Mumbai Subway Stations Classification

Coursera Project

Introduction

The subrban rail network in mumbai commonly known as local railway is known as lifeline of mumbai as it connects most of the commercial office locations where rent is as high as skyscrappers to suburbs from where majority of workforce travels to work. There are mainly 3 sub network of this network, namely central, western and harbour. So any suburb neighbourhood in mumbai is defined mostly by nearest train station. In this notebook we will try to find out based upon facilities which suburbs are similar to each other so that if someone's looking for apartment on rent he can decide out of similar suburbs which neighbourhood to pick based upon similarity and trade off between rent and proximity from city center

Data

1.List of suburb stations : obtained from this wikipedia page:(https://en.wikipedia.org/wiki/List_of_Mumbai_Suburban_Railway_stations) and we will extract data to pandas dataframe using beautifulsoup

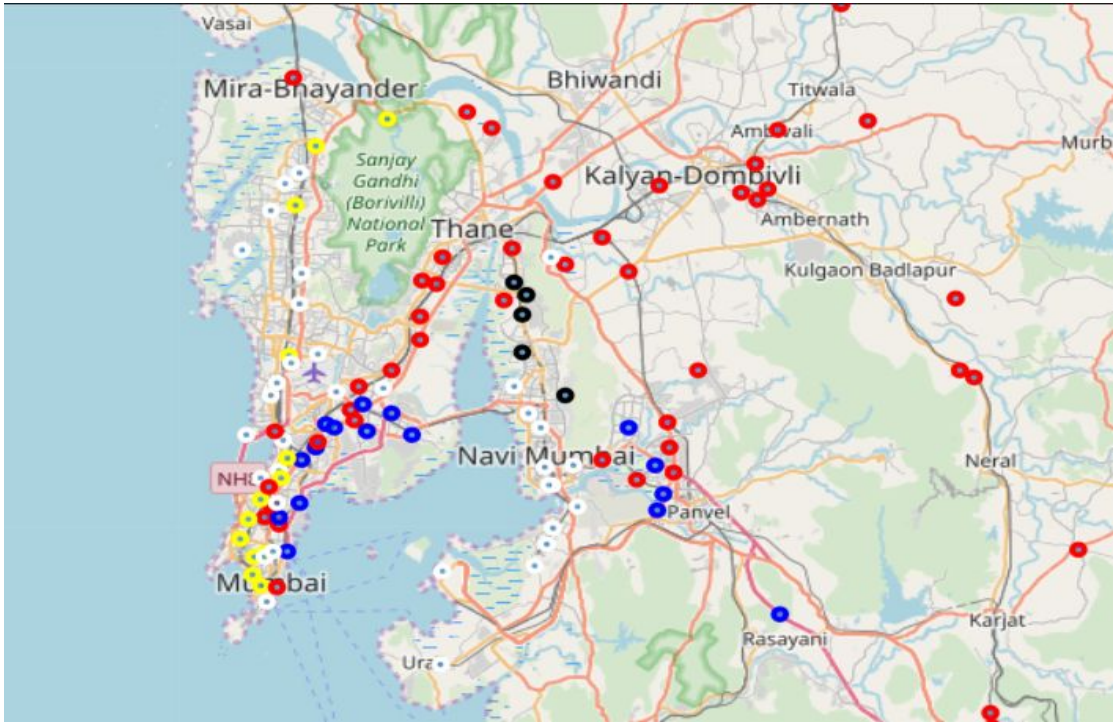
2.Coordinates for stations obtained from python geocoder library using here map api and open street map api

3.Neighbourhood places Data obtained from foursuare api

Data

	Name	Marathi Name	Line	Fast Local	Long Distance Trains	Latitude	Longitude
0	Airoli	ऐरोली	Trans-Harbour Line	N	N	19.155590	72.998160
1	Ambarnath	अंबरनाथ	Central Line	Y	Y	19.143607	73.295535
2	Ambivli	आंबिवली	Central Line	N	Y	19.266080	73.175360
3	Andheri	अंधेरी	None	Y	Y	19.103930	72.866980
4	Asangaon	आसनगांव	Central Line	Y	Y	19.018430	73.105180
...	...	...	...	...	...	...	...
129	Sewri	शिवडी	Harbour Line	N	N	18.996360	72.853880
130	Taloja	तळोजे पंचानंद	Central Line	N	Y	19.092000	73.121570
131	Pen	पेण	Harbour Line	Y	Y	18.783000	73.095720
132	Rasayani	रसायनी	Harbour Line	Y	Y	18.915630	73.176820
133	Nagothane	नागोठणे	Harbour Line	Y	Y	18.541700	73.138630

