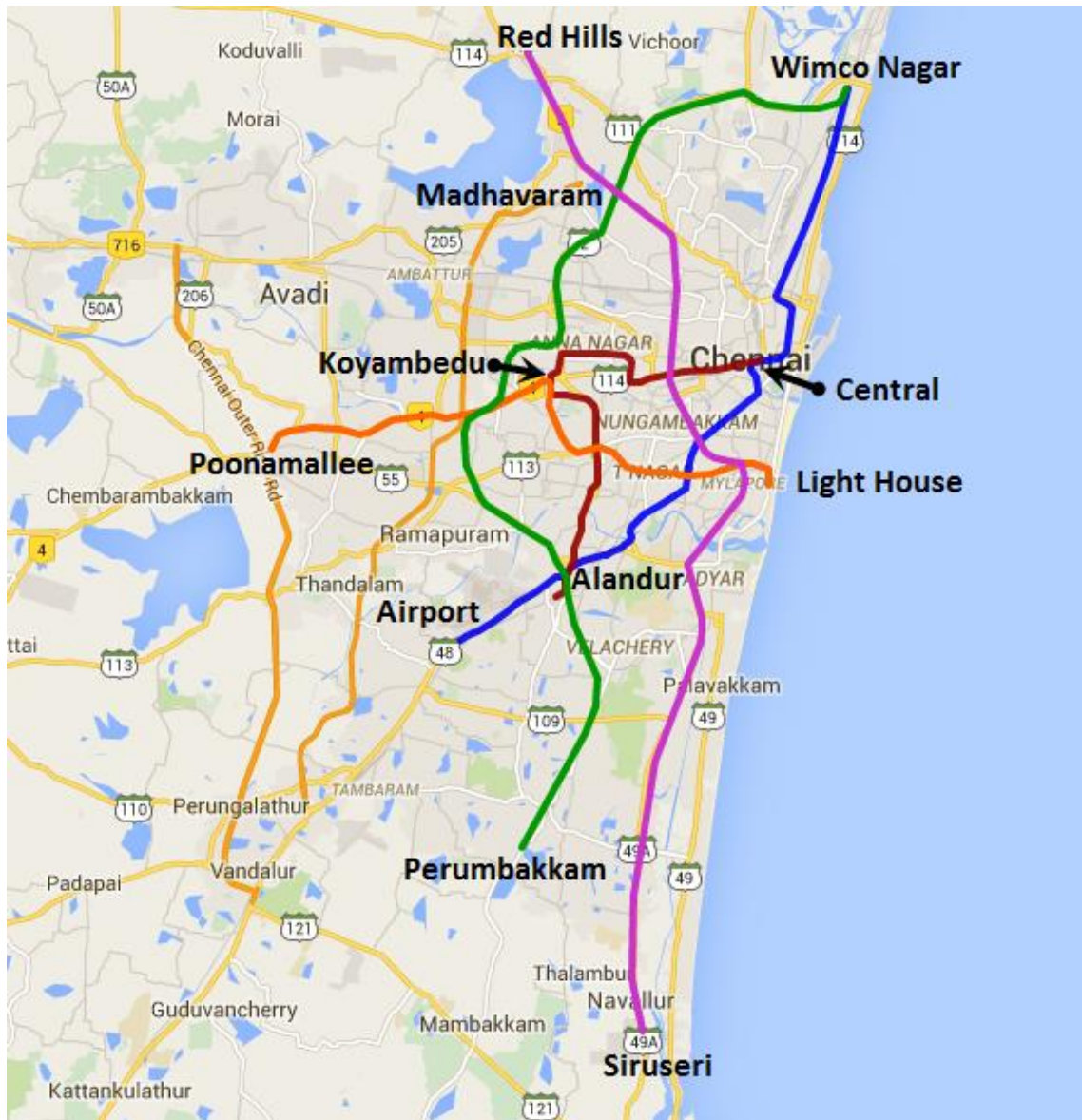


Planning Dockless e-Bike Sharing Service for Chennai: Segmenting and Clustering Neighbourhoods



Capstone Project Report

Planning Dockless e-Bike Sharing Service for Chennai: Segmenting and Clustering Neighbourhoods

1 Introduction to Chennai City Transportation Systems

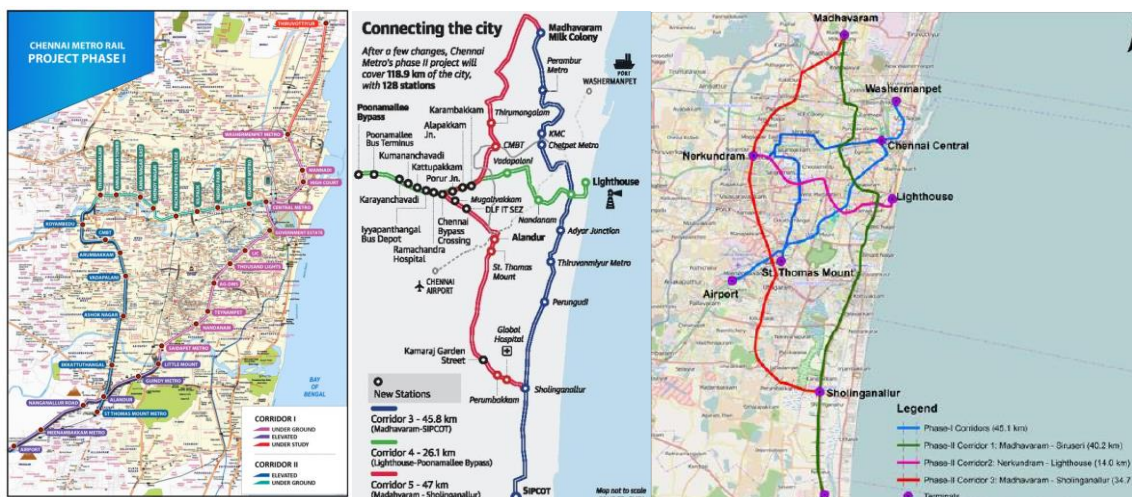
The metropolitan city of Chennai (capital of Tamilnadu state, India) with a population of more than 10 million, is the 4th most populous metro in India, 22nd in Asia and 40th in the world. Chennai metropolitan area (also referred to as Greater Chennai Corporation) is spread across 426 sq. kms. To address increasing needs of urban transportation within the city, several modes of public transportation systems are currently in operation:

