Clustering of Pune Metro Line -1 Neighbourhood's

Business Problem

Pune is a city in India. Pune Metro is a metro rail based rapid transit system under construction to serve Pune Central and the areas of Pimpri & Chinchwad. The system comprises 3 lines with a total length of 54.58 km.

The company KKK Ltd. Is a utility store chain and plans to set up stores near to metro station. It wants to take advantage of the increase in footfall near to metro station location. In order to decide on the most suitable location , the company wants a preliminary investigation about the venues around the to-be metro stations. The initial investigation is limited to metro stations along the Line 1. The decision will be forwarded to marketing department for decision on further studies for final decision of product mix to be sold through the metro station outlets.

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

We have used the Foursquare location data to get the venues around the metro stations.

Methodology

The methodology adopted for the analysis is as below,

1. Actual metro stations are sourced from official website by web scrapping
2. Latitude and Longitude of the metro station locations are uploaded from a flat file.
3. A Folium map of Pune is generated with metro locations superimposed on it.
4. Foursquare location data is used to find 100 venues within a radius of 1500m around the metro stations
5. The distribution of unique categories to Metro Locations is mapped.
6. There are few Venue categories , which are available only in certain locations. Unique locations of few such indicative categories are identified.
7. The top 10 most common venue for each metro station location is derived.
8. Elbow method is used to know the optimal clusters . Elbow method of K-means clustering helps to choose the optimum value of 'k' (number of clusters) by fitting the model with a range of values of 'k'.
9. The metro station locations are grouped into individual clusters and a folium map is generated to show the metro stations that belongs to same cluster.
10. The clusters generated are mapped against the range of 'Unique Venue categories" for analysis. This is to understand the cluster characteristics.

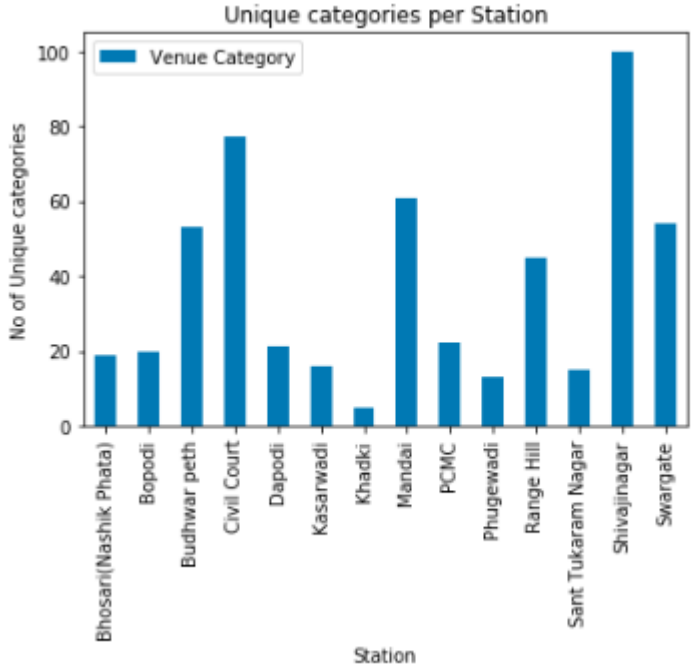
Results

1. The Pune metro Line 1 consists of 14 stations

Station
Bhosari(Nashik Phata)
Bopodi
Budhwar peth
Civil Court
Dapodi
Kasarwadi
Khadki
Mandai
PCMC
Phugewadi
Range Hill
Sant Tukaram Nagar
Shivajinagar
Swargate

2. A total of 521 locations found for 91 unique categories within the 1500m radius of those 14 stations
3. The no , of unique categories are distributed as given in the figure below. This implies that shops near to some of the locations trade in widespread items compared to others.

