Ideal Spot

For LondonCryo Expansion



Sanjay Navgale

TAble of  
Contents

Executive Summary 1

Description & Discussion of the Background 1

## Executive Summary

LondonCryo is new, unique and growing business in London. LondonCryo specialises in Sports Recovery, Pain Relief, Cryo>S 3.0 Slimming, Toning & Facials, X°Cryo – Local Cryo, Massage and IV Drips & Vitamin Shots.

LondonCryo is currently situated close to London Liverpool Street station and looking to expand in other areas within Greater London boroughs.

This report describes the whole process from problem description, mythology used, results and recommendation made based on detailed data analysis, visualisation and machine learning techniques to come up selection of post codes and nearest London public transport most suitable for LondonCryo to expand.

## Description & Discussion of the Background

Greater London is a ceremonial county of England that is located within the London region. This region forms the administrative boundaries of London and is organised into 33 local government districts—the 32 London boroughs and the City of London, which is located within the region but is separate from the county. The City of London Corporation is the principal local authority for the City of London, with a similar role to that of the 32 London borough councils.

For any business to expand the selection of location and proximity with public transport, friendly businesses, population and spending capacity available in nearby population is very important.

For type of services LondonCryo provides having gym or similar sports and fitness related venues/businesses around is more beneficial than having fast food or convenience store. Higher population and population with higher net earnings is more likely to result into higher footfall for LondonCryo

## Data available and usage

With the problem discussed the consideration was given to what data we can use and will be beneficial from the vast available data set

1. London\_postcodes.csv

source (https://www.cdrc.ac.uk/geodatalondon/)

This has data related to each postcode in greater London area. The attributes available in this data are - Postcode', 'In Use?', 'Latitude', 'Longitude', 'Easting', 'Northing', 'Grid Ref', 'County', 'District', 'Ward', 'District Code', 'Ward Code', 'Country', 'County Code', 'Constituency', 'Introduced', 'Terminated', 'Parish', 'National Park', 'Population', 'Households', 'Built up area', 'Built up sub-division', 'Lower layer super output area', 'Rural/urban', 'Region', 'Altitude', 'London zone', 'LSOA Code', 'Local authority', 'MSOA Code', 'Middle layer super output area', 'Parish Code', 'Census output area', 'Constituency Code', 'Index of Multiple Deprivation', 'Quality', 'User Type', 'Last updated', 'Nearest station', 'Distance to station', 'Postcode area', 'Postcode district', 'Police force'.

This has enough information required to identify unique stations and corresponding postcode, latitude, longitude required to foursquare venues.

1. MSOA\_earningsdata.csv

Source: (<https://data.london.gov.uk/dataset/earnings-place-residence-borough>)

The subset of data from original excel data was converted into csv for this purpose. The net earnings after housing expenses were considered for this exercise.

## Foursquare venue data and usage:

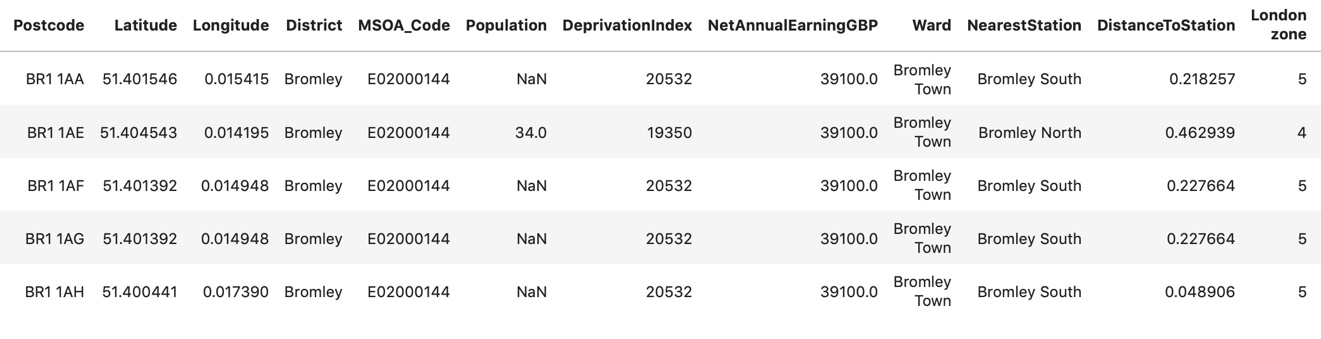
## Foursquare venues are used to get list of all available venues around the public transport station. As Foursquare does not have filter available for venue categories, the results are later filtered to remove not so important venues from the results. The top 10 venues are used to create 10 K-mean clusters and each of the station is marked on the map.

