DATA SCIENCE PROJECT WEEK- 4

A problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute.

INTRODUCTION

This is a capstone project for IBM Data Science Professional Certificate.In this project, I am creating a hypothetical scenario for a concept that there may not be enough Indian Restaurants in Toronto Area. Therefore it might be a great opportunity for an entrepreneur who is based in Canada. As the Indian food is popular among Asian community, so this entrepreneur might think of opening its business in areas where asian community resides.With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

BUSINESS PROBLEM

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Indian Restaurant in Toronto,Canada. By using data science methods and tools along with machine learning algorithms such as clustering, this project aims to provide solutions to answer the business question : In Toronto, if an entrepreneur wants to open an Indian Restaurant, where should they consider opening it?

TARGET AUDIENCE

The entrepreneur who wants to find the location to open authentic Indian restaurant.

DATA

To solve this problem, we will need below data: List of neighborhoods in Toronto, Canada Latitude and Longitude of these neighborhoods Venue data related to Indian restaurants. This will help us find the neighborhoods that are more suitable to open an Indian Restaurant.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide, King, Richmond	100	100	100	100	100	100
Berczy Park	56	56	56	56	56	56
Brockton, Exhibition Place, Parkdale Village	23	23	23	23	23	23
Business Reply Mail Processing Centre 969 Eastern	16	16	16	16	16	16
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	16	16	16	16	16	16
Cabbagetown, St. James Town	43	43	43	43	43	43
Central Bay Street	87	87	87	87	87	87
Chinatown, Grange Park, Kensington Market	100	100	100	100	100	100
Christie	16	16	16	16	16	16
Church and Wellesley	84	84	84	84	84	84
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	35	35	35	35	35	35
Davisville North	8	8	8	8	8	8
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West	15	15	15	15	15	15
Design Exchange, Toronto Dominion Centre	100	100	100	100	100	100
Dovercourt Village, Dufferin	14	14	14	14	14	14
First Canadian Place, Underground city	100	100	100	100	100	100

EXTRACTING THE DATA

 Scrapping of Toronto neighborhoods via Wikipedia Getting Latitude and Longitude data of these neighborhoods via Geocoder package. Using Foursquare API to get venue data related to these Neighbourhoods.for extracting data we use clustering.

```
array(['Trail', 'Health Food Store', 'Pub', 'Pizza Place', 'Neighborhood',
       'Greek Restaurant', 'Cosmetics Shop', 'Ice Cream Shop',
       'Italian Restaurant', 'Brewery', 'Yoga Studio', 'Juice Bar',
       'Fruit & Vegetable Store', 'Dessert Shop', 'Restaurant',
       'Bubble Tea Shop', 'Bookstore', 'Furniture / Home Store', 'Spa',
       'Grocery Store', 'Diner', 'Indian Restaurant',
       'Caribbean Restaurant', 'Coffee Shop', 'Bakery', 'Café',
       'American Restaurant', 'Sports Bar', 'Liquor Store',
       'Burger Joint', 'Gym', 'Fish & Chips Shop', 'Park',
       'Sushi Restaurant', 'Pet Store', 'Burrito Place', 'Steakhouse',
       'Fast Food Restaurant', 'Movie Theater', 'Sandwich Place',
       'Board Shop', 'Food & Drink Shop', 'Fish Market', 'Gay Bar',
       'Seafood Restaurant', 'Cheese Shop', 'Middle Eastern Restaurant',
       'Thai Restaurant', 'Comfort Food Restaurant', 'Chinese Restaurant',
       'Stationery Store', 'Coworking Space', 'Latin American Restaurant',
       'Bar', 'Gastropub', 'Gym / Fitness Center', 'Convenience Store',
       'Bank', 'Clothing Store', 'Thrift / Vintage Store', 'Lawyer',
       'Swim School', 'Bus Line', 'Breakfast Spot', 'Hotel',
       'Asian Restaurant', 'Sporting Goods Shop', 'Salon / Barbershop',
       'Mexican Restaurant', 'Health & Beauty Service', 'Gift Shop',
       'Bagel Shop', 'Rental Car Location', 'Toy / Game Store',
       'Gourmet Shop', 'Pharmacy', 'Farmers Market', 'Indoor Play Area',
       'Fried Chicken Joint', 'Discount Store', 'Dance Studio',
       'Summer Camp', 'Playground', 'Supermarket',
       'Vietnamese Restaurant', 'Light Rail Station',
       'Japanese Restaurant', 'General Entertainment', 'Jewelry Store',
       'Butcher', 'Deli / Bodega', 'Taiwanese Restaurant', 'Market',
       'Beer Store', 'Snack Place', 'Theme Restaurant',
       'Ramen Restaurant', 'Tea Room', 'Creperie', 'Ethiopian Restaurant'],
      dtype=object)
```

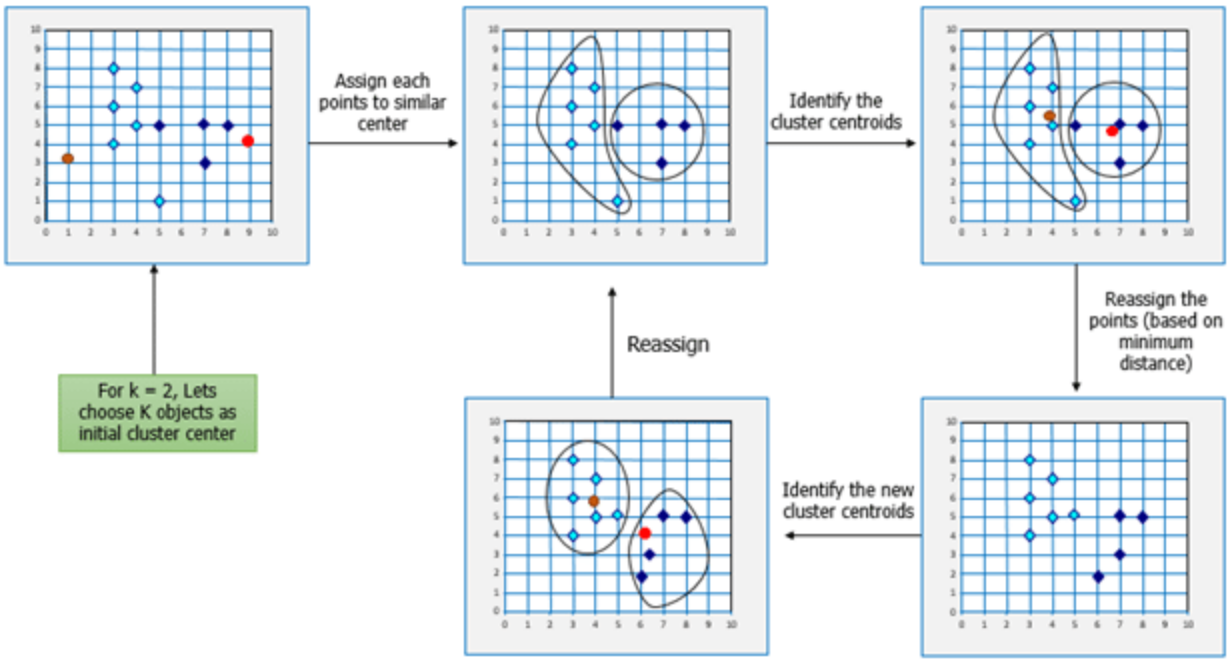
CLUSTERING

K-means Clustering Method:

If k is given, the K-means algorithm can be executed in the following steps:

- Partition of objects into k non-empty subsets
- Identifying the cluster centroids (mean point) of the current partition.
- Assigning each point to a specific cluster
- Compute the distances from each point and allot points to the cluster where the distance from the centroid is minimum.
- After re-allotting the points, find the centroid of the new cluster formed.

THE STEP BY STEP PROCESS



RESULTS

- ● Cluster 0: Neighborhoods with the less number of Indian
- restaurants.
- ● Cluster 1: Neighborhoods with no Indian restaurants.
- ● Cluster 2: Neighborhoods with a more number of Indian restaurants

DISCUSSION

Most of the Indian restaurants are in cluster 2 which is around Central Bay Street, Church and Wellesley, Berczy Park, Union Station, Richmond, lowest in Cluster 1 areas which are in North Toronto West and Parkade areas.