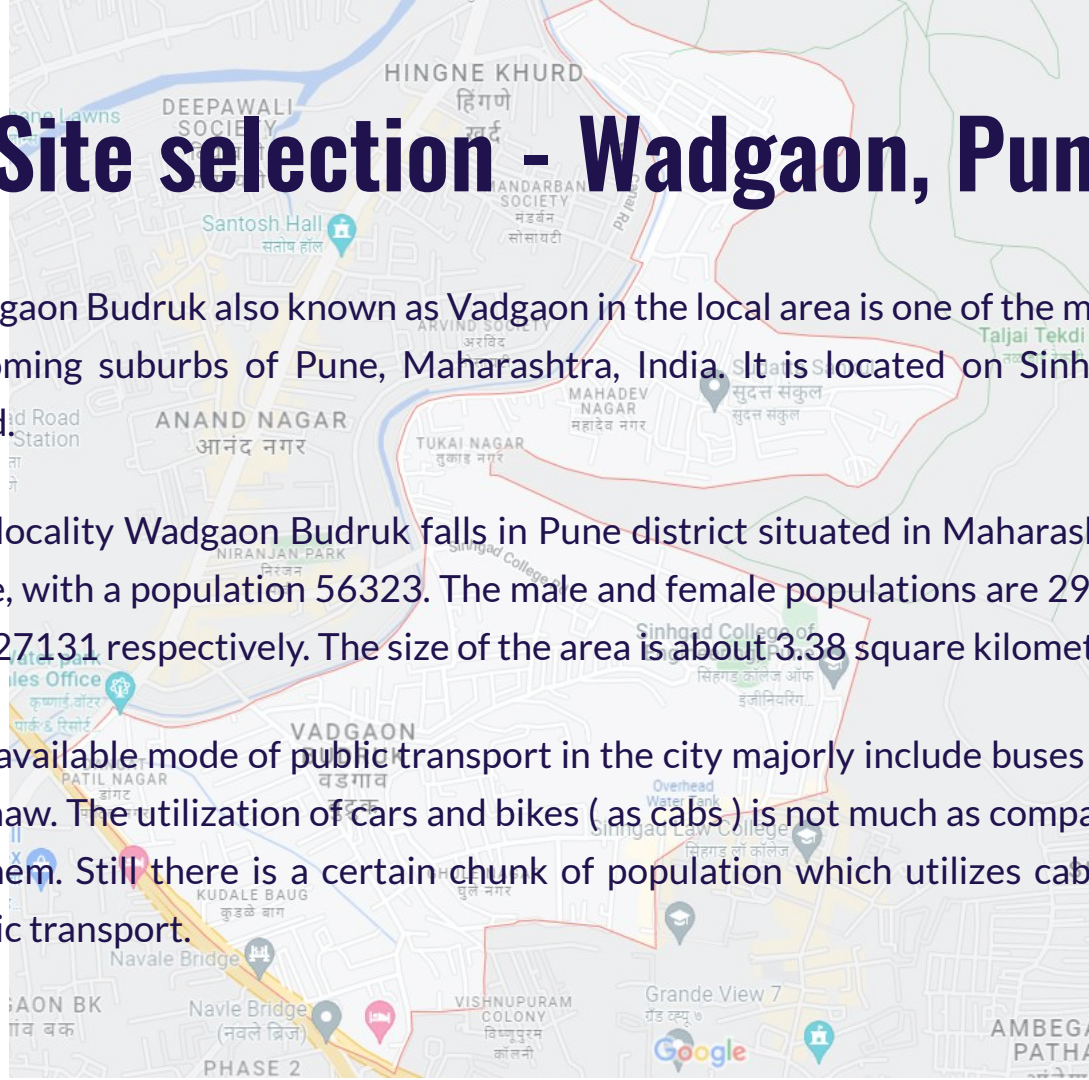
Data Used

We have used a sample data for the Wadgaon district of Pune

- It is a survey data with information spread across 3 different tables.

- The data is collected for the Wadagao region of Pune.

- It contains various columns (features), many of which are redundant to be used for mode choice modeling. Some of these redundant features are - Address, Household Number, Telephone number, Name of Members, Person ID, etc.

Link to Survey Data

Data Preprocessing

- Data preprocessing is one of the most important steps of our project. The collected data has to be preprocessed to make it suitable for feeding as an input to our model.

- Then it can be analyzed to infer important information like relationship of features on each other, redundant features, etc. It can also give insights about fine tuning our model for better prediction accuracy.

