UNIVERSITY OF WOLLONGONG AUSTRALIA

Developing Trip Distribution using High-Frequency Proxy Data

Thivya P Amalan

Supervisors:

Dr Bo Du

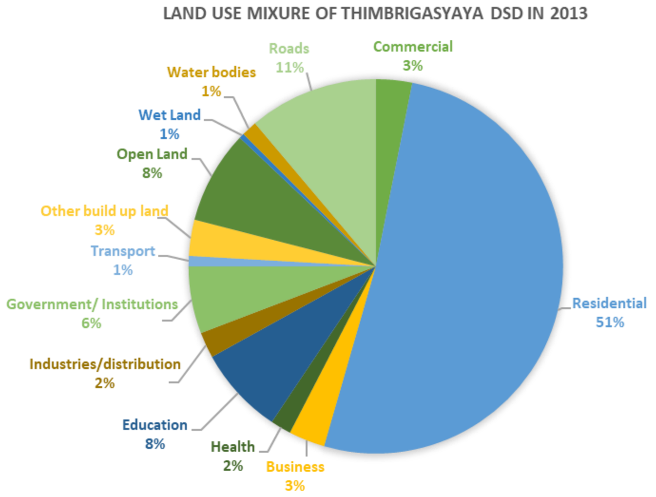
Introduction

- As the first and essential step of the classic four-step transport planning model, trip generation aims to predict the number of trips generated from each traffic zone.
- Cross classification and Multiple Linear Regression (MLR) are widely used to estimate trip generation.
- Data availability is one of the most critical constraints when estimating independent variables for traditional trip generation models.
- In many cases, traditional trip generation models require survey data, which can be time-consuming and expensive.
- There has been an increase in interest in developing trip generation models using alternative data sources and methods in recent years.
- As an alternative to traditional data collection methods such as the Household Travel Survey (HTS), High Frequency Data (HF data) offers several advantages.

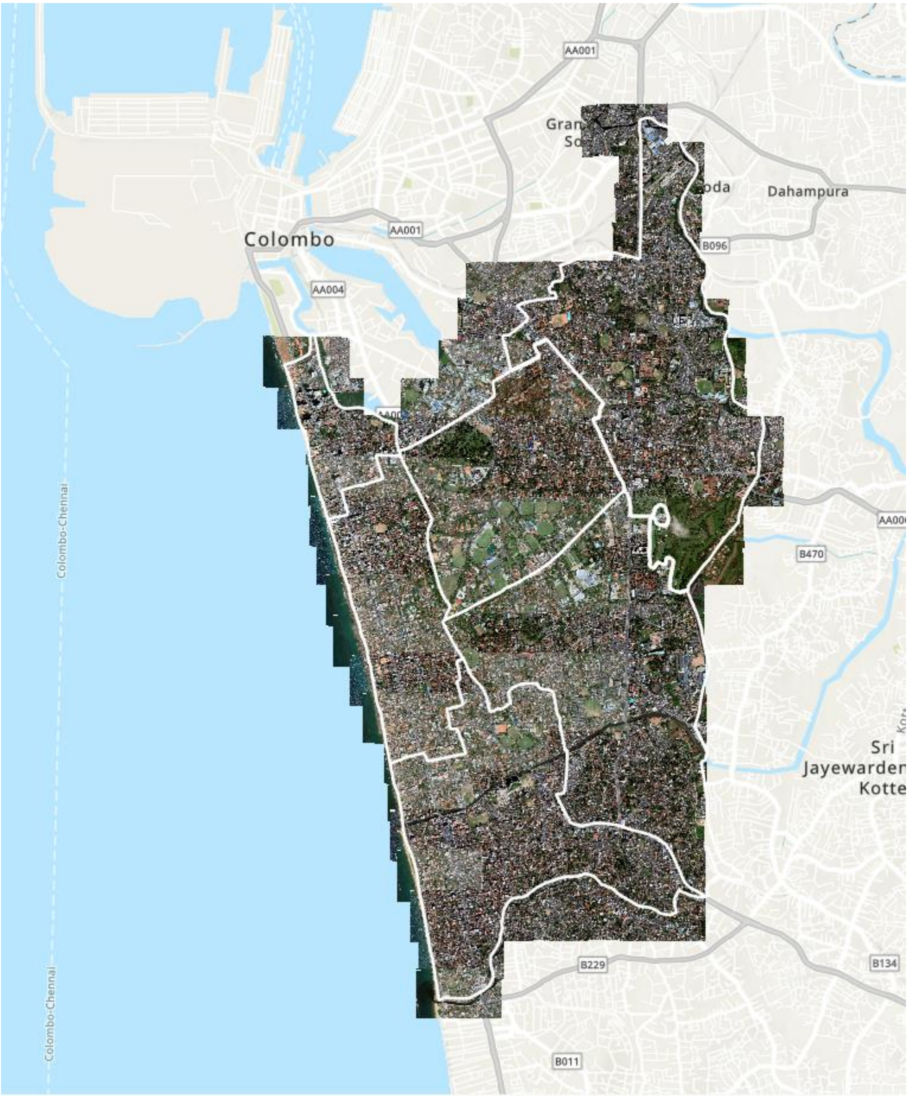
Thimbirigasyaya DSD

- Area 22km^2 (in 2012 50% of land is utilised by residential purpose)

