Battle of Neighbourhoods - Delhi Metro Transit

Report created by Abhijit Vasili
(for the purpose of Capstone Project for the course completion requirement.)

Introduction

In this project we will be analyzing the neighbourhoods on the basis of the metro trains stations located in the city of Delhi.

There are 137 stations in the metro transit in the city of Delhi which is the most densely populated place in India. In this project the stations will be used to understand the city and its neighbourhoods.

Later on in the project K Means is used to cluster the data.

Here the locations and names of the metro station are loaded into a dataframe from a JSON file. The data is refered from https://raw.githubusercontent.com/dhirajt/delhi-metro-stations/master/metro.json, I would like to thank 'dhirajit' for creating the data and publishing it on open source GitHub for others to use the data.

Data Preparation and Cleaning

Here from the previous visualization, a few potential error stations have been dropped because of problematic latitude value.

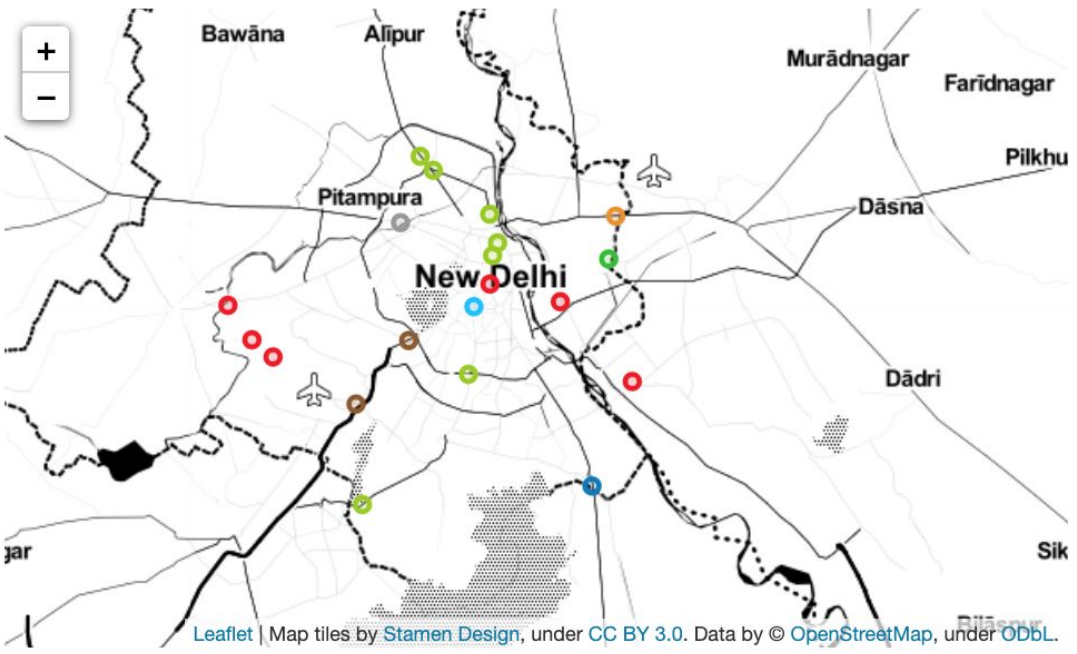
And only 20 stations have been selected because I was facing problems with the foursquare site where the limit for the day was consumed.

Out[8]:

	Line	English name	Layout	Coordinates
0	[Yellow Line]	Adarsh Nagar	Underground	28.71642,77.17046
1	[Yellow Line]	AIIMS	Underground	28.56892,77.20771
2	[Blue Line]	Akshardham	Elevated	28.61806,77.27869
3	[Blue Line branch]	Anand Vihar	Elevated	28.64695,77.31603
4	[Yellow Line]	Arjan Garh	Elevated	28.48076,77.12583
5	[Green Line]	Ashok Park Main	Elevated	28.67153,77.15527
6	[Yellow Line]	Azadpur	Underground	28.70696,77.18053
7	[Violet Line]	Badarpur	Elevated	28.49334,77.30307
8	[Blue Line]	Barakhambha Road	Underground	28.63003,77.22436
9	[Blue Line]	Botanical Garden	Elevated	28.56409,77.3342
10	[Yellow Line, Violet Line]	Central Secretariat	Underground	28.61474,77.21191
11	[Yellow Line]	Chandni Chowk	Underground	28.65785,77.23014
12	[Yellow Line]	Chawri Bazar	Underground	28.64931,77.22637
13	[Yellow Line]	Civil Lines	Underground	28.67726,77.2241
14	[Airport Express]	Delhi Aerocity	Underground	28.54881,77.12092
15	[Airport Express]	Dhaula Kuan	Elevated	28.59178,77.16155
16	[Red Line]	Dilshad Garden	Elevated	28.67592,77.32142
17	[Blue Line]	Dwarka	Elevated	28.61564,77.02197
18	[Blue Line]	Dwarka Sector 10	Elevated	28.58068,77.05682
19	[Blue Line]	Dwarka Sector 12	Elevated	28.59232,77.04051

