**Data**

Data Requirements

As stated, the data has to be from the Manhattan borough of New York, NY, and Inner London.

Two sets of data will be needed for each city, one with the price of the average rent for two-bedroom apartments, and the other with the venues per neighborhood.

The data from both cities must be from comparable periods. Because the relative small inflation in both countries, the rates can be from periods separated by more than a year, but a year will be used as the limit for precision sake. Given the tendency and speed of change of neighborhood character and identity, the data has to be no older than 2017, so the results are still valid at the moment, and can be considered contemporary between the different data sets.

The rates will all be presented in United States dollars, so the appropriate exchange rate will be used to convert the pound sterling. An average during the relevant period will be used to have a more accurate result.

The values will be rounded to have a more clean and clear presentation and can be easier to the user to interpret.

Since the interest is in the neighborhood as a cultural and economic center, the geographical specific limits are not important, since those things tend to change gradually across the borders, therefore the neighborhoods will be represented by a circle of 500mts of radius centered on the geographical center. The venues inside that circle will be used as representative of all the venues in the neighborhood.

To this end, a list of the geographical coordinates of all the centers of the neighborhoods have to be gathered.

Data Collection

The data from the venues will be gather from Foursquare because it provides a wide selection of venues, not only places of business, but also recreational non-profit areas like parks, to have a more complete understanding of the neighborhood.

The ten most common venues of each neighborhood will be used to compare them, since the intention is not to have a comprehensive list, but a general “feeling” of the character. Any venue with very few instances would be not only irrelevant, it can be prejudicial for this project, since it can give false positives. An example would be a pet store in a heavily business oriented neighborhood, or a single disco in a residential area.

The ten selected venues in each neighborhood will be inspected to determine if it’s a valid instance. For example there had been cases of the actual neighborhood appearing as a venue.

For the data on the rent rates of London, the official information provided by the Government of the UK through the Valuation Office Agency will be used. A yearly inform is released with all the data needed. It is already presented in an Excel file table, it is free and reliable, therefore is the best option. Only the average rates of the neighborhoods will be used, discarding the rest of the data on the table (number of people rent, for example).

The selection of neighborhoods of Inner London will be according to the ones expressed in the rent rate table. For the geographical coordinates of each neighborhood, a table provided by the Government will also be used.

In the case of New York city, no official records were found, therefore private information will have to be used. A Business Insider article was found that states the average rates for 28 neighborhood, citing as source the apartment listing company RentCafé, which has a comprehensive database of rental data from all the USA.

The data will have to be scraped from the website of Business Insider and use a parsing tool like Beautiful Soup to be used and converted into a pandas dataframe. The 28 neighborhood will be the ones used in this project.

For the coordinates of the Manhattan neighborhoods, a list provided by the course will be used.