**Criteria of Data Evaluation**

We will explore the neighborhoods within the key commercial areas in these cities, i.e. City of London & Westminster in London and Manhattan in New York. The scope of this report is limited to neighborhoods within these areas.

Findings of this report will be around suitability of places, we will analyze:

* Recommended venues within these places and how they are spread around the location
* Primary Criteria:
  + Gen Y & Gen Z prefer comfort living, we will look at the availability of Residential Accommodation (High Priority) within the neighborhoods as this is to do with moving to these cities
* Secondary Criteria:
  + Presence of Schools
  + Access to Medical & Dental Facilities
  + Gym and Fitness Centers
  + Parks & Playgrounds
  + Childcare Services
  + Laundromat and Laundry Services
  + Presence of Lifestyle Venues
    - Shopping Malls
    - Night Clubs and Lounge
    - Pubs and Sports Bars
    - Movie Theaters and Performing Arts Venues
    - Restaurants

We want to pinpoint neighborhoods that provide a consistent quality of facilities and features within a short distance of the neighborhoods

**Data Source and Method**

1. We will obtain information about the neighborhoods via various sources, this will include wikipaedia, Google Map and .json files provided during the course.
2. Data required at this stage is Post Code, Borough, Neighborhood, Latitude and Longitude, information will be taken for various wikipaedia, Google Map and .json files and doogal.co.uk has information about Latitude and Longitude for City of London and City of Westminster
3. Data Wrangling methods will be applied to standardize the data and bring it into the right format for usage
4. We will use Foursquare API’s, to get information about these neighborhoods. We will use the:
   1. Venue Explore Option: Returns a list of recommended venues near the current location. This result will let us know what are the popular and what the people living within these locations like doing.
   2. Venue Search Option: This will Returns a list of venues near the current location, matching a search category ID that will be provided based on the criteria mentioned above. This will let us conclude if the comfort living that Gen Y and Gen Z are looking for is possible.
5. We will adopt K-means Clustering as a way of further exploring the data and segmenting it into clusters with similar features. This will be done for all data received for both the Explore and the Search options
6. The Clusters will be displayed on Folium Maps to ensure that the data is still intact and relevant to the analysis.
7. We will explore that data in each neighborhood further, but, looking into businesses and services.
8. Data will be visualized in bar charts to identify if the best clusters which meet the criteria of data evaluation. Clusters from the Search query will be selected, based on how strongly they meet the criteria of data evaluation
9. Clusters from the Explore query will be selected based on size.
10. The neighborhoods from each selected cluster will be made into Sets
11. An intersection of the Neighborhood’ selected from the Search query cluster and a single Explore query cluster will be the recommended cluster

**Limitations of approach and analysis**

The explore query returns recommended venues and these are then selected. As part of this project we will be identifying the intersection of the selected ‘Search’ query cluster and the largest ‘Explore’ query cluster to recommend neighborhoods that offer a certain level of features uniformly. This will be the best fit result.

The limitation is because we can have intersections between the selected ‘Search’ query cluster, and any ‘Explore’ query cluster. Assuming the intersection is not a null set it’s a potential target neighborhood to consider that offers a different level of features uniformly.

This will depend on individual target audience and for the purpose of this report the other option are ignored.