- Population 236,903 (2012)

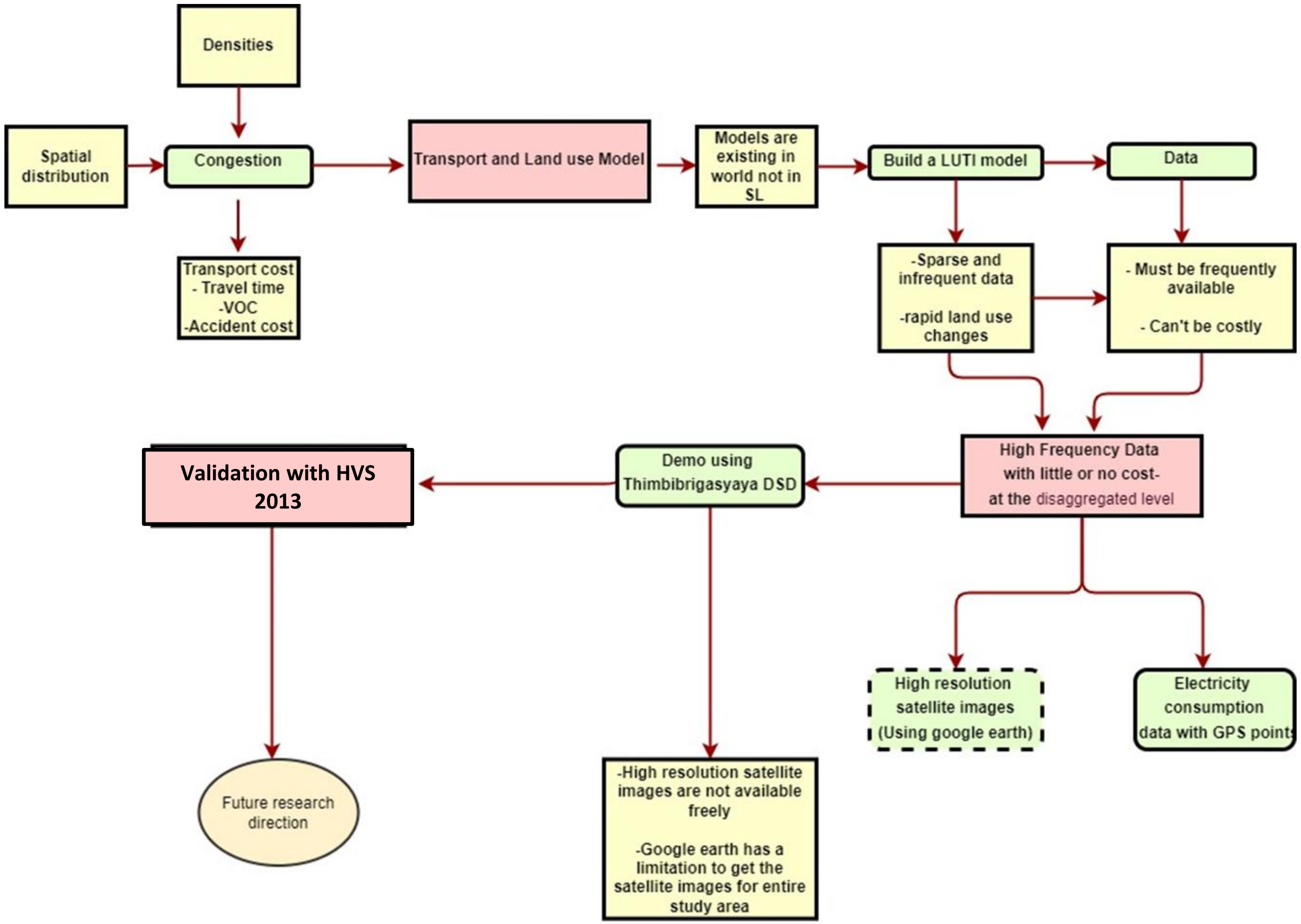


Source: CoMTrans, 2013



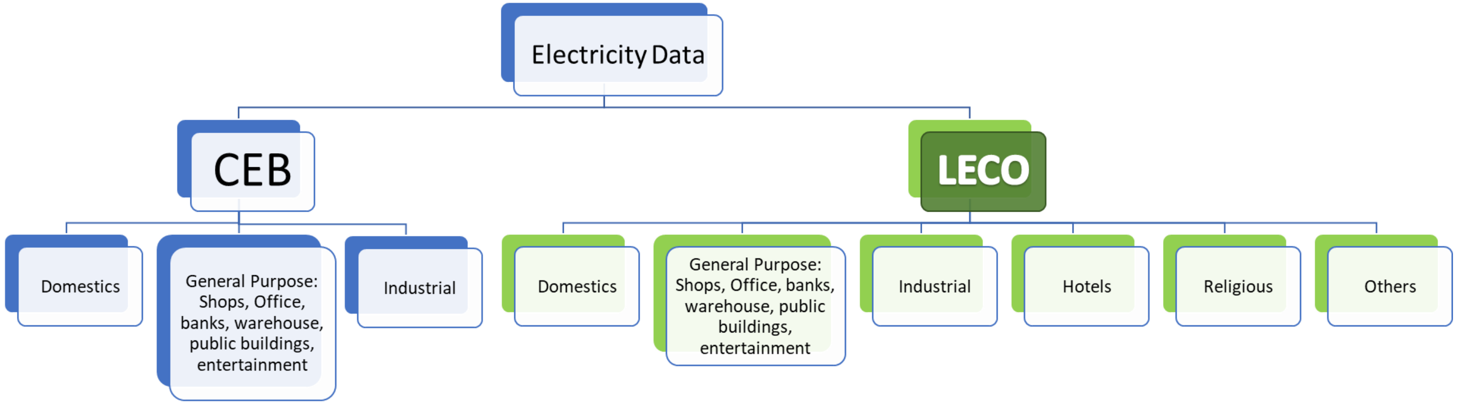
Research Gap and Overall Methodology

Using the available HF data such as household monthly electricity consumption, GPS locations of each household, and satellite image data, this study attempts to develop a home-based work trip generation model based on fuzzy logic, where the MLR model shows very low precision (R^2 is 8%),



