MODE ESTIMATION : TOUR BASED APPROACH

Team

Akhil Johnson (M2)
Melna Jose (M2)
Shanaraj V.K. (M2)
Omkar Deepak Karmarkar (D2)

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY, INDIA

Determinants of Mode choice

- Socio – demographic variables: Age, Gender
  - Women walk and bicycle more while men use car more
  - Car usage increases with age
- Trip related variables: start time, purpose, travel time
  - Large work trips in morning while shopping trips are after that
  - Shopping and leisure trips are of shorter durations and dominated by specific modes
- Activity variable: Stay time
- What role does tours' attributes play?

Details of PT Data

- A total of 1.46 Million data points (trips)
- More than 511,000 unique 'trip chains' in the data
  - Corresponds to set of trips per person in a day (3 AM to 11:59 PM)
- Variable details:
  - Socio - economic variables: Age (groups of 5 years) and Gender
  - Trip related variables: Purpose, start time, end time, mode
  - Activity characteristic: Stay time at trip destination

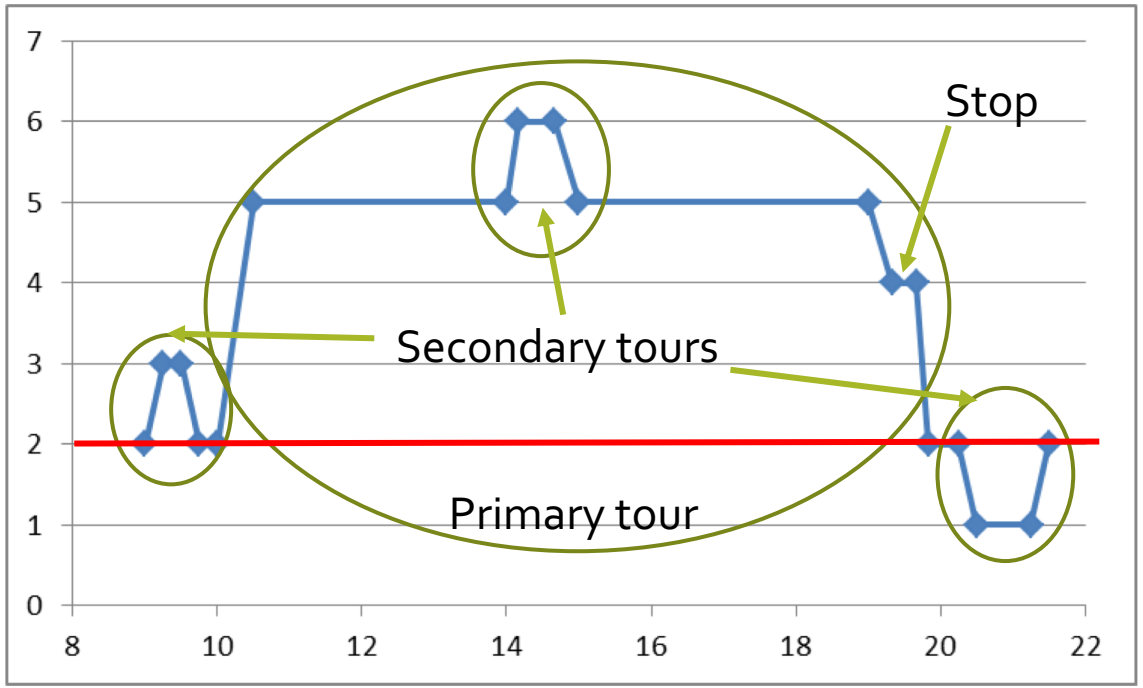
Trip Chain and its Pattern

- Trip chains - unique numbers correspond to series of trips performed by a person in a day
- Are all of them a single tour? - Not necessary
- How are we defining a tour then?

Time	Place
9:00:00 AM	Home
9:15:00 AM	Shop 1
9:30:00 AM	Shop
9:45:00 AM	Home
10:00:00 AM	Home
10:30:00 AM	Work
2:00:00 PM	Work
2:10:00 PM	Lunch
2:40:00 PM	Lunch
3:00:00 PM	Work
7:00:00 PM	Work
7:20:00 PM	Shop 2
7:40:00 PM	Shop 2
7:50:00 PM	Home
8:15:00 PM	Home
8:30:00 PM	Dinner
9:15:00 PM	Dinner
9:30:00 PM	Home

