

Identifying A Suitable Location for Business in Toronto, Canada

- Applied Data Science Capstone
- IBM Data Science Professional Certificate
- June 2019
- Gary N. Thomas



Image source: [google.com/images](https://www.google.com/images)

Identifying A Suitable Location for Business in Toronto, Canada



Image source: google.com/images



Research Scenario

Introduction-Problem Description:

The premise of this research paper will center on investigating and determining the ideal locale in Toronto, Canada, for the establishment of a new business-dining experience, a restaurant targeting people whose taste palette are unconventional and outside the box. In this study, I endeavor to identify the right location by finding the right cluster of people and amenities, whose income, age, education, household size combined with the right competitive mixture create an enabling environment to support business and those who depend on it.

Research Scenario

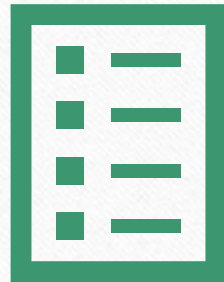
Data Requirements:

The research top will require household data, cross-sectional and time dimensional. It will also rely on geo-locational information about that areas of interests, i.e. the boroughs and neighborhoods of focus.

I will assume it is “the environs of Toronto.” Most of the information publicly available, online and elsewhere.



Research Scenario



Research Scenario: Data Requirements(cot'd)



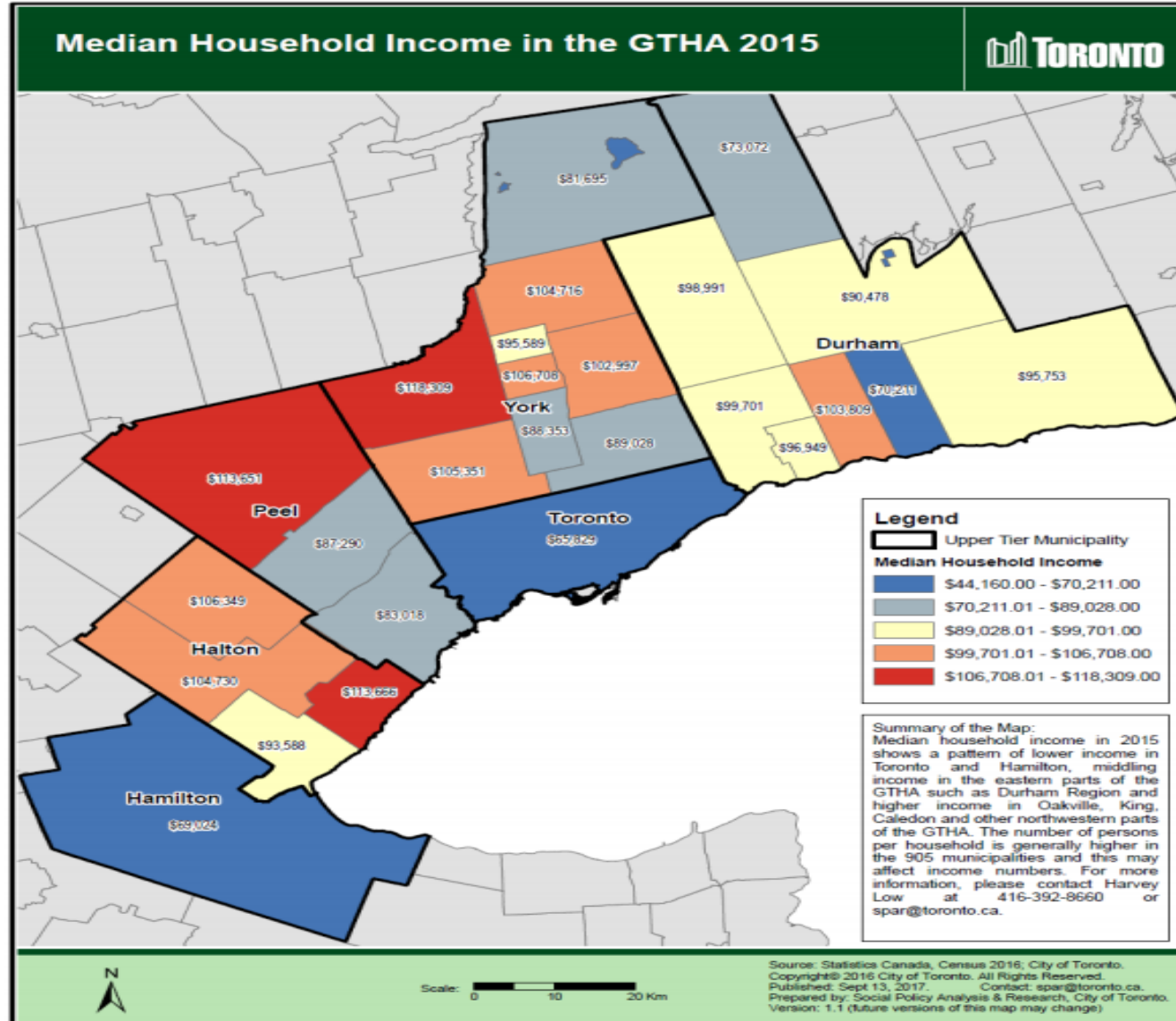
We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. A typical request from Foursquare will provide us with the following information:

