IBM Applied Data Science

Capstone Project – The Battle of the Neighborhoods

**Introduction**

We have been approached by our client to research possible locations in London for their new organic coffee and juice café. They currently have one café located near Hackney Central station which has proved very successful, so they are looking to open a second café located in or near a London station where the new location matches their current cafés location in relation to the amount of people who pass through the station each day and the makeup of the neighborhood surrounding the station.

**Data**

To try an identify a similar location to the one their current café resides in, we will first identify suitable stations with a similar through put of people each day. The Wikipedia page List of London railway stations (<https://en.wikipedia.org/wiki/List_of_London_railway_stations>) provides a full list of London stations along with a geographic location (latitude and longitude). It also contains a categorisation for each station which is based on the number of trips per annum that are made from the station. We will use this category to match stations with a similar amount of trips per annum to the station their current café is located near. Once we have identified suitable stations we will then use the Foursquare data to profile the neighborhoods surrounding each station to find a similar makeup of shops and businesses that best match their current location.