FOOD VENUES IN CANADA

Mehul Prajapati

IBM Certification

Introduction

Background

Food venues are one of the most numerous and most diverse businesses in any major City. Following economic recessions, competition among these small businesses is fierce. Information on where different types of food restaurants are aggregated in major Canadian cities may help business owners understand both who their competition is but also where local population demographics may favor certain restaurant types.

Problem

How do geographic areas within a city and between cities compare, with regard to concentrations of different categories of food businesses?

Audience

This project is aimed at restauranteurs, realtors, small business owners, or even consumers, who might benefit from more insight into the distribution of different types of food businesses between postal code locations and cities.

Data

Canadian Postal Codes

I examined the three largest Canadian municipalities, using postal code lists collected on Wikipedia. These tables were scraped using pandas read HTML and reshaped as needed.

Foursquare Location Data

Foursquare's location dataset is one of the most comprehensive available. Built from 13+ billion crowd-sourced "check-ins" since 2009, the data is used by over 150,000 developers such as Apple, Samsung, Twitter, and Uber.[3] Foursquare's data is freely accessible via its Places API. The API's Search endpoint returns up to 50 venues within a radius of a location, including basic venue data such as name, categories, and location.

Method

Overview

All data collection and analysis were performed in Jupyter Notebook running Python v.3.7.3.

After scraping 261 FSAs from Wikipedia, I used them to iteratively query the Foursquare Search endpoint to return 50 venues of the umbrella category 'Food' within 1 km of each postal code. This

resulted in a dataset of 5798 venues distributed throughout each city, to represent the characteristics of each postal code area and each city.

I used Foursquare's categories hierarchy to further categorize each venue in broader categories. (E.g. 'Sushi Restaurant' may also be categorized as 'Japanese Restaurant' and 'Asian Restaurant'.)

For analysis, I used one-hot encoding on the categories, followed by k-means clustering machine learning to cluster the FSAs according to their characteristic food categories. The top categories per FSA, per city and per cluster were then determined, as well as the breakdown of clusters and categories between cities.

I then created map visuals of the postal codes, venues and clusters.

Modules he python modules imported for use are listed be

Postal Code Data

I scraped the following Wikipedia tables of postal codes in Canada:

1. Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. Montreal: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_H
3. Calgary: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T

Web scraping was achieved using:

```
html = requests.get(url).content df = pd.read_html(html)
```

Tables were re-structured to obtain a list of 261 FSAs.

Inapplicable FSAs were excluded such as those designated ‘Not Assigned’. The letter ‘T’ FSAs include all of the province of Alberta. Cities other than Calgary were filtered out.

Foursquare Data

I used a function to iteratively call the Foursquare API with a search of the 50 closest venues within a radius 1km of each postal code. The Foursquare Search endpoint returns the nearest venues to the point of origin, matching the search criteria: the ‘Food’ top level category id. For simplicity I only included the primary category of each venue, as some venues have multiple categorizations, but this is rare.

The following data elements were extracted from the JSON into a Pandas DataFrame:

```
response['geocode']['feature']['geometry']['center']['lat']
response['geocode']['feature']['geometry']['center']['lng'] For v in
requests.get(url).json()['response']['venues']:

v['id'],
v['name'],
```

```
v['location']['lat'],
v['location']['lng'],
v['categories'][0]['name'], # primary category only v['categories'][0]['id']
```

Similarly, the Foursquare categories hierarchy was extracted from JSON by looping through each hierarchy level to create a DataFrame with additional rows for each parent category, as below.

The ‘Coffee Shop’ and ‘Café’ categories were excluded as they dominated the data at 19% of total venues. The remaining categories were merged to the original venues data.

Analysis

Categories of each venue were one-hot encoded to produce a wide DataFrame counting the applicable categories for each venue.

```
pd.get_dummies(df[['Categories']])
```

After dropping venues with category mismatches, the final count of unique venues was 5798.

The encoded data was then grouped by FSA, calculating the mean, as the proportion of nearby venues that matched each category.

```
df.groupby('Place').mean()
```

KMeans Clustering

These values were used for KMeans Cluster analysis. Clustering is a machine learning technique to segment data points into mutually exclusive clusters. It is an unsupervised technique, meaning pre-existing labels are not provided, rather the inferences are made based on data similarities. K-means is a partition-based clustering algorithm which produces sphere-like clusters and is relatively efficient for medium and large datasets.

```
KMeans( n_clusters=5).fit( df.iloc[:,1:]
```

The cluster labels were then merged with the previous data to generate several summary tables breaking down the clusters, cities and FSAs with their most frequent food categories, as below.

City

Montreal, Quebec

Toronto, Ontario

Montreal, Quebec

Montreal, Quebec

Montreal, Quebec

...

Place H1Y M9R H0M H1E H1G

Place Latitude

45.5486 43.6898 45.6986 45.6342 45.6109

Place Longitude

-73.5788 -79.5582 -73.5025 -73.5842 -73.6211

Cluster 0
0
0

0 0

1st Most Frequent

Fast Food Restaurant

Pizza Place

Fast Food Restaurant

Italian Restaurant

Fast Food Restaurant

2nd Most Frequent

Asian Restaurant

Asian Restaurant

Italian Restaurant

Fast Food Restaurant

Asian Restaurant

3rd Most Frequent

Breakfast Spot

Sandwich Place

Xinjiang Restaurant

Pizza Place Restaurant

4th Most Frequent

Italian Restaurant

Chinese Restaurant

Falafel Restaurant

Restaurant

Sandwich Place

5th Most Frequent

Pizza Place

American Restaurant

Food Court Diner

Breakfast Spot

Mapping

The geopy nominatim module was used to obtain coordinates of each city.

```
Nominatim(user_agent="explorer") .geocode(city)
```

Maps were generated using the Folium module. By looping through each city, the coordinates were used to centre the map and corresponding dataframes used to build map markers. In this way I generated maps Toronto, Montreal and Calgary display postal codes alone, venues alone, and postal codes coded by KMeans Cluster.

```
folium.Map(location=[c_lat, c_lng]) folium.CircleMarker([p_lat, p_lng]) ).add_to(map)
```

To display the large amount of venue markers, the FastMarkerCluster plugin was used.

```
map.add_child(FastMarkerCluster(marker_data.values.tolist()))
```

Color coding was assigned using the Matplotlib qualitative colormap ‘Set1’.

```
matplotlib.cm.Set1(np.linspace(0, 1, 9))
```

A legend for the colormap was also added in html.

```
map.get_root().html.add_child(folium.Element(legend_html
```

Asian restaurants, fast food and pizza were common across all clusters and cities. Asian restaurants however are a fairly broad umbrella category for many food subcategories.

Reviewing the clusters revealed some notable differences between Montreal and Toronto, despite the two cities being of similar size and density, and geographically close to each other. Clusters 0 and 1 were predominant in Montreal and revealed that Dessert shops may be relatively common in Montreal.

Each city also had a distinct 5th most frequent food venue near the FSA centers: Vietnamese in Calgary, dessert shops in Montreal, and Japanese in Toronto.

Conclusion

Despite the confound of differences in FSA density, concrete similarities and differences between clusters and cities emerged when analyzing food venue categories near FSA centers.

This exploratory analysis paves the way for a closer look comparing Montreal and Toronto and which food businesses thrive in each of these two large Canadian cities.

References

1. https://en.wikipedia.org/wiki/List_of_the_100_largest_municipalities_in_Canada_by_population 2. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
3. https://foursquare.com/about