Data Collection

- Electricity Consumption data for Oct 2013 and Oct 2019 with GPS

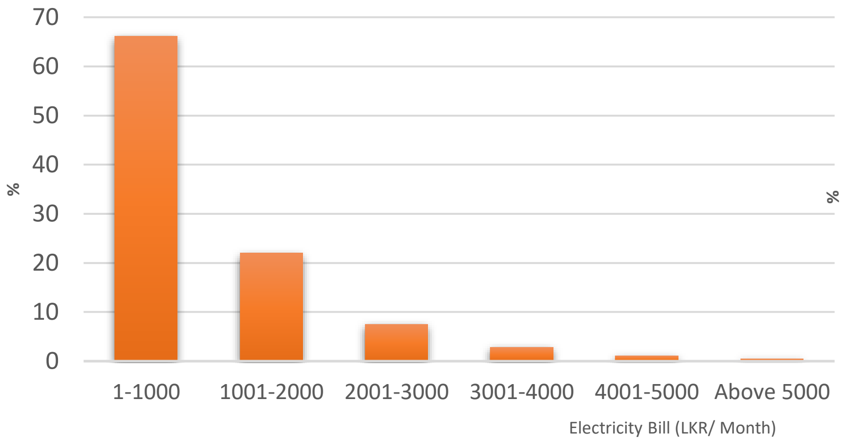


- Google Earth Images for 2013 and 2019

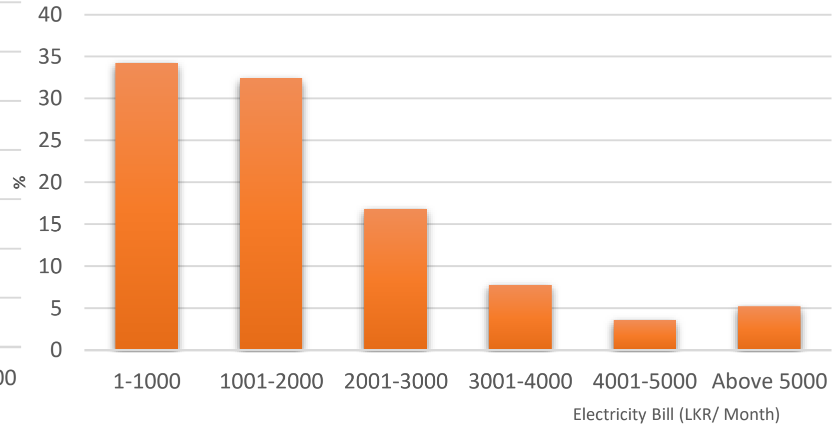
- Home Visit Survey (HVS) 2013 Survey data for validation

Income vs Electricity Based on HVS 2013

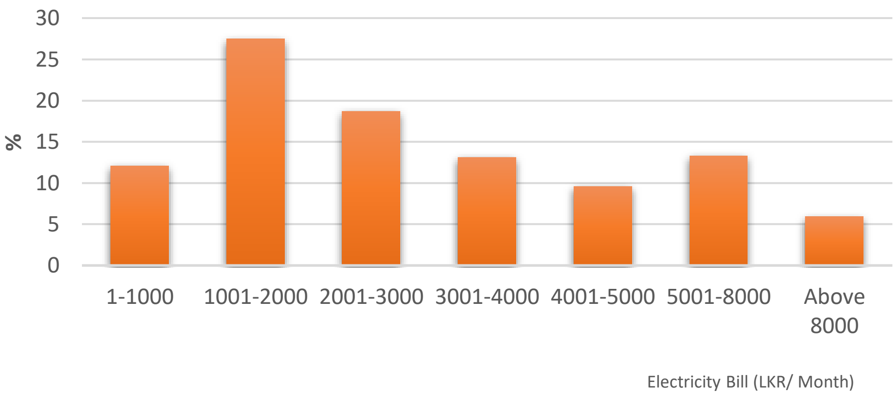
Income Level	Mean	Stand. Error	Avg Electricity unit (kWh)
Low	731.4	3.9	78
Middle	1347.8	8.02	97
High	2053.1	16.9	118