Data


Vasai
Mira-Bhayander
Bhiwandi
Titwala
Ambivali
Murb
Sanjay Gandhi (Borivilli) National Park
Kalyan-Dombivli
Thane
Ambernath
Kulgaon Badlapur
Navi Mumbai
Neral
Panvel
Karjat
Rasayani

MethodoLogy

We will use Foursquare's places api to get places nearby to the stations.Using geolocation data to get places within 1.5km radius of station with limit set to getting top 20 nearest stations.After that we will get all categories of places available places near all stations and one hot encode frequency of each category within stations radius and then use K means clustering on this data while making use of scikit learn to classify stations within 4 categories

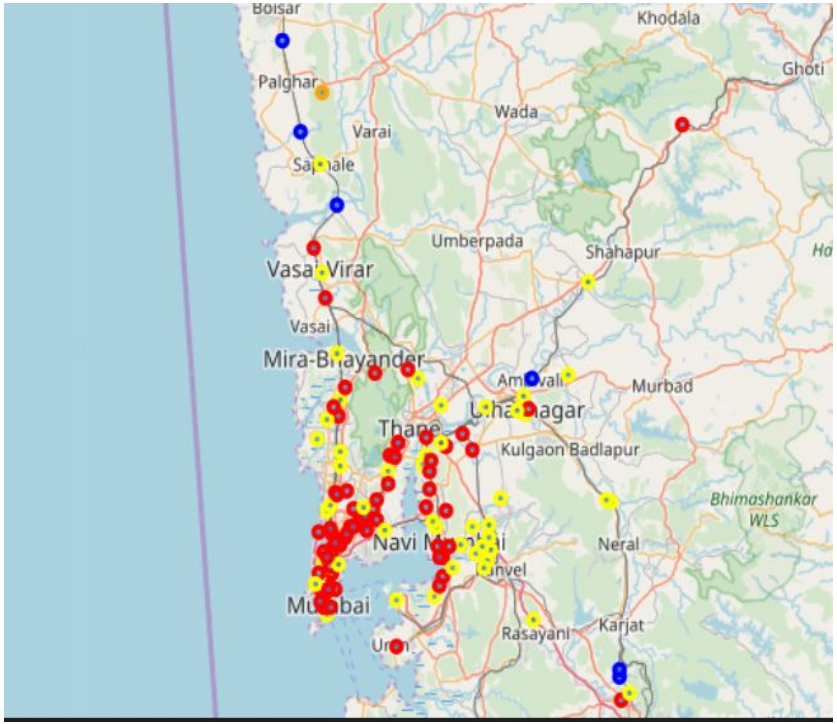
Methodology

	Station	ATM	Afghan Restaurant	Airport Service	American Restaurant	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...	Theme Park	Thrift / Vintage Store
0	Chhatrapati Shivaji Maharaj Terminus	0.00	0.0	0.00	0.0	0.0	0.05	0.0	0.000000	0.0	...	0.0	0.0
1	Airoli	0.00	0.0	0.00	0.0	0.0	0.00	0.0	0.052632	0.0	...	0.0	0.0
2	Ambivli	0.25	0.0	0.00	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.0	0.0
3	Andheri	0.00	0.0	0.05	0.0	0.0	0.00	0.0	0.050000	0.0	...	0.0	0.0
4	Asangaon	0.00	0.0	0.00	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.0	0.0
5	Atgaon	0.00	0.0	0.00	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.0	0.0
6	Badlapur	0.00	0.0	0.00	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.0	0.0
7	Baman Dongari	0.00	0.0	0.00	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.0	0.0
8	Bandra	0.00	0.0	0.00	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.0	0.0
9	Bhandup	0.00	0.0	0.00	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.0	0.0

Result and Conclusion

	Station	cluster
0	Chhatrapati Shivaji Maharaj Terminus	1
1	Airoli	1
2	Ambivli	2
3	Andheri	0
4	Asangaon	1
...	...	...
123	Vikhroli	0
124	Vile Parle	0
125	Virar	0
126	Vithalwadi	1
127	Wadala Road	0

Result and Conclusion



Result and Conclusion

After carefully examining map we can see our 4 clusters:

1.red: mostly concentrated around mumbai and navi mumbai's main corporate areas

2.Yellow: Nearby corporate areas but not exactly corporate ares themselves.mostly nearby resedential suburbs

3.Blue and Orange:Outliers and far from city areas