Visualization of Stations on a Map

Out[51]:



Methodology

Exploring a 1000 meter radius around each station to explore the venues and count them and place them in the table.

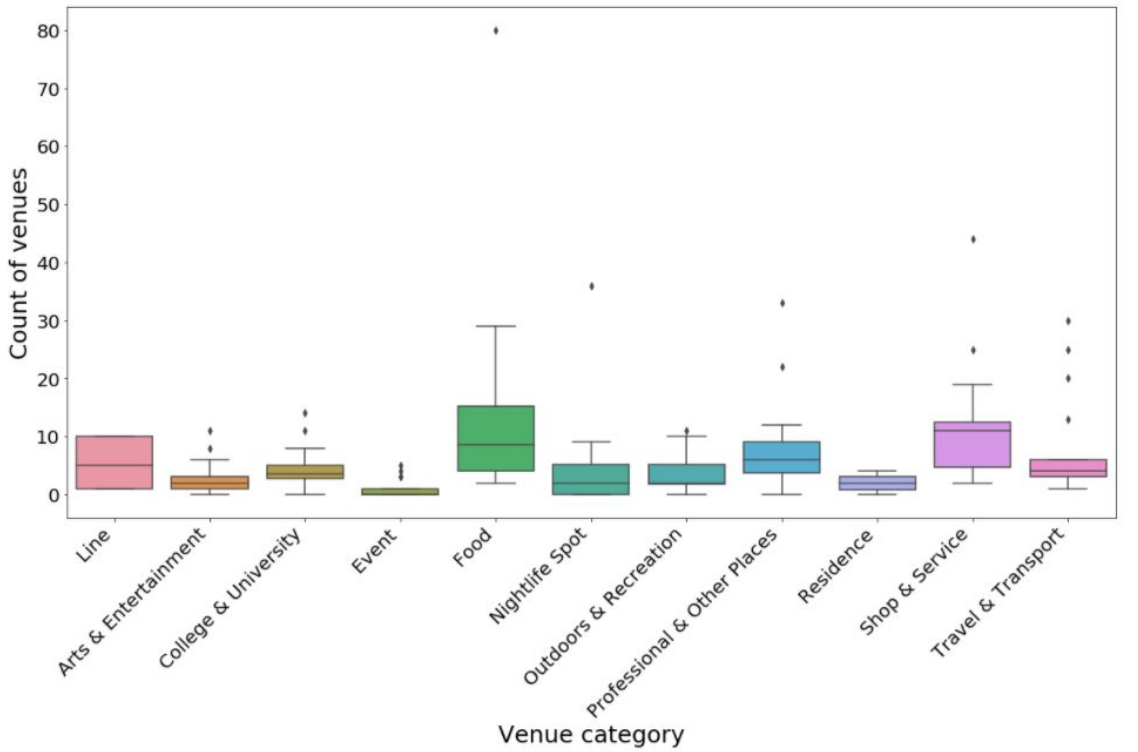
Out[24]:

	Line	English name	Layout	Coordinates	Arts & Entertainment	College & University	Event	Food	N
0	1	Botanical Garden	Elevated	28.56409,77.3342	4	4	1	27	
1	1	Dwarka	Elevated	28.61564,77.02197	0	1	0	2	
2	1	Dwarka Sector 10	Elevated	28.58068,77.05682	3	3	1	6	
3	1	Barakhambha Road	Underground	28.63003,77.22436	11	3	3	80	
4	1	Dwarka Sector 12	Elevated	28.59232,77.04051	1	3	1	12	

Out[24]:

Arts & ıment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
4	4	1	27	7	3	4	2	25	4
0	1	0	2	0	1	0	0	14	1
3	3	1	6	3	2	3	4	19	6
11	3	3	80	36	8	22	3	44	20
1	3	1	12	5	5	3	2	12	4

Lets understand the data. Visualizing the data in the form of Box Plots.



