**IBM Data Science Capstone Project**

**Data description**

In order to build a neighbourhood recommendation model, the following data classified in two blocks is required:

* Data gathered from the user – (The user is the person who is looking for a neighbourhood to live)
  + Rent willing to pay
  + % of variation of rent to pay, i.e. 20% variation
  + List of top 5 venue categories ranked, i.e. 1) Coffee Shop, 2) Japanese Restaurant, etc.
  + The office where the user will be relocated. It could be the office description, address, or location
* The average amount of rent per neighbourhood
* Venues data – Gathered from Foursquare
  + List of venues and their category per neighbourhood in a radius to be specified
* Neighbourhoods geolocation data
  + Latitude and longitude of all neighbourhoods considered
* List of the top five boroughs to live in Mexico City, based on statistical information or research studies
  + Used to narrow down the set of neighbourhoods and data required associated with them

**Data usefulness**

Data gathered from the user will be used to get her preferences, which are the key differentiator from other recommender solutions. In addition to this, the rent amount and allowed variation will be used to filter neighbourhoods within the budget range.

It is essential to have a list of the average rent per neighbourhood in order to identify which neighbourhoods are the range specified by the user and discard neighbourhoods that are out of the budget range.

The complement of having the user’s venue categories preferences is to have the list of venue categories per neighbourhood. With all this information, a recommender system can be developed to search for neighbourhoods that are closer to the user’s preferences.

Mexico City is a place known for its dense traffic, due to this fact, the distance from home to the workplace is critical. Here is where the geolocation of each neighbourhood comes into play, in order to calculate the distance to the workplace by using the latitude and longitude.

Last but not least, Mexico City is also well-known for its size, this means that there are a lot of neighbourhoods dispersed in a bit less than 1,500 . Considering this, it is important to narrow down the list neighbourhoods. Hence, the data to handle. One way to do this by reviewing studies or research papers to limit the options to safe zones or boroughs that provide a better quality of life.