## 

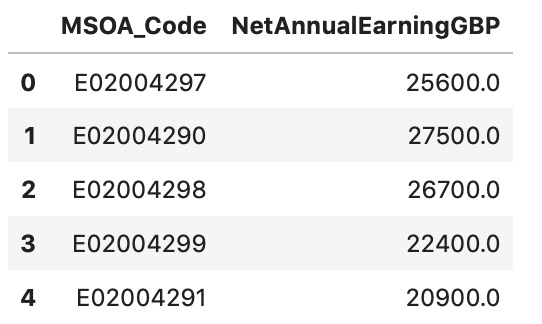
## Methodology Used

I used python jupyterLab notebook on [IBM Cognitive Lab](https://labs.cognitiveclass.ai/). The source code and data was stored on GitHub repository.

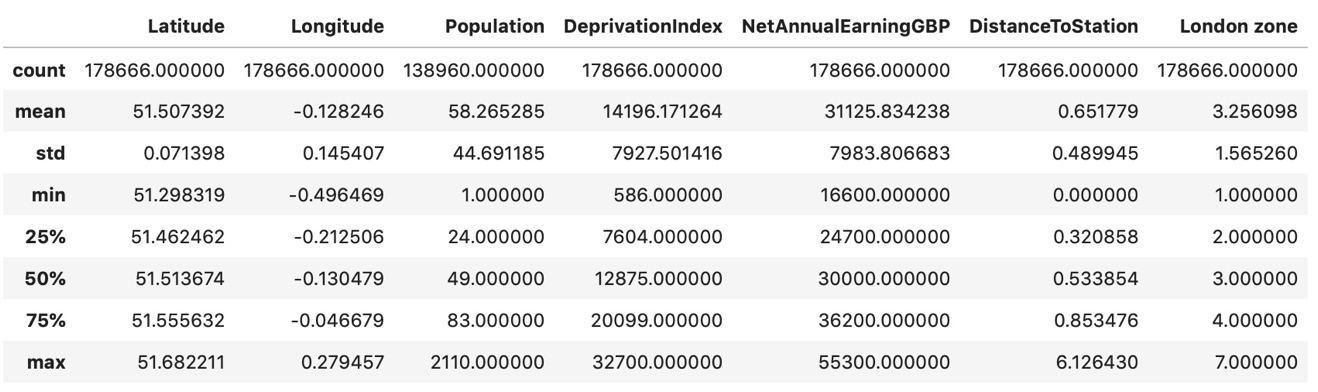
The starting point was the London postcode data with Postcode, London Zone, Lattitue, Longitude as key data elements.



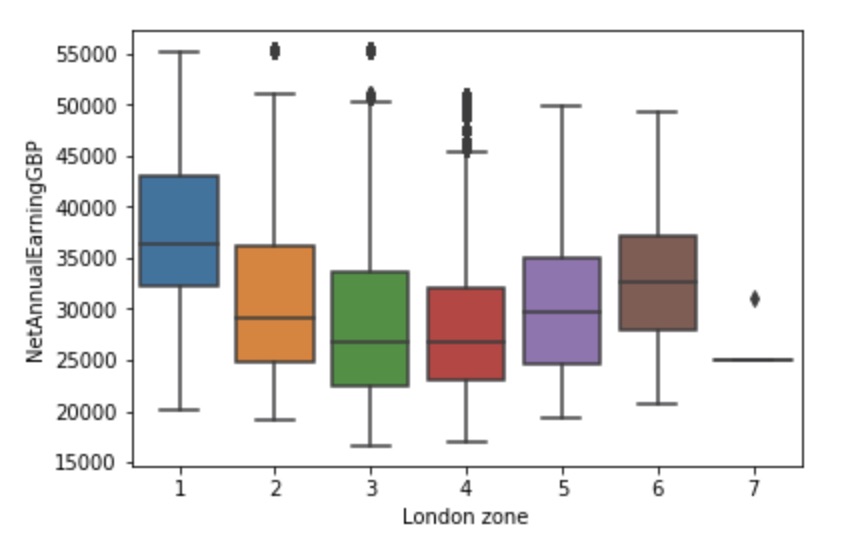
Earnings data was taken and merged with London postcode data.



