Clustering based Itinerary Planner

Rohan Shukla
August 11, 2019

Scope

- A lot of people wish to step out of their boundaries and go out exploring, learning new cultures, tasting different delicacies and connecting with various people around the globe
- It is not very easy to take time out of your tight schedule and plan trips to a place and prepare your travel itinerary for your travel
- The itineraries available might not always match the constraints of the travelers and people often tend to miss some places on their tour
- The major stakeholders of this project are the tourists and travelers who wish to enjoy their vacations to the fullest!

Data Acquisition and cleaning

- I have scraped almost all the location data including the nearby attractions and other hangout points nearby with travel distances between places from the FourSquare API data

- the following data fields from the Foursquare location data will be helpful in the analysis :
  - 1. tourist destinations - this will give data about the top locations in and around the city
  - 2. location coordinates - this data is used to calculate the distance between places
  - 3. location reviews - this data will be used to select top locations

Data Acquisition and Cleaning (2)

- The project requires a few inputs from the user as mentioned below:
    - 1. Location details - Name to the city to be visited and the country code where the city resides
    - 2. Duration of trip - The total number of days on the trip
    - 3. Max visits per day - This number of places the user wishes to visit per day
    - 4. Off days - the number of days, the user wants to take a break for other planned activities or relaxation.

Sample - Toronto City

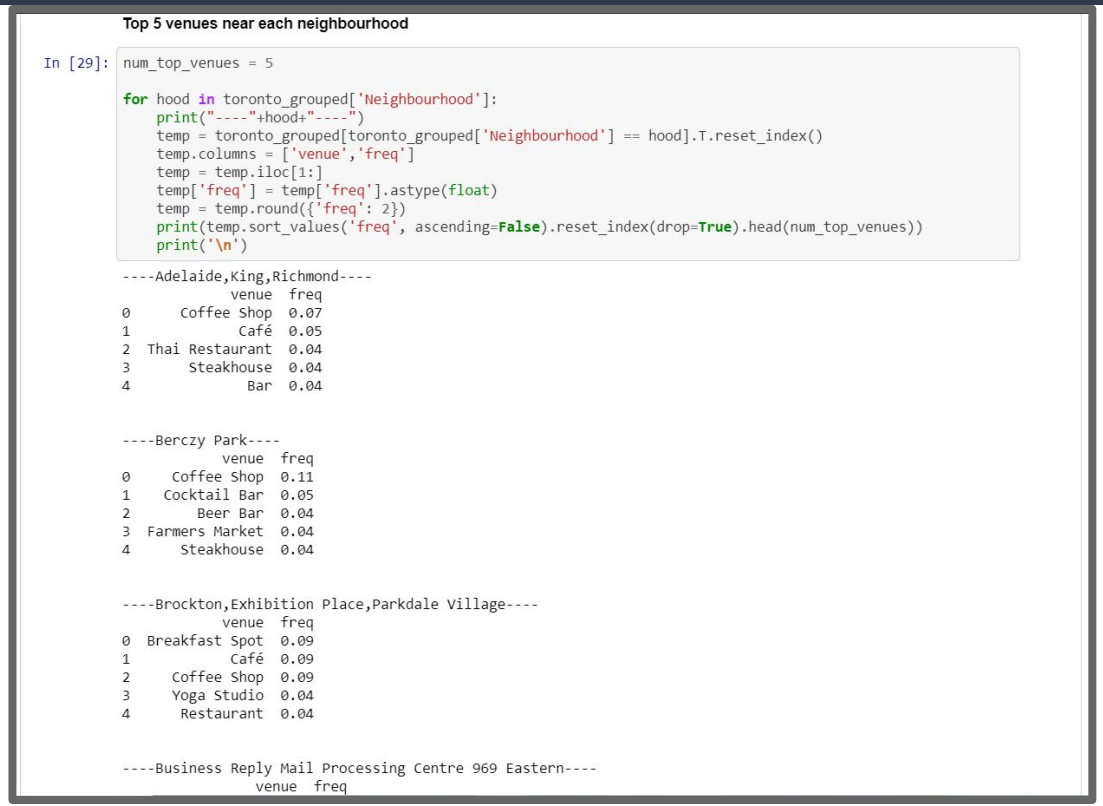
Getting nearby venues for neighbourhoods in Toronto

```
In [16]: toronto_venues = getNearbyVenues(names=toronto_data['Neighbourhood'],
                                          latitudes=toronto_data['Latitude'],
                                          longitudes=toronto_data['Longitude']
                                         )
```

```
The Beaches
The Danforth West,Riverdale
The Beaches West,India Bazaar
Studio District
Lawrence Park
Davisville North
North Toronto West
Davisville
Moore Park,Summerhill East
Deer Park,Forest Hill SE,Rathnelly,South Hill,Summerhill West
Rosedale
Cabbagetown,St. James Town
Church and Wellesley
Harbourfront,Regent Park
Ryerson,Garden District
St. James Town
Berczy Park
Central Bay Street
Adelaide,King,Richmond
Harbourfront East,Toronto Islands,Union Station
Design Exchange,Toronto Dominion Centre
Commerce Court,Victoria Hotel
Roselawn
Forest Hill North,Forest Hill West
The Annex,North Midtown,Yorkville
Harbord,University of Toronto
Chinatown,Grange Park,Kensington Market
CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway Lands,South Niagara
Stn A PO Boxes 25 The Esplanade
First Canadian Place,Underground city
Christie
Dovercourt Village,Dufferin
Little Portugal,Trinity
Brockton,Exhibition Place,Parkdale Village
High Park,The Junction South
Parkdale,Roncesvalles
```