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

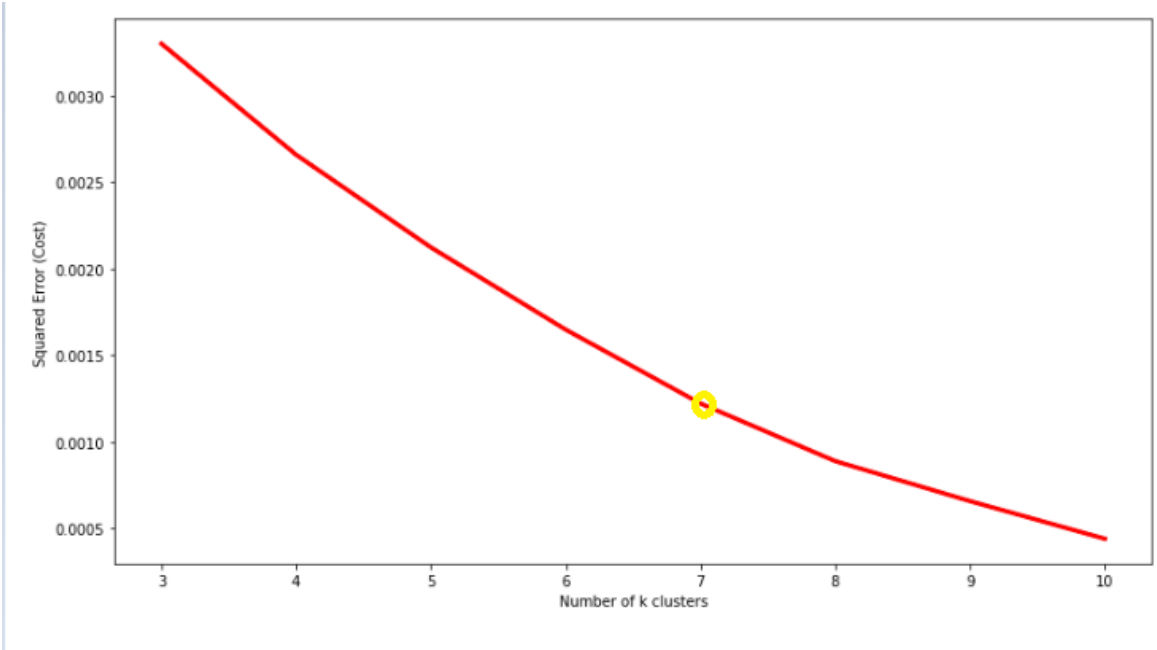


4. It is also observed that , some of the categories are unique to certain metro station locations only. Example of few such observations are given below:

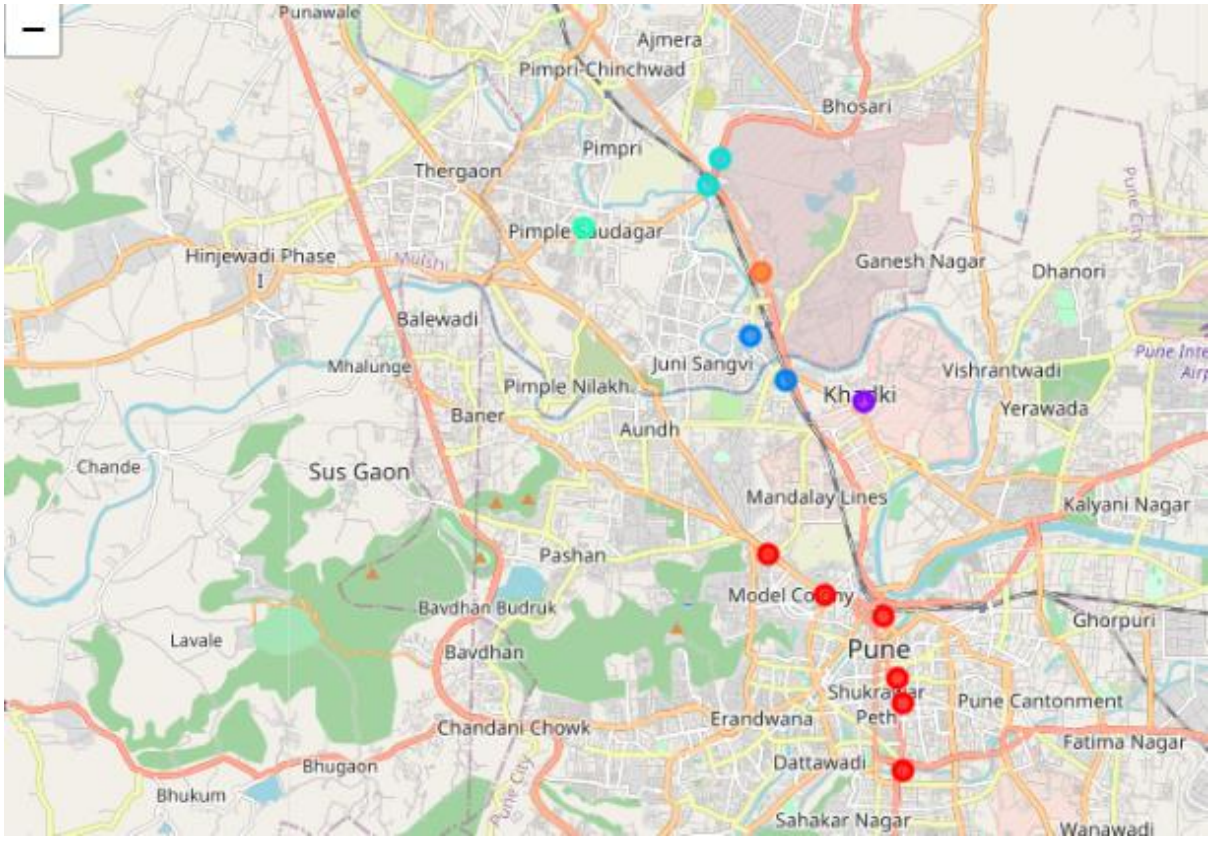
Venue Category	Station
American Restaurant	1
Fried Chicken Joint	1
Frozen Yogurt Shop	1
Fruit & Vegetable Store	1
Furniture / Home Store	1
Gastropub	1
Gift Shop	1
Hookah Bar	1
Wine Shop	1
IT Services	1

	Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
8	PCMC	18.598609	73.797170	KFC Restaurant	18.593440	73.786083	American Restaurant
124	Bopodi	18.570910	73.836426	Smiley House	18.562297	73.829573	Vietnamese Restaurant
338	Civil Court	18.527608	73.855151	LTI	18.534989	73.851935	IT Services
347	Civil Court	18.527608	73.855151	Marrakesh	18.524349	73.841362	Middle Eastern Restaurant
480	Swargate	18.499020	73.859070	Aufside	18.501410	73.872157	Gastropub
519	Swargate	18.499020	73.859070	Timber Market	18.509267	73.868149	Furniture / Home Store

5. We have chosen 7 clusters for our analysis from the result of Elbow method.



6. The folium map of individual clusters is shown below



Discussions

The metro stations are clustered into 7 clusters ranging from 0-6 as below:

Cluster -0

	Station	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
8	Range Hill	Lounge	Indian Restaurant	Coffee Shop	Chinese Restaurant	Café	Asian Restaurant	Italian Restaurant	Gym	Hotel	Multiplex
9	Shivajinagar	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Café	Sandwich Place	Multiplex	Chinese Restaurant	Restaurant	Ice Cream Shop	Dessert Shop
10	Civil Court	Indian Restaurant	Sandwich Place	Ice Cream Shop	Seafood Restaurant	Hotel	Bar	Café	Fast Food Restaurant	Bakery	Historic Site
11	Budhwar peth	Indian Restaurant	Snack Place	Vegetarian / Vegan Restaurant	Ice Cream Shop	Seafood Restaurant	Fast Food Restaurant	Convenience Store	Dessert Shop	Tea Room	Café
12	Mandai	Indian Restaurant	Snack Place	Vegetarian / Vegan Restaurant	Ice Cream Shop	Jewelry Store	Dessert Shop	Coffee Shop	Seafood Restaurant	Convenience Store	Men's Store
13	Swargate	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Men's Store	Southern / Soul Food Restaurant	Multiplex	Shopping Mall	Snack Place	Mobile Phone Shop