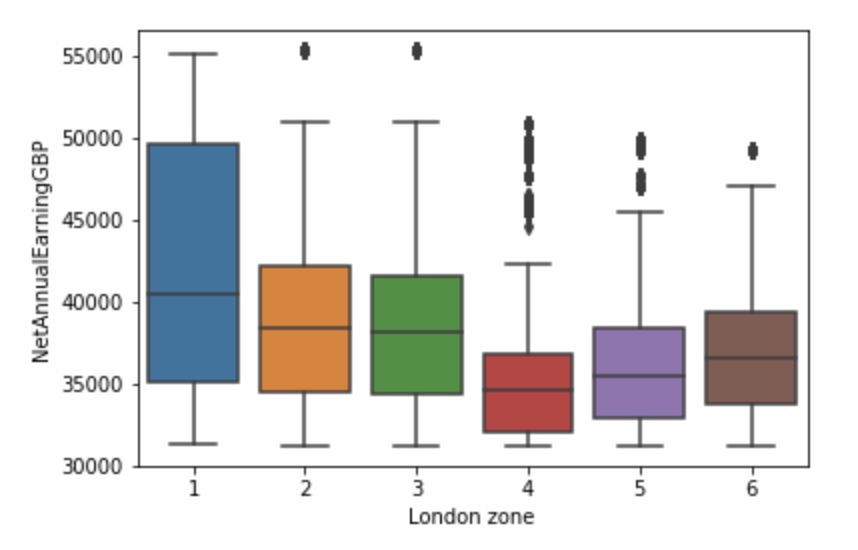
The python pandas dataframe.describe is useful simple command to understand the numerical data available within your dataset.

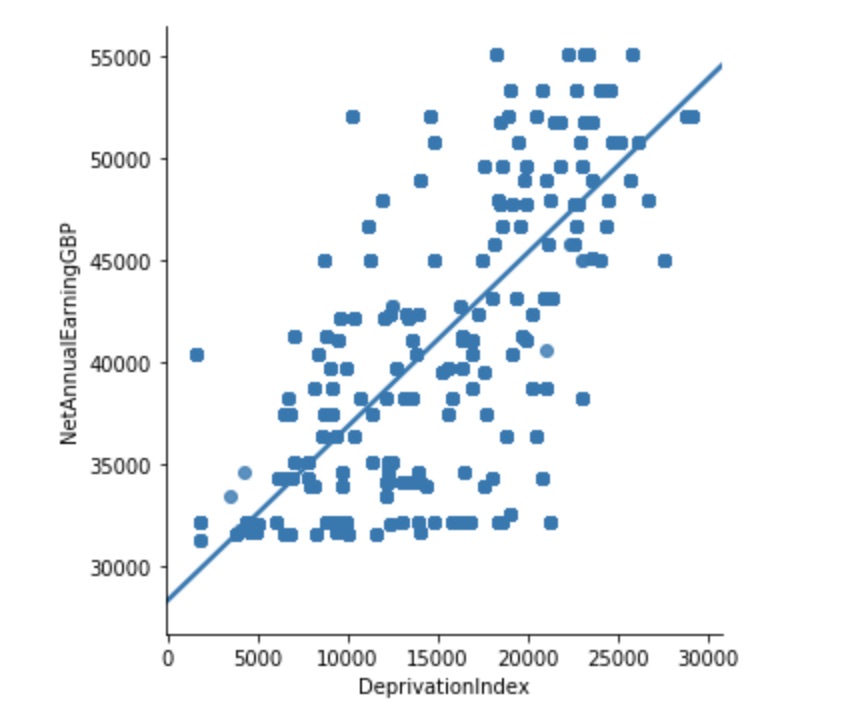


Python Matplotlib and seaborn was used to visualize the data before doing any further analysis with data. The box plot below shows the earnings pattern across London zones and it shows net average earnings reduces as you move towards zone 2, 3 and 4 and again starts increasing from zone 5 and 6.