Data Preprocessing

- Many rows of the data were incomplete, i.e, the feature information was not present. Here, we either removed such rows (if a huge amount of data was present, and removing some rows did not affect our model), or we filled these empty columns using some statistical measure, such as mean, median, or mode of the available column data.

- Some important features may be present as categorical data. To use them in our model, we need to converte them into numerical data using methods such as one-hot encoding.

Link to Data Spreadsheet

Data Preprocessing

The 3 tables of our data have to be merged into one so that information can be extracted. We have used the common columns, namely Form Number and Household numbers for merging the tables.

- The form number is a unique number which is given to each form which is used to collect survey data.

- The Household number is a unique number given to a household. A household can have more than one individual filling different forms.

Selecting Mode Alternatives

We have selected these four mode choice alternatives:

- Rickshaw
- Scooter
- Car
- Bus

From the available data, we can calculate the average speed of these mode choices. For the rows in which this mode is an alternative, we added the expected travel time using the average speed and distance between travel zones.

Using Supplementary Sheets Data

1. Distance sheet
2. Car Mileage
3. Bus and Rickshaw Fare data

Using Supplementary Sheets Data

Distance:

The "Distance" table contains the information of average trip distance between any two zones. Using this table, we can find the distance of the trip in Km by mapping the distance between zone codes and add the column to our data.

From the travel time and distance values, the average speed during the trip can be calculated.

$$\text{Speed} = \frac{\text{Distance}}{\text{Time}}$$

Using Supplementary Sheets Data

Bus and rickshaw fare data:

When mode choice is a bus or a rickshaw, we cannot use the mileage data to compute the cost, as these are public transportation vehicles, which have non linear fare rates based on the traveled distance.

- For rickshaws, the fare rates in Pune region are Rs 18 for the first Km, and Rs 12.31 per Km thereafter.
- For Bus, we have collected a table which defines the fare for different distance ranges. Using interpolation, we can forecast the cost for each trip based on the distance.

Link to Data Spreadsheet

Using Supplementary Sheets Data

Car Mileage:

When the travel mode choice is a Car or a Scooter, the average mileage can be used to calculate the cost of the trip. To find the average mileage, we have collected the mileage data of top ten most used cars, and took the weighted average.

Using the average mileage(Km/L) and the cost of fuel(Rs/L), we calculated the average cost per kilometer(Rs/Km).

The trip cost is then calculated as:
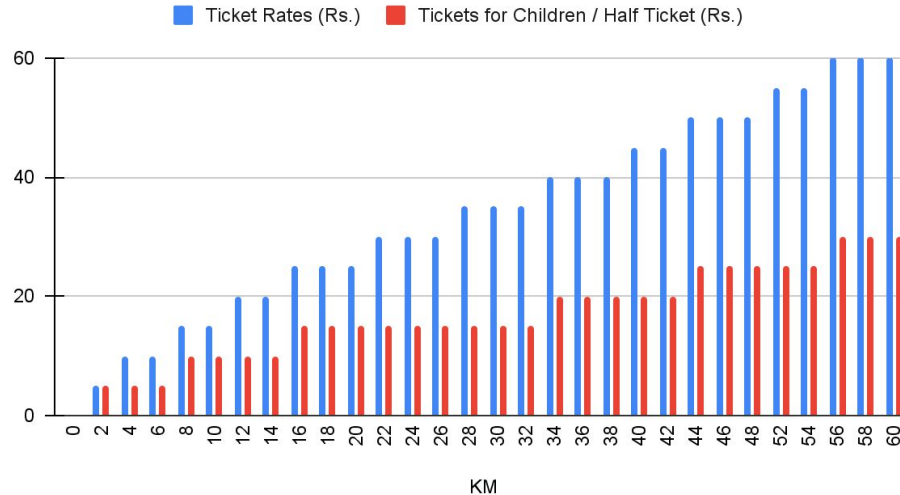
Cost per km = (Cost of fuel) / (Mileage)

Cost = (Distance) x (Cost per Km)

Using Supplementary Sheets Data

Bus fare data:

The graph below shows the distribution of fare rate based on distance:
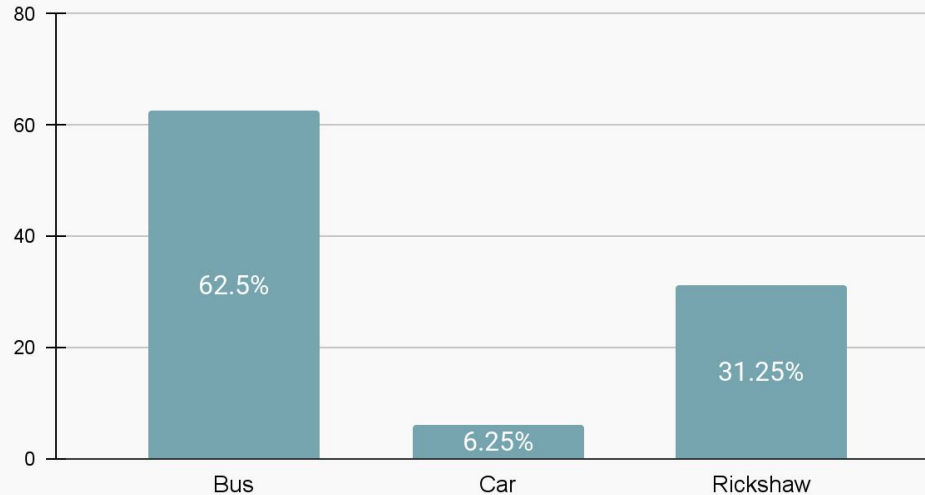
Data Analysis

Out of the available sample data, bus was chosen, 62.5% times, Rickshaw was chosen 31.25% of the times and the lowest chosen vehicle was Car with a percentage of 6.25. But as the sample size is very small, we cannot conclude that Car is rarely preferred as a Travel Mode.
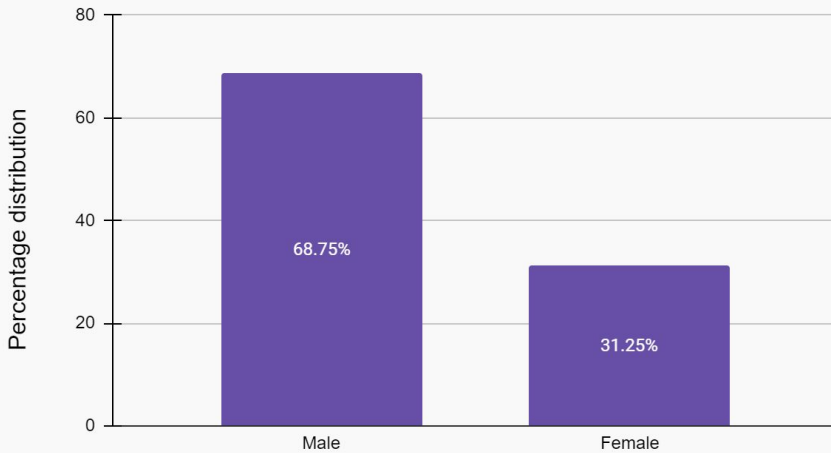
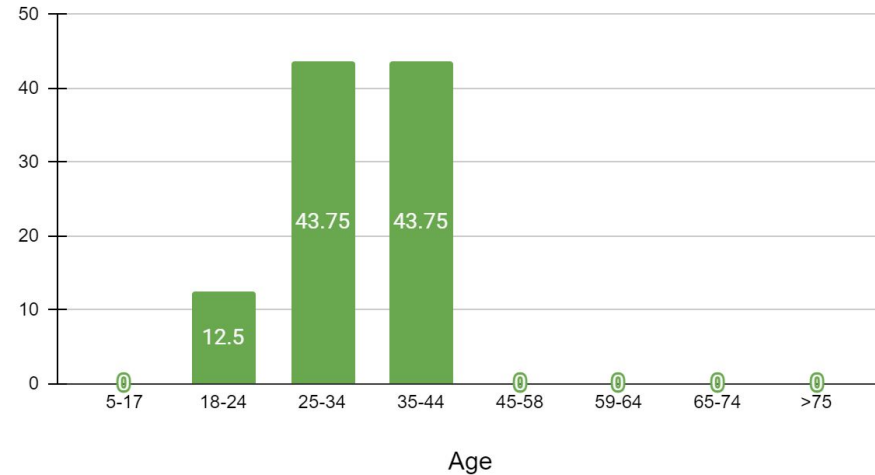Frequency of chosen Travel Modes