Electricity Bill for a Month Low Income



Electricity Bill for a Month Middle Income



Electricity Bill for a Month High Income

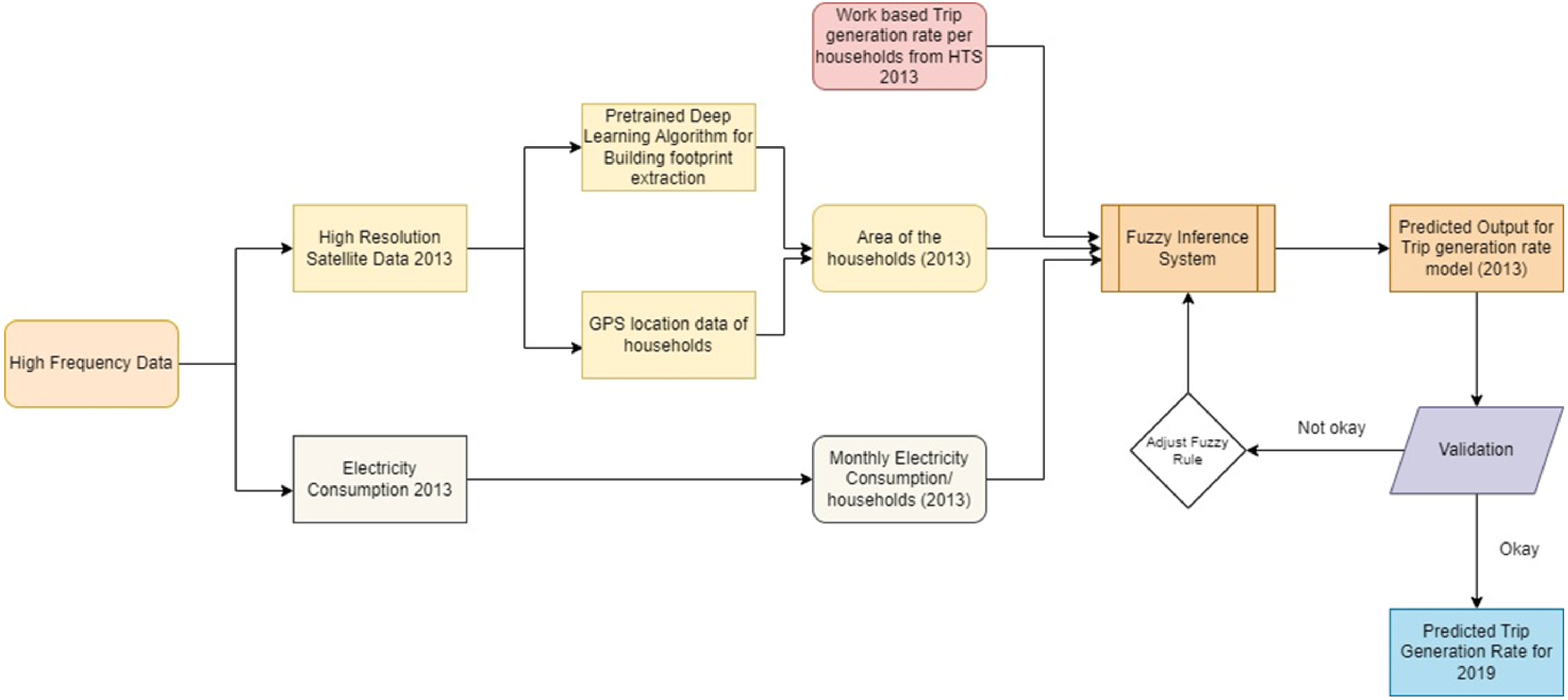
Low income: Household monthly income is less than 40,000 Sri Lankan Rupees (LKR)
Middle income: Household monthly income is between 40,000 to 60,000 LKR
High income: Household monthly income is above 60,000 LKR

Validation of Electricity Domestic Customers' GPS Location Data

Location	Total HHS from Census 2013	Total HHS from GPS points 2013	%
Colombo District	572,475	553,713	96.7
Thimbirigasyaya DSD	52,285	53,628	102.6

