

# Application of Machine Learning in Restaurant Industry

by

Yukun Zhao

A report  
presented for Coursera  
in fulfillment of the requirement for  
IBM Data Science Professional Certificate

## Abstract

# Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
<b>2</b>	<b>DATA</b>	<b>2</b>
2.1	Population Data . . . . .	2
2.2	Foursquare Location Data . . . . .	3
<b>3</b>	<b>METHODOLOGY</b>	<b>4</b>
<b>4</b>	<b>RESULTS</b>	<b>5</b>
<b>5</b>	<b>DISCUSSION</b>	<b>6</b>
<b>6</b>	<b>CONCLUSION</b>	<b>7</b>
	<b>Bibliography</b>	<b>8</b>

# Chapter 1

## INTRODUCTION

Anyone who works in the restaurant industry knows that it takes time and dedication to be successful. Even still, food industry in Canada has some of the lowest survival rates for new business [2]. For newly opened restaurants, 60% of them could survive beyond their second year and only 22% could survive beyond eight years [2]. This is why opening and maintaining a successful restaurant is not an easy task. It requires proof of concept, dedicated staffs, targeted marketing and realistic financial expectations. Moreover, with the average restaurant profit margins falling between 2.7-7.8% [3], it's vital to set the business up for success from the beginning by having a strategic plan. Without considering the financial aspect in this study, we will come up with a recommendation for a client who is looking to open a Chinese restaurant in York Region.

York Region is a regional municipality in Southern Ontario, Canada, and is part of the Greater Toronto Area. The 2016 census population in York Region was 1,109,909 with a growth rate of 7.5% from 2011 to 2016 [1]. The Government of Ontario expects its population to surpass 1.5 million residents by 2031 [4]. With this large number of potential customers, we need to find a location that is accessible. We should consider things like ease of parking, local events and nearby attractions. Although prime locations could mean higher costs and rent, it could also lead to higher anticipated sale volume. In this study, we will look at population and location data to find a neighborhood that is suitable for the client to open the Chinese restaurant. In addition, we should also look into the competitions in the neighborhood. This could be done using the location data to find out if there are opportunities for the client to provide something that is not yet available.

## Chapter 2

# DATA

The data used in this study consist of the 2016 census data from Statistics Canada and the location data from Foursquare. The 2016 Canadian Census is the most recent detailed enumeration of Canadian residents. The 2016 census reinstated the mandatory long-form census which collects information on the demographic. Three in four households (75%) received the short-form census questionnaire, while one in four households (25%) received the long-form questionnaire. With a response rate of 98.4%, this census is said to be the best one ever recorded.

The location data from Foursquare come from the Foursquare Places API which offers real-time access to Foursquare’s global database. This allows us to perform geotagging and venue search. In addition, we could leverage the 900+ venue categories, sourced by the Foursquare consumer community, to find the location of interest for our client to open the restaurant.

### 2.1 Population Data

We will use the visible minority data from the census. Figure 2.1 illustrates the top five rows of the data. These data have information on the number of people who are identified as visible minority in each city of York Region, from which we could determine the city that has the largest Chinese community. As we have mentioned above, the demographic data were only collected from one out of four households. Therefore, we have to assume that the sample data were representative of the entire population in York Region. Moreover, the growth in each visible minority group is assumed to be the same, so the data are still relevant in 2020.

	CITY	SOUTH_ASIAN	CHINESE	BLACK	FILIPINO	LATIN_AMERICAN	ARAB	SOUTHEAST_ASIAN	WEST_ASIAN	KOREAN	JAPANESE
0	Vaughan	55.0	130.0	15.0	70.0	0.0	0.0	10.0	25.0	125.0	0.0
1	Vaughan	95.0	210.0	0.0	35.0	0.0	0.0	0.0	45.0	45.0	0.0
2	Vaughan	925.0	505.0	225.0	335.0	105.0	120.0	85.0	135.0	45.0	20.0
3	Vaughan	280.0	70.0	35.0	15.0	0.0	0.0	20.0	10.0	0.0	0.0
4	Vaughan	70.0	0.0	0.0	0.0	15.0	0.0	10.0	10.0	0.0	0.0

Figure 2.1: Example of visible minority data from 2016 census.

## 2.2 Foursquare Location Data

Depending on our findings from the census data, we will use the postal code data and divide the city of interest into neighborhoods. Foursquare Places API utilizes the geographic coordinates to explore nearby venues. We could collect the geographic coordinates of each neighborhood using GeoPy which is a geocoder service. With the latitude and longitude information, we could then use Foursquare to investigate Chinese restaurants in the neighborhoods, as well as other deciding factors such as local attractions and ease of parking. This information will help us decide where the restaurant should be located and what type of food the restaurant should serve.

## Chapter 3

# METHODOLOGY

## Chapter 4

# RESULTS



## Chapter 5

# DISCUSSION

## Chapter 6

# CONCLUSION

# Bibliography

- [1] *Census Profile, 2016 Census*. URL: <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CD&Code1=3519&Geo2=PR&Code2=35&Data=Count&SearchText=York&SearchType=Begins&SearchPR=01&B1=Population&TABID=3> (visited on 04/23/2020).
- [2] *Failure Rates for New Canadian Firms: New Perspectives on Entry and Exit*. URL: <https://www150.statcan.gc.ca/n1/en/pub/61-526-x/61-526-x1999001-eng.pdf?st=vvCq0IHS> (visited on 04/23/2020).
- [3] *Financial performance - Canadian Industry Statistics*. URL: <https://strategis.ic.gc.ca/app/scr/app/cis/performance/7225;jsessionid=0001gMVaxQRJbd-Djb3Rt9YAtmh:27I1RR0T9> (visited on 04/23/2020).
- [4] *York Region Official Plan - December 2009*. URL: <https://web.archive.org/web/20110706213558/http://www.york.ca/Departments/Planning+and+Development/Growth+Management/default+Growth+Management.htm> (visited on 04/23/2020).