The battle of neighborhoods: finding a place for your gym

Week 1 - intro

SEPTERMBER 2019

Authored by Gabriele Favia

coursera

IBM®

Table of the contents

Contents

Introduction

Using the data to know what to do

In the previous decade, the explosive diffusion of data management and retrieval system developed numerous applications of “Business Intelligence” and “Decision Support Systems”: software born to help the business owners to make effective choices based on past data digested in a statistical report form.

Nowadays, machine learning research is continuously rising the level of reliability of the previsions, lowering at the same time the costs of these systems.
Indeed, a series of factors such as the spreading of high-speed internet connection, bigger data storage capacity and improved computational performance is leading to massive amount of data, also accessible in public form, that are super-charging the robustness of the models.

> ***“It is a capital mistake to theorize before one has data.”***
> ***Sherlock Holmes, A Study in Scarlett (Arthur Conan Doyle).***

In this document we will try to navigate through the sea of data available in the bay of Foursquare to get useful insights on what place can be the best one to open a gym in an Italian medium-sized city such as Bari.

Definition of the problem

Where to place your business?

In a world where open a new business is pressed by the risk of failure due to bad management decision, let's take a step back and see the head of the chain of good business practices: choosing the right place.
Given a city and a possible type of shop/business/POI, is it possible to estimate which one of the vacant premises is the most suitable for a new opening?

Of course, the quantity of variables to make a solid decision is unconceivable, but for a practicing purpose the idea is looking at the other gyms in similar cities to the target one, extracting the ratings given by the clients, and put them in correlation with the quantity of POI/venues in their surroundings.

In the end it should be possible to estimate, for pure fun, which one of the neighborhoods composing the target city is the most suitable to host the next gym.

In this case, the clue will be the output of a supervised machine learning algorithm, where the suggested places to set a new business should have the same surrounding situation of the gym with rating above a certain threshold.