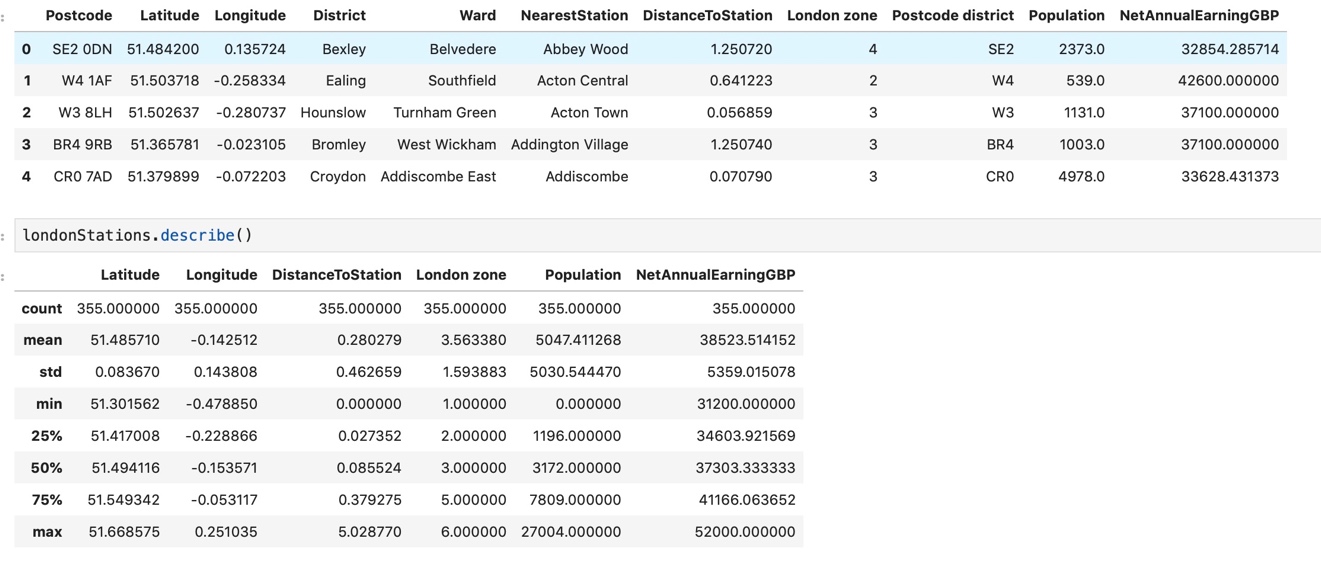


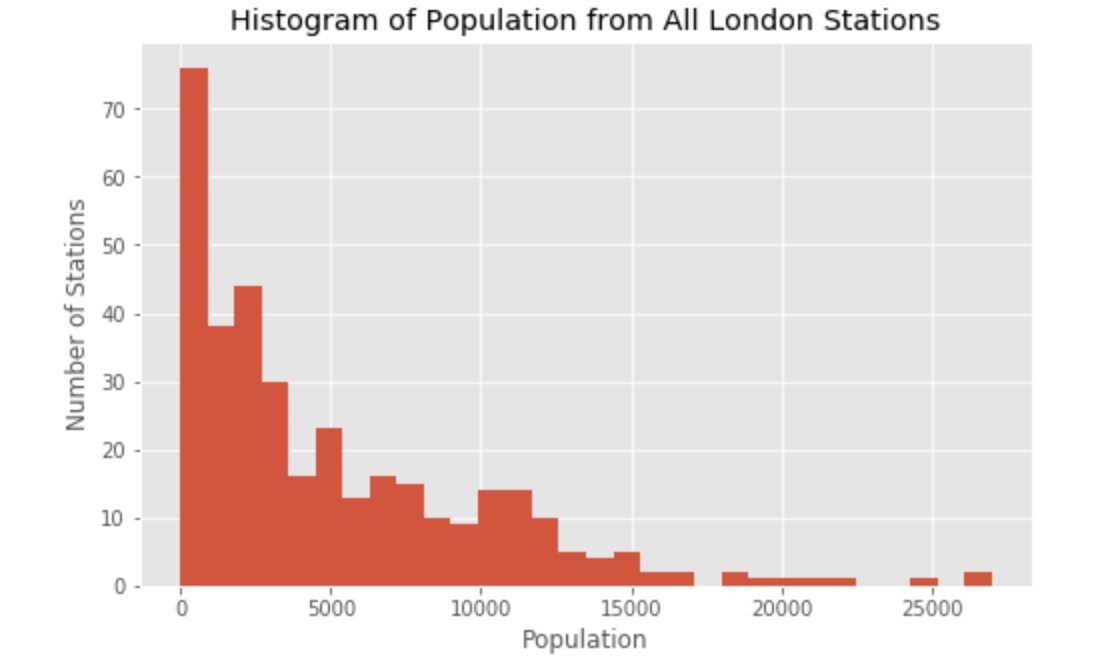
Data was reduced by considering the Net Annual Earnings and considering only above mean level of earnings across the greater London area. The next box plot shows the box plot for reduced dataset.



