Identifying

Proper location

for

Restaurants

Sonal Chourasia

08-May-2019

**Table of Contents:**

[1 Introduction 3](#_Toc8189959)

[1.1 Background: 3](#_Toc8189960)

[1.2 Problem:. 3](#_Toc8189961)

[1.3 Interest:. 3](#_Toc8189962)

[2 Data Acquisition and Cleaning 3](#_Toc8189963)

[2.1 Data Sources: 3](#_Toc8189964)

[2.2 Data Cleaning: 3](#_Toc8189965)

[3 Methodology 4](#_Toc8189966)

[3.1 Relationship between population density and restaurant location 4](#_Toc8189967)

[3.2 Relationship between average income and restaurant location 4](#_Toc8189968)

[3.3 Relationship between Transit commuting % and restaurant location 4](#_Toc8189969)

[4 Results 5](#_Toc8189970)

[5 Discussion & Future directions 5](#_Toc8189971)

[6 Conclusion 5](#_Toc8189973)

# Introduction

## Background: We see plenty of restaurants being getting inaugurated every month but only few of them make profit and survive. There are various factors affecting the sustainability of eatery business like – food quality, pricing, customer service and location.

## Problem: If the location is proper then restaurant owner can not only survive but maximize profit based on his food quality and proper pricing. We see that location can be a crucial factor for deciding the success of any business.

## Interest: This reports aims at deciding the appropriate location of restaurant considering different factors like – population density, income, presence of other restaurants etc.

# Data Acquisition and Cleaning

## Data Sources: For our analysis we will need the data relating to city with different locations, population density, per capita income as well as information relating to surroundings.

Considering the city as Toronto,

1. We will be using wiki page <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> to get information about all postal codes and borough of Canada. This data has to be filtered based on Borough like ‘Toronto’ to get information regarding Toronto.
2. Above data has to be joined with ‘**http://cocl.us/Geospatial\_data’** to get information the latitude and longitude coordinates for each postal code.
3. Since latest data is not available relating to population density, average income, we will be using older data from the wiki page - <https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods>.
4. We will use Four square API to explore nearby neighbourhoods and get information about nearby cafe, restaurants etc.

## Data Cleaning:

This involves removing unwanted columns from existing dataset obtained after scrapping ‘<https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods>’. We are mainly interested in Name, Density, Average income and Transit commuting % and so we will drop the remaining columns from this data set and then join the same to Toronto location & neighbourhood dataset as obtained in step 1, 2 and 4 above.

# Methodology

Basically we see that data obtained in step 1 and 2 containing all postal codes along with latitude, longitude values for Toronto doesn’t join fully with demographic details obtained in step 3. One solution could be to take mean of population density & average income for a particular postal code and then populate the same wherever the values are Null. But we observe that within same postal code there is a high disparity between population density and average income and so we decide to drop the rows where these columns are null instead of populating average values.

#In code

# Dropping the rows with Null values

df3.dropna(how='any', inplace=True)

## Relationship between population density and restaurant location

The idea is to explore the relationship between population density and restaurant- it is seen that highly populated areas tend to have greater number of restaurants. The more the population the greater is the probability of eateries- cafe, restaurants etc being the most common 1st venue.

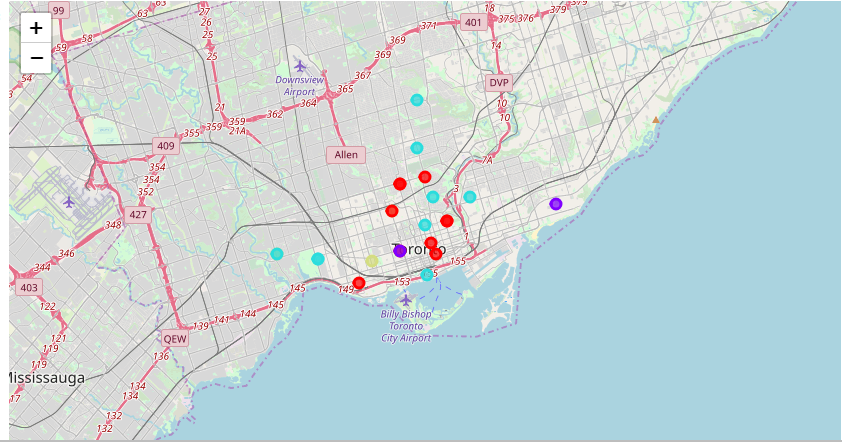
## Relationship between average income and restaurant location

Our analysis suggests that eateries are most common at areas where average income is below or around 50,000. Higher income group people may be too busy to go out for eating.

## Relationship between Transit commuting % and restaurant location

Higher Transit commuting % should mean more people on road and it should promote more eatery – restaurant business. But we were not able to establish this relationship using exploratory data analysis.

# Results



Of the 4 clusters for Toronto we see that cluster 3 is safest for ‘Restaurant’ business and there is no competition from peers. Similarly, data can be analyzed and other eatery business location can be decided based on different clusters.

# Discussion & Future directions

# Other than population density alone we could consider population density based on age range and then check if presence of particular age population leads to more eatery business. Also, nearby spaces can be grouped into official spaces, commercial spaces as well as residential spaces and then we could check which spaces nearby are more favorable for restaurant business.

# Conclusion

In this study, we have analyzed the location for restaurant based on population density, average income, commuting transition % and presence of other restaurants nearby. Once restaurant /eatery business has been setup at proper location other factors like food quality, pricing, customer service will decide the success of business.