IBM Capstone Peer-graded Assignmengt

**Best Locations to apen an Asian restaurant in Toronto**



Vinay Kumar Makina

IBM Data Science Professional Certificate course

Coursera

Contents

[Problem Statement 2](#_Toc31479585)

[Problem Description/Objective 2](#_Toc31479586)

[Target Audience 2](#_Toc31479587)

[Description of the data 2](#_Toc31479588)

[List of Required Datasets 2](#_Toc31479589)

[Discussion Points 3](#_Toc31479590)

[Methodology 3](#_Toc31479591)

[Data Collection 3](#_Toc31479592)

[Data cleaning 3](#_Toc31479593)

[Data transformation 3](#_Toc31479594)

[Data visualization 3](#_Toc31479595)

[Model Building 4](#_Toc31479596)

[Results 4](#_Toc31479597)

[Recommendations 4](#_Toc31479598)

[References. 4](#_Toc31479599)

# Problem Statement

In Toronto-Canada, if someone is looking to open an Asian restaurant, where would you recommend that they open it?

# Problem Description/Objective

The objective of this project is to find the best suitable location(s) to open a new Asian restaurant in Toronto, Canada. To be more specific, this project tries to find the locations where there is high popularity for the Asian food. There might me many other factors to be considered to open a restaurant but in this project we are considering only the number of Asian restaurants already present in that area to recommend a location.

# Target Audience

Any person or organization who wants to open an Asian restaurant in Toronto Canada.

# Description of the data

To solve this problem, we need the location related data of the existing Asian restaurants. As this data is not available directly we need to use/join the 3 below data sets.

## List of Required Datasets

1. Neighborhood-Postcode data from Wikipedia.

2. Postcode-coordinates (Latitude and Longitude) from Geospatial data.

3. venue-locations using Foursquare API.

The variable of interest in Neighborhood-Postcode data are:

1. Postcode – postal code of an area

2. Borough - a town or district which is an administrative unit.

3. Neighborhood – a small area

The variable of interest in Postcode-coordinates data are:

1. Postal Code – postal code of an area

2. Latitude

3. Longitude

The variable of interest in Venue-locations data are:

1. Neighborhood

2. Venue

3. Venue Latitude

4. Venue Longitude

5. Venue Category

# Discussion Points

The idea behind this Capstone project is to solve an imaginary scenario of finding a suitable location to open an Asian restaurant in Toronto Canada. Hypothetically assume that there is an individual or organization who wants help to research in opening a restaurant.

To be more specific the idea of this model is to determine the areas in Toronto where the Asian food has good demand, by finding the neighborhoods where there are more Asian restaurants already.

To solve this problem in hand, the below methodology is designed and implemented.

# Methodology

## Data Collection

The venue details with latitude, longitude information are the requirement. This is achieved by joining the 3 datasets listed above.

## Data cleaning

Drop the records with Borough value “not assigned” and fill the unassigned/blank neighborhood values with corresponding Borough values.

## Data transformation

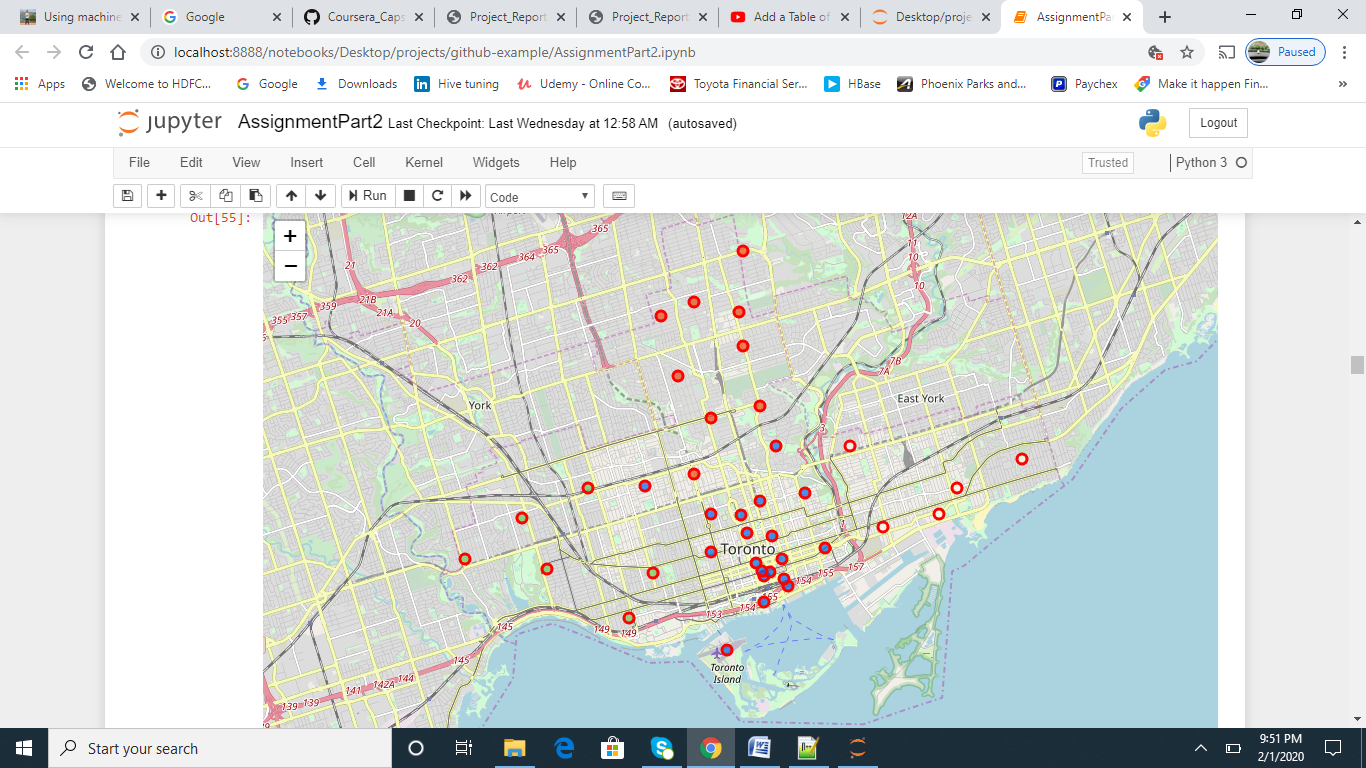
Join the Neighborhood-Postcode data with Postcode-coordinates data to get coordinates for all the neighborhoods.

