

Coursera Capstone
IBM Applied Data Science Capstone
Opening a New Shopping Mall in Belgrade, Serbia

By: Vojin Tadic

December 2019

Data

To solve the problem, we will need the following data:

- List of suburbs in Belgrade. This defines the scope of this project which is confined to the city of Belgrade, the capital city of the country of Republic of Serbia in East Europe.
- Latitude and longitude coordinates of those suburbs. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the suburbs.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Belgrade) contains a list of suburbs in Belgrade, with a total of 133 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and “beautiful soup” packages. Then we will get the geographical coordinates of the suburbs using Python Geocoder package which will give us the latitude and longitude coordinates of the suburbs.

After that, we will use Foursquare API to get the venue data for those suburbs. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, I will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.