Problem Focus

| | Borough | Neighbourhood | Latitude | Longitude |
|---|-----------------|---|-----------|------------|
| 0 | East Toronto | The Beaches | 43.676357 | -79.293031 |
| 1 | East Toronto | Riverdale, The Danforth West | 43.679557 | -79.352188 |
| 2 | East Toronto | India Bazaar, The Beaches West | 43.668999 | -79.315572 |
| 3 | East Toronto | Studio District | 43.659526 | -79.340923 |
| 4 | Central Toronto | Lawrence Park | 43.728020 | -79.388790 |
| 5 | Central Toronto | Davisville North | 43.712751 | -79.390197 |
| 6 | Central Toronto | North Toronto West | 43.715383 | -79.405678 |
| 7 | Central Toronto | Davisville | 43.704324 | -79.388790 |
| 8 | Central Toronto | Moore Park, Summerhill East | 43.689574 | -79.383160 |
| 9 | Central Toronto | Deer Park, Forest Hill SE, Rathnelly, South Hi... | 43.686412 | -79.400049 |

- Identifying the Best Location for a New Business, using location data, via Postal Codes [Neighborhoods in Toronto]

Map 1: Median Household Income by Census Subdivision in the GTHA, 2015



Problem Focus

- A view of the economic and financial capital of Canada, Toronto.
- The area is colored based on household income.

Problem Focus



Connecting to Foursquare and Retrieving Locational Data for Each Venue in Every Neighborhood



After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 1000 meter. It means that we have asked Foursquare to find venues that are at most 1000 meter far from the center of the neighborhood.

Problem Focus

```
toronto_onehot.shape
```

```
(827, 187)
```

```
toronto_onehot.describe()
```

| | Airport | Airport Food Court | Airport Gate | Airport Lounge | Airport Service | Airport Terminal | American Restaurant | Antique Shop | Aquarium |
|-------|------------|--------------------------|-----------------|-------------------|--------------------|---------------------|------------------------|-----------------|------------|
| count | 827.000000 | 827.000000 | 827.000000 | 827.000000 | 827.000000 | 827.000000 | 827.000000 | 827.000000 | 827.000000 |
| mean | 0.001209 | 0.001209 | 0.001209 | 0.002418 | 0.002418 | 0.002418 | 0.012092 | 0.001209 | 0.001209 |
| std | 0.034773 | 0.034773 | 0.034773 | 0.049147 | 0.049147 | 0.049147 | 0.109362 | 0.034773 | 0.034773 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 50% | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 75% | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| max | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 |

- **Data Sourcing, Cleaning and Wrangling**

When the data is completely gathered, we will perform processing on that raw data to find our desirable features for each venue. Our main feature is the category of that venue.

After this stage, the column "Venue's Category" will be One-hot encoded and different venues will have different feature-columns.

After On-hot encoding we will integrate all restaurant columns to one column "Total Restaurants" and all food joint columns to "Total Joints" column.