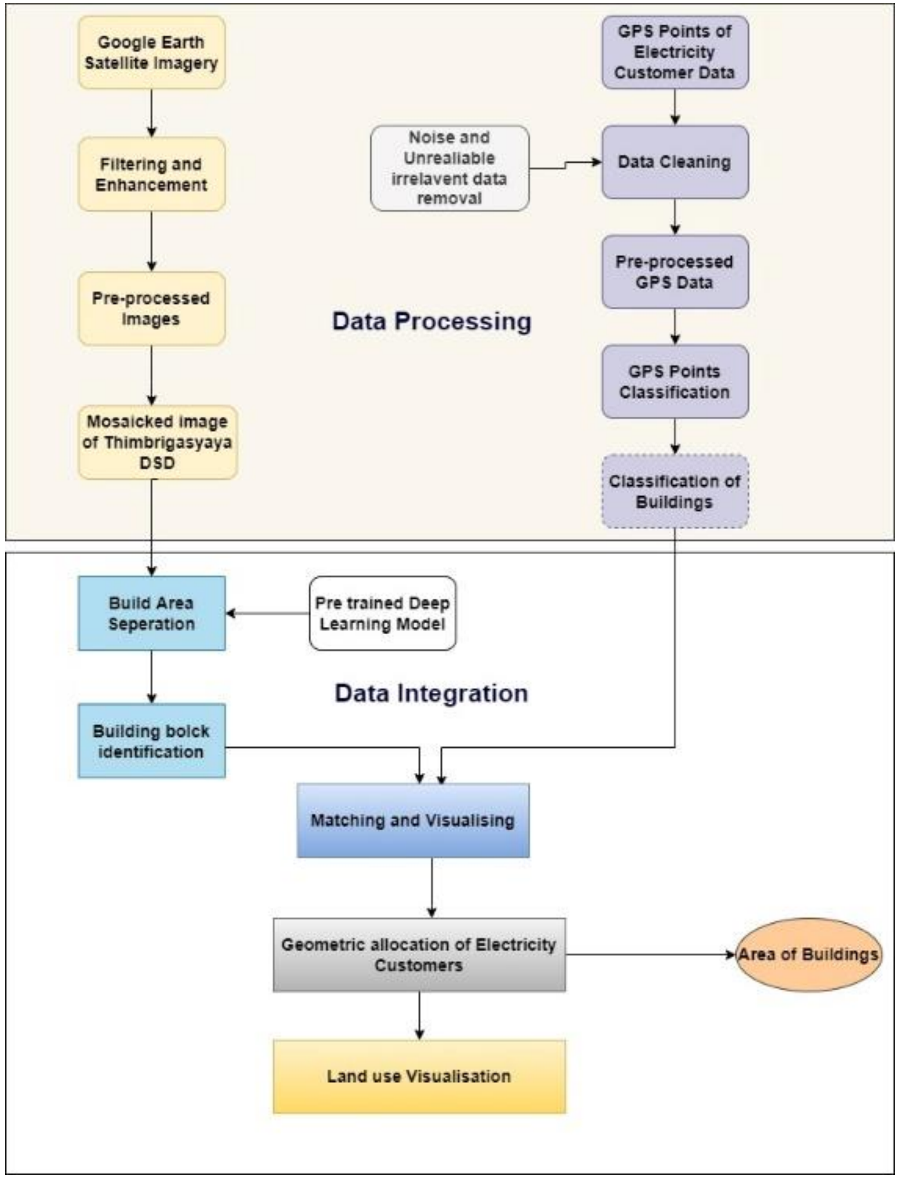
Detected Buildings Using Deep Learning

1. Extracting a google earth image and geo referencing
2. Detect footprints of buildings using deep learning method and map matching with Electricity GPS data
3. Developing fuzzy theory based on the extracted data



Methodology of Building Footprint Extraction and Map Matching of GPS

- Used a pre-trained deep learning model developed for the United States building footprint extraction
  - This model used the Mask RCNN model architecture implemented using ArcGIS API for Python.
  - this model requires 8-bit, 3-band high resolution (10-40 cm) imagery
  - has an average precision score of 71.8 %.
- Due to the limitation and high cost of acquiring high-resolution satellite imagery for Sri Lanka, the study uses freely available GE images (high resolution)
- Images are retrieved manually from Google Earth Pro 7.3.4.
- Images of 2013 are gathered between Dec 2012 to Jan 2014, and 2019 are acquired between Dec 2018 to Jan 2020 due to the cloud coverages and availability of historical satellite images.
- Total of 88 images at 760 m eye elevation levels were collected for Thimbirigasyaya DSD for each year. These images are geo-referenced and mosaicked using Arc GIS Pro software.



Detected using USA model – Deep Learning

Category	GPS points of building 2013	Detected Building 2013-DL
Households	57,141	24,397 (45%)
General Purpose	6,448	2769 (42.94%)
Industrial	21	16 (76.19%)

Category	GPS Points	Detected using DL	%
Households- 2013	57,141	24,397	45
Households-2019	60,306	24,608	45.88

Developing fuzzy theory based on the extracted data

- Research used Mamdani to develop working trips.

- Mamdani systems used as an inference system to develop the fuzzy model for working trips, which is well suited for human inputs and more versatile to handling complex fuzzy relationships between inputs and outputs. .

- The fuzzy rules are controlled by AND operation and OR operation and a decision tree was used to determine the relationship between these variables, the parameters range, and fuzzy rules.

