Location for Mediterranean restaurant in Chicago

Faisal Farooqi

February 2021

1. Introduction

1.1 Background

I am a big fan of Mediterranean food. More than a decade ago I introduced a colleague at work to this cuisine and he immediately took a liking to it. So much so that a few months later he decided to open a restaurant and asked me for help with the menu and selection of a location.

I was more than willing guidance on the menu but had no idea how to recommend a location. Our office was located in the suburbs of Chicago, IL and there were multiple restaurants with similar offering within 10-minute driving distance. Similar situation existed in the vicinity of my home where I could pick up a kebab dinner within a short driving distance. This was a time when data science had not become popular and using data to make a selection never crossed mind. We resorted to scouting locations the hard way which was driving to them and assessing the neighborhood. After wasting considerable time and resources my colleague dropped the idea of opening a restaurant.

Many years later I now feel that I am in a better position to suggest a location for opening a restaurant if someone asks me for help. Data science tools have evolved, python libraries for data handling are well developed and a multitude of data sources are available to make an informed decision.

1.2 Problem

Restaurant industry has been severely impacted by the covid pandemic. Local regulations in almost all countries have limited in-store dining to some extent and the state of Illinois (where Chicago is located) is no different. In the current situation anyone looking to open a restaurant would prefer a location with good foot traffic. The usual spots like city centers, sports arenas and movie theaters are not good candidates either due to covid related restrictions on attendance.

The aim of this report is to not only suggest a location for a Mediterranean restaurant in the city of Chicago but to also consider the influence of covid related factors and select a location which is suitable for customers to carry out the food, for example close to a bus or train station.

2. Data sources and usage

2.1 Sources of data

The primary source of data is the Foursquare location data available via *Place API* here. This API allows query of data based on location and other optional parameters.

City of Chicago can be considered a grid due to the road network generally aligned in East-West and North-South directions. This grid can be used to define regions and evaluate restaurants within each region but would require extensive work to achieve evenly distributed regions and to calculate latitude/longitude of region centers. A better source for the purpose of creating geographical regions is the zip codes used by the United States Postal service. List of zip codes used in United States is available here. This data includes longitude and latitude of each zip code along with GeoJSON format map description.

A map of zip codes which belong to city of Chicago can be found here and the list of Chicago zip codes is here.

Chicago Transit Authority (CTA) maintains a website which contains route maps, station location and rider statistics. Data can be downloaded in various formats.

2.2 Data cleaning

Multiple Foursquare API sub-categories within the *Food* category can be considered as Mediterranean cuisine. To narrow response results the following sub-categories need to be selected:

- Mediterranean Restaurant
- Kebab Restaurant
- Turkish Restaurant
- Middle Eastern Restaurant
- Greek Restaurant
- Falafel Restaurant

Zip code longitude/latitude are available via *opendatasoft* need to be filtered to select Chicago zip codes. There are multiple ways to achieve this, first is to filter on state and city. Second method is to merge the table from *opendatasoft* with list of Chicago zip code obtained from other sources listed above. 'Timezone' and 'Daylight savings time flag' fields provided by *opendatasoft* are redundant for the analysis of this report and can be dropped.

Within data available from CTA we shall only consider train stations and ignore bus stops for this project. Train stations cover a larger surface area which results in better chances of finding an open lease when compared to bus stops. From the data available for train stations only 'STATION_NAME' and 'Location' fields are relevant, the remaining columns can be dropped. There are duplicate entries for some stations

due to the fact they have entrances on different streets, these need to be removed so that only unique entries are left.

2.3 Data usage

Once data collected from various sources is cleaned up the first step is to setup a grid of geographical regions covering city of Chicago via zip codes boundaries. After that general restaurant density and Mediterranean restaurant density should be computed for each zip code. This can be done based on distance or by matching the zip code from results returned by Foursquare API. A heatmap can be used to visualize the results.

Train station locations can be matched up with zip codes via location information to enhance the heat maps to select candidate areas suitable for opening a new restaurant. Use of clustering may or may not be required depending on the results obtained via heat map examination. If no areas stand out then clustering can be used to incorporate additional features returned by Foursquare API (e.g ratings) to form regions which can serve as candidate for a new Mediterranean restaurant.