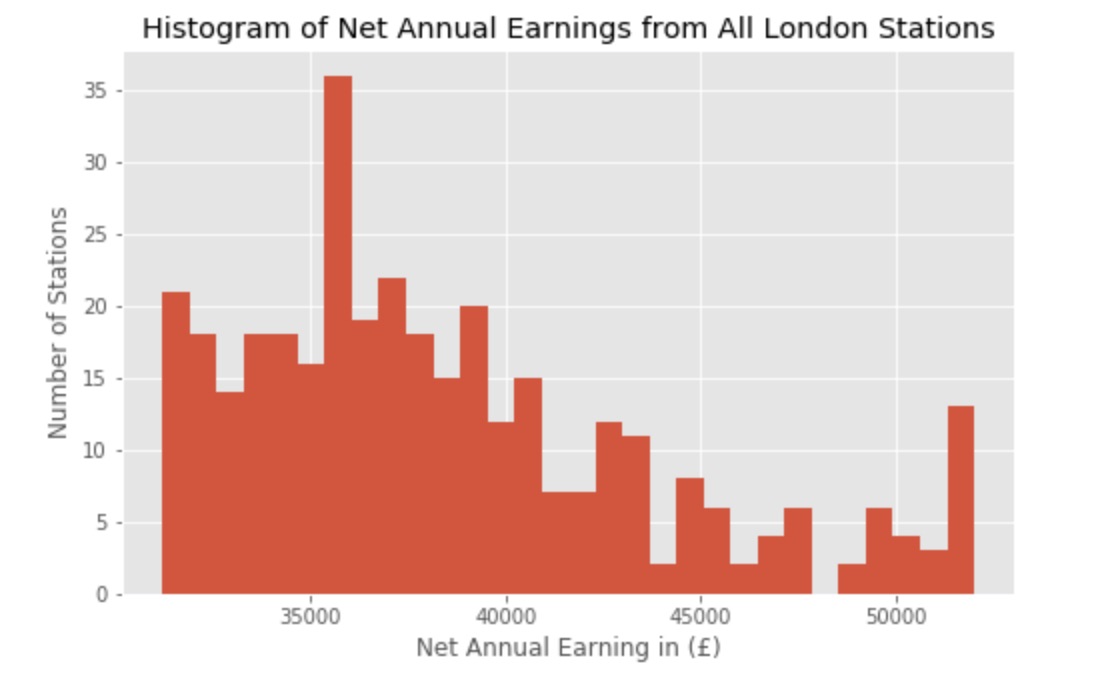
The plot below shows the correlation between net earnings within zone 1 and deprivation index available in data.

The combined dataset with postcodes data and earnings data against each London public transport station is as below.



The histograms below shows population served by each of the London station. 

The next histogram shows average net earnings against the number of stations.