Trip Chain Pattern, Activity & Tour

	Time	Place	
9	9:00:00 AM	Home	2
9.25	9:15:00 AM	Shop 1	3
9.5	9:30:00 AM	Shop	3
9.75	9:45:00 AM	Home	2
10	10:00:00 AM	Home	2
10.5	10:30:00 AM	Work	5
14	2:00:00 PM	Work	5
14.167	2:10:00 PM	Lunch	6
14.667	2:40:00 PM	Lunch	6
15	3:00:00 PM	Work	5
19	7:00:00 PM	Work	5
19.33	7:20:00 PM	Shop 2	4
19.66	7:40:00 PM	Shop 2	4
19.833	7:50:00 PM	Home	2
20.25	8:15:00 PM	Home	2
20.5	8:30:00 PM	Dinner	1
21.25	9:15:00 PM	Dinner	1
21.5	9:30:00 PM	Home	2



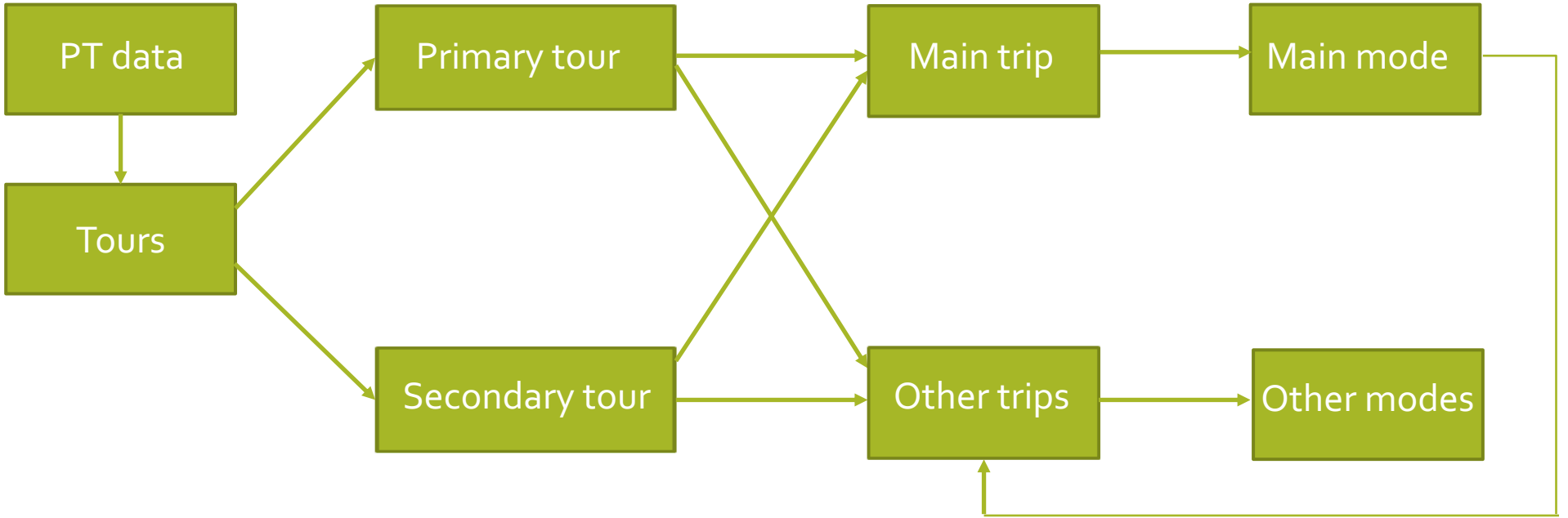
Understanding the tour

1. Tour type - Primary tour or secondary tour
2. Stop pattern - tour without stop v/s with one stop v/s multiple stops
3. Tour's main purpose - work or shopping or socializing and leisure
4. Trip in the tour - main trip v/s return trip v/s intermediate trip
5. Trip characteristics:
   - Start time - morning peak v/s evening peak v/s off peak
   - Intermediate trip purpose - shopping or leisure

Multinomial Logistic Regression

- The most widely used mathematical form for choice probabilities in behavioural travel demand analyses
- *Key Strengths:*
- Simple to perform
- Computationally efficient
- Permits a simple behavioural interpretation of its parameters
- *Key Weaknesses:*
- Independence of Irrelevant Alternatives (IIA) property
- No correlation between error terms (i.i.d. errors)
- Random taste variation can not be represented