- Chennai metro bus network with a fleet of about 4000 buses operating across 1000 routes criss-crossing the city, serving more than 5 million commuters per day.
- Chennai sub-urban rail network with more than 1000+ services every day and serving more than 1 million commuters per day.
- Chennai mass-rapid-transit train network serving more than 100K commuters every day.

Besides these public transportation systems, there are several privately-operated services comprising of 4-wheeler (call-taxis, cabs, share-autos) and 3-wheelers (autos) serving more than 10% of the requirement of population, either for end-to-end commute or for purpose of first-and-last-mile connectivity to above mentioned public transportation services.

2 Chennai Metro Rail System

One of the biggest concerns of these transportation systems is ever-increasing levels of pollution across the city. To holistically address this concern, government had planned for overhauling of urban transportation system, through Chennai metro rail network. Chennai metro phase-I project was commenced in 2010 and the staggered-rollout of the service commenced from 2015. Phase-II project has commenced in 2020 and staggered-rollout of the service is expected from 2026.



Map-1: Chennai Metro Phase-I; Map-2: Chennai Metro Phase-II; and Map-3: Combined Map
(src: <https://chennaiemrtrail.org/>)

- Phase-I: 45 Kms, 32 Stations, Launched in 2019, Average Commuters/Day: 1 Lakh
- Phase-2: 120 Kms, 100 Stations, expected to be launched by 2026. Expected Commuters/Day: 25 Lakh/Day

3 Business Problem Description (Need for e-Bike Sharing Service)

While this large network of electric metro train network (100+ stations spanning 150+ KMs) has well addressed the pollution concern of the city, it has opened another concern viz. last-mile connectivity to/from these 100+ locations (metro stations). All the current modes of last-mile connectivity are fossil-fuel-based modes (2/3/4-wheelers, privately-owned/operated).

To overhaul this last-mile connectivity based on green-modes, government is planning for dedicated cycle/e-bike tracks along the city roads. This initiative has created interest in mobility service providers to plan for rolling out e-bike services across the city.

Key business problem to be solved for these prospective e-bike service providers is to identify zones/locations within the city to profitably-deploy this e-bike based shared mobility service, and thereby enabling switch-over the last-mile connectivity to green-mode of transportation.

4 Data Required for Analysis

Data required for solving this business problem (segmenting and identification of potentially profitable locations within the city) needs to be fetched from several sources, as listed below:

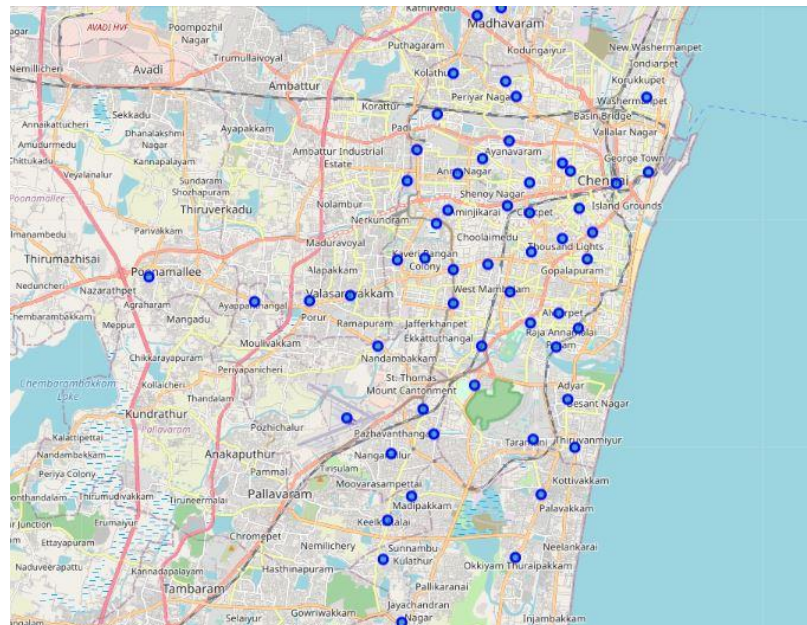
- **Metro station data:** List of metro stations and estimated commuters boarding and alighting at each station can be fetched from the official portal of Chennai Metro Rail Limited (CMRL; <https://chennaiemtrorail.org/>)
- **Venues and venue categories data:** Exploring neighbourhood around each metro station for most common/popular venues and their categorization would help analyse commuting needs/patterns. This would in turn help segmenting the locations into potentially profitable locations for operating e-bike shared mobility service (as last-mile connectivity solution) between stations and these venues. Foursquare API-based data access would help fetch this data.
 - Searching for specific venue categories:
https://api.foursquare.com/v2/venues/search?client_id=CLIENT_ID&client_secret=CLIENT_SECRET&ll=LATITUDE, LONGITUDE&v=VERSION&query=QUERY&radius=RADIUS&limit=LIMIT
 - Exploring specific locations:
https://api.foursquare.com/v2/venues/explore?client_id=CLIENT_ID&client_secret=CLIENT_SECRET&ll=LATITUDE, LONGITUDE&v=VERSION&limit=LIMIT
- **PinCode data:** Distance between two adjacent stations for most of the metro stations are less than 1 km. We can choose to explore neighbourhoods around stations with a radius for 1 Km, as e-bike services are generally convenient for these short distance rides. Therefore, we can prune the list of stations by considering one station per pin-code.