- Collected all nearby venues and places around *Toronto city*
- The city name is a parameter accepted from the user
- All the nearby attractions are listed with coordinates

Deciding nearby venues



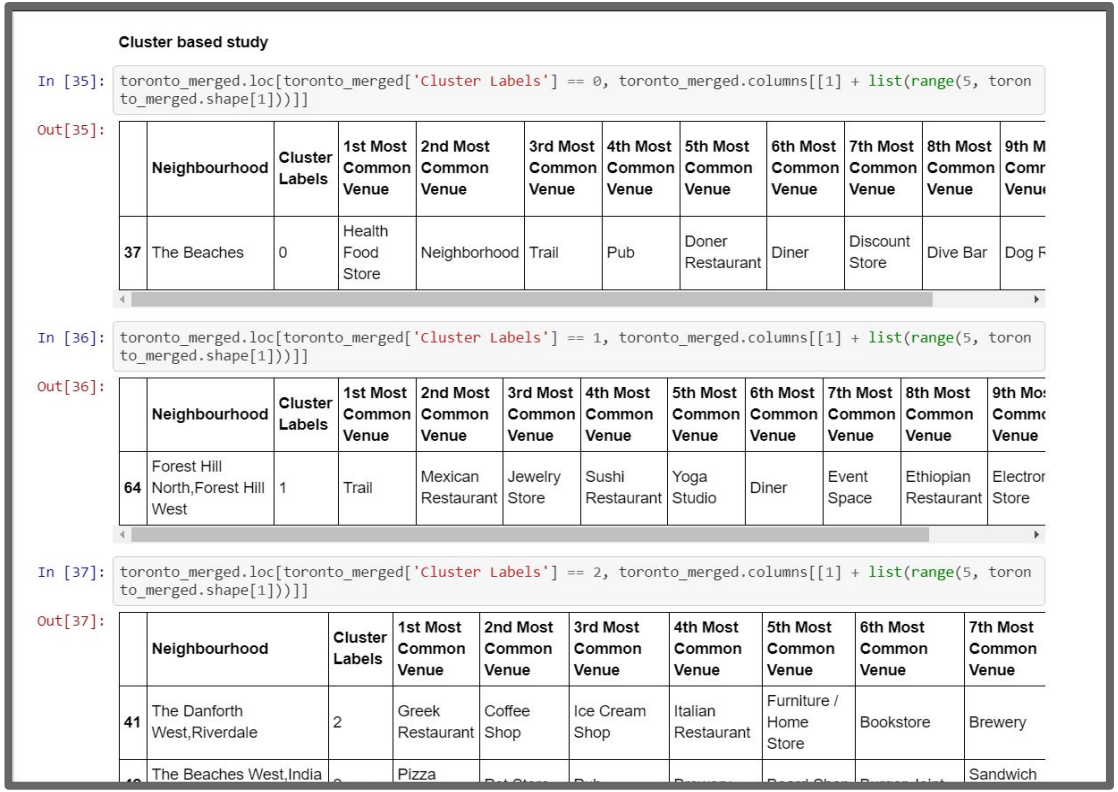
Top 5 venues near each neighbourhood

In [29]:
```
num_top_venues = 5

for hood in toronto_grouped['Neighbourhood']:
    print("----"+hood+"----")
    temp = toronto_grouped[toronto_grouped['Neighbourhood'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----Adelaide,King,Richmond----
              venue  freq
0       Coffee Shop  0.07
1              Café  0.05
2  Thai Restaurant  0.04
3        Steakhouse  0.04
4               Bar  0.04


----Berczy Park----
            venue  freq
0     Coffee Shop  0.11
1    Cocktail Bar  0.05
2        Beer Bar  0.04
3  Farmers Market  0.04
4      Steakhouse  0.04


----Brockton,Exhibition Place,Parkdale Village----
            venue  freq
0  Breakfast Spot  0.09
1           Café  0.09
2     Coffee Shop  0.09
3     Yoga Studio  0.04
4      Restaurant  0.04


----Business Reply Mail Processing Centre 969 Eastern----
              venue  freq
```

- Based on the user ratings and frequency of nearby shops I have decided *N* places that the tourists would love to visit
- *N* is a user parameter which signifies the number of places a person wishes to cover in a particular day
- Here is a list of attractions and nearby venues

Final Clustering of places



Cluster based study

In [35]:
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[1] + list(range(5, toron
to_merged.shape[1]))]]
```

Out[35]:

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th M Comr Venu
37	The Beaches	0	Health Food Store	Neighborhood	Trail	Pub	Doner Restaurant	Diner	Discount Store	Dive Bar	Dog F

In [36]:
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(5, toron
to_merged.shape[1]))]]
```

Out[36]:

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Mo Comm Venue
64	Forest Hill North,Forest Hill West	1	Trail	Mexican Restaurant	Jewelry Store	Sushi Restaurant	Yoga Studio	Diner	Event Space	Ethiopian Restaurant	Electror Store

