Leveraging Foursquare API and Hamilton Neighborhood Dataset for Auto-workshop location

Abiola Obembe
February 21, 202

1. Introduction

1.1 Background

Hamilton is a port city in the Canadian province of Ontario. An industrialized city in the Golden Horseshoe at the west end of Lake Ontario, Hamilton has a population of 536,917, and its census metropolitan area, which includes Burlington and Grimsby, has a population of 747,545. The city is 58 kilometres (36 mi) southwest of Toronto, with which the Greater Toronto and Hamilton Area (GTHA) is formed. On January 1, 2001, the current boundaries of Hamilton were created through the amalgamation of the original city with other municipalities of the Regional Municipality of Hamilton–Wentworth. Residents of the city are known as Hamiltonians. Since 1981, the metropolitan area has been listed as the ninth largest in Canada and the third largest in Ontario. With such a significant population lies opportunities and the potential for establishing businesses to cater for the populace. Take for instance, it will be beneficial for prospective entrepreneurs to be able to determine the type of business to establish and the best location to setup their office.

1.2 Problem statement

After residing a few months in the city of Hamilton, a wealthy contractor has decided to establish a business in the local area. Based on his previous work experience in the auto industry, he has suggested that his preference will be to set up an auto business in Canada. However, after a careful study of the city, he is confident that establishing an automobile repair workshop will be profitable. However, he is unsure of the best location for his workshop. The aim of this project is to utilize available data from the city of Hamilton and Foursquare API to guide the contractor on choice location(s) to setup his workshop.

1.3 Interest

Using a data-driven approach to determine a business type, the best location(s) to setup businesses will be highly beneficial to Entrepreneurs as it ensures that their business will receive traction. In addition, this will help the government to develop the required infrastructure in such neighborhoods to attract more businesses which will increase employment and provide revenue for the government through taxes from businesses and people.

2. Data acquisition and cleaning

2.1 Data sources

Data to be employed for this project will be obtained from the city of Hamilton website from the site: http://open.hamilton.ca/datasets/ac6fc684043341f6b1d6298c146a0bcf_1. Specifically, the webpage has data on the distinct municipal addresses in Hamilton. This dataset exists as both CSV files and GEOJSON files. Besides, the original dataset consists of 253,876 rows of data and 12 columns (features).

2.2 Data cleaning and feature engineering

The GEOJSON file was downloaded and subsequently converted to a pandas dataframe for exploratory analysis. For convenience and to reduce cost of computation, only the first 10,000 samples rows of data was employed for modeling in this capstone project. The assumption being that the listed addresses are the preferred locations for establishing an auto business based on preliminary market survey. Additionally, inspection of the dataset indicated a few redundant columns which may be neglected from further analysis. These attributes include "Number Complete", "Unit Number Complete", "Settlement", "Community", "Province", "Country" some of which also had missing values as entries. After cleaning the dataset and performing feature engineering only five features were selected as presented in Table 1.

Table 1 First five entries of Hamilton neighborhood dataset

	ID	Address	Longitude	Latitude	Community
0	1001	Thornton Trail	-79.986884	43.260717	Dundas
1	1002	North Service Road	-79.708260	43.235548	Stoney Creek
2	1003	Frances Avenue	-79.723719	43.240725	Stoney Creek
3	1004	Eastview Avenue	-79.754216	43.233877	Hamilton
4	1005	Concession 2 West	-80.047481	43.264434	Flamborough

3. Exploratory Data Analysis

Due to the nature of the dataset, little or no exploratory analysis or statistical inference can be performed. Analysis of the dataset indicates that the city of Hamilton is divided into six main communities namely Hamilton, Dundas, Stoney Creek, Ancaster,

Flamborough, and Glanbrook. Figure 1 shows the map of the city of Hamilton with 1000 samples from the dataset superimposed.