- The result from the fuzzy logic for 2013 shows that the predicted value for home based work from the fuzzy model is closer to the relevant value calculated from the HTS data. Based on the output from the fuzzy model trip rate and the total households detected from the deep learning model, the home based work trips were calculated for the entire households using the below formula,

$$TG_{Zone\ i} = \left(\frac{TG_{Fuzzy\ i}}{\sum_i HHS_{DL}}\right) X \sum_i HHS_{Total}$$

wherein

$TG_{Zone\ i}$ = total trip generation for a zone i

$TG_{Fuzzy\ i}$ = total trip generation calculated fuzzy model for zone i

$\sum_i HHS_{DL}$ = total number of households detected by deep learning modal for the zone i

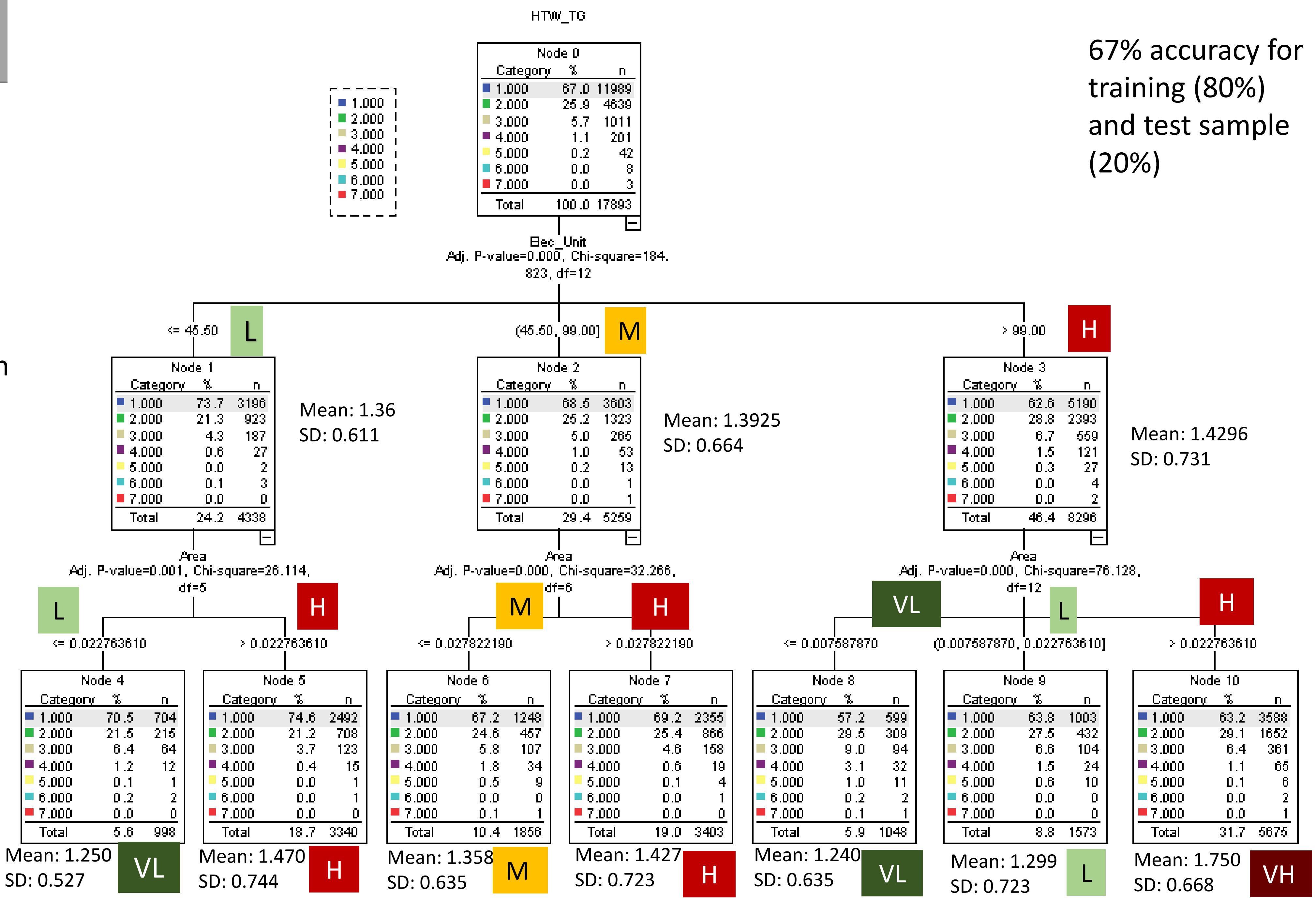
$\sum_i HHS_{Total}$ = total number of GPS points of HHS from electricity data for zone i

Decision Tree: HVS 2013 data

67% accuracy for training (80%) and test sample (20%)

Electricity Unit/Month

Area (Ha)



Fuzzy Rules

Electricity Unit/ Month	Criteria
<45.5	Low
45.5-99	Medium
>99	High

Area of HHS (in Ha)	Criteria
<0.0075	Very Low
0.0075-0.0227	Low
0.0227-0.0278	Medium
>0.0278	High

Rule 1: if the area of HHS is VL AND the electricity unit is L, THEN TG is VL
Rule 2: if the area of HHS is L AND the electricity unit is L, THEN TG is VL*
Rule 3: if the area of HHS is M AND the electricity unit is L, THEN TG is L*
Rule 4: if the area of HHS is H AND the electricity unit is L, THEN TG is H
Rule 5: if the area of HHS is VL AND the electricity unit is M, THEN TG is VL*
Rule 6: if the area of HHS is L AND the electricity unit is M, THEN TG is L*
Rule 7: if the area of HHS is M AND the electricity unit is M, THEN TG is M
Rule 8: if the area of HHS is H AND the electricity unit is M, THEN TG is H
Rule 9: if the area of HHS is VL AND the electricity unit is H, THEN TG is VL
Rule 10: if the area of HHS is L AND the electricity unit is H, THEN TG is L
Rule 11: if the area of HHS is M AND the electricity unit is H, THEN TG is M*
Rule 12: if the area of HHS is H AND the electricity unit is H, THEN TG is VH
Rule 13: if the area of HHS is zero OR the electricity unit is zero, THEN TG is zero

