Clustering boroughs based on cuisines

Pranav Gurditta

15 January 2020

Introduction where you discuss the business problem and who would be interested in this project.

Problem statement: Suppose a company wants to know which boroughs of Canada have the same type/cuisines of restaurants so that
it can know the number of types of stores of raw food it should open in the whole of Canada.
For example, if a borough in Canada is a part of cluster that offers Chinese and Japanese food the company would open a store
in that area and all the areas in that cluster that offer raw food material of Chinese and Japanese dishes.

Benefits of solving this:

This would help the restaurants also as it would bring down the transportation costs of raw food items and hence provide cheaper
food to the customers.

The company can gain a lot of profit by becoming the nearest seller as restaurants would need these items everyday.
The company can also filter which raw materials to send to a store on a daily basis.

Data where you describe the data that will be used to solve the problem and the source of the data

We will be extracting neighborhoods and boroughs of Canada using wikipedia page.
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Finding all the venues in each borough and then find the venue category of it using ForeSquare API.

Then filter the columns based on type of restaurant

Then grouping data based on different type of boroughs in Canada , the type is based on the type of restaurants present in that borough.

Different clusters of boroughs based on the type/cuisines of restaurants present in that area.

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

We do one hot encoding of the venue category .

We use K-Means to cluster the boroughs and the elbow method to find the best K.

Results section where you discuss the results.

The appropriate number of clusters are determined using elbow method and hence we get a K value of 4.

We cluster the boroughs into 4 clusters based on cuisines present in the boroughs.

The result is that we have clustered the boroughs of Canada based on the cuisine offered by the restaurants in that area, hence we now know the raw materials to supply in different boroughs. This helps the company to know which store has to be supplied with which kind of raw material so as to increase the profit of the company.

Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

1. The clustering of boroughs also tells us about the origin of citizens residing in that area. For example a cluster of Downtown Toronto,West Toronto and Central Toronto has a large number of Italian cuisine restaurant hence a lot of Italian origin citizens resides in these boroughs.
2. The clustering of boroughs also tell us about the language preferred in a particular borough. For example: Greek cuisine is popular in East Toronto , hence greek language is more popular in this cluster.