The Battle of Neighborhoods

1. Introduction

Business problem: London is the capital of England, UK, with a population of almost 9 million. It's a global financial and commercial hub, so it's a real challenge for entrepreneurs from all over the world to invest in any field of industry.

A very well-known Asian healthy Food and Supplements Stores-chain, wants to expand its businesses in Europe by opening, firstly, a new store in London. Despite their online-shops, company's success is based, mainly, on small stores providing healthy food-snacks and supplements to people who have adopted or want to adopt a healthy lifestyle. Board of directors of the company has already decided to try to apply the same recipe of success in London and they need us to investigate some good locations for that purpose.

We are going to collect all the appropriate data, first, and then we will analyze them with the help of K-Means algorithm in order to finally give our suggestions to directors.

2. Data

In order to build our recommender system, we need lots of data and, especially, specific data defining what venues exist in each neighborhood and where they are. In this way, we can examine which places can attract more of our providing products, for example venues like gyms and fitness centers. So, first, we need to find how many neighborhoods London has and what their names are. This data will be acquired from wikipedia. Secondly, we need to find their exact locations(Latitude, Longitude) from Google Maps Geocoding API. We need this info, beacause by using these coordinates in Foursquare, we can find nearest veneus for each neighborhood. The result will be a table containing info of most common venues in each neighborhood and, thus, our data is ready for further process.