**THE BATTLE OF NEIGHBORHOODS**

**FIND OUT BEST LOCATION TO OPEN INDIAN RESTAURANT IN EDMONTON, CANADA**

**­**

**Peer-graded Assignment: Capstone Project**

**BY**

**PRABIN DAHAL**

**12/3/2019**

**FIND OUT BEST LOCATION TO OPEN INDIAN RESTAURANT IN EDMONTON, CANADA**

**TABLE OF CONTENTS**

1. **INTRODUCTION**
2. **DATA SECTION**
3. **METHODOLOGY**
4. **RESULTS**
5. **DISCUSSION**
6. **CONCLUSION**

**FIND OUT BEST LOCATION TO OPEN INDIAN RESTAURANT IN EDMONTON, CANADA**

1. **INTRODUCTION**
   1. **BACKGROUND**

***Being a Canadian citizen and living in Edmonton for more than 7 years, I want to focus my project work within same city. My project will help for Restaurant business owners who wants to start their business career in Edmonton city and looking for better business location within this city to open an Indian restaurant.***

* 1. **BUSINESS PROBLEM**

***Anyone who wants to start business in new city is challenging if they are not familiar with the location and they are not aware if business is targeted for the community within the neighborhood. Searching a best location for Restaurant is not an easy task in today’s world as there are several requirements needed to fulfill to get this task done. On top of that, we need to find better place where targeted customers are interested in your business. At present scenario, there are multiple restaurant businesses operating in almost each location. Customer are so selective that they want to visit/order from restaurant where they can find food of their choice.***

* 1. **Target**

***As a student of data science, I want to provide some research so that it will be easier for anyone who wants to start their career in this field as Restaurant owner and overcome this extra headache to find proper location for business.*** *I will provide recommendation based on finding that they can consider before opening new restaurant.*

1. **DATA SECTION**
   1. **Data section Introduction**

*I will provide recommendation based on finding that new restaurant owners can consider before opening new Indian restaurant, my goal will be to find location that is most suitable to open new Indian restaurant in Edmonton. For the purpose, I will collect data with highest number of Indian community and perform data analyses to find the result. Also, I will describe the data that will be used to solve the problem and the source of the data.*

* 1. **Background**

*Edmonton has a total of 12 borough and 400 neighborhoods with total population of 97,2223 (2019). In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 12 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.*

* 1. **Data mining**

*From the data obtained, I will convert addresses into their equivalent latitude and longitude values. Also, will determine Indian community within each location. After merging these files and finding most dense Indian community I will select this denser location and use the Foursquare API to explore neighborhoods in Edmonton city. I will also use the****explore****function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. I will use the*k*-means clustering algorithm to complete this task. Finally, I will use the Folium library to visualize the neighborhoods in Edmonton City and their emerging clusters.*

* 1. **Import libraries**

*For purpose of data analysis, I will import several libraries to perform the project that includes but not limited to below:*

*Numpy to handle data in vectorized manner*

*Pandas for data analysis*

*Json to handle json files*

*Nominatim to convert address in latitude and longitude values*

*Requests to handle requests*

*Json normalize to tranform JSON file into a pandas dataframe*

*Matplotlib and associated plotting modules*

*import k-means from clustering stage*

*folium as map rendering library*

* 1. **Data collection**

*For the purpose of required data of location, Neighborhood, ward, household, particular community information, I will get grab the data from “2016 Census - Dwelling Unit by Language (Neighborhood/Ward)” with html address* [*https://data.edmonton.ca/Census/2016-Census-Dwelling-Unit-by-Language-Neighbourhoo/jsc3-gmwb*](https://data.edmonton.ca/Census/2016-Census-Dwelling-Unit-by-Language-Neighbourhoo/jsc3-gmwb) *data provided by City of Edmonton and API Endpoint* [*https://data.edmonton.ca/resource/jsc3-gmwb.json*](https://data.edmonton.ca/resource/jsc3-gmwb.json)*. I will select Indian community information from this data.*