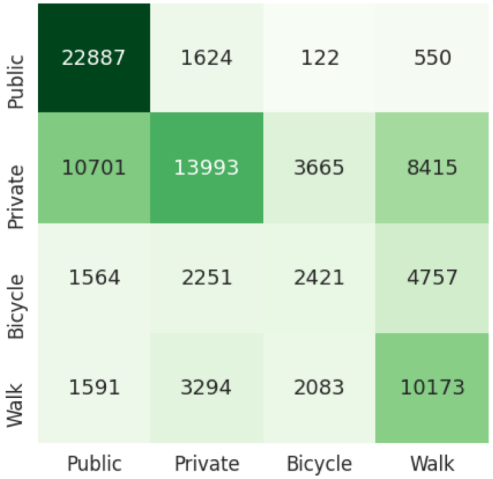
Multinomial Logistic Regression



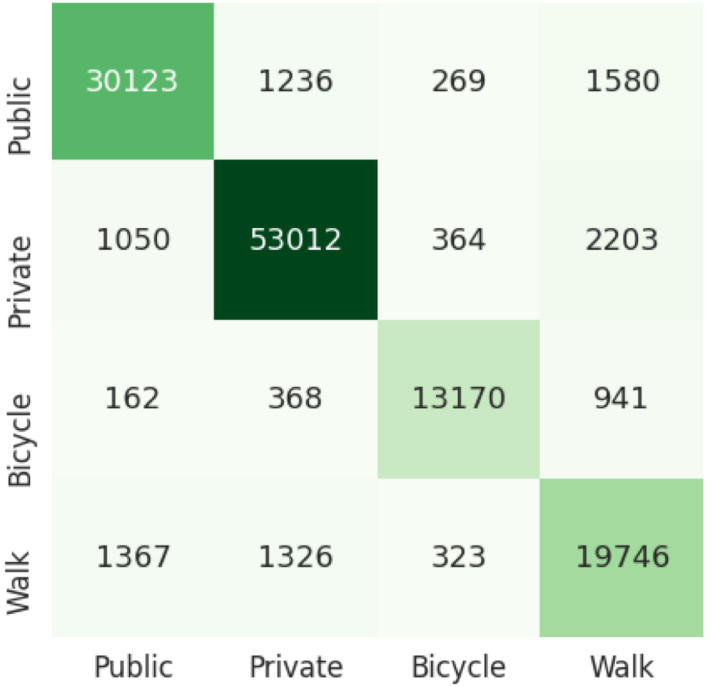
Multinomial Logistic Regression

Primary Multinomial Logit	
Pseudo R^2	0.31
Log-Likelihood	-472620
Log-Likelihood null	-684900

Secondary Multinomial Logit	
Pseudo R^2	0.72
Log-Likelihood	-274610
Log-Likelihood null	-970730



Primary (main trip)



