Predicting the Chances for Opening an Italian Cuisine Restaurant IN CANADA

SOWMYA.C

31 DEC 2020

PROBLEM STATEMENT

Italian Cuisine is the most loved cuisine in the world. This Cuisine is loved by the Canadians as well so we have to predict the perfect place for the setup and central place where the crowd will get attention on the hotel and there should be limited Italian restaurants around using the machine learning techniques, four square API and the capstone.

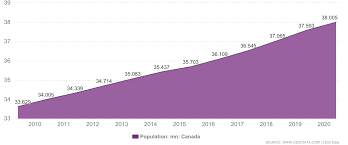
DATA

1. Total restaurants in Canada

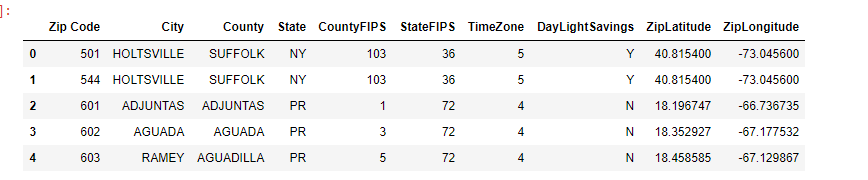
2. Area that has less Italian restaurant.

3. Centralized area for setup

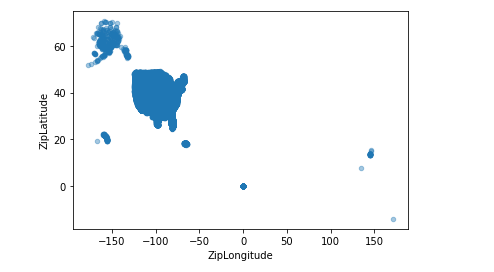
4. Population in particular area

5. Most loved Italian Cuisines

The above data represents the population chart of Canada.

****

The above table represents the dataset for the city of Canada.



The above data are to collected using **geocoding** and **foursquare API.**

METHODOLOGY

In this project we will direct our efforts on detecting areas of Canada that have low restaurant density, particularly those with low number of Italian restaurants. We will limit our analysis to area ~9km around city center.

In first step we have collected the required data: location and type (category) of every restaurant. We have also identified Italian restaurants (according to Foursquare categorization).

Foursquare is a social location service that allows users to explore the world around them. Users can download the Foursquare application to their I phone, Blackberry, or Android phone and sign up for free, then connect their Foursquare accounts to their other social media accounts.

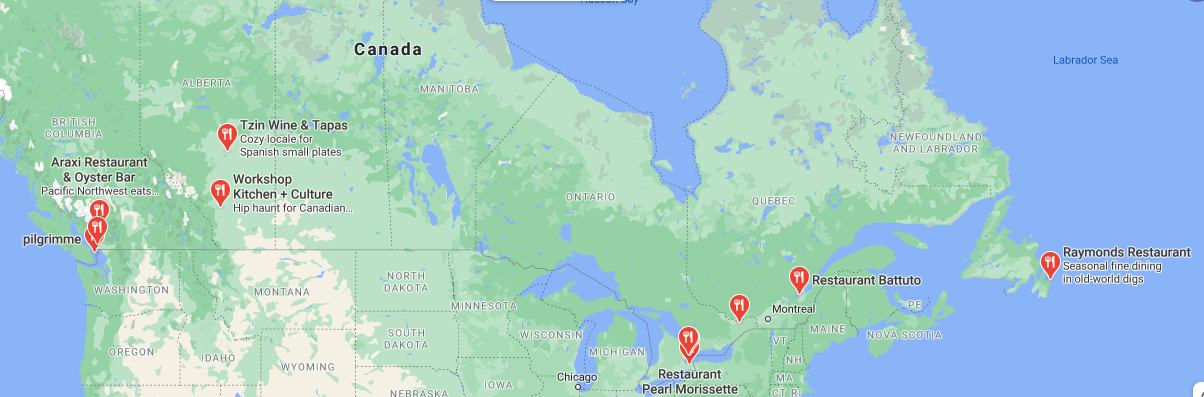
Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Canada - we will use heat maps to identify a few promising areas close to center with low number of restaurants in general and focus our attention on those areas.

A geocode is the set of latitude and longitude coordinates of a physical address. Geocoding is the process of turning a physical address into a set of latitude and longitude coordinates which can then be plotted or displayed on a map.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Italian restaurants in radius of 400 meters by k-means clustering.

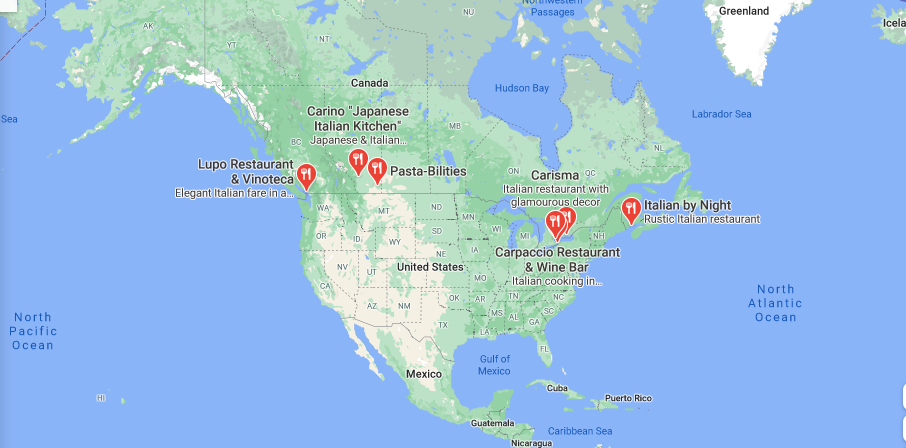
k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.

ANALYSIS

The total number of restaurants in the central city are counted.

Then the Italian restaurants are then counted along with the population and the heat map is used to locate them. 

The top running hotels are taken and then the specialty is noted and areas with less Italian restaurants are noted.



RESULTS AND DISCUSSIONS

Here based on the population density and central areas some restaurants are selected and then based on the facts worst running Italian restaurants and places with no Italian restaurants some places are selected.

Result of all this is 10 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant.

Purpose of this analysis was to only provide info on areas close to Canada center but not crowded with existing restaurants (particularly Italian) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

CONCLUSION

The area suitable for a Italian restaurant in Canada has to be chosen with less number of competitive restaurants, population, people taste, quality and so on. The central part that satisfies all the constraints has been selected through k-means clustering, Geocoding , Foursquare API and Google API.  
The analysis has been made completely based on machine learning and clustering whereas the setup include other factors like design, construction, money, quality, noise and so on which has to be decided by stakeholders