*I will get geographical location information from City of Edmonton - Neighborhoods (Centroid Point), data provided by City of Edmonton and add latitude and longitude information based on that information.*

* 1. **Download and explore dataset**

*Download data files as json files and or csv files and merge files together to create a data frame we were looking for with information of ward numbers, neighborhoods, latitude and longitude of location and community within each location. Transform the data into a pandas data frame. Create a map of Edmonton with neighborhoods superimposed on top.*

* 1. **Segmenting and clustering neighborhoods in Edmonton city.**
  2. **Exploring neighborhoods.**

*Explore the neighborhoods that has highest number of Indian communities. Get latitude and longitude of the neighborhoods and explore top venues falls within that neighborhoods.*

* 1. **Analyze each neighborhood.**

*Get neighborhoods info with top venues information.*

* 1. **Cluster neighborhoods.**

*Run*k*-means to cluster the neighborhood*

*Analyze top venues for each neighborhood.*

*Visualize clusters*

* 1. **Examine clusters**

*Examine each cluster and determine the discriminating venue categories that distinguish each cluster*

1. **METHODOLOGY**

*As mentioned in data section, From the data obtained, all addresses are converted into their equivalent latitude and longitude values. Also, Indian community within each location is determined. After merging these files and finding most dense Indian community, ward belonging to these community are selected and the Foursquare API is used to explore neighborhoods in Edmonton city. Also, the****explore****function was used to get the most common venue categories in each neighborhood, and then used this feature to group the neighborhoods into clusters. The*k*-means clustering algorithm is to complete this task. Finally, The Folium library is to visualize the neighborhoods in Edmonton City and their emerging clusters. Most relevant steps during coding are as belows.*

***Downloaded libraries.***

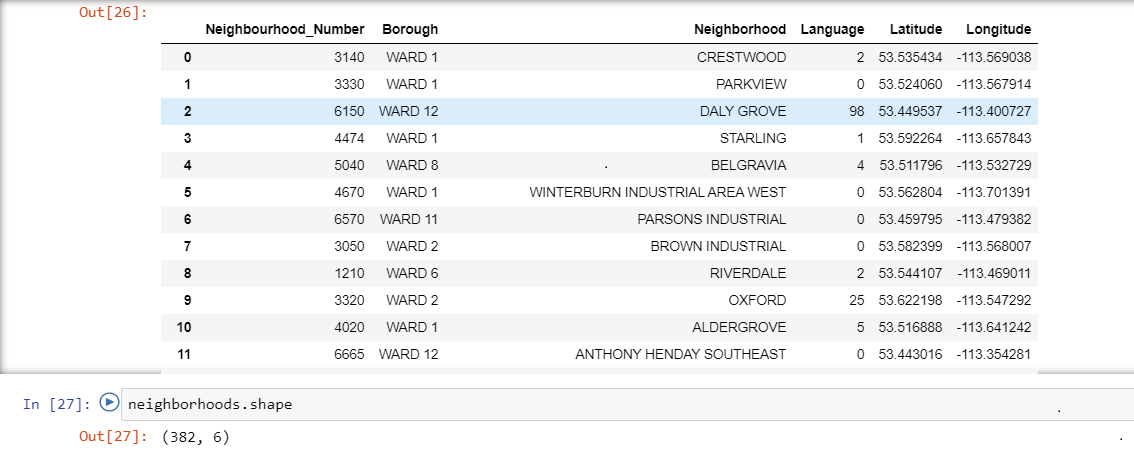
***Downloaded data file (File number 1) containing (neighborhoods name, ward numbers and household language spoken information) json file and opened it.***

***Loaded and explored the data.***

***Transformed the data into a pandas data frame.***

***Another data file (File number 2) containing latitude and longitude downloaded for geographical positioning.***

***Merged 2 data files (File number 1 and File number 2) to create single data files for assessment.***

***Picture1: Data frame with ward number and neighborhood.***

***The data frame had 12 boroughs and 382 neighborhoods.***

***Data frame with Borough and language are separated to find number of Punjabi speaking household in given neighborhoods.***

