Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Table of contents

Introduction: Business Problem

In this project, we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Italian restaurant** in **Ho Chi Minh City**, Vietnam.

Since there are lots of restaurants in Ho Chi Minh City, we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Italian restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on definition of our problem, factors that will influence our decision are:

* Number of existing restaurants in the neighborhood (any type of restaurant)

* Number of and distance to Italian restaurants in the neighborhood (if any)

* Distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

* Coordinate of Ho Chi Minh City center will be obtained using GeoPy **Nominatim API geocoding**

* Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using GeoPy **Nominatim API reverse geocoding**

* Number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**