

Clustering of Pune Metro Line -1 Neighbourhood's

Data

The official website of Pune Metro is <http://punemetrorail.org/> . The name of the metro stations will be sourced from the website.

```
from bs4 import BeautifulSoup
import requests
import pandas as pd
web_url = requests.get('http://punemetrorail.org/').text
soup=BeautifulSoup(web_url,features="html5lib")
#soup.prettify()
```

Thereafter , we will manually load a flat file with the location co-ordinates of the metro stations.

```
## Let us load the Latitude and Longitudes of only Line 1 stations.
latlon = pd.read_csv("Line1Metrostations.csv")
latlon.head()
```

	Station	latitude	longitude
0	PCMC	18.598609	73.797170
1	Sant Tukaram Nagar	18.622561	73.820851
2	Bhosari(Nashik Phata)	18.611349	73.823623
3	Kasarwadi	18.606410	73.821294
4	Phugewadi	18.590464	73.831599

As per the requirement we have used the Foursquare location data to get the venues around the metro stations.

A total of 521 locations found for 91 unique categories within the 1500m radius of those 14 stations

	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Station						
Bhosari(Nashik Phata)	19	19	19	19	19	19
Bopodi	20	20	20	20	20	20
Budhwar peth	53	53	53	53	53	53
Civil Court	77	77	77	77	77	77
Dapodi	21	21	21	21	21	21
Kasarwadi	16	16	16	16	16	16
Khadki	5	5	5	5	5	5
Mandai	61	61	61	61	61	61
PCMC	22	22	22	22	22	22
Phugewadi	13	13	13	13	13	13
Range Hill	45	45	45	45	45	45
Sant Tukaram Nagar	15	15	15	15	15	15
Shivajinagar	100	100	100	100	100	100
Swargate	54	54	54	54	54	54

```

: #Let's find out how many unique categories can be curated from all the returned venues
print('There are {} uniques categories.'.format(len(pune_venues['Venue Category'].unique())))

There are 91 uniques categories.

```

Thereafter , we identified the most common venues for each of the stations. The data may also be explored for certain indicative categories during the analysis. The metro stations will be clustered based on similarities of venue.