5 Dataset

A list of 59 metro stations considered for this analysis is tabulated here:

S. No.	PostalCode	Neighborhood	Latitude	Longitude	S. No.	PostalCode	Neighborhood	Latitude	Longitude
1	600102	Anna Nagar East	13.092	80.2236	31	600004	Mandaveli	13.0279	80.2605
2	600040	Anna Nagar Tower	13.0863	80.2138	32	600028	Greenways Road	13.0208	80.252
3	600016	Alandur	12.9975	80.2006	33	600020	Adyar	13.0012	80.2565
4	600106	Arumbakkam	13.0724	80.2102	34	600041	Thiruvannmiyur	12.983	80.2594
5	600083	Ashok Nagar	13.0373	80.2123	35	600113	Taramani Link Road	12.9863	80.2432
6	600027	Chennai Airport	12.9941	80.1709	36	600096	Perungudi	12.9654	80.2461
7	600008	Egmore	13.0732	80.2609	37	600097	Thoraipakkam	12.9416	80.2362
8	600032	Guindy	13.0067	80.2206	38	600119	Sholinganallur	12.901	80.2279
9	600002	LIC	13.0642	80.266	39	600130	Navalur	12.8459	80.2265
10	600104	High Court	13.0867	80.2877	40	603103	Siruseri	12.8352	80.2011
11	600010	Kilpauk	13.0828	80.2417	41	600018	Alwarpet	13.0335	80.2531
12	600107	CMBT	13.0675	80.2056	42	600017	T. Nagar	13.0418	80.2341
13	600015	Saidapet	13.0213	80.2231	43	600024	Kodambakkam	13.0521	80.2255
14	600035	Nandanam	13.03	80.2421	44	600093	Saligramam	13.0545	80.2011
15	600061	Nanganallur	12.9807	80.1882	45	600087	Valasaravakkam	13.0403	80.1723
16	600030	Pachaiyappas College	13.0741	80.2332	46	600116	Porur	13.0382	80.1565
17	600003	Central Station	13.0825	80.2755	47	600056	Iyyapanthangal	13.0381	80.1354
18	600050	Thirumangalam	13.0835	80.1945	48	600123	Poonamallee	13.0473	80.0945
19	600006	Thousand Lights	13.0617	80.2544	49	600060	MMBT	13.1456	80.2215
20	600026	Vadapalani	13.05	80.2121	50	600099	Kolathur	13.124	80.2121
21	600021	Washermanpet	13.1148	80.2872	51	600049	Villivakkam	13.1086	80.2061
22	600051	Madhavaram	13.1488	80.2306	52	600101	Anna Nagar Depot	13.0952	80.1981
23	600011	Sembiyan	13.1154	80.2367	53	600092	Elango Nagar	13.0537	80.1906
24	600012	Perambur	13.121	80.2326	54	600089	Manapakkam	13.0213	80.1832
25	600023	Ayanavaram	13.0986	80.2337	55	600088	Adambakkam	12.988	80.2047
26	600007	Doveton	13.0872	80.2575	56	600091	Madipakkam	12.9647	80.1961
27	600084	Purasaiwakkam	13.0902	80.2543	57	600117	Kilkattalai	12.9556	80.1869
28	600031	Chetpet	13.0714	80.2417	58	600129	Kovilambakkam	12.9409	80.1851
29	600034	Nungambakkam	13.0569	80.2425	59	600100	Medavakkam	12.9171	80.1923
30	600014	Royapettah	13.054	80.2641					

These 59 stations are shown on the Chennai city map as depicted below:



With this dataset as the starting point, we propose to carry out exploratory analysis and then use k-means clustering method to segment these neighbourhoods and identify a sub-set of stations/locations that are potentially profitable for rolling out e-bike shared mobility service, as a solution to the last-mile connectivity problem.

6 Methodology

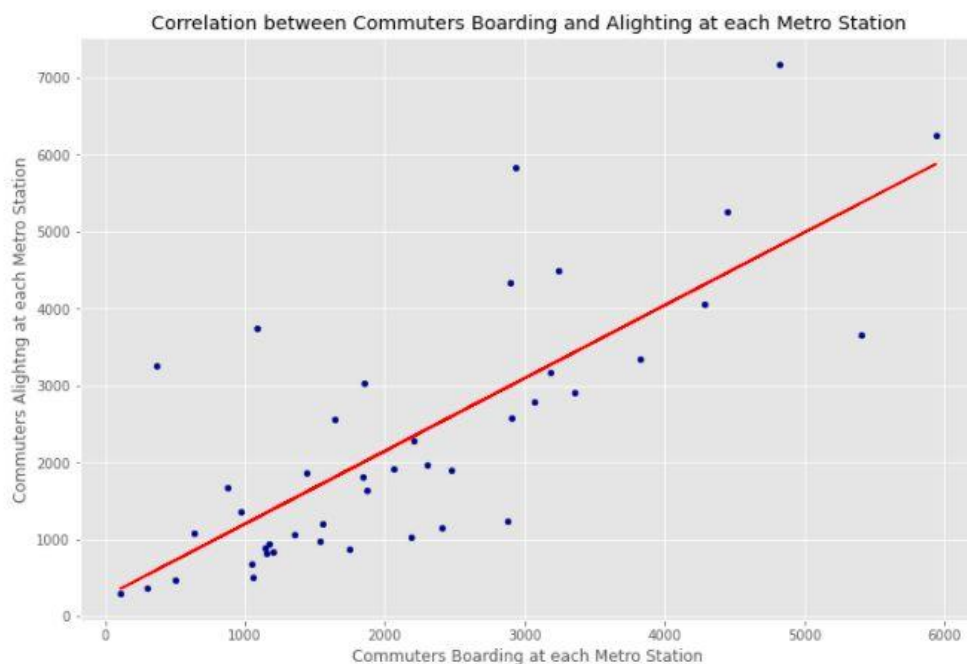
6.1 Exploratory Analysis

Profitability-case for e-bike shared mobility service providers (as a last-mile connectivity solution) depends on commuter-miles clocked per day for each of the e-bikes deployed into service. These commuter-miles in-turn would depend on constant flow (boarding + alighting) of commuters at each metro station.