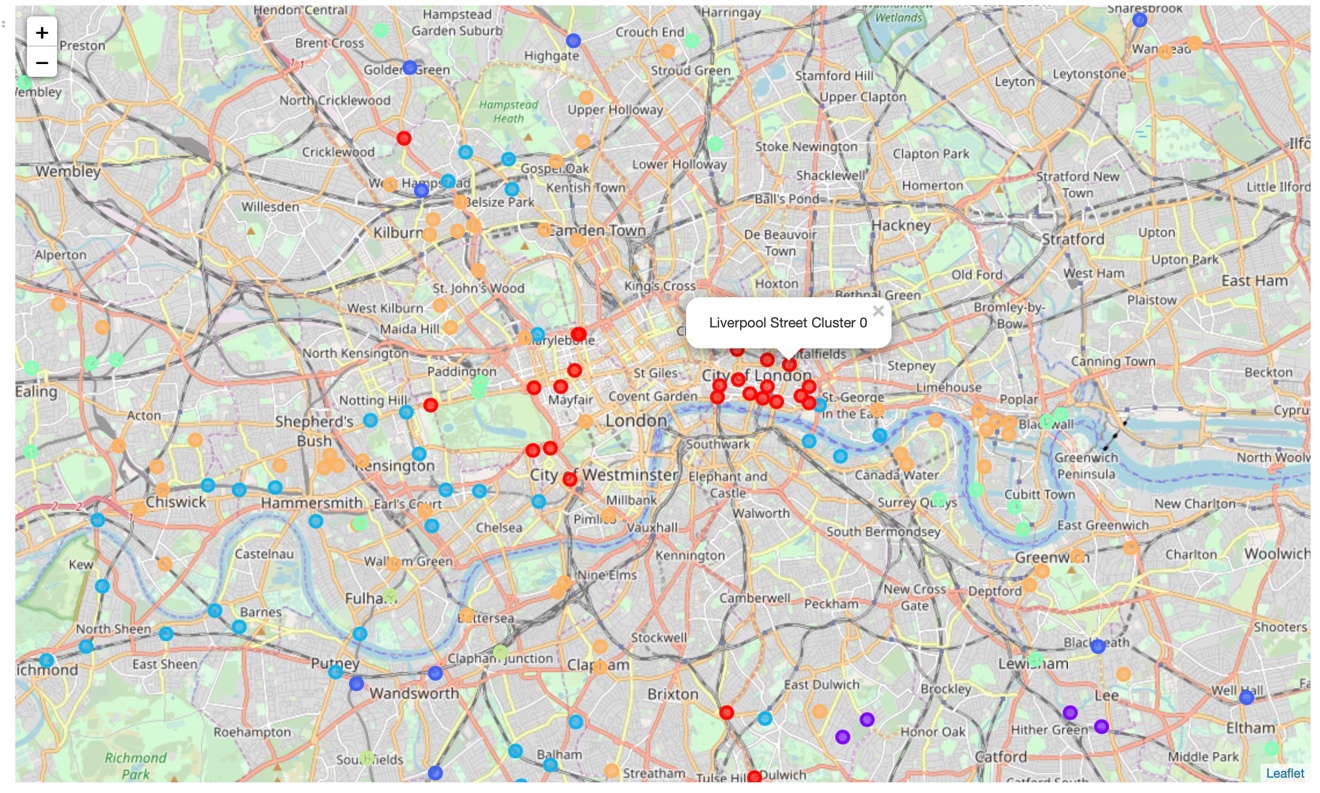


The foursquare APIs are used to extract list of 100 venues and venue categories within 750 meters of each London public transport station. These venues then arranged into top 10 most common venues.