***Count of each ward in given data assessed.***

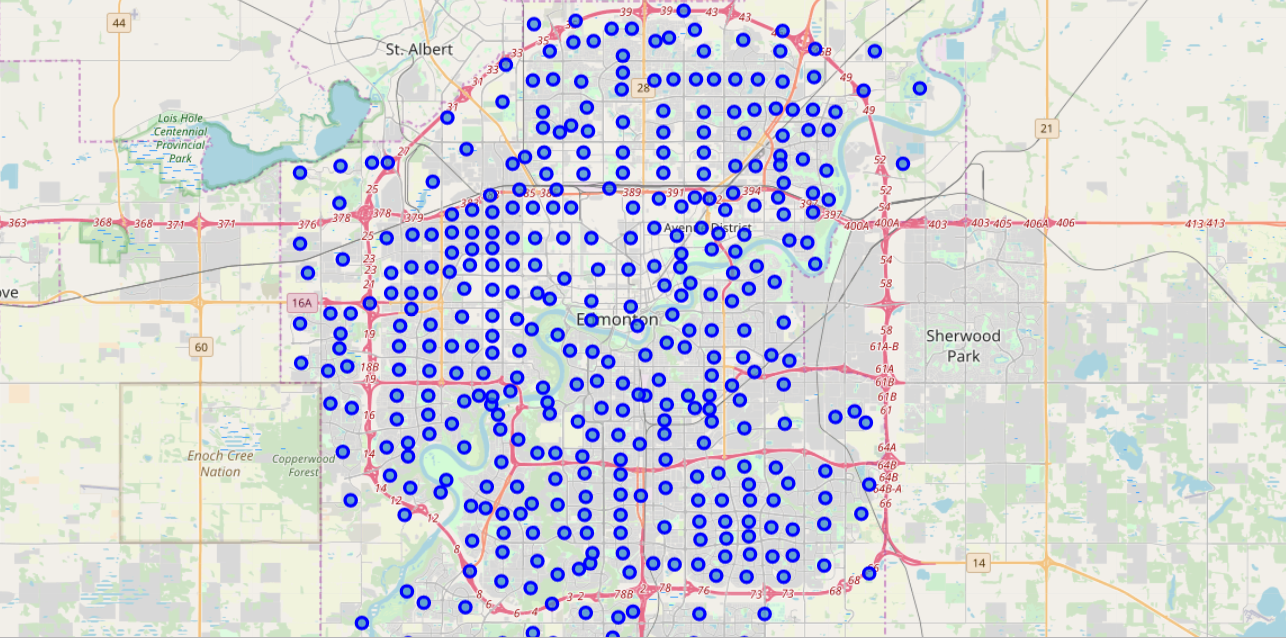
***Punjabi speaking households for each ward is separated by group by code.***