From the plot we can gather the idea that "Food" and "Shop & Service" are the most popular categories. Let's drop "Event" as it is not being a value for exploratory analysis.

Also normalizing the data for better analysis.

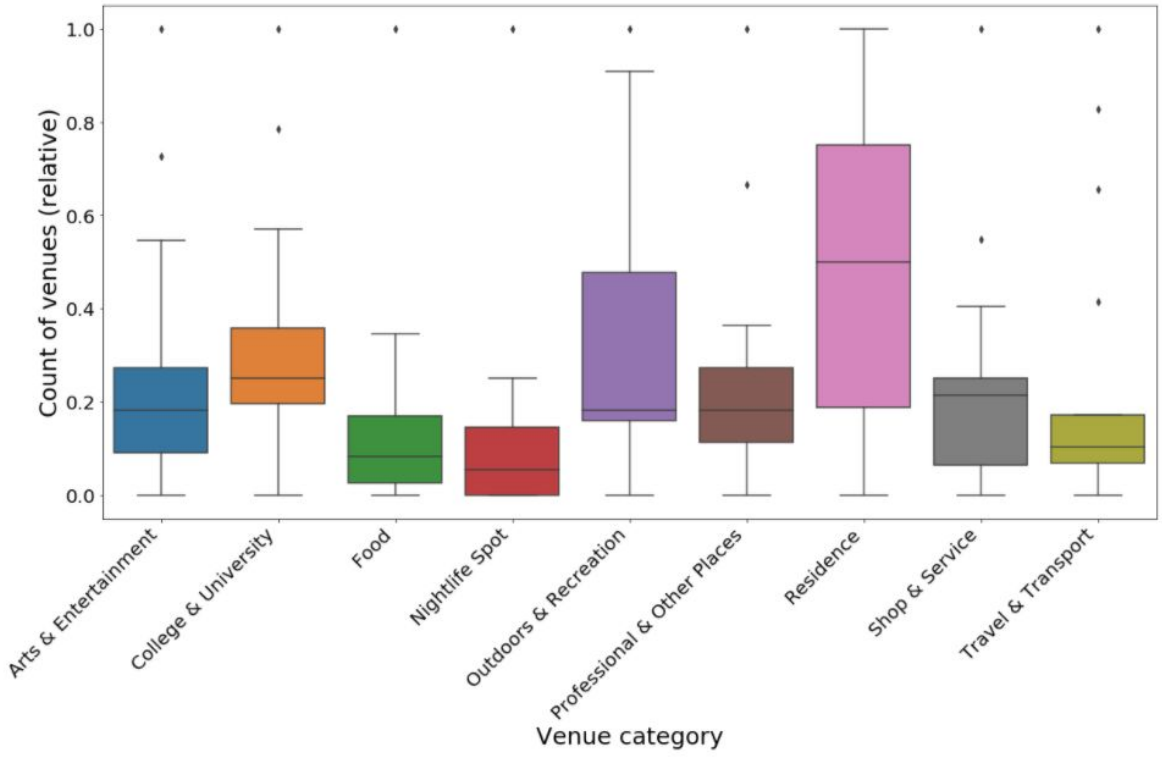
In [29]:
```
from sklearn.preprocessing import MinMaxScaler

X = stations_venues_df.values[:,4:]
cluster_dataset = MinMaxScaler().fit_transform(X)
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/sklearn/utils/vali
dation.py:595: DataConversionWarning: Data with input dtype object was
converted to float64 by MinMaxScaler.
  warnings.warn(msg, DataConversionWarning)
```

