Introduction and Business Problem

London Underground has 270 stations and is one of the largest networks in the world. It is used by about 5 million people on an average day.

This project attempts to classify the neighbourhoods surrounding these stations. Some neighborhoods are mostly residential, others more commercial areas surrounding them. The venues closest to a station determine why and how people use it. For instance if there are no commercial areas in a neighborhood its residents are likely to travel to other areas for work. This creates daily travel and movement of people.

By analyzing this data we can classify stations by their primary usage. This data can be useful for urban and city planners to determine travel motives for work or leisure, and plan further development of the underground network and find places for new activity.

Data

Analysis is based on location of underground station data and areas close to stations.
List of stations and their geographical coordinates is scrapped from
https://wiki.openstreetmap.org/wiki/List_of_London_Underground_stations.

Public Domain Tube Stations List, located using a GPS

Name	Latitude	Longitude	Platform / Entrance	Collected By	Collected On	Line
Acton Town	51.502500	-0.278126	Platform	User:Gagravarr	24/11/06	District, Piccadilly
Acton Central	51.50883531	-0.263033174	Entrance	User:Firefishy	08/05/2007	London Overground
Acton Central	51.50856013	-0.262879534	Platform	User:Firefishy	08/05/2007	London Overground
Aldgate	51.51394	-0.07537	Aldgate High Street entrance	User:Morwen	28/4/2007	Metropolitan
Aldgate East	51.51514	-0.07178	Entrance	User:Parsingphase	(2006)	District, Hammersmith & City
Alperton	51.54097	-0.30061	Platform	User:Mattwestcott	28/01/2007	Piccadilly
Amersham	51.67435	-0.60732	Entrance	User:Kake	2007-03-06	Metropolitan
Angel	51.53253	-0.10579	Entrance	User:Kake	2007-04-14	Northern
Archway	51.56536	-0.13474	Junction Road entrance	User:Kake	2007-04-24	Northern
Arnos Grove	51.61625	-0.13355	Entrance	User:Kake	2007-04-24	Piccadilly
Arsenal	51.55847	-0.10561	outside Gillespie Road exit; the platforms are a long walk west in tunnel	User:Morwen	5/5/2007	Piccadilly
Baker Street	51.52265	-0.15704	Circle/Hammersmith & City Entrance (probably)	User:Steve	11/2/06	Circle, Hammersmith & City

Foursquare API to explore area types surrounding each station. Foursquare outlines these high-level venue categories with more sub-categories.

- Art Gallery 4bf58dd8d48988d1e2931735
- Football Stadium 4bf58dd8d48988d189941735
- College & University 4d4b7105d754a06372d81259
- Food 4d4b7105d754a06374d81259
- Fish & Chips Shop 4edd64a0c7ddd24ca188df1a
- Nightlife Spot 4d4b7105d754a06376d81259
- Shop & Service 4d4b7105d754a06378d81259
- Residence 4e67e38e036454776db1fb3a

We will query the number of areas in each category in a 500m radius around each station. This radius was chosen because 500m is reasonable walking distance in London.

Methodology

We can use the Foursquare explore API with category ID to query the number of venues of each category in a specific radius. The response contains a totalResults value for the specified coordinates, radius and category.

