IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

CAPSTONE PROJECT

Opening an Indian Restaurant in Calgary

1. Introduction to business problem
2. Data sources
3. Exploring data set
4. Cleaning data and creating new data frame
5. Methodology
6. Observations
7. Results
8. Conclusion & future developments

1.Introduction

Calgary is one of the Canadian cosmopolitan cities famous for its status as the center of Canada's oil industry. It has consistently been recognized for its high quality of life. Calgary is a home for opportunities. People migrate to this beautiful city to establish new career and businesses for a better lifestyle.

Business Problem

The aim of this capstone project is to analyze the data of various neighborhoods in Calgary and determine best locations to set up a new business "Indian Restaurants".

Target audience for this project

Calgary is one of the fastest growing cities in Canada, with a population over 1.3 million. It is constantly developing new neighborhoods and inviting the investors to contribute to the growth of the city. This project would assist the investors and stakeholders to determine their industry for future investments.

2.Data Sources

Data required

Data needed to provide the targeted insights:

* List of neighborhoods in Calgary including new communities.
* Geographical data of these neighborhoods, like latitude and longitude information to generate map for visualization.

* Data of various businesses(venues) to apply machine learning methodology to produce insights to determine which businesses would be profitable to launch.

Sources of Data

1. Calgary Wikipedia: https://en.wikipedia.org/wiki/Calgary
This data is used in listing the existing neighborhoods.
2.Opendata-CityofCalgary:
https://data.calgary.ca/browse?tags=open+data
This site is used in obtaining new neighborhoods and their communities.
3. Foursquare API: https://foursquare.com/
Foursquare API is used to obtain latitude, longitudes and venue data for the different neighborhoods

Data Cleaning

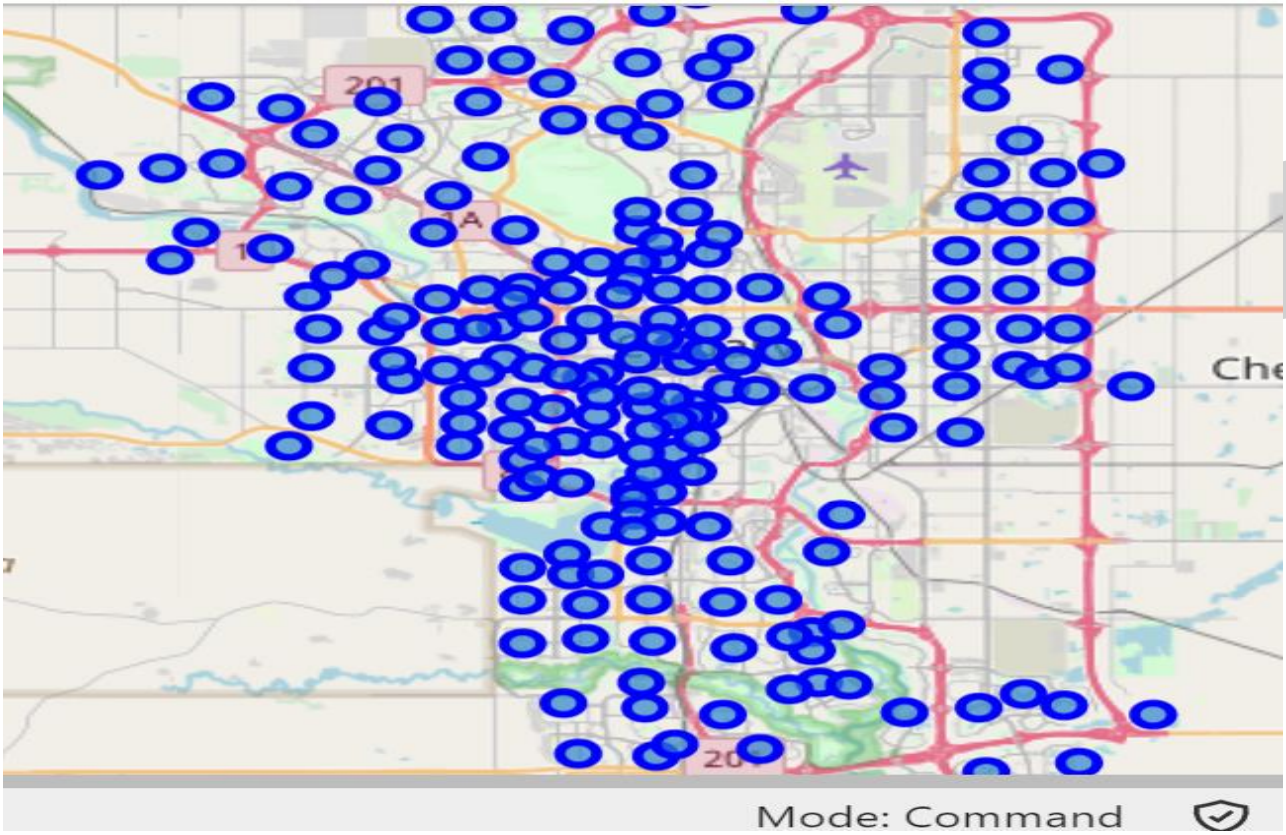
The necessary libraries are imported into the Jupyter notebook. The required data is then extracted from the websites and loaded into the pandas data frame.
Data is organized in the required format for further analysis.

