

IBM Applied Data Science Capstone

Coursera

Week 4

Problem and Data discussion

Giampaolo Tumminello

Gym market in San Francisco

Introduction

The rise in awareness among millennials about health issues has resulted in an increase in expenditure on healthy lifestyle and fitness activities, which is encouraging them to join fitness clubs. Revenue of the fitness industry is expected to show an annual growth rate (CAGR 2020-2024) of 0.6%, resulting in a projected market volume of US\$23,127m by 2024.

As a result, new fitness clubs are rapidly emerging in the streets of every city worldwide.

Problem

Given that, for entrepreneurs who wants to enter this market, the knowledge of where to open a new gym is crucial for the success of the investment.

In this project, the geographical distribution of gyms in the City of San Francisco will be analysed, to see **where the “good spots” are hided** in the city.

The definition of a “good spot” is fundamental for the analysis, thus it’s important to find the right characteristics:

- Great potential client base.
- No strong direct competition.
- Good geographical position overall.

The first condition means to have a big amount of people living in the neighbourhoods surrounding the spot; the second means that no other gyms (or very few) are surrounding the spot and the third refers to all other geographical characteristics which can result in a competitive advantage, for instance a good building, a good view from the window, nice entrance and many more.

The analysis is intended to be useful for entrepreneurs who wants to enter in the fitness industry in San Francisco.

Data

The Data needed for the analysis regards the neighborhoods of San Francisco, which are easily retrievable from the web and from government websites in particular. Since this is a geographical analysis, the coordinates of each Neighbourhood are required and would be acquired using the Geocoder library in python programming language.

The second category of data regards the venues for each neighbourhood, divided into categories, so as to identify the fitness centre and gyms present in each area. These data can be retrieved using Foursquare API. The advantage of this method is that venues will be already divided for each neighborhood of belonging.