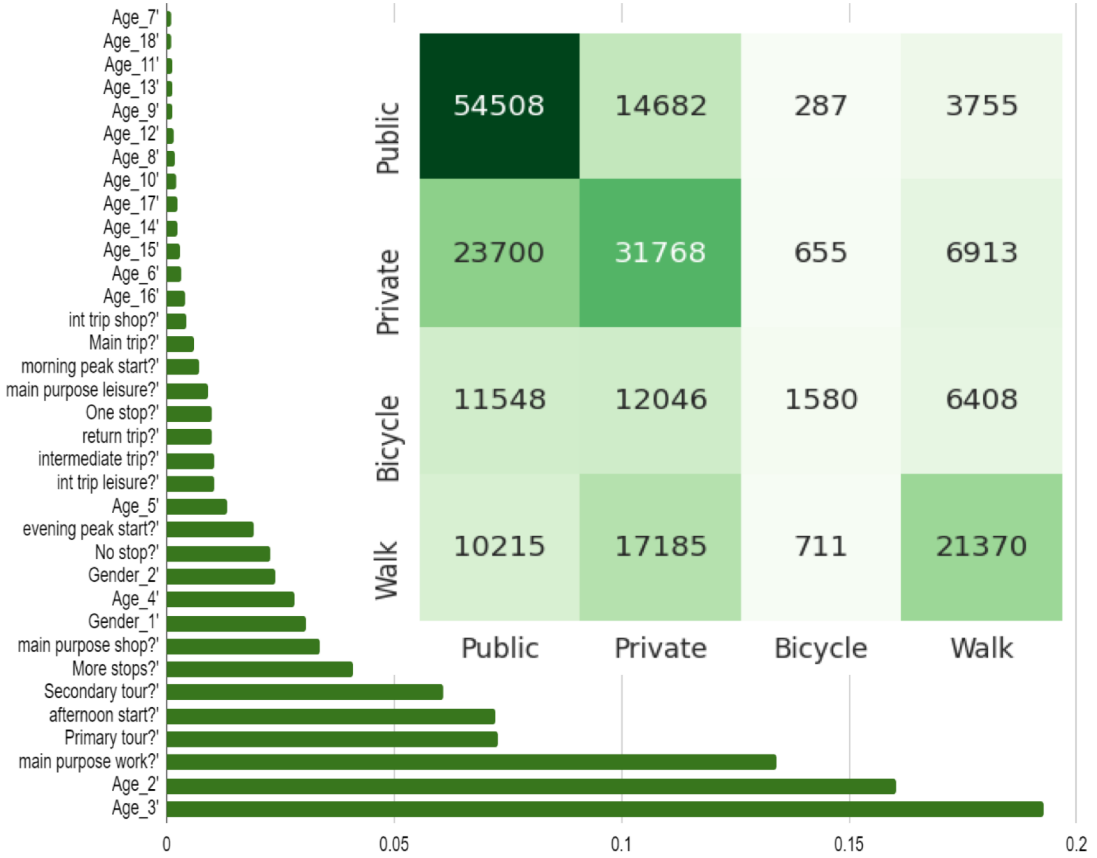
Secondary

Random Forest

- A decision tree is a decision support tool that uses a tree-like graph or model of decisions.
- RFs train each tree independently, using a random sample of the data.
- *Key Strengths:*
- RF is much easier to tune. Only two hyperparameters (a) depth of trees (10) and (b) number of estimators (20).
- More robust than a single decision tree, and less likely to over fit on the training data.
- Does not require preparation of the input data.
- Works with unscaled data and missing values.
- Provides information on Feature importance.
- *Key Weaknesses:*
- For data including categorical variables with different number of levels, RFs are biased against attributes with more levels.

Random Forest

	precision	recall	f1-score
Public	0.55	0.74	0.63
Private	0.42	0.5	0.46
Cycle	0.46	0.06	0.11
Walk	0.55	0.44	0.49
accuracy			0.5
macro avg	0.5	0.44	0.42
weighted avg	0.5	0.5	0.47