To explore for any relationship between commuters boarding and alighting at each metro station, we fetched the estimated peak-hour commuter data for 43 stations of Phase-II project from a document published by CMRL (src: <https://chennai-metro-rail.org/wp-content/uploads/2015/11/Comprehensive-Detailed-Project-Report-for-Chennai-Metro.pdf>). The data is tabulated here:

S. No.	PostalCode	Neighborhood	Latitude	Longitude	Boarding	Alighting	Total Commuters
1	600040	Anna Nagar Tower	13.0863	80.2138	1559	1206	2765
2	600016	Alandur	12.9975	80.2006	1085	3742	4827
3	600010	Kilpauk	13.0828	80.2417	3823	3338	7161
4	600107	CMBT	13.0675	80.2056	5940	6252	12192
5	600035	Nandanam	13.03	80.2421	4280	4052	8332
6	600050	Thirumangalam	13.0835	80.1945	1847	3033	4880
7	600026	Vadapalani	13.05	80.2121	2304	1964	4268
8	600051	Madhavaram	13.1488	80.2306	2184	1023	3207
9	600011	Sembayan	13.1154	80.2367	2881	1234	4115
10	600012	Perambur	13.121	80.2326	2404	1155	3559
11	600023	Ayanavaram	13.0986	80.2337	2899	4330	7229
12	600007	Doveton	13.0872	80.2575	973	1357	2330
13	600084	Purasaiwakkam	13.0902	80.2543	1842	1809	3651
14	600031	Chetpet	13.0714	80.2417	1439	1870	3309
15	600004	Mandaveli	13.0279	80.2605	5401	3650	9051
16	600028	Greenways Road	13.0208	80.252	2901	2578	5479
17	600020	Adyar	13.0012	80.2565	4446	5257	9703
18	600041	Thiruvanniyur	12.983	80.2594	1637	2566	4203
19	600113	Taramani Link Road	12.9863	80.2432	637	1082	1719
20	600096	Perungudi	12.9654	80.2461	1869	1634	3503
21	600097	Thoraipakkam	12.9416	80.2362	2206	2291	4497
22	600119	Sholinganallur	12.901	80.2279	877	1681	2558
23	600130	Navalur	12.8459	80.2265	501	473	974
24	603103	Siruseri	12.8352	80.2011	305	363	668
25	600018	Alwarpet	13.0335	80.2531	1043	683	1726
26	600017	T. Nagar	13.0418	80.2341	4821	7178	11999
27	600024	Kodambakkam	13.0521	80.2255	3183	3163	6346
28	600093	Saligramam	13.0545	80.2011	2471	1902	4373
29	600087	Valasaravakkam	13.0403	80.1723	110	298	408
30	600116	Porur	13.0382	80.1565	2934	5836	8770
31	600056	Iyyapanthangal	13.0381	80.1354	1145	893	2038
32	600123	Poonamallee	13.0473	80.0945	2066	1909	3975
33	600060	MMBT	13.1456	80.2215	1201	840	2041
34	600099	Kolathur	13.124	80.2121	1750	872	2622
35	600049	Villivakkam	13.1086	80.2061	3070	2779	5849
36	600101	Anna Nagar Depot	13.0952	80.1981	3240	4493	7733
37	600092	Elango Nagar	13.0537	80.1906	3357	2904	6261
38	600089	Manapakkam	13.0213	80.1832	1350	1067	2417
39	600088	Adambakkam	12.988	80.2047	1172	946	2118
40	600091	Madipakkam	12.9647	80.1961	1054	508	1562
41	600117	Kilkattalai	12.9556	80.1869	365	3257	3622
42	600129	Kovilambakkam	12.9409	80.1851	1533	975	2508
43	600100	Medavakkam	12.9171	80.1923	1152	812	1964

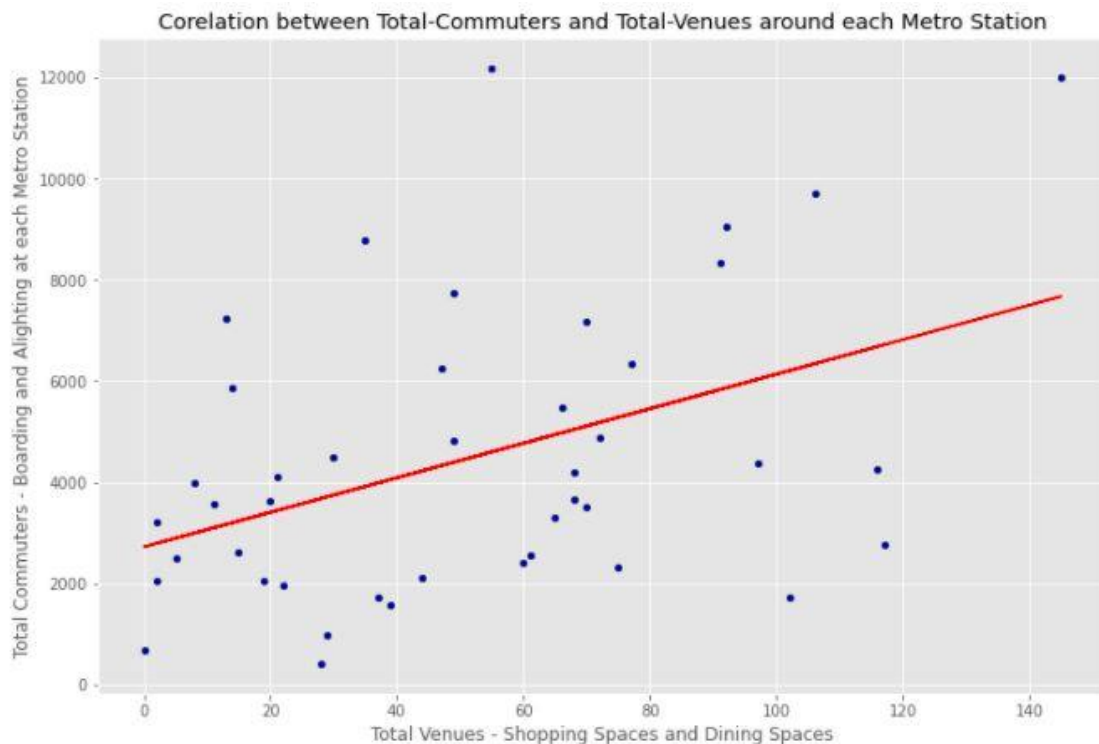
It can be observed that the estimated count of commuters boarding at each station is almost proportional to the count of commuters alighting at the respective station. This correlation can be seen in the scatter plot as depicted below. This two-way proportional-flow of estimated commuters increases the likelihood of e-bikes getting utilized continuously.