Then after getting all the venues from Foursquare API locate and map them to corresponding neighborhoods. Remap all the different Asian food restaurants to “Asian Restaurants”.

Then filter the Asian restaurant per each neighborhood.

## Data visualization

Locate the different neighborhoods with Asian restaurants on the Toronto Map.



## Model Building

To get to the solution, I want to create a k-means cluster of the neighborhoods into 3 clusters, depending on the Asian restaurant count per neighborhood, with the intention of finding the cluster with neighborhoods which are popular for Asian restaurant.

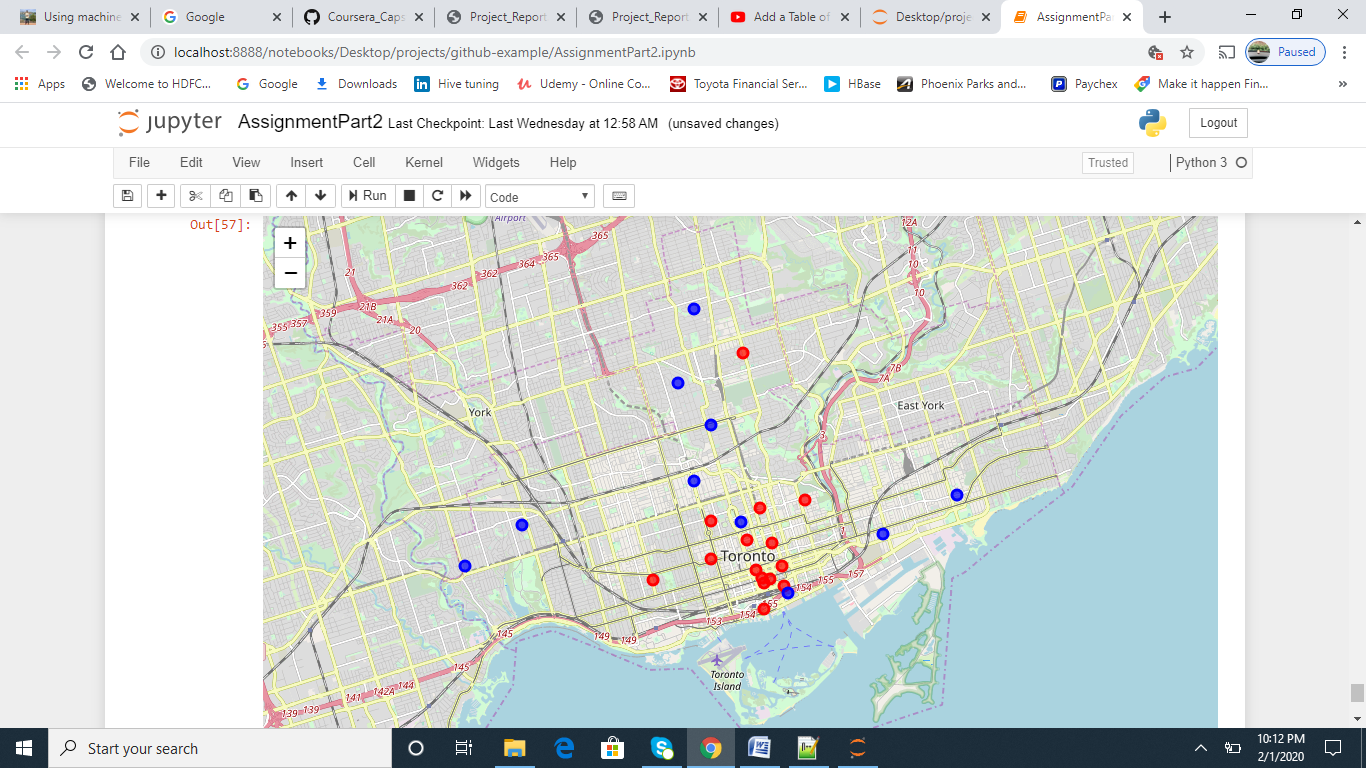
# Results

Looking at the resultant clusters it is clear that

