Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Gym / Fitness Center in Moscow, Russia

By: Sergei Dyachenko

February 2020

Introduction

Many people have extra weight. It's a worldwide problem. Sedentary work, unhealthy diet are causes of obesity. And need some place where people can spend some energy to fight obesity.

Also many people lead a healthy lifestyle and that's why it's important to keep fit for them too. And I'm among them.

So having good fitness center in your area is really useful.

Most of the visitors are adult people and they prefer a comfortable place for trainings with good conditions where they can relax or exercise after the end of the working day or on holiday.

I chose Moscow for analysis, the city where I live, so I could use my first-hand experience.

Target Audience of this project

This project is particularly useful to property developers and investors because it's good business project in a city with big population (According to stats, Moscow's population was 12.6 million people in 2019).

Also this project might be useful for other people. For example, if someone moves to another area and one of criterion of choosing new place of residence is good fitness center near his future home. This project will help in this situation.

Business Problem

Of course, as with any business decision, opening a new fitness center requires serious consideration and is a lot more complicated than it seems.

The objective of this capstone project is to analyze and select the best locations in the city of Moscow, Russia to open a Fitness Center because location is one of the most important decisions that will determine whether the gym will be a success or a failure.

Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: *in the city of Moscow, Russia, if a property developer is looking to open a new Fitness Center, where would you recommend to do it?*

Data

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

- List of neighborhoods in Moscow. This defines the scope of this project which is confined to the city of Moscow.
- Latitude and longitude coordinates of those neighborhoods. It's required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to fitness centers. It will help to know about current fitness centers in every district. This data will be used to perform clustering on the neighborhoods to answer the business question.

Sources of data and methods to extract them

Geojson file which contains a list of neighborhoods in Moscow (with a total of 146 districts) and their coordinates.

We will use Foursquare API to get the venue data for those neighborhoods.

Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers.

Foursquare API will provide many categories of the venue data; we are particularly interested in the Gym/Fitness Center 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: working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).