The Battle of Neighbourhoods

Introduction

This capstone projected aims to identify promising location for business like restaurants, bars and snack shops. For the success of such a business the quality of food and servers is essential success. However, the location of such business is also significant for their long-term success. For example, it is unlikely that a restaurant flourishes in an industrial industrial area even if it serves exceptional food. In contrast, a mediocre restaurant can be well visited if it is located next to famous tourist attraction. Hence a tool that could evaluate a location on its suitability for a certain establishment could be very valuable for people who contemplate the idea to become self-employed as well as for big changes who want to extent by opening new brunches.

As discussed above the envisioned tool is in general useful, however due to the current COVID-19 pandemic its benefit is quadrupled. The drastic emergency acts, which were ratified to enforce social distancing, hitting gastronomy especially hard. Hence it is expected that many independent establishments have to close permanently. For example, according to the New York Times as many as 75% of the independent restaurants of New York may not survive the COVID-19 pandemic. While such a cataclysmic event is extremely tragic for the millions of entrepreneurs who may lose the fruits of decades of work, it also sets the stage for new success stories. In other words, at the end of the COVID-19 crisis the shortage of gastronomy businesses of all kinds will present tremendous business opportunities. Furthermore, large numbers of prime location will be available. This unprecedented situation makes tools capable of reliably evaluate the suitability of a location for a particular type of gastronomy business more valuable than ever.

Dataset

The target of this project is to develop a tool which calculates a score for any location in the US, specified by longitude and latitude, and a particular type of business, where the score represents who suitable the location is for opening a new business of the given type. For this purpose, we will leverage geospatial data gather via the Foursquare API. In particular we query the recorded venues in Foursquare for each postcode of the US. To enrich our database, we use census data provided by the US government. In particular, we utilize the number of employees in a postal code, the total income of these employees, and the number of registered business to further characterize economic strength of the particular post code area.

In addition, we query the details of all business of interest. For example, if a client is interested in opening an ice cream shop, we query the details off all US-based ice cream shops in the Foursquare API. To build the training and test sets we rank the business based on the number of likes they received on Foursquare. Then, we label the post-codes of the top-30% business as promising locations. Postal codes without a business of interest (e.g., ice-cream store) are labelled as not promising locations.

This rich database, will then be used to train a logistic regression model to categorize locations into promising and non-promising location. Logistic regression is employed since it provides not only a label (‘promising’ or ‘not promising’) but also an estimate how reliable the classification was. Hence, the tool cannot only be used to determine if a given location is promising, but we can also compare different location against each other.