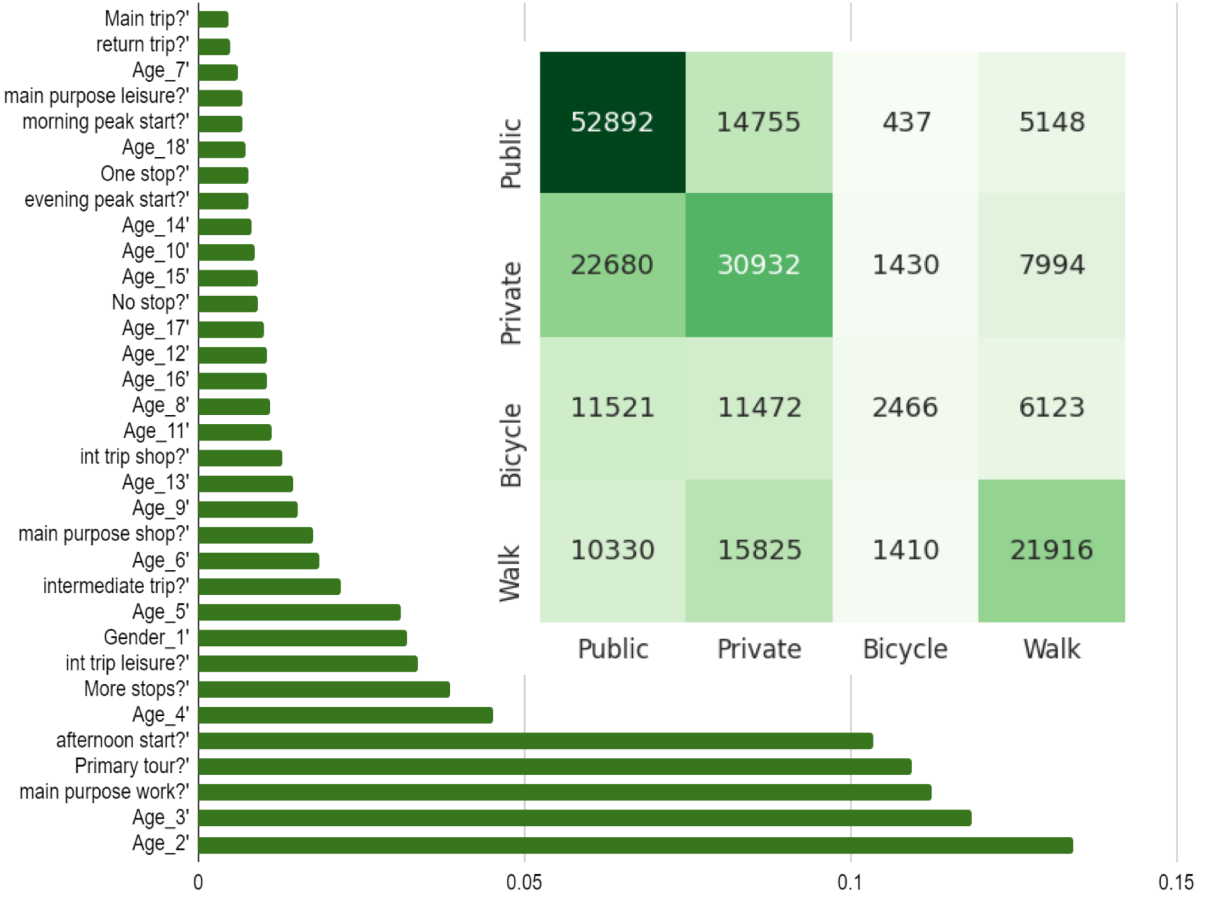


XG Boost

- XGB build trees one at a time, where each new tree helps to correct errors made by previously trained tree.
- *Key Strength:*
- It performs the optimization which makes the use of custom loss functions much easier.
- Boosting focuses on unbalanced datasets by strengthening the impact of the positive class.
- Provides information on Feature importance.
- *Key Weaknesses:*
- Training generally takes longer because of the fact that trees are built sequentially.
- XGB is harder to tune than RF.

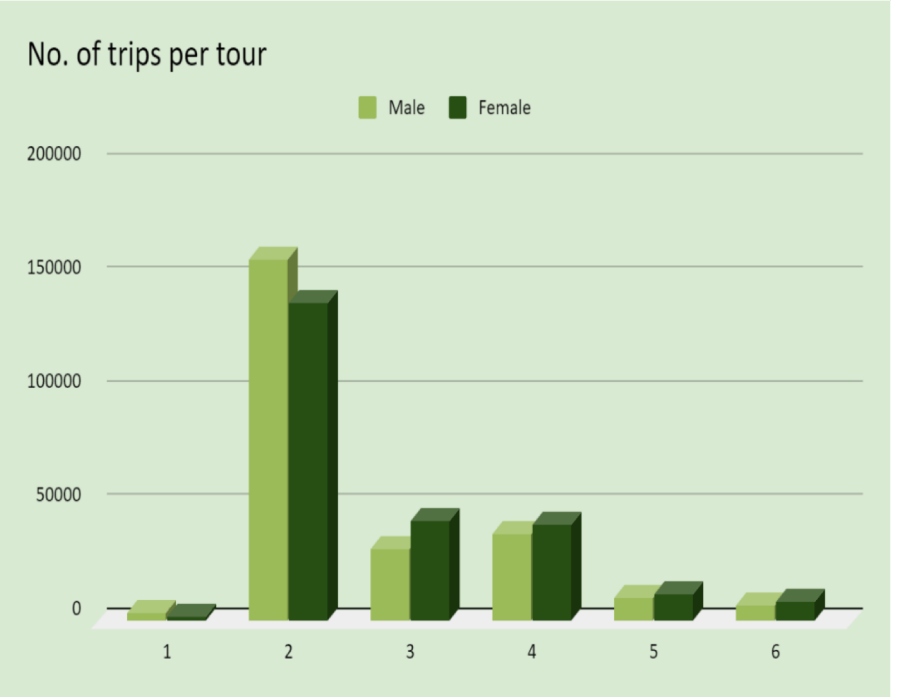
XG Boost

	precision	recall	f1-score
Public	0.54	0.72	0.62
Private	0.42	0.49	0.45
Cycle	0.43	0.08	0.13
Walk	0.53	0.44	0.48
accuracy			0.5
macro avg	0.48	0.43	0.42
weighted avg	0.49	0.5	0.47



Potential Use & Policies

- Potential use
  - primary mode detection and future mode change prediction for same tour
  - Usage of home check - in for mode prediction through time of the day and purpose
  - Baseline framework for activity based model (can be enhanced with companionship and temporal data)
- Policies
  - Promoting multiple trips per hour – reduced carbon emission
  - Decrement of exclusive shopping tours in future



Thank You!