Further to this, also explored for any relationship between total commuters at each station and the most common venue categories around each station/neighbourhood. We used foursquare API calls and accessed data for two popular categories of venues, viz. food-venues/dining-spaces (restaurants, hotels and cafes) and shop-venues/shopping-spaces (stores, shops and markets). We fetched this data for all 43 stations and tabulated the count of these venues as shown below (with first 10 rows).

Neighborhood	Latitude	Longitude	Boarding	Alighting	TotalCommuters	Restaurant	Hotel	Cafe	Store	Shop	Market	FoodVenues	ShopVenues	TotalVenues
Anna Nagar Tower	13.0863	80.2138	1559	1206	2765	20	18	19	27	26	7	57	60	117
Alandur	12.9975	80.2006	1085	3742	4827	6	18	2	4	15	4	26	23	49
Kilpauk	13.0828	80.2417	3823	3338	7161	6	17	11	16	13	7	34	36	70
CMBT	13.0675	80.2056	5940	6252	12192	6	27	8	2	9	3	41	14	55
Nandanam	13.0300	80.2421	4280	4052	8332	16	22	18	9	24	2	56	35	91
Thirumangalam	13.0835	80.1945	1847	3033	4880	11	12	8	18	17	6	31	41	72
Vadapalani	13.0500	80.2121	2304	1964	4268	9	45	11	17	31	3	65	51	116
Madhavaram	13.1488	80.2306	2184	1023	3207	1	0	0	0	1	0	1	1	2
Sembian	13.1154	80.2367	2881	1234	4115	4	4	2	5	6	0	10	11	21
Perambur	13.1210	80.2326	2404	1155	3559	2	2	1	2	4	0	5	6	11

We could observe correlation between total-commuters and total-venues (dining + shopping spaces). This linear relationship (increase in commuters for an increase in venues) can be seen in the scatter plot as depicted below. This trend further builds the case for profitable operations of e-bike service if the dock-less e-bike zones can be located around those stations where both the categories of venues exist among its common venues.



With these insights, we moved onto the next section of modeling the data using unsupervised machine learning model (K-means clustering), for the purpose of segmenting and clustering the 59 neighbourhoods/stations into similar clusters, and thereby identifying the cluster that may have venue categories from both the types (dining spaces and shopping spaces) among its 10-most common venue categories.

6.2 Segmentation and Clustering Neighbourhoods/Stations

Using the Foursquare “Explore Venues” API call, fetched “NearbyVenues” within a radius of 1 Km, for each of the 59 stations.

```
def getNearbyVenues(names, latitudes, longitudes, radius=1000):  
  
    venues_list=[]  
    for name, lat, lng in zip(names, latitudes, longitudes):  
        print(name)  
  
        # create the API request URL  
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(  
            CLIENT_ID,  
            CLIENT_SECRET,  
            VERSION,  
            lat,  
            lng,  
            radius,  
            LIMIT)
```

Created a dataframe called “Chennai_venues” that has NearbyVenues along with VenueCategories for all the stations/neighbourhoods. There are 1476 venues in total across these 59 stations within a radius of 1 Km, and these venues come under 171 unique venue categories.

```
print(chennai_venues.shape)  
chennai_venues.head()
```

(1476, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Anna Nagar East	13.092	80.2236	Pantry D'Or	13.092122	80.218893	Café
1	Anna Nagar East	13.092	80.2236	Coco Jaunt	13.090318	80.217500	Café
2	Anna Nagar East	13.092	80.2236	Cream Centre	13.090306	80.216317	Vegetarian / Vegan Restaurant
3	Anna Nagar East	13.092	80.2236	Quality Fast Food	13.090326	80.218606	Fast Food Restaurant
4	Anna Nagar East	13.092	80.2236	A2B Restaurant Chinthamani	13.084275	80.224705	Indian Restaurant

After this, used one-hot-encoding method to identify most common venue categories for each of the 59 stations/neighbourhoods. Filtered 10-most common venues to create a 59x10 matrix (59 stations/neighbourhoods x 10 venue-categories) to be fed to K-means clustering algorithm.