Cluster 0: group of the neighborhoods with medium popularity for Asian food  
**Cluster 1: group of the neighborhoods with high popularity for Asian food**

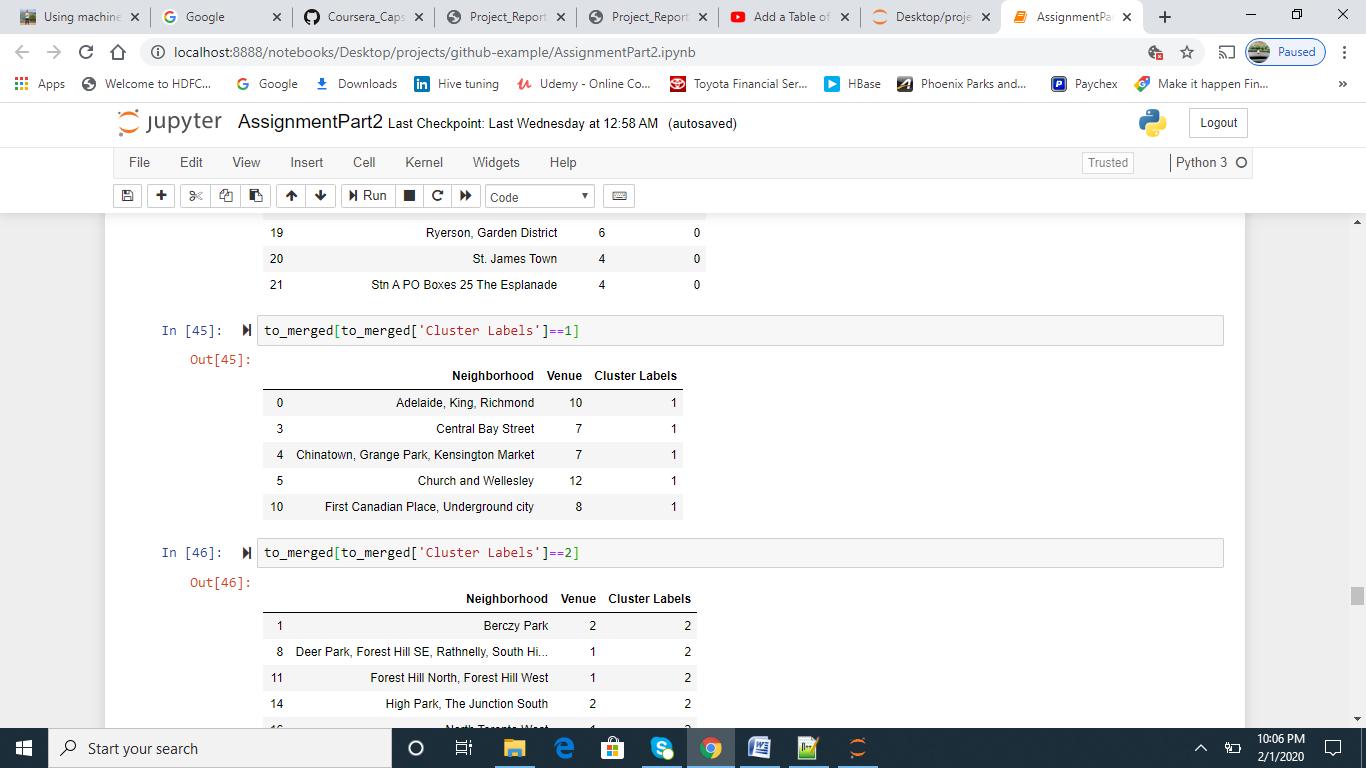
Cluster 2: group of the neighborhoods with low popularity for Asian food

From cluster 1 these are the list of neighborhoods.(Marked Red in the figure)



# Recommendations

Cluster 1 is the list of neighborhoods with high popularity for Asian food. So, we can recommend that these neighborhoods will be a good choice to open an Asian restaurant.



References.  
Wiki for postal code details.

Geospatial data for location details.

Foursquare API for venue details.