Problem Focus

- Applying one of Machine Learning Techniques (K-Means Clustering)

```
# Run k-means to cluster the neighborhood into 6 clusters
# set number of clusters
kclusters = 10

toronto_grouped_clustering = toronto_grouped.drop('Neighbourhood', 1)
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=1).fit(toronto_grouped_clustering)

# check cluster labels generated for each row in the dataframe
print(kmeans.labels_[0:20])
print(len(kmeans.labels_))

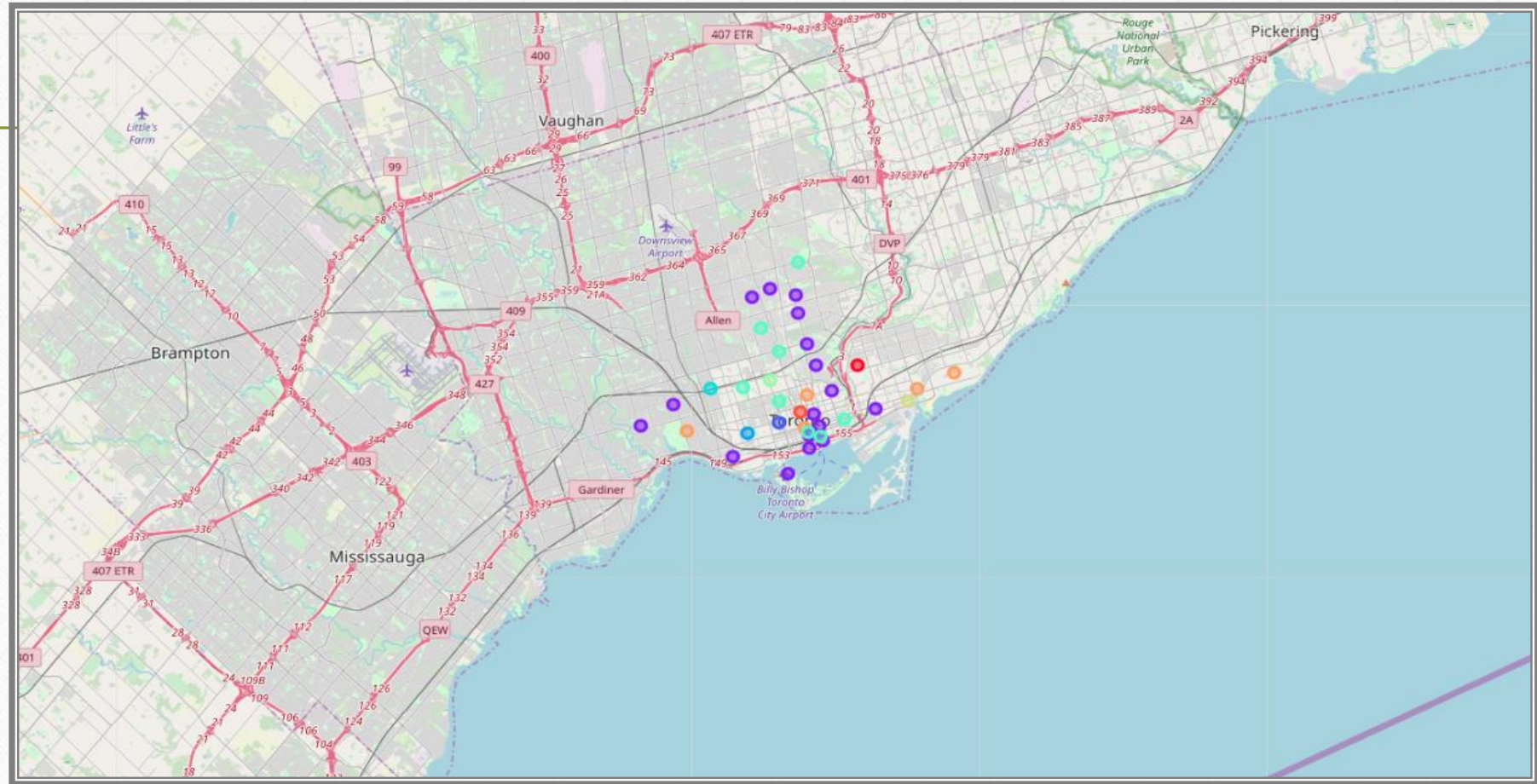
[8 0 8 1 5 1 1 1 1 5 1 1 8 5 1 1 1 9 8 1]
38
```


Decision Making and Reporting Results

- This report may be helpful for someone planning on opening a restaurant in Toronto, as any similar such type of business.
- They can be confident that by comparing the current offers and neighborhoods profiles, however it may not cover all variables such as access to public transportation or even the restaurants profiles, so it shall not be used as a single decision-making tool.
- Reports of this nature have a shelf life and one should always appreciate and contextualize their process going forward. However, I have no doubt that businesses will be aware of all these factors.

Decision Making and Reporting Results

- The illustration below shows the final neighborhood cluster, East Toronto, in red as the best location.



Problem Focus

- **Conclusion and Relevance**

After examining the above 5 clusters, I believe and would recommend that cluster label (8), East Toronto is the best Neighborhood to set up shop for a new business venture. While there are great locations like downtown, Lawrence Park and Riverdale, East Toronto presents the best opportunity to be a first mover as well an early entrant into a community that is poised to experience phenomenal growth.

East Toronto is the most suitable locale as it has an expanding population, colleges, a dependable public transport system, supporting entities for the productive population a business will need and attract.



This Photo by Unknown Author is licensed under CC BY-NC-ND

The End

- Grateful and appreciative of all comments, suggestions, sited errors etc.
- email: garynthomas@gmail.com