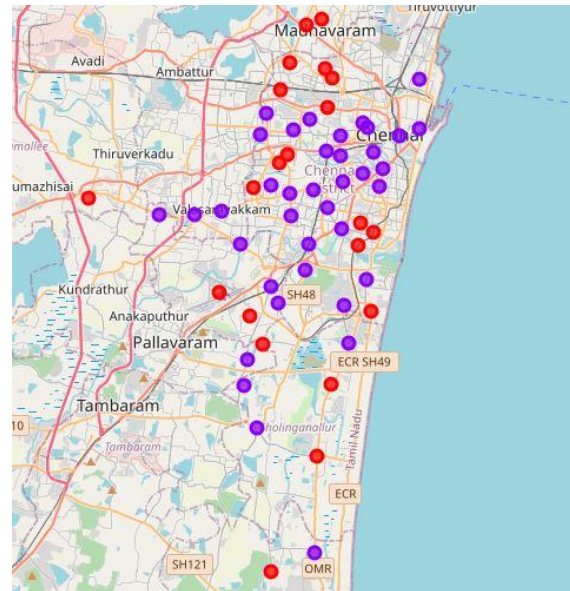
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adambakkam	Department Store	Train Station	Pizza Place	Burrito Place	Bakery	Asian Restaurant	Indian Restaurant	American Restaurant	Food	Furniture / Home Store
1	Adyar	Indian Restaurant	Café	Pizza Place	Department Store	Asian Restaurant	Fast Food Restaurant	Ice Cream Shop	Electronics Store	North Indian Restaurant	Rock Club
2	Alandur	Indian Restaurant	Pizza Place	Train Station	Breakfast Spot	Metro Station	Arts & Crafts Store	Asian Restaurant	Bakery	Food	Fruit & Vegetable Store
3	Alwarpet	Café	Restaurant	Indian Restaurant	Hotel	Italian Restaurant	Coffee Shop	Bakery	Department Store	Chinese Restaurant	Ice Cream Shop
4	Anna Nagar Depot	Café	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Bus Station	Ice Cream Shop	Juice Bar	Park	Market	Sporting Goods Shop

Ran 3 iterations of k-means-clustering (for values of K as 2, 5 and 10 respectively), to study the variation in the similarity/dis-similarity of cluster members for each iteration. In the subsequent sections, we have captured the results for these 3 iterations and presented observations and recommendations to prospective shared mobility service providers, with an objective of profitably operating e-bike services within the city of Chennai, and thereby providing a green/sustainable solution to Chennai Metro’s last-mile connectivity problem.

7 Results

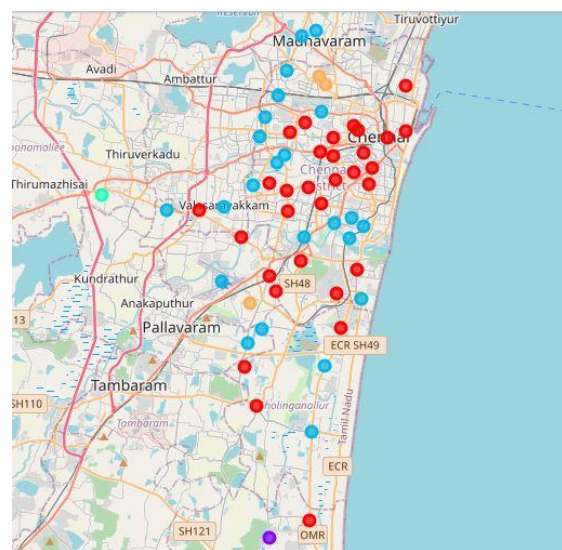
In the first iteration, with $K = 2$, K-means-clustering returned two clusters with 38 and 21 stations/neighbourhoods, respectively. List of these stations and distribution of food-venues and shop/other-venues among the 10-most common venue categories are tabulated below, along with the colour-coded cluster map.

Iteration #1: No. of Clusters = 2					
Cluster #1: 38 Neighborhoods			Cluster #2: 21 Neighborhoods		
S. No.	Neighborhood	Food Venues	S. No.	Neighborhood	Food Venues
1.1	Anna Nagar East	8	2.1	Arumbakkam	7
1.2	Anna Nagar Tower	7	2.2	Chennai Airport	4
1.3	Alandur	6	2.3	CMBT	6
1.4	Ashok Nagar	8	2.4	Nanganallur	5
1.5	Egmore	8	2.5	Madhavaram	3
1.6	Guindy	6	2.6	Sembiyar	6
1.7	LIC	7	2.7	Perambur	6
1.8	High Court	3	2.8	Ayanavaram	6
1.9	Kilpauk	8	2.9	Mandaveli	9
1.10	Pachaiyappas College	7	2.10	Greenways Road	8
1.11	Central Station	4	2.11	Thiruvanniyur	7
1.12	Thousand Lights	9	2.12	Thorapakkam	8
1.13	Vadapalani	7	2.13	Sholinganallur	9
1.14	Washermanpet	6	2.14	Siruseri	5
1.15	Doveton	7	2.15	Alwarpet	9
1.16	Purasaiwakkam	6	2.16	Poonamallae	5
1.17	Chetpet	9	2.17	MMBT	4
1.18	Nungambakkam	10	2.18	Kolathur	2
1.19	Royapettah	7	2.19	Villivakkam	3
1.20	Adyar	7	2.20	Elango Nagar	5
1.21	Taramani Link Road	8	2.21	Madipakkam	2
1.22	Perungudi	7	2.22	Average Count	5
1.23	Navalur	7			
1.24	T. Nagar	7			
1.25	Kodambakkam	6			
1.26	Saligramam	6			
1.27	Porur	7			
1.28	Manapakkam	9			
1.29	Adambakkam	7			
1.30	Kovilambakkam	6			
1.31	Medavakkam	10			
1.32	Saidapet	7			
1.33	Nandanam	7			
1.34	Thirumangalam	5			
1.35	Valasaravakkam	5			
1.36	Iyyapanthangal	8			
1.37	Anna Nagar Depot	6			
1.38	Kilkattalai	5			
1.39	Average Count	7			



