IBM/Coursera´s Applied Data Science Capstone course

Assignment: Capstone Project - The Battle of Neighborhoods

_____

Project Title

Candidate neighborhoods for opening new restaurants - Methodology and case study.

Author: Paschoal Molinari

Date: 2021-02-18

Introduction

Welcome!

This is the report of the project "Neighborhoods candidate to the opening of new restaurants - Methodology and case study" presented as the final project of the course "Applied Data Science Capstone" given by IBM on the online platform Coursera.

The problem to be solved is finding neighborhoods that are good candidates for opening new restaurants.

A specific methodology for restaurants is developed and applied to an Italian restaurant case study in the city of Toronto, Canada. The project is developed using Data Science tools such as the Python language, tools and libraries, and important geospatial data from Foursquare.

The audience for this Project are companies and entrepreneurs in the restaurant business involved in opening new venues.

Business Problem

The business problem in question is finding neighborhoods in a specific city that are potentially good candidates for opening new restaurants. This problem affects many companies and entrepreneurs in the restaurant industry who need timely information for decision-making when choosing a new business location.

Usually finding a good neighborhood for a new business is a time and resource-consuming effort and has some blind spots of factual data about demand and competition. Foursquare geospatial data can be used to uncover these spots.

The solution to this problem aims to:

• Seize business opportunities with high demand and low competition.

• Increase return on investment (ROI).

• Maximize the chances of a successful restaurant opening.

• Add the value of proximity to potential customers of new restaurants.

• Reduce the risk, time, and cost of neighborhood selection.

The problem is complex and the methodology developed selects potential good candidates neighborhoods for more detailed analysis in the field and final business decision.

The methodology developed can be extended to other lines of business.

Data Description

For this project, Foursquare's geospatial data are used.

The geospatial features are:

- Neighborhood

- Neighborhood Latitude

- Neighborhood Longitude

- Venue

- Venue Latitude

- Venue Longitude

- Venue Category

The geospatial data obtained from Foursquare about the different venues (not just restaurants) are engineered, tabulated and new indicators, which are part of the methodology, are elaborated.

As methodology´s case study we use geospatial data from the City of Toronto, as it is well known and has a wealth of details.

In the case study, additional features are generated:

1. Top 10 venues by neighborhood.
2. Total count of venues per neighborhood
3. Total count of restaurants by neighborhood
4. Total count of Italian restaurants by neighborhood.
5. Mean of restaurants by neighborhood.
6. Mean of Italian restaurants by neighborhood.

Itens 5 and 6 are specially important. The lower the mean the better is the chance of success.

Finally, machine learning techniques are applied to the new features, generating more information about the neighborhoods candidates.

The case study data are available in csv format files.