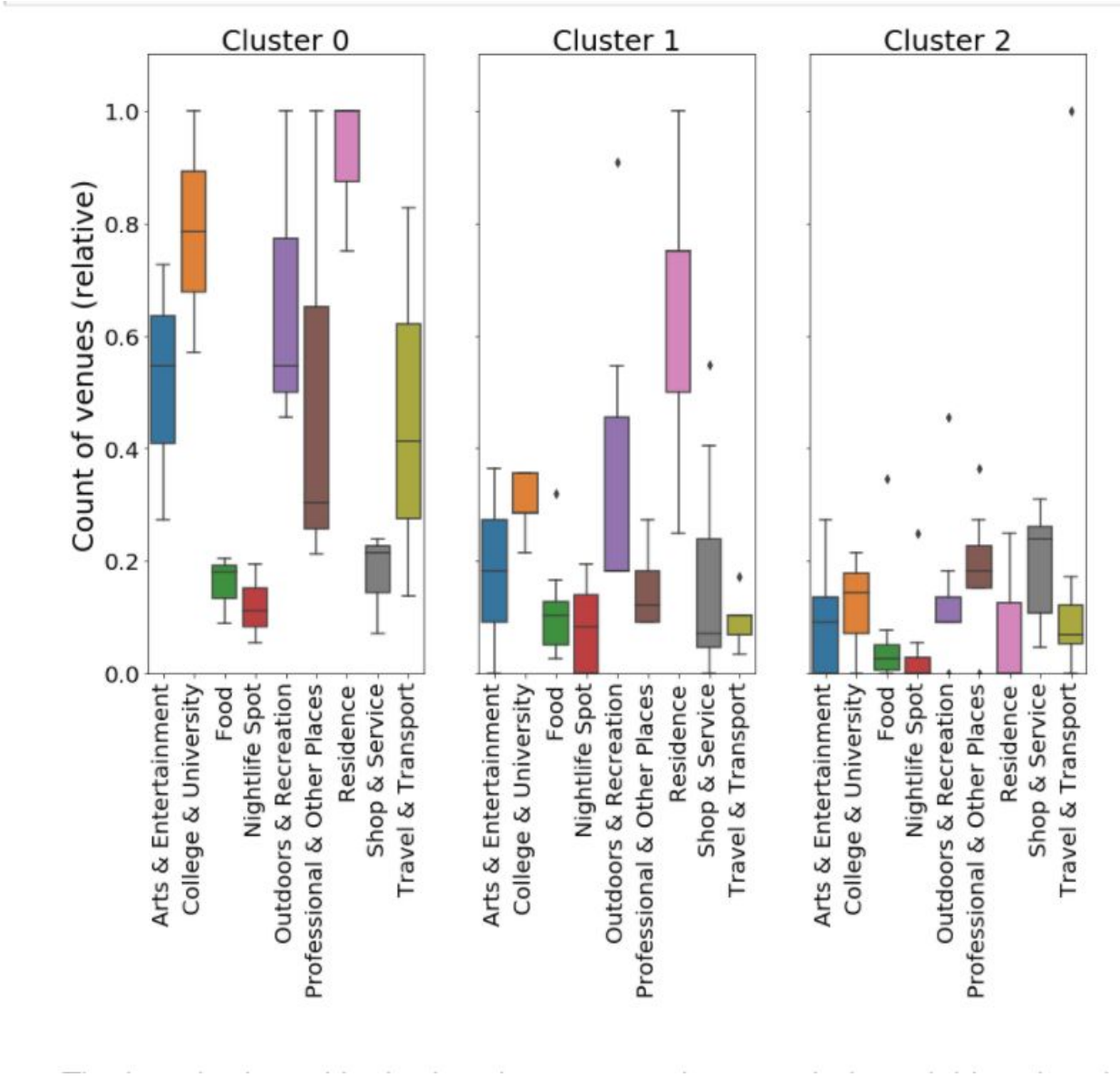
In [30]:
```
cluster_df = pd.DataFrame(cluster_dataset)
cluster_df.columns = [c[0] for c in categories_list]
cluster_df.head()
```

Out[30]:

Arts & ertainment	College & University	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transpor
0.363636	0.285714	0.320513	0.194444	0.272727	0.121212	0.50	0.547619	0.10344
0.000000	0.071429	0.000000	0.000000	0.090909	0.000000	0.00	0.285714	0.00000
0.272727	0.214286	0.051282	0.083333	0.181818	0.090909	1.00	0.404762	0.17241
1.000000	0.214286	1.000000	1.000000	0.727273	0.666667	0.75	1.000000	0.65517
0.090909	0.214286	0.128205	0.138889	0.454545	0.090909	0.50	0.238095	0.10344



Results



1) Cluster 0 has an overall high score for all venue categories. This is the most diversely developed part of the city. (Represented in Blue Blobs)

2) Cluster 1 has the highest marks for Residence Places and next highest score was outdoor and recreation. This must be a residential area.(Represented in Green Blobs)

3) Cluster 2 has an overall low mark. Best scores are observed in shop & service.(Represented in Orange Blobs)