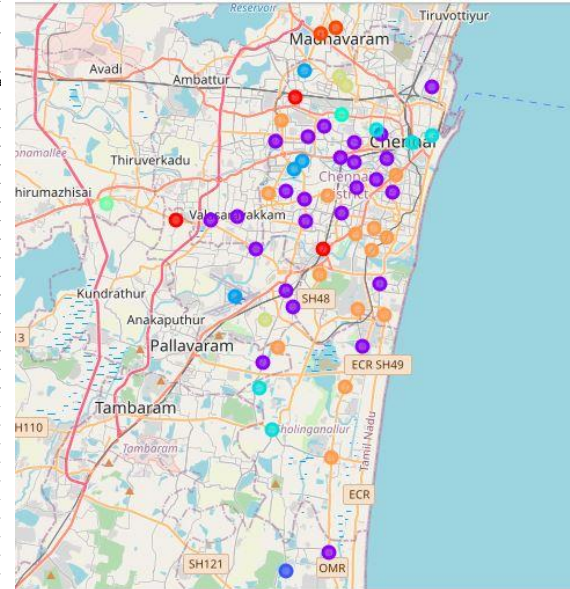
In the second iteration, with $K = 5$, K-means-clustering returned two big-sized clusters with 31 and 23 stations/neighbourhoods, respectively.

Iteration #2: No. of Clusters = 5					
Cluster #1: 31 Neighborhoods			Cluster #2: 23 Neighborhoods		
S. No.	Neighborhood	Food Venues	S. No.	Neighborhood	Food Venues
1.1	Anna Nagar East	8	2.1	Arumbakkam	7
1.2	Anna Nagar Tower	7	2.2	Chennai Airport	4
1.3	Alandur	6	2.3	CMBT	6
1.4	Ashok Nagar	8	2.4	Saidapet	7
1.5	Egmore	8	2.5	Nandanam	7
1.6	Guindy	6	2.6	Thirumangalam	5
1.7	LIC	7	2.7	Madhavaram	3
1.8	High Court	3	2.8	Ayanavaram	6
1.9	Kilpauk	8	2.9	Mandaveli	9
1.10	Pachaiyappas College	7	2.10	Greenways Road	8
1.11	Central Station	4	2.11	Thiruvanniyur	7
1.12	Thousand Lights	9	2.12	Thorapakkam	8
1.13	Vadapalani	7	2.13	Sholinganallur	9
1.14	Washermanpet	6	2.14	Alwarpet	9
1.15	Doveton	7	2.15	Valasaravakkam	5
1.16	Purasaiwakkam	6	2.16	Iyyapanthangal	8
1.17	Chetpet	9	2.17	MMBT	4
1.18	Nungambakkam	10	2.18	Kolathur	2
1.19	Royapettah	7	2.19	Villivakkam	3
1.20	Adyar	7	2.20	Anna Nagar Depot	6
1.21	Taramani Link Road	8	2.21	Elango Nagar	5
1.22	Perungudi	7	2.22	Madipakkam	2
1.23	Navalur	7	2.23	Kilkattalai	5
1.24	T. Nagar	7	2.24	Average Count	6
1.25	Kodambakkam	6			
1.26	Saligramam	6			
1.27	Porur	7			
1.28	Manapakkam	9			
1.29	Adambakkam	7			
1.30	Kovilambakkam	6			
1.31	Medavakkam	10			
1.32	Average Count	7			



In the third iteration, with K =10, K-means-clustering again returned only two big-sized clusters with 25 and 14 stations/neighbourhoods, respectively.

Iteration #3: No. of Clusters = 10							
Cluster #1: 25 Neighborhoods				Cluster #2: 14 Neighborhoods			
S. No.	Neighborhood	Food Venues	Shop/Other Venues	S. No.	Neighborhood	Food Venues	Shop/Other Venues
1.1	Anna Nagar East	8	2	2.1	Guindy	6	4
1.2	Anna Nagar Tower	7	3	2.2	LIC	7	3
1.3	Alandur	6	4	2.3	Nandanam	7	3
1.4	Ashok Nagar	8	2	2.4	Mandaveli	9	1
1.5	Egmore	8	2	2.5	Greenways Road	8	2
1.6	Kilpauk	8	2	2.6	Thiruvannmiyur	7	3
1.7	Pachaiyappas College	7	3	2.7	Taramani Link Road	8	2
1.8	Thirumangalam	5	5	2.8	Thoralpakkam	8	2
1.9	Thousand Lights	9	1	2.9	Sholinganallur	9	1
1.10	Vadapalani	7	3	2.10	Alwarpet	9	1
1.11	Washermanpet	6	4	2.11	Kodambakkam	6	4
1.12	Doveton	7	3	2.12	Anna Nagar Depot	6	4
1.13	Chetpet	9	1	2.13	Elango Nagar	5	5
1.14	Nungambakkam	10	0	2.14	Madipakkam	2	8
1.15	Royapettah	7	3	2.15	Average Count	7	3
1.16	Adyar	7	3				
1.17	Perungudi	7	3				
1.18	Navalur	7	3				
1.19	T. Nagar	7	3				
1.20	Saligramam	6	4				
1.21	Valasaravakkam	5	5				
1.22	Porur	7	3				
1.23	Manapakkam	9	1				
1.24	Adambakkam	7	3				
1.25	Kilkattalai	5	5				
1.26	Average Count	7	3				



8 Observations and Recommendation to e-Bike Mobility Service Providers

Upon observing the results of all the three iterations, the second cluster of first iteration (with K=2) had equal-distribution of venue-categories (food-venues and shop/other-venues) among its 10-most common venue categories, at an aggregate-level.

S. No.	Iteration	Cluster No.	Average Count of:	
			Food Venues	Shop/Other Venues
1.1	No. of Clusters = 2	1	7	3
1.2		2	5	5
2.1	No. of Clusters = 5	1	7	3
2.2		2	6	4
3.1	No. of Clusters = 10	1	7	3
3.2		2	7	3

It can be recommended that this cluster with 21 neighbourhoods/stations are the candidates among 59 stations considered for analysis (or 132 metro stations in all), for rolling out e-bike shared mobility service, that can potentially turn profitable to the prospective service provider. Upon these stations/neighbourhood locations turning profitable, service provider(s) can choose to scale-out the shared mobility services to other locations in a phased manner.