***Ward number 12 has maximum numbers of Punjabi speaking households.***

***Use geopy library to get the latitude and longitude values of Edmonton***

***The geographical coordinate of Edmonton are 53.535411, -113.507996.***

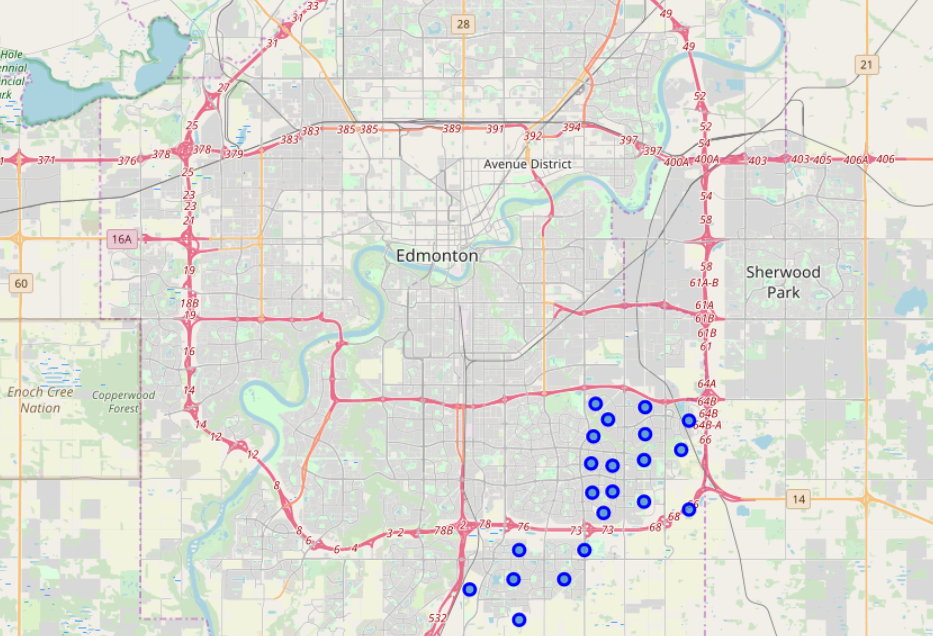
***Created a map of Edmonton with neighborhoods superimposed on top.***

******

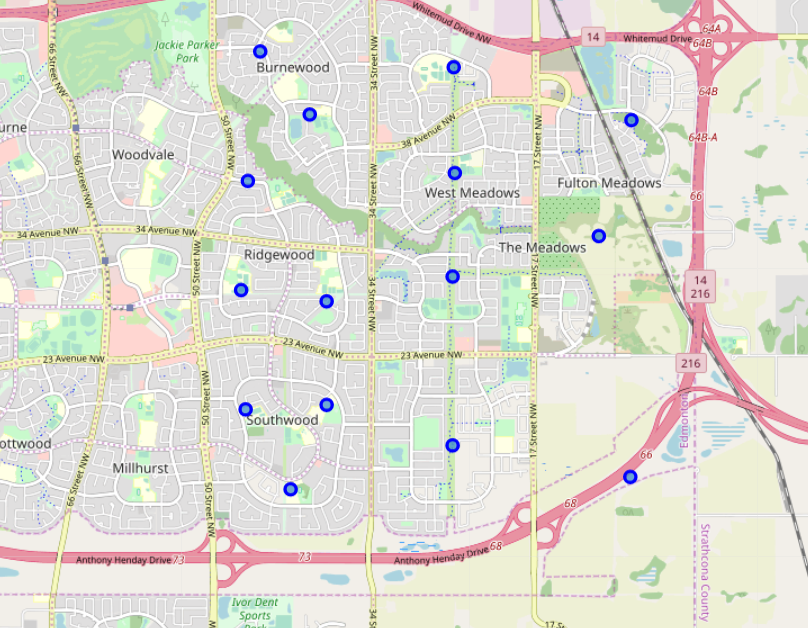
***Picture2: Map of Edmonton with neighborhoods superimposed on top.***