3.Exploring Data Set

Calgary dataset:

- The data frame has 8 sectors and 213 neighborhoods.
- Using Four Square API to obtain latitude, longitude and venue information for Calgary

Calgary Map showing all the neighborhoods



4.Cleaning data and creating new data frame

Extracting venue data for all the neighborhoods of Calgary within a radius of 1000 and limit of 100 with the help of Four-Square API. It is found that there are 260 unique categories of venues(businesses).

After the data is extracted, new data frame consisting of data related to Indian Restaurant in each neighborhood is created. A unique venue list is printed to observe the different businesses in the city.

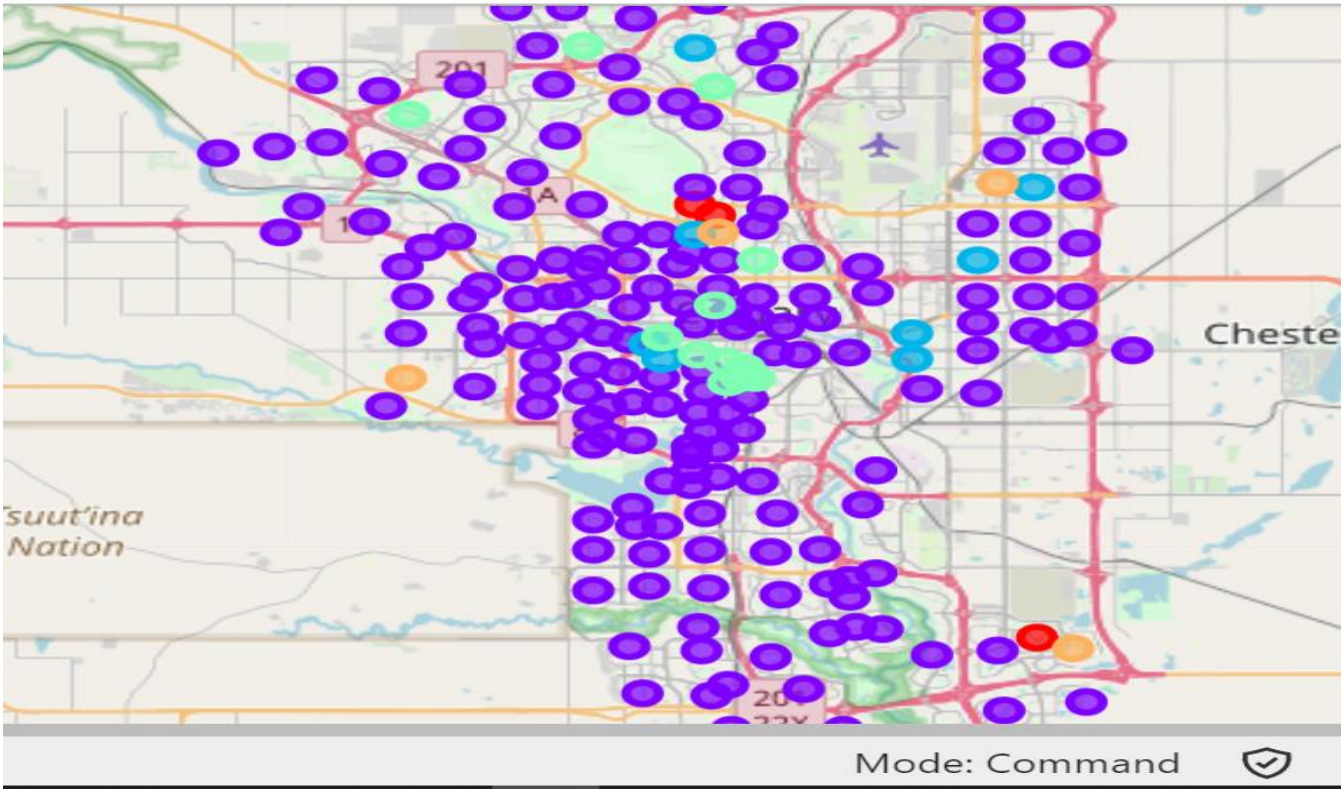
Output of python code showing unique venue categories

```
Array: (['Grocery Store', 'Coffee Shop', 'Garden Center', 'Gastropub',
'Bank', 'Fast Food Restaurant', 'Pharmacy', 'Indian Restaurant',
'Mexican Restaurant', 'Construction & Landscaping',
'Movie Theater', 'Supermarket', 'Café', 'Recreation Center',
'Liquor Store', 'Athletics & Sports', 'Hotel', 'ATM',
'Health & Beauty Service', 'Gym', 'Playground', 'Park',
'Ice Cream Shop', 'Food & Drink Shop', 'Brewery', 'Restaurant',
'Vietnamese Restaurant', 'Music Store', 'Korean Restaurant', 'Pub',
'Bagel Shop', 'Smoke Shop', 'Snack Place', 'Sushi Restaurant',
'Bar', 'American Restaurant', 'Shopping Mall', 'Comic Shop',
'Rental Car Location', 'Wine Shop', 'Golf Course',
'Convenience Store', 'Pizza Place', 'Lake', 'Sandwich Place',