The list of these 21 identified stations is tabulated below.

```
chennai_merged.loc[chennai_merged['ClusterLabels'] == 0, chennai_merged.columns[[1] + list(range(5, chennai_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Arumbakkam	Hotel	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Scenic Lookout	Metro Station	Bakery	Food Court	Pub	Chinese Restaurant	Pool
5	Chennai Airport	Women's Store	Airport Lounge	Airport Service	Airport Terminal	Metro Station	Hotel	Fast Food Restaurant	Restaurant	Yoga Studio	Food
11	CMBT	Hotel	Bus Station	Vegetarian / Vegan Restaurant	Metro Station	Pool	Flower Shop	Pub	Food Court	Fast Food Restaurant	Chinese Restaurant
14	Nanganallur	Chinese Restaurant	Department Store	Pizza Place	Multiplex	Grocery Store	Gym	Flower Shop	Fried Chicken Joint	Food Truck	Food Court
21	Madhavaram	Food & Drink Shop	Bookstore	Farm	Bus Station	Moving Target	Yoga Studio	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	Food Truck
22	Sembiyan	Multiplex	Department Store	Chinese Restaurant	Shopping Mall	Snack Place	Pizza Place	Falafel Restaurant	Food	Frozen Yogurt Shop	Diner
23	Perambur	Department Store	Multiplex	Snack Place	Pizza Place	Donut Shop	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	Food Truck	Food Court
24	Ayanavaram	Department Store	Tennis Court	Café	Diner	Indie Movie Theater	Hotel Bar	Flea Market	Fried Chicken Joint	Food Truck	Food Court
30	Mandaveli	Café	Indian Restaurant	Restaurant	Snack Place	Dessert Shop	Vegetarian / Vegan Restaurant	Italian Restaurant	Chinese Restaurant	Coffee Shop	Department Store
31	Greenways Road	Chinese Restaurant	Dessert Shop	Department Store	Indian Restaurant	Restaurant	Café	Coffee Shop	Hotel	Boat or Ferry	Mediterranean Restaurant
33	Thiruvanniyur	Beach	Hotel	Indian Restaurant	Chinese Restaurant	Coffee Shop	Juice Bar	Gym	Bakery	Scenic Lookout	Burger Joint
36	Thoraipakkam	Café	Food Court	Bakery	Indian Restaurant	Food	Bus Station	Vegetarian / Vegan Restaurant	Hotel	North Indian Restaurant	Gift Shop
37	Sholinganallur	Food	Middle Eastern Restaurant	Restaurant	Pizza Place	Hotel	Vegetarian / Vegan Restaurant	Ice Cream Shop	Sandwich Place	Café	Sports Club
39	Siruseri	ATM	Tea Room	Gas Station	Furniture / Home Store	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop
40	Alwarpet	Café	Restaurant	Indian Restaurant	Hotel	Italian Restaurant	Coffee Shop	Bakery	Department Store	Chinese Restaurant	Ice Cream Shop
47	Poonamallee	Movie Theater	Breakfast Spot	Bus Station	Diner	Furniture / Home Store	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	Food Truck	Food Court
48	MMBT	Bus Station	Department Store	Bookstore	Bed & Breakfast	Food & Drink Shop	Furniture / Home Store	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	Food Truck
49	Kolathur	Intersection	Indie Movie Theater	Hotel	Department Store	Airport Service	Donut Shop	Gaming Cafe	Furniture / Home Store	Fruit & Vegetable Store	Frozen Yogurt Shop
50	Villivakkam	Light Rail Station	Bus Station	Food Truck	Multiplex	Bakery	Smoke Shop	Yoga Studio	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint
52	Elango Nagar	Pizza Place	Fast Food Restaurant	Movie Theater	Café	Sandwich Place	Multiplex	Gym	Chinese Restaurant	Farmers Market	Fish Market
55	Madipakkam	Indian Restaurant	Indie Movie Theater	Department Store	Convenience Store	Coffee Shop	Lake	Grocery Store	Yoga Studio	Flower Shop	Frozen Yogurt Shop

These 21 stations (as mentioned under column: neighbourhood) are:

Arumbakkam, Chennai Airport, CMBT, Nanganallur, Madhavaram, Sembiyan, Perambur, Ayanavaram, Mandaveli, Greenways Road, Thiruvanniyur, Thoraipakkam, Sholinganallur, Siruseri, Alwarpet, Poonamallae, MMBT, Kolathur, Villivakkam, Elango Nagar and Madipakkam.

9 Conclusion

This project is an application of data science methodology (using unsupervised machine learning model, K-means-clustering) to solve a business problem relevant to mobility service providers.

With the help of data accessed from location service providers (such as Foursquare in our case), we have demonstrated this application of data analytics in the business context of “Identification of locations to roll-out e-bike shared mobility services” in the city of Chennai, and thereby provide solution to last-mile connectivity problem of Chennai Metro service.

10 References

- Comprehensive Detailed Project Report published by Chennai Metro Rail Limited (CMRL), for station data and estimated commuter data, <https://chennaietrorail.org/wp-content/uploads/2015/11/Comprehensive-Detailed-Project-Report-for-Chennai-Metro.pdf>
- Chennai Metro Rail Limited (CMRL) official website for station maps depicting phase-I and phase-II projects, <https://chennaietrorail.org/>
- Foursquare API-based data access for location-based venues and venue-category data, <https://api.foursquare.com/>
- Location data (latitude and longitude details) from Google, <https://www.google.com/search>
- PInCode data for Chennai, from Wikipedia, https://en.wikipedia.org/wiki/Areas_of_Chennai