



Capstone Project:

The Battle of Neighborhoods



Alejandro Sócrates Pérez López

Student of the Coursera – IBM Data Science Professional Certification Program

Contact: Alejandro_s_perez@hotmail.com

OCTOBER 16, 2019

Contents

Introduction	2
Business Problem	2
Data	2

Introduction

The use of location data to explore geographical locations allows us to analyze options for business in cities around the world.

This is possible by gathering information about existing business in specific area, so the details can be classified and clustered according different criteria in order to assemble a solidly funded recommendation.

Business Problem

What location is recommended to open a business in Seattle, WA?

The location is an important factor in the success of a new business. So, it is required to provide a set of alternatives for opening a business in Seattle, WA.

The recommendations will include location and kind of pre-existing business in the surrounding area, so, it can be identified what business will successfully fit in that neighborhood.

Foursquare location data will be used as source of information about the market places in the different neighborhoods in the city.

This research should be interesting to investors looking for opportunities with a comprehensive argumentation and details about the different options, so their decisions will lead to successful results.

Data

The data to be used will be extracted, mainly, from Foursquare.com. This will provide quantitative and qualitative information of pre-existing business, in order to identify the cluster where a new business could fit.

However, to be able of requesting the information from Foursquare.com, it will be necessary to obtain the coordinates of the different neighborhoods in Seattle. This information will be extracted from <http://www.discoverseattle.net/zipcodes.php>, where the postal codes are listed along with their geographical coordinates.

The queries to Foursquare will be focused on venues for different kind of business within the different postal codes that will be considered as neighborhoods.

The analysis will be handled in data frames similar to the following, from where maps will be generated to obtain visualizations that will help in identifying the existing business so, the more convenient new business can be determined.

Business name	category	latitude	longitude
----------------------	-----------------	-----------------	------------------

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
---------------------	----------------------------------	-----------------------------------	--------------	---------------------------	----------------------------	---------------------------

Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
------------------------	------------------------------	----------------	-----------------------------------	-------------------------	---------------------------	----------------------------	-----------------------------	--------------------------------