***sliced the original data frame and created a new data frame of the ward number12.***

***As we did with all of Edmonton City, let's visualize WARD 12 the neighborhoods in it.***

******

***picture 3- visualize WARD 12 the neighborhoods in it.***

******

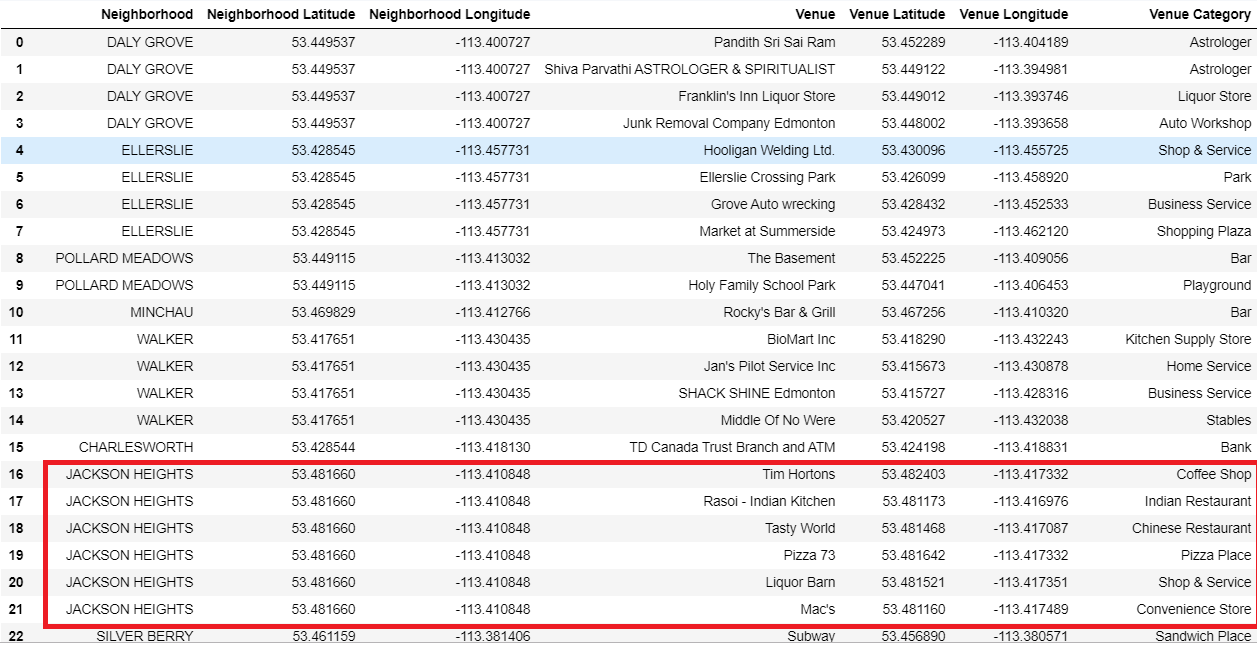
***Picture 4 -visualize WARD 12 the neighborhoods in it. Zoom***

***Utilized the Foursquare API to explore the neighborhoods and segment them.***

***Defined Foursquare Credentials and Version***

***Got the top 100 venues that are in neighbor Daly grove.***

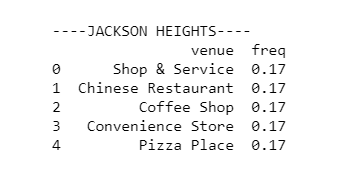
***Explored Neighborhoods in ward 12***

******

***Picture 5: Neighborhoods in ward 12***

***Analyzed Each Neighborhood***

***print each neighborhood along with the top 5 most common venues***

******

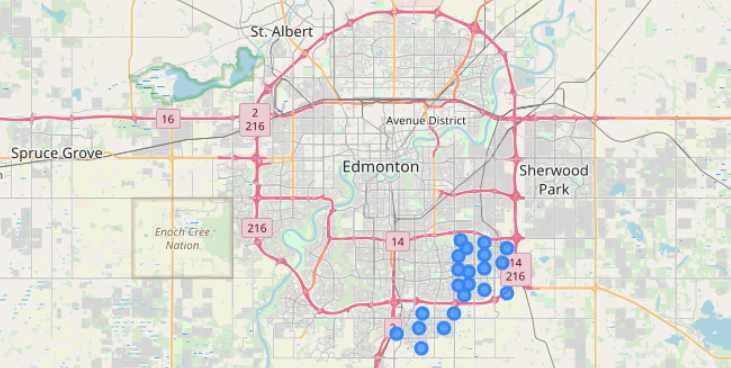
***picture6: neighborhood along with the top 5 most common venues***

