

# **Applied Data Science Capstone Project**

---

## **Recommending location for a new Restaurant through Data Analysis**

**Nithin John Jacob**

**24 April 2020**

This report presents the details of recommending an ideal location for an Indian restaurant in Toronto, Canada by analyzing the demographics and neighborhood of various boroughs of toronto

## **Contents**

1.	Introduction .....	2
2.	Literature Review .....	2
3.	Data.....	3
3.1.	Toronto Demographics Data.....	3
3.2.	Geographic data of Toronto neighbourhoods.....	3
3.3.	Foursquare data.....	3
4.	Methodology .....	4
5.	Results.....	4
6.	Discussion .....	4
7.	Conclusion .....	4
8.	References .....	4
9.	Acknowledgement.....	4
10.	Appendices.....	5

## **1. Introduction**

Setting up a new restaurant is a huge investment. If you're thinking about opening a restaurant, location is everything. Experts often argue a bad location is one of the biggest, if not THE biggest, reason restaurants fail. Accessibility, Parking facility, Transportation, Size of restaurant, Competition, Labor Costs, the Minimum Wage and target customers are some of the parameters to be checked.

We will be focusing on the "target customers " part, i.e. the demographics of various neighbourhoods of Toronto for this. We will be using various machine learning tools to analyse the data to identify a location. This report maybe useful for all stakeholders who would like to open a restaurant serving Indian cuisine in Toronto.

### **1.1. Business Problem**

This project aims to find an ideal location for a new Indian restaurant in Toronto, Canada by interpreting the data acquired from various sources.

### **1.2. Target stakeholders**

All restaurateurs interested in opening an Indian restaurant in Toronto. The data acquisition and research done in this project may provide useful insights in determining a location for their new restaurant.

## **2. Literature Review**

Various researches have emphasized the importance of location over and again.

Some of the studies can be found at

<https://study.com/academy/lesson/location-analysis-for-the-restaurant-industry.html>

<https://www.thebalancesmb.com/choosing-restaurant-location-2888543>

<https://www.restaurantindia.in/article/5-smart-ways-to-choose-best-location-for-your-restaurant.9444>

[https://www.researchgate.net/publication/301578175\\_Restaurant\\_Location\\_and\\_Price\\_Fairness\\_as\\_Key\\_Determinants\\_of\\_Brand\\_Equity\\_A\\_Study\\_on\\_Fast\\_Food\\_Restaurant\\_Industry](https://www.researchgate.net/publication/301578175_Restaurant_Location_and_Price_Fairness_as_Key_Determinants_of_Brand_Equity_A_Study_on_Fast_Food_Restaurant_Industry)

<https://www.webstaurantstore.com/article/81/restaurant-environmental-analysis.html>

### **3. Data**

Different data sets are used in this project

#### **3.1. Toronto Demographics Data**

This data is obtained from the open Toronto portal. The csv file tabulates various statistics of all 140 neighbourhoods. Like, ethnicity, population, area , average income, etc. we use the average income ethnicity distribution and area of all the 140 neighbourhoods. This is downloadable from this link. <https://open.toronto.ca/dataset/neighbourhood-profiles/>.

#### **3.2. Geographic data of Toronto neighbourhoods**

This file was made from the Neighbourhood information from Namara.io and Toronto's Open Data portal. This Geojson file contains boundaries of City of Toronto Neighbourhoods in latitude and longitudes.. The Centroid of the neighbourhoods and distance of centroid to each of the boundary points of the neighbourhood is calculated and taulated

This data is available at <https://github.com/jasonicarter/toronto-geojson> .

#### **3.3. Foursquare data**

The information regarding the restaurants in the neighbourhood is obtained using the foursquare venue search API. Since Foursquare AP returns only 50 venues at a time , in order to collect maximum data , the request is made three times with varying(near, average, far) search radius for each neighbourhood. The results are combined and cleaned, ensuring that there is no repetition of venues and that each venue belonged to their corresponding neighbourhood. This data is made to a data frame containing, location information like

latitude, longitude, address, name and category of each venue and their corresponding neighbourhood details.

#### **4. Methodology**

To be filled

#### **5. Results**

To be filled

#### **6. Discussion**

filled

#### **7. Conclusion**

filled

#### **8. References**

filled

#### **9. Acknowledgement**

filled

## **10.Appendices**

filled