Data Analysis

Demographic data

Gender Percentage Distribution
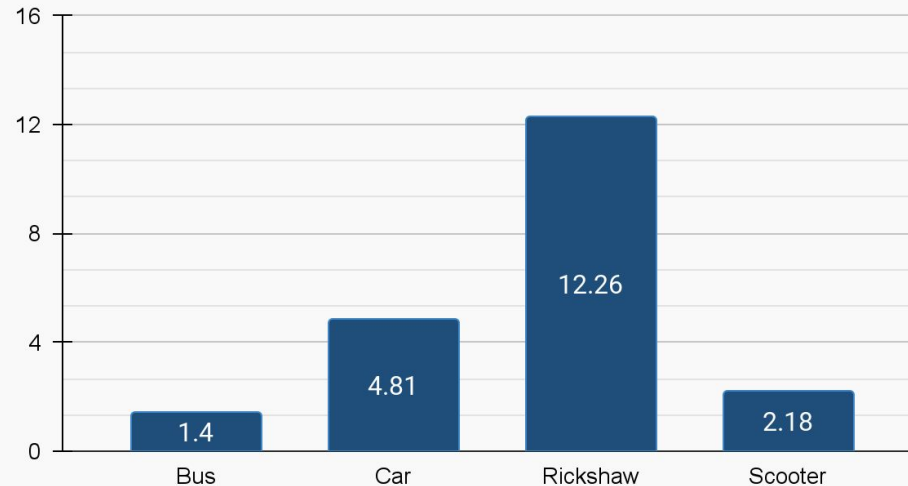

Age Percentage Distribution

Data Analysis

The order of cost per Km of different mode choices is:

Bus < Scooter < Car < Rickshaw

Rickshaw has the highest cost per Km of travel

Average Cost per Km (Rs/Km)

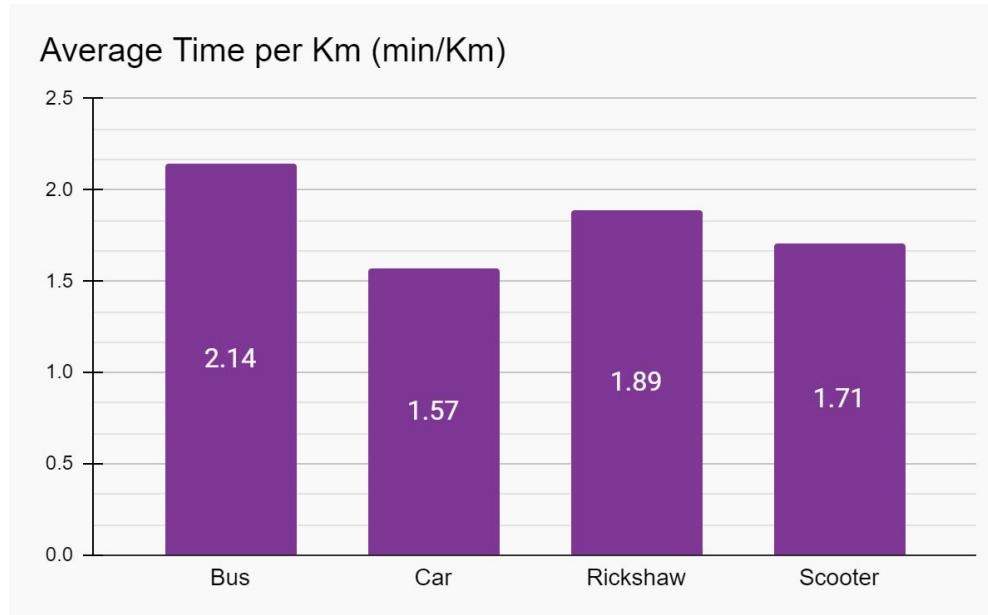| Mode	Cost
Bus	1.4
Car	4.81
Rickshaw	12.26
Scooter	2.18

Data Analysis

The order of average Time per Km of different mode choices is:

Bus > Rickshaw > Scooter > Car

Car takes the least average time per Km of the trip.

Average Time per Km (min/Km)

Bus	Car	Rickshaw	Scooter
2.14	1.57	1.89	1.71

Data Analysis

- Based on the above analysis, we can say that the cost and travel time are among the most important factors that determines the mode choice preference.

- Rickshaw has the highest cost, but has lower travel times per Km as compared to bus. Also due to greater mobility through traffic, people generally choose this mode when travelling lesser distances, and when personal mode of travel is not available.

- Whereas for greater distances, bus is the most preferred mode.

Results

- Using the sample survey data, we have tried to compile a dataset which can be used as an input to a mathematical mode choice model. We have seen how the individual choice preferences vary based on the the cost and travel time tradeoffs.

- Furthermore, we were able to generate some Demographic insights which can be used to arrange our data in suitable orders while extending the calculations to larger data set.

Future Tasks (for Capstone II)

1. Our next goal is to use these techniques that we have learned so far on a large dataset containing thousands of examples of individual mode choices. Such large and refined dataset can be used to make an extremely accurate model.

2. Furthermore, we aim to Predict the mode choice preferences of decision makers using the selected model which is trained on the large preprocessed dataset.