

Decision making for entrepreneurs

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

IBM Data Science Professional Certificate

Edward Ortiz

2019

Content

1. Background
2. Problem
3. Data Sources
4. Data cleaning and acquisition
5. Data visualization
6. Result

Background

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.

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.

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.

Data Sources



Dataset from venues by category in
Toronto



WIKIPEDIA
The Free Encyclopedia

Dataset from neighbor in Toronto

The IBM Watson logo is displayed on a black rectangular background. The word 'IBM' is in a light blue, sans-serif font, and 'Watson' is in a darker blue, sans-serif font, with a small 'TM' trademark symbol to the right.

Data visualization tool

Data cleaning and acquisition

1. Setting up foursquare developer API fro data extraction.
2. Extracting data set from Wikipedia
3. Creating map of Toronto placing neighbors

Data cleaning and acquisition

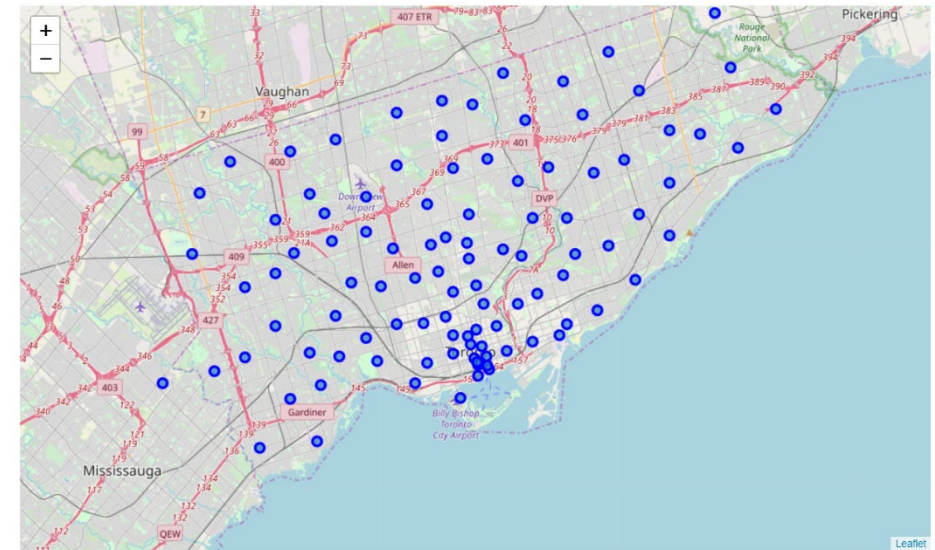
Data set from Wikipedia extraction

```
df_can = pd.read_csv(body)
df_can.head()
```

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

Creation of map



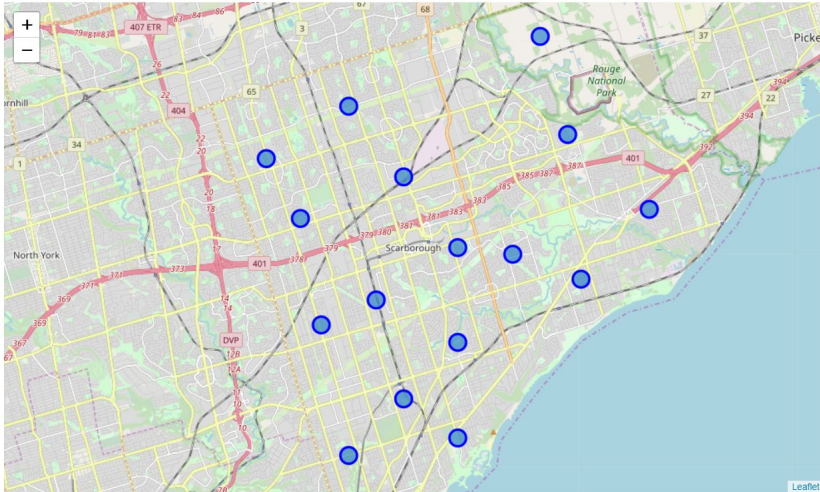
Data cleaning and acquisition

4. Segmentation of the map

5. Data mining from square space according category venue

Data cleaning and acquisition

Creation of segmented map



Data mining of segmented neighbor

```
import pickle
with open("neigh_foursquare_dataset.txt", "wb") as fp: #Pickling
    pickle.dump(neigh_foursquare_dataset, fp)
print('Received Data from Internet is Saved to Computer.')
```

Received Data from Internet is Saved to Computer.

```
with open("neigh_foursquare_dataset.txt", "rb") as fp: # Unpickling
    neigh_foursquare_dataset = pickle.load(fp)

print(type(neigh_foursquare_dataset))

neigh_foursquare_dataset
```

Data cleaning and acquisition

- 6. Segmentation of dataset by venue category in the selected neighbor
- 7. Grouping database

Data cleaning and acquisition

Segmentation of dataset by venue category in the selected neighbor

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

	Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Distance	African Restaurant	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	Airport
0	M1B	Rouge ,Malvern	43.806686	-79.194353	Images Salon & Spa	This spot is popular	595	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
2	M1B	Rouge ,Malvern	43.806686	-79.194353	Wendy's	This spot is popular	600	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
4	M1B	Rouge ,Malvern	43.806686	-79.194353	Harvey's	This spot is popular	796	0	0	0	0	0	0

Grouping database

```
neigh_onehot = neigh_onehot[important_list_of_features].drop(
    columns = ['Neighborhood Latitude', 'Neighborhood Longitude']).groupby(
    'Neighborhood').sum()
```

```
neigh_onehot.head()
```

[illegible]

Data train and visualization

8. K-means clustering setup

7. K-means data set

Data train and visualization

K-means clustering setup

```
from sklearn.cluster import KMeans  
  
kmeans = KMeans(n_clusters = 5, random_state = 0).fit(neigh_onehot)
```

Data train and visualization

Applying K-means into the database

```
means_df = pd.DataFrame(kmeans.cluster_centers_)
means_df.columns = neigh_onehot.columns
means_df.index = ['G1', 'G2', 'G3', 'G4', 'G5']
means_df['Total Sum'] = means_df.sum(axis = 1)
means_df.sort_values(axis = 0, by = ['Total Sum'], ascending=False)
```

	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

Result

Battle of neighbors

Cluster 5 is the best options to place the warehouse, following

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

	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

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. Next there are 5 other neighborhoods in which the warehouse can be placed