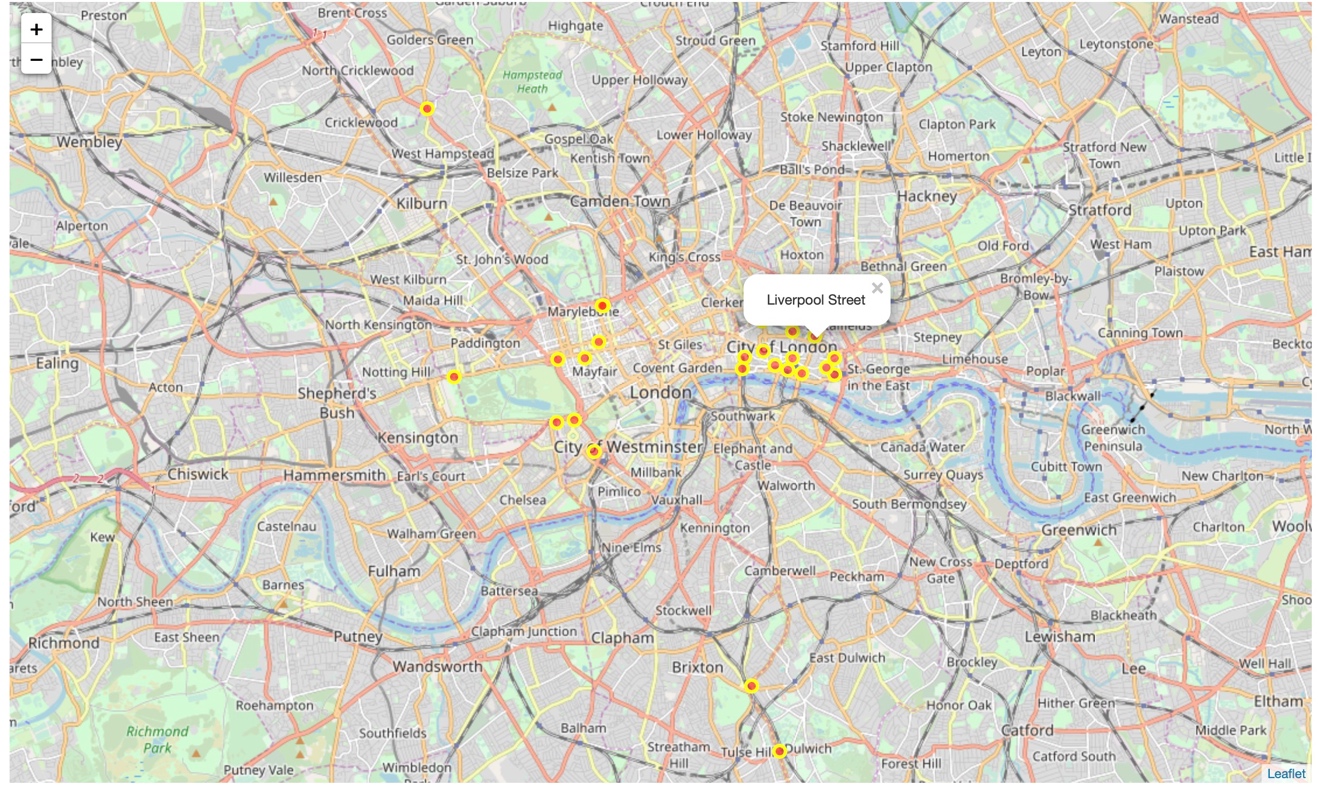
The population and earning data was normalized before it was put through K-nearest Neighbors algorithm to create 10 clusters based on top 10 common venues, population and Net annual earnings.

The first Map below shows all 10 clusters.



## Result

The selected cluster based on current location of LondonCryo and friendly venues was chosen and below map shows all the locations from the chosen cluster.



The table below shows the list selected for recommendation for LondonCryo to consider for expansion which are expected to provide similar financial performance as current LondonCryo as well as the most preferred locations within greater London area.



## Discussion

The result clearly shows that London Zone 1 is most lucrative for LondonCryo which still considered to be niche and new business offerings.

## Conclusion/Recommendations

I recommend LondonCryo to consider this list of 27 postcodes/London public transport station for their next expansion. All these locations are expected to perform similar in terms of performance. The final selection of the location can be reduced by further considering the investment cost across these locations to evaluate return on investments.

## Reference:

* [1] [London — Wikipedia](https://en.wikipedia.org/wiki/Greater_London)
* [2] [London Postcode data](https://www.cdrc.ac.uk/geodatalondon/))
* [3] [London Earnings data](https://data.london.gov.uk/dataset/earnings-place-residence-borough)
* [3] [Forsquare API](https://developer.foursquare.com/" \t "_blank)
* [4] [LondonCryo](https://www.londoncryo.com/)