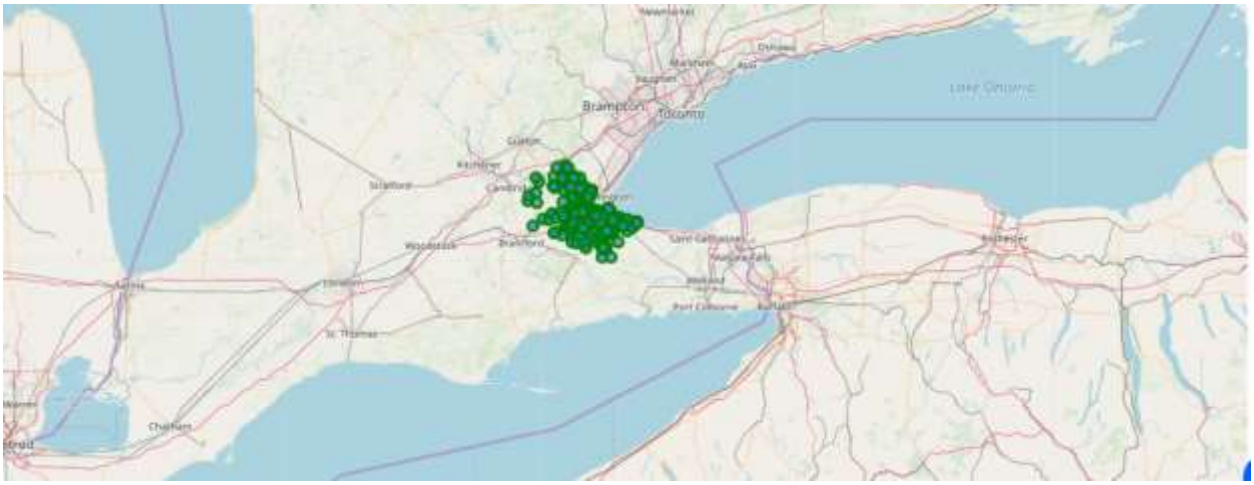

Figure 1: A map of city of Hamilton including a subset of dataset

4. Predictive Modeling

Herein, both the Foursquare API and K-means clustering algorithm were utilized for predictive modeling of the dataset. The k-means clustering is a method of vector quantization that aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The *n* observations were obtained by making Foursquare API calls on the Hamilton address dataset (Table 2). The Foursquare API calls returned results of the venues within 500 meters of each address. Besides, to ensure the venues returned in the search results were not redundant for the aim of the project, a search query "Auto" was supplied as part of the search query of the Foursquare API call. In addition, the venues returned from each address was limited to 50.

The Foursquare API call returned a total of 2132 venues and 17 unique venue categories including 'Automotive Shop', 'Auto Workshop' ,'Auto Dealership', 'Hardware Store' , 'Locksmith' 'Auto Garage', 'Auditorium' , 'Urgent Care Center' , 'Shoe Repair', 'College Engineering Building' , 'Electronics Store' , 'Car Wash', 'Dry Cleaner', 'Design Studio', 'Insurance Office' etc. as listed in Table 2.

Table 2 Subset of venues returned from Foursquare API search query

	Cluster Labels	Address	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	3	Aberdeen Avenue	Automotive Shop	Auto Garage	Urgent Care Center	Shoe Repair	Locksmith	Insurance Office	Hardware Store	Electronics Store	Dry Cleaner	Design Studio
1	0	Academy Street	Auto Garage	Urgent Care Center	Shoe Repair	Locksmith	Insurance Office	Hardware Store	Electronics Store	Dry Cleaner	Design Studio	College Engineering Building
2	2	Ackland Street	Auto Dealership	Urgent Care Center	Shoe Repair	Locksmith	Insurance Office	Hardware Store	Electronics Store	Dry Cleaner	Design Studio	College Engineering Building
3	3	Adair Avenue South	Automotive Shop	Urgent Care Center	Shoe Repair	Locksmith	Insurance Office	Hardware Store	Electronics Store	Dry Cleaner	Design Studio	College Engineering Building
4	4	Adeline Avenue	Hardware Store	Urgent Care Center	Shoe Repair	Locksmith	Insurance Office	Electronics Store	Dry Cleaner	Design Studio	College Engineering Building	Car Wash

Subsequently, the k-means clustering algorithm was utilized to cluster the resulting dataset into 5 distinct clusters as shown in Figure 2.

5. Conclusions

Conclusion section where you conclude the report.