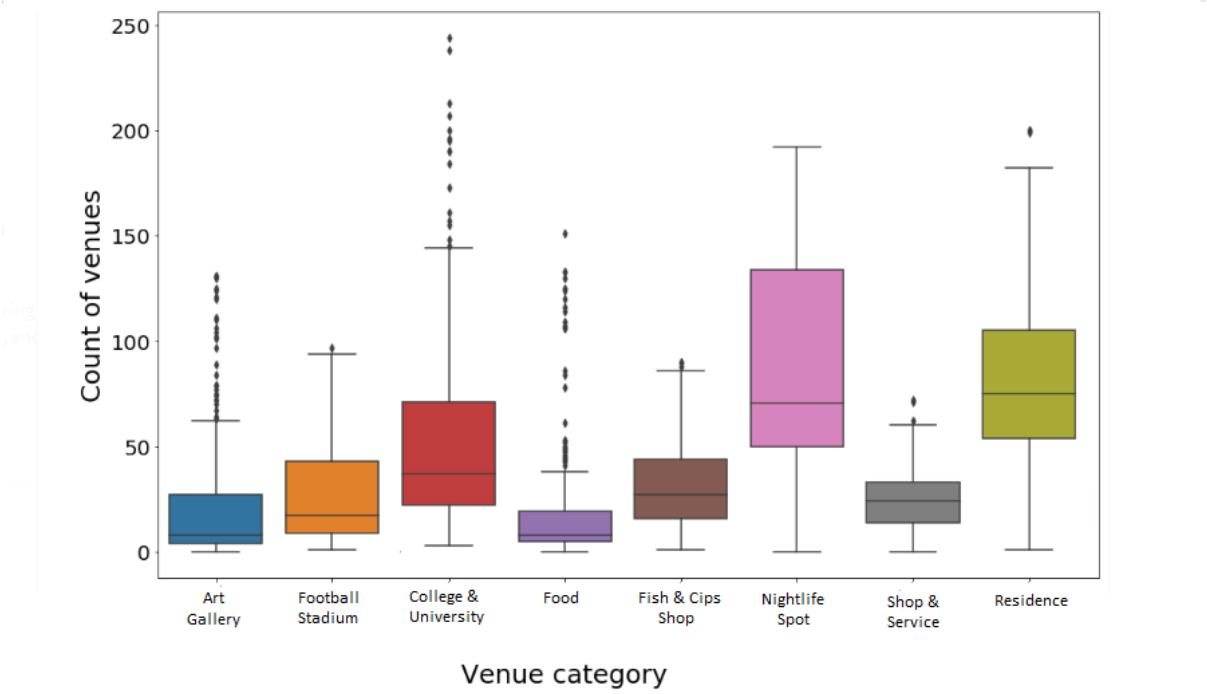
GET
https://api.foursquare.com/v2/venues/explore?client_id={{client_id}}&client_secret={{client_secret}}&v={{v}}&ll=51.51394,-0.07537&radius=500&categoryId= 4d4b7105d754a06374d81259

Response:


```
{
"meta": {
"code": 200,
"requestId": "5cfec0e31ed21914c1db7dd9"
},
"response": {
"suggestedFilters": {
"header": "Tap to show:",
"filters": [
{
"name": "Open now",
"key": "openNow"
}
]
},
"headerLocation": "Aldgate",
"headerFullLocation": "Aldgate, London",
"headerLocationGranularity": "neighborhood",
"query": "Food",
"totalResults": 102,
<...>
}
```

Analysis

Display of number of venues as boxplots. The most frequent venue categories are Nightlife Spot and Residence followed by College & University.



Using K-Means Clustering, the results with different clusters are below.

Visual presentation of cluster profiles of boxplots

Results

Summary of characteristics of clusters by venue scores:

- Cluster 0 (Blue) has consistently high scores for all venue categories. This is the most artistic and diversely developed part of the city. This is the oldest and central part of the city.
- Cluster 1 (Orange) has most football stadiums and nightlife. This is the recreational part of the city.
- Cluster 2 (Red) has low marks across the board. This appears to be underdeveloped areas.

Discussion

Cluster 0 is the central most part of the city.
Cluster 1 is the middle area of the city between central core and outer boundaries.
Cluster 2 is the boundary areas of the city.

The highest number of venues are in the Food and Fish & Chip Shop categories. Data doesn't take into account a venue's size and unique characteristic features necessarily. The central part of city is oldest with plenty of history, so has most art and culture attractions, this is clearly visible from a comparison of the three different clusters which when viewed on a map show differing characteristics.

Conclusion

The above data and analysis provides an insight into the city's growth and development indicating areas for future extension in line with inhabitants' preferences. This data should be combined with other insightful data like population and demographics to provide more accurate results.