Cluster-1

	Station	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
7	Khadki	Department Store	Train Station	Bakery	Food Truck	Fried Chicken Joint	Dessert Shop	Diner	Donut Shop	Dumpling Restaurant	Electronics Store

Cluster-2

	Station	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
5	Dapodi	Snack Place	Indian Restaurant	Ice Cream Shop	Department Store	Fast Food Restaurant	Food Court	Coffee Shop	Restaurant	Chinese Restaurant	Cafeteria
6	Bopodi	Indian Restaurant	Chinese Restaurant	Train Station	Bakery	Hotel	Italian Restaurant	Cafeteria	Food Court	Breakfast Spot	Diner

Cluster-3

```
pune_merged.loc[pune_merged['Cluster Labels'] == 3, pune_merged.columns[[0] + list(range(5, pune_merged.shape[1]))]]
```

	Station	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
2	Bhosari(Nashik Phata)	Bus Station	Gift Shop	Snack Place	Fast Food Restaurant	Italian Restaurant	Indian Chinese Restaurant	Hotel Bar	Hotel	Gas Station
3	Kasarwadi	Fast Food Restaurant	Bus Station	Platform	Cosmetics Shop	Gas Station	Hotel Bar	Indian Chinese Restaurant	Italian Restaurant	Snack Place

Cluster-4

	Station	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	PCMC	Fast Food Restaurant	Pizza Place	Indian Restaurant	Chinese Restaurant	Sandwich Place	Snack Place	Gym	Wine Shop	Ice Cream Shop	Lounge

Cluster-5

	Station	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
1	Sant Tukaram Nagar	Bus Station	Indian Restaurant	Sandwich Place	Chinese Restaurant	Café	Snack Place	Food Truck	Dessert Shop	Cosmetics Shop	Gas Station

Cluster-6

	Station	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
4	Phugewadi	Indian Restaurant	Fast Food Restaurant	Food Truck	Dessert Shop	Restaurant	Coffee Shop	Snack Place	Department Store	Hotel	Train Station

Clusters	Observations	Venue categories range
0	The cluster is central area of Pune city. This cluster is characterized by higher range of venue categories. There are certain unique categories that belong to this cluster e.g. American Restaurants, Middle eastern Restaurants , Gastropubs, Furniture Market etc. This indicates high local and outstation traffic in those metro stations.	45-100
1	This is a cantonment area with limited business opportunities.	5
2	The cluster has small snack places and restaurants and a growing area for future investment.	20-21
3	The suburbs are industrial areas of Pune. The cluster has Bus Stations for intercity travel. These are city exit points with Gas Stations.	16-19
4	This is an intra city transport hub with small fast food restaurants around.	22
5	There is a busy Bus Station and the locality has small local food and snack shops and has educational institutes.	15
6	Suburban area with lower footfall and characterized by variety of small restaurants.	13

Conclusion

KKK Ltd has received the initial analysis of the clusters and the first cluster identified as Cluster -0 is the most suitable place to open their first utility store. The Cluster-0 has multiple locations . Shivajinagar is the most attractive location with high 'Venue categories". Considering the high footfall and high venue categories , it can utilize the complete product mix for revenue generation for locations of Cluster-0.

Reference :

1. www.punemetrorail.org
2. Foursquare API.

The entire exercise is a part of Capstone project for IBM Data Science Professional Certificate. The problem is an imaginary in nature and this analysis is only a data science exercise.