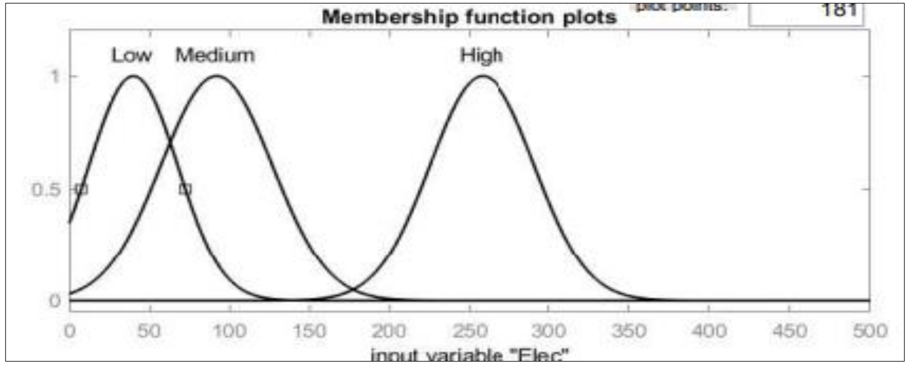
		Electricity Unit/ Month		
		L	**M**	**H**
Area of HHS	**VL**	VL	VL*	VL
	L	VL*	L*	L
	M	L*	M	M*
	H	H	H	VH

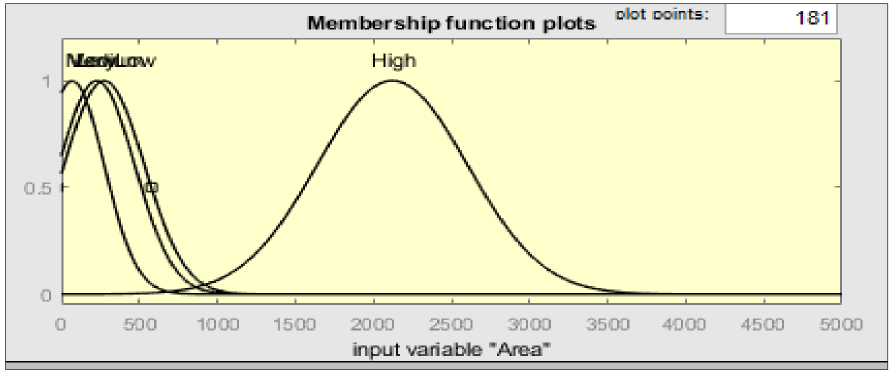
* Assumed rules

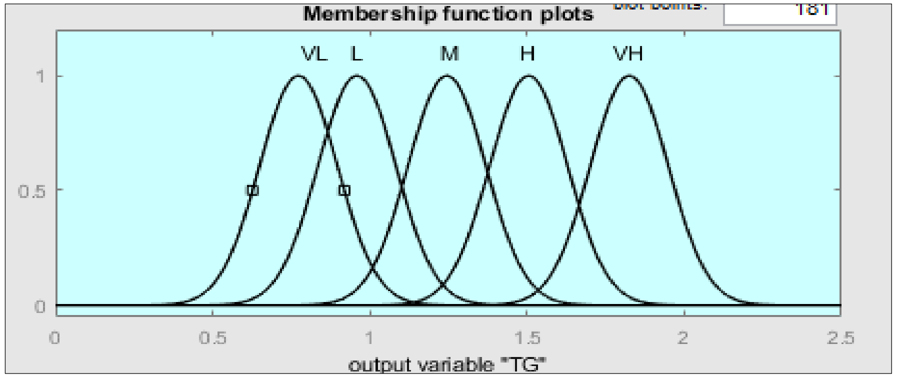
Brief Description of Mean and Standard Deviation (SD)

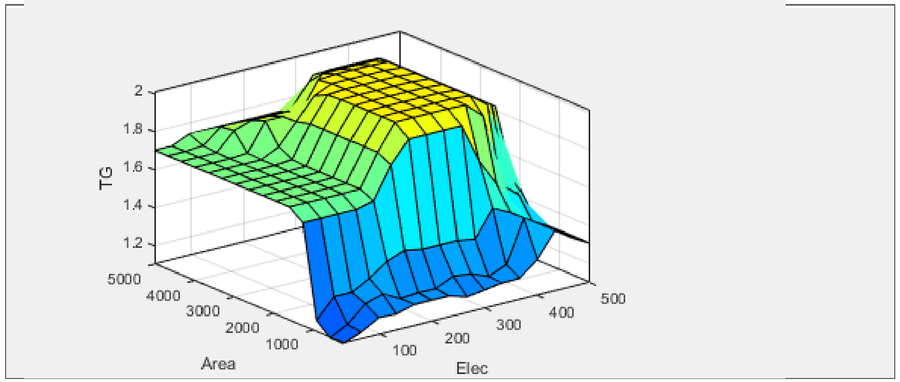
Variables	Classification	Mean	Standard Deviation
Electricity Unit (kWh)	Low	40	27.6
	Medium	91.94	34.7
	High	258.1	33
Area of Households (m^2)	Very Low	70	205
	Low	227	243
	Medium	274.7	257
	High	2120	4805
Trip Generation (Trip/HHS)	Very Low	0.8354	0.124
	Low	1.085	0.124
	Medium	1.344	0.124
	High	1.64	0.124
	Very High	1.931	0.124

