Applied Data Science Capstone Project
Leslie Lui
4/26/20

A Data-Driven Approach to Restaurant Location Selection

1. INTRODUCTION

1.1 Background

Around the world, there are few social institutions that typify societies than the restaurant. Members of every society desire places they can go to enjoy fine meals and socialize with friends, family, and strangers, and without the need to do any of the cooking and cleaning themselves. However, the restaurant business is a competitive industry to operate in. Every restaurant must compete with numerous other nearby restaurants for the attention and money of prospective customers, and provide a service that meets the cost considerations of their customers while still being profitable.

One recent trend in restaurant industry is the growing popularity of gastropubs. These restaurants fuse characteristics of restaurants with bars in a premium dining and entertainment experience. They offer the quality dining and comfort food menus of upscale and traditional neighborhood restaurants, a wide and premium variety of alcoholic beverages associated with some higher end bars, pubs, and microbreweries, and an upscale, yet comfortable and accessible environment for dining and socializing. When consumers go to a restaurant, they increasingly want more than to just simply eat a prepared meal – they want to have a complete experience with dining, socializing, and entertainment, and they are willing to pay more of a premium for it. Gastropubs are one of the newer venues offering this combination. A new gastropub could be a lucrative business opportunity by offering customers an attractive alternative to traditional restaurants, and provide attractive profits by cashing in on the premium experience associated with gastropubs.

The increased popularity of gastropubs is an international trend. Toronto is a very international city, with a large and diverse population, so it is not surprising that there are already dozens of gastropubs in Toronto.

1.2 Problem

The presence of numerous gastropubs in Toronto proves that they are a viable restaurant concept in this market. However, selecting a suitable location for a gastropub is still important; a good location can allow a gastropub to thrive while a bad location can mean an early end. Instead of merely taking a guess, a data-driven approach was used in this study to identify suitable locations for a gastropub, based on the attributes of locations in Toronto that currently have gastropubs.

2. METHODOLOGY

2.1 Overview

The study investigated existing gastropubs in the Toronto area, identified traits about the neighborhoods around them, built a model to determine if any traits are common to neighborhoods with gastropubs, and then used that model to identify other similar neighborhoods that could be good candidates for locating a new gastropub.

A cluster analysis will be used to group the neighborhoods. When a cluster is found that contains most of the gastropubs, it can be assumed that the neighborhoods in the cluster have attributes that generally make them more favorable locations for gastropubs, so they would be the recommended locations for new gastropubs.

Regression and classification were not considered suitable methods to apply to this study for the following reasons:
1. The intent of this study is to determine which existing Toronto neighborhoods would be the most suitable for a gastropub. Regression and classification would need to use existing Toronto neighborhoods as training data, and so they would not be helpful for predicting the suitability of existing neighborhoods. They would be more useful for predicting whether new samples outside the training data set (ie. neighborhoods outside Toronto) would be appropriate for gastropubs.
2. The goal of regression and classification would be to create models that accurately predict whether a neighborhood should have a gastropub. If the model was very accurate, it would predict that the only the neighborhoods that should have gastropubs are the neighborhoods that already have existing gastropubs, so regression and classification models would not be helpful for suggesting other potential neighborhoods.

2.2 Data

The study data included data on the Toronto neighborhoods and business venues in them.

Two Toronto neighborhood datasets from the Toronto Open Data were used:

1. Toronto neighborhood profiles dataset [1]. This dataset contains demographic data on each neighborhood, taken from census reports. Data include general population attributes of each neighborhood, such as overall population and land area, as well as more specific population attributes such as distributions of age, income, family statuses, living conditions, ethnicities, languages spoken, etc. Only a select portion of this data was used for the analysis. This data was included to factor in the effects of a neighborhood's demographics on the likelihood of a gastropub to be located there.
2. Toronto neighborhood boundaries dataset [2]. This dataset contains geographic coordinates on the each of the 140 neighborhoods that comprise Toronto. The central coordinates were used for venue searches (described below), and the boundary coordinates were used for plotting neighborhood boundaries on a map.

The business venue information was provided by the Foursquare API service [3]. The information was obtained in queries sent to the API service, through a Jupyter notebook. The information obtained included the following:

1. Gastropub venue search results for Toronto. A search was done in each Toronto neighborhood to identify as many gastropubs as possible. This information was included in the analysis in order to create profile of the venues around each gastropub.
2. Exploration search results, starting at the central coordinates of each Toronto neighborhood. This provides information popular venues around each neighborhood center, and thus a profile of the popular venues in each neighborhood.

2.3 Data Analysis Process

2.3.1 Data Analysis Procedures

1. Gather geographic information on the Toronto neighborhoods from the Toronto Open Data website. The key information to obtain will be the names of each neighborhood, their central coordinates, and the coordinates of their boundaries.
2. Gather basic demographic information on the Toronto neighborhoods from the Toronto Open Data website. Key information to obtain will be the neighborhood names and high level neighborhood and demographic information, such as total population, population density, and age, income, and family distributions.
3. Perform a search for gastropubs in the Toronto neighborhoods using the Foursquare API. Key information to retrieve are venue names, unique ID's, coordinates, and venue category.
4. Assign a neighborhood to each gastropub. The Foursquare API cannot determine the neighborhood that each gastropub officially resides in, so instead, each gastropub will be assigned to the neighborhood whose central coordinates are closest to it.
5. Use the Foursquare API to get venues located around the center of each Toronto neighborhood. Key information to retrieve are venue names, unique ID's, coordinates, and venue category.
6. One-hot encode the venue category information for the venues around each Toronto neighborhood. Combine this information with the neighborhood basic demographic data to create a general profile of each neighborhood.
7. Perform cluster analysis on the neighborhoods. The end goal is to arrive at a cluster that contains most of the neighborhoods with gastropubs.
8. Review the attributes of the neighborhoods in the cluster with most of the gastropubs to determine if what common attributes they have, and how they might differ from neighborhoods in other clusters.
9. Use the result to recommend neighborhoods that would be good candidates to locate a gastropub.

References

[1] https://open.toronto.ca/dataset/neighbourhood-profiles/
[2] https://open.toronto.ca/dataset/neighbourhoods/
[3] https://developer.foursquare.com/