***Clustered Neighborhoods***

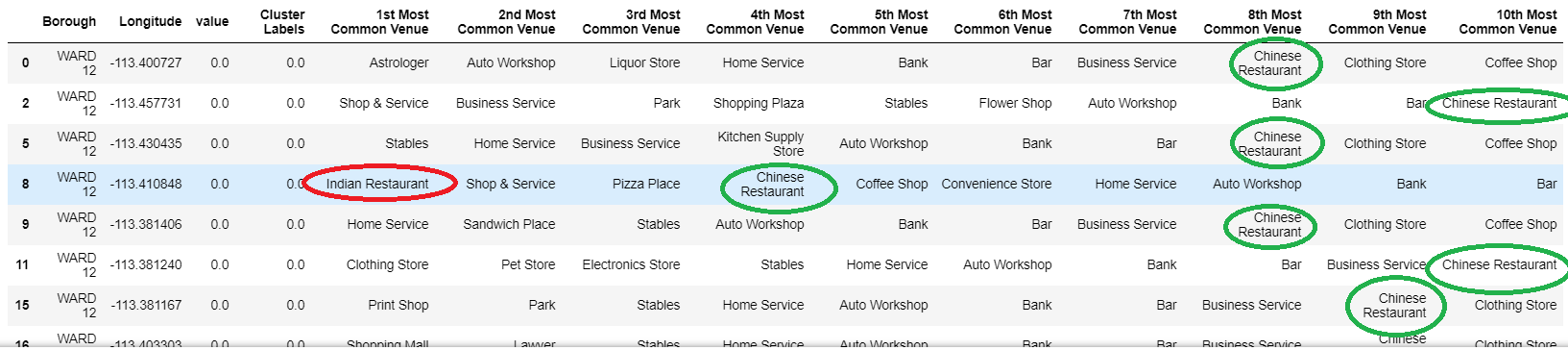
***Ran k-means to cluster the neighborhood into 5 clusters.***

***visualized the resulting clusters***

***Examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories***

******

***Picture 7: discriminating venue categories that distinguish each cluster.***

****

***picture 8: determine the discriminating venue categories that distinguish each cluster***

1. **RESULTS**

***Jackson Height Neighborhood that is in ward number 12 of Edmonton city shows best place to open Indian restaurant.***

***From the data, it’s clear that majority of Indian community are in ward number 12 and within that community Jackson height neighborhood has only one Indian restaurant and majority of top venues in this neighborhood is food service.***

1. **DISCUSSION**

***From the project work, we can recommend to open Indian restaurant in Jackson Height Neighborhood which is in ward number 12 of Edmonton.***

***We had received data from City of Edmonton census 2016 with neighborhoods containing household of different language. We selected data from Punjabi speaking household to figure out neighborhood with Indian communities. It shows out of 12 wards, Ward number 12 had household that speaks Punjabi languages and thus most Indian communities are in this category.***

***We selected ward number 12 as our data to access more information as it had most Indians and our project work is related to open Indian restaurant. In this ward, we can see there are many more Chinese restaurants but there are lesser number of Indian restaurants. This may be because of lack of enough data provided by foursquare. We may need to get more relevant data to assess exact number of Indian restaurants with this community.***

***Most of the neighborhood has several types of venues of business that is ranked as top 5 but Jackson Height Neighborhood that is within ward number 12 is the only one that shows top 5 venues of food service and there is no Indian restaurant among this top 5. So, based on the resources of data, I would recommend this location. We can further research in future if there is any other better location when we will collect more relevant data that include crime rate, rent cost etc.***

1. **CONCLUSION**

***From the project, we received data from city of Edmonton and using foursquare data, I would suggest opening Indian restaurant in Jackson Height Neighborhood of Edmonton city which is ward number 12.***

1. **REFERENCE**

[*https://developer.foursquare.com/*](https://developer.foursquare.com/)

[*https://data.edmonton.ca/Census/2016-Census-Dwelling-Unit-by-Language-Neighbourhoo/jsc3-gmwb*](https://data.edmonton.ca/Census/2016-Census-Dwelling-Unit-by-Language-Neighbourhoo/jsc3-gmwb)

[*https://data.edmonton.ca/City-Administration/City-of-Edmonton-Neighbourhoods-Centroid-Point-/3b6m-fezs/data*](https://data.edmonton.ca/City-Administration/City-of-Edmonton-Neighbourhoods-Centroid-Point-/3b6m-fezs/data)