Fuzzy Logis Model Input and Output





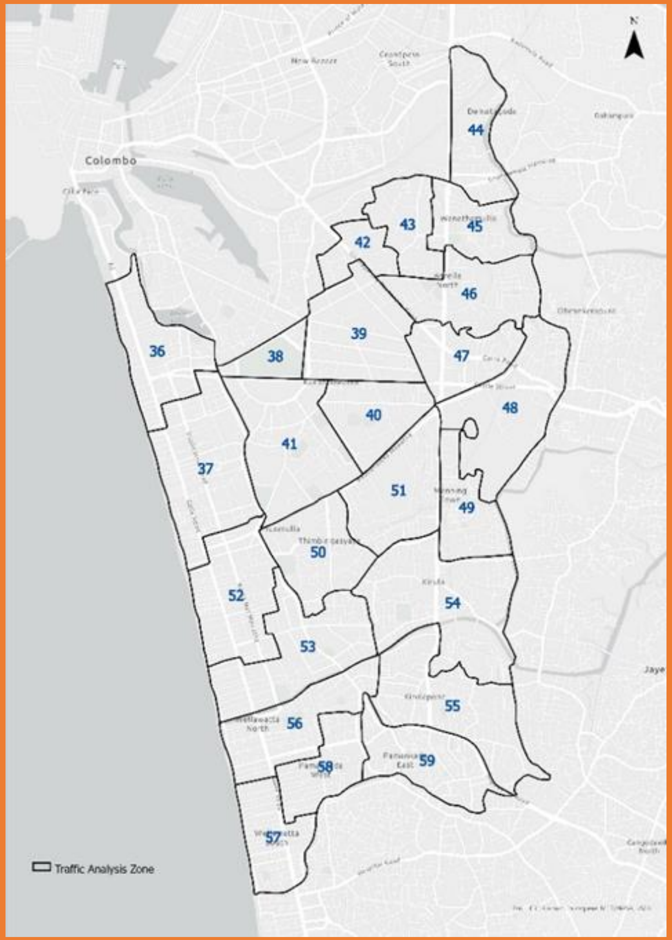




Data Used:
Area from Deep learning and Electricity consumption

Source	HTW	No of HHS	HTW_TG/HHS
HVS-2013	69,790	63,746	1.0947
Fuzzy Model- 2013	26,400.59	24,397	1.0821
Aggregated Model - 2013	61,917	53,629	1.15
Fuzzy Model-2019 (Predicted)	28,436	24,608	1.155

Fuzzy Model Output for 2013 and 2019



Source	HTW	No of HHS	HTW_TG/HHS
HTS-2013	69,790	63,746	**1.0947**
Fuzzy Model- 2013	26,401	24,397	**1.0821**
Fuzzy Model-2019 (Predicted)	31,376	29,468	1.0647

TAZ	TG (Fuzzy)	Total HHS Detected	Trip Rate/HHS	Total HHS in CEB	% Detected HHS	TG for the whole HHS	Trip Rate/HHS (Population)
36	522	554	0.9429	1232	45%	1162	0.9429
37	1273	1188	1.0718	2589	46%	2775	1.0718
38	108	87	1.2394	156	56%	193	1.2394
39	996	839	1.1870	1447	58%	1718	1.1870
40	389	344	1.1305	552	62%	624	1.1305
41	345	197	1.7496	477	41%	835	1.7496
42	367	748	0.4911	1375	54%	675	0.4911
43	670	854	0.7843	2255	38%	1769	0.7843
44	906	809	1.1201	2873	28%	3218	1.1201
45	1185	1102	1.0756	3298	33%	3547	1.0756
46	1104	980	1.1261	2562	38%	2885	1.1261
47	682	620	1.1004	1220	51%	1342	1.1004
48	742	678	1.0943	1432	47%	1567	1.0943
49	910	838	1.0857	2314	36%	2512	1.0857
50	818	740	1.1050	1318	56%	1456	1.1050
51	597	522	1.1441	1256	42%	1437	1.1441
52	1485	1307	1.1362	2749	48%	3123	1.1362
53	1467	1302	1.1269	3144	41%	3543	1.1269
54	2394	2115	1.1319	4841	44%	5480	1.1319
55	2368	2207	1.0728	4961	44%	5322	1.0728
56	1877	1629	1.1523	3859	42%	4447	1.1523
57	1681	1520	1.1058	3820	40%	4224	1.1058
58	1855	1695	1.0943	3695	46%	4043	1.0943

Thank you