'Chinese Restaurant', 'Skating Rink', 'Video Store', 'Theme Park',
'Electronics Store'], dtype=object)
```

5.METHODOLOGY

- Clustering the neighborhoods in Calgary
- Run k-means to cluster the neighborhoods in Calgary into 5 clusters.

Calgary map indicating the 5 clusters



6.OBSERVATIONS

The following observations have been made:

- Cluster 0: This cluster 3 neighborhoods and has equal number of Indian restaurants.
- Cluster 1: It can be observed that about 179 neighborhoods in Calgary don't have a signification number of Indian restaurants. Starting a new Indian restaurant in these neighborhoods is a good option for new investors.
- Cluster 2: This cluster 9 neighborhoods and has Indian restaurants.
- Cluster 3: This cluster 12 neighborhoods and has Indian restaurants.
- Cluster 4: This cluster 4 neighborhoods and Indian restaurants are present in these neighbourhoods.

7.RESULTS

❖ Neighborhoods in Calgary that can become a potential area to start new Indian Restaurants.

❖ Data shows that

- 40 neighborhoods in Central Calgary
- 25 in the West
- 8 in the East
- 16 in the North
- 38 in the South
- 9 in southeast
- 23 in northwest
- 18 in northeast

for starting new Indian restaurant in Calgary.

8.CONCLUSION & FUTURE DEVELOPMENTS

- Data shows that about 178 neighborhoods in Calgary have a potential to launch new Indian Restaurants and improve the economy.
- This dataset can be combined with dataset consisting of immigrant information, employment and transportation to make more specific assumptions.