Decision Making for Entrepreneurs

case of study of the use of data analytics for help entrepreneur's strategic decisions

Edward Ortiz

October 6, 2019

1. Introduction

1.1 Background

Part of the success for entrepreneurs are their performance and time result when taking strategic decisions. Today the market volatility and strict demands require that any entrepreneur or big companies take decisions based on the data available so the accuracy provide for the organization a value performance and can step ahead their competitors. Data visualization has been a concept that through the years has been perfected the way people can gather the data, process it and manage it so they can create proper strategies either for a financial plan, marketing plan or whatever need the company requires.
This paper, will present a random case of study based from the point of view of an entrepreneur in the industry of food logistics who needs to make a decision for improve their logistics process by using data science techniques for take that decision.

1.2 Problem

An entrepreneur from Toronto, Canada wants to build a warehouse to store the raw products that he acquires from rural farmers and that supply them for local restaurants, groceries, bakeries, etc. The entrepreneur wants to set up the warehouse in a central location so the deliveries can take less time than their competitors and target more potential clients.

1.3 Interest

The entrepreneur wants to use data science techniques for determine which is the better location to place the warehouse.

2. Data acquisition and cleaning

2.1 Data sources

```
In [6]: # for the city Toronto, latitude and longtitude are manually extracted via google search
        toronto_latitude = 43.6932; toronto_longitude = -79.3832
        map_toronto = folium.Map(location = [toronto_latitude, toronto_longitude], zoom_start = 10.7)

        # add markers to map

        for lat, lng, borough, neighborhood in zip(df_can['Latitude'], df_can['Longitude'], df_can['Borough'], df_can['Neighbourhood']):
            label = '{}, {}'.format(neighborhood, borough)
            label = folium.Popup(label, parse_html=True)
            folium.CircleMarker(
                [lat, lng],
                radius=5,
                popup=label,
                color='blue',
                fill=True,
                fill_color='#3186cc',
                fill_opacity=0.7).add_to(map_toronto)


        map_toronto
```

Figure 1 API extraction method

Using foursquare API it was extracted the data of shops by category in Toronto, Canada. Once the dataset was mined, by using Wikipedia, it was extracted the data of postal code and neighborhoods so by preprocessing both data sets a final dataset could be used to have the location b latitude and longitude of each grocery, restaurant, food shop, etc. that is related with entrepreneur activity.

2.2 Data cleaning and preprocessing

Once the data from foursquare was scrapped by using the module “pickle”, a segmentation was made in order to create a small scenario. It was chosen a random neighborhood.

The radius of extraction of shops was of 1000 meters from the central point of the neighborhood. The final dataset to make train the model contained the categories of postal code, neighborhood, latitude, longitude, venue name, popularity of venue, distance for central point and category.

Out[4]:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge ,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek ,Rouge Hill ,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood ,Morningside ,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Table 1 Borough selected randomly

```
In [27]: neigh_onehot = pd.get_dummies(data = neigh_venues, drop_first  = False,
                                       prefix = "", prefix_sep = "", columns = ['Venue Category'])
         neigh_onehot.head()
```

Out[27]:

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Distance	African Restaurant	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	Automotive Shop	BBQ Joint	Badminton Court	Bakery	Bank	Bar	Beach	Beer Store	Bo
0	M1B	Rouge ,Malvern	43.806686	-79.194353	Images Salon & Spa	This spot is popular	595	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	M1B	Rouge ,Malvern	43.806686	-79.194353	Staples Morningside	This spot is popular	735	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	M1B	Rouge ,Malvern	43.806686	-79.194353	Wendy's	This spot is popular	600	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	M1B	Rouge ,Malvern	43.806686	-79.194353	Wendy's	This spot is popular	387	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	M1B	Rouge ,Malvern	43.806686	-79.194353	Harvey's	This spot is popular	796	0	0	0	0	0	0	0	0	0	0	0	0	0	

Table 2 Data preprocessed

3. Selecting data visualization model

3.1. Choosing data analysis method

In order to solve the problem in this case of study, it was chosen the ML technique K-mean clustering. The selection was due by having a dataset which contained information of position of many food venues, the data can be trained to be clustered in different groups the one which were near the center of the neighborhood and based on the popularity. The popularity index gathered from foursquare, provided insights about these frequent shops in which by having a large demand of customers, it had a better probability to demand raw food which is the product offered by the entrepreneur. So based on that 5 clustered were made and the result will provide a neighborhood who can accomplish the problem solution.

Out[38]:

	Bakery	Breakfast Spot	Diner	Fish Market	Food & Drink Shop	Fruit & Vegetable Store	Grocery Store	Noodle House	Pizza Place	Sandwich Place	Total Restaurants	Total Joints	Total Sum
G1	2.000000	1.000000	0.000000	0.0	0.000000	0.0	1.000000	1.0	1.000000	2.000000	21.000000	1.000000	30.000000
G3	2.000000	0.000000	0.000000	0.0	0.000000	0.0	0.666667	1.0	1.666667	1.000000	11.666667	1.666667	19.666667
G4	4.000000	0.000000	0.000000	0.0	0.000000	0.0	1.000000	0.0	0.000000	0.000000	7.000000	3.000000	15.000000
G2	0.400000	0.400000	0.000000	0.2	0.000000	0.2	1.200000	0.2	1.000000	0.800000	6.200000	0.800000	11.400000
G5	0.333333	0.166667	0.333333	0.0	0.166667	0.0	0.500000	0.0	1.166667	0.333333	2.833333	0.666667	6.500000

Table 3 Using clustering for dataset

Once the dataset was clustered in 5 groups and by applying the k-means technique, the data showed that the Agincourt neighbor was the most suitable one to set up a warehouse.

In [43]:

```
neigh_summary = pd.DataFrame([neigh_onehot.index, 1 + kmeans.labels_]).T
neigh_summary.columns = ['Neighborhood', 'Group']
neigh_summary
```

Out[43]:

	Neighborhood	Group
0	Agincourt	1
1	Agincourt North ,L'Amoreaux East ,Milliken ,St...	3
2	Birch Cliff ,Cliffside West	5
3	Cedarbrae	4
4	Clairlea ,Golden Mile ,Oakridge	5
5	Clarks Corners ,Sullivan ,Tam O'Shanter	3
6	Cliffcrest ,Cliffside ,Scarborough Village West	5
7	Dorset Park ,Scarborough Town Centre ,Wexford ...	3
8	East Birchmount Park ,Ionview ,Kennedy Park	2
9	Guildwood ,Morningside ,West Hill	5
10	Highland Creek ,Rouge Hill ,Port Union	5
11	L'Amoreaux West	2
12	Maryvale ,Wexford	2
13	Rouge ,Malvern	2
14	Scarborough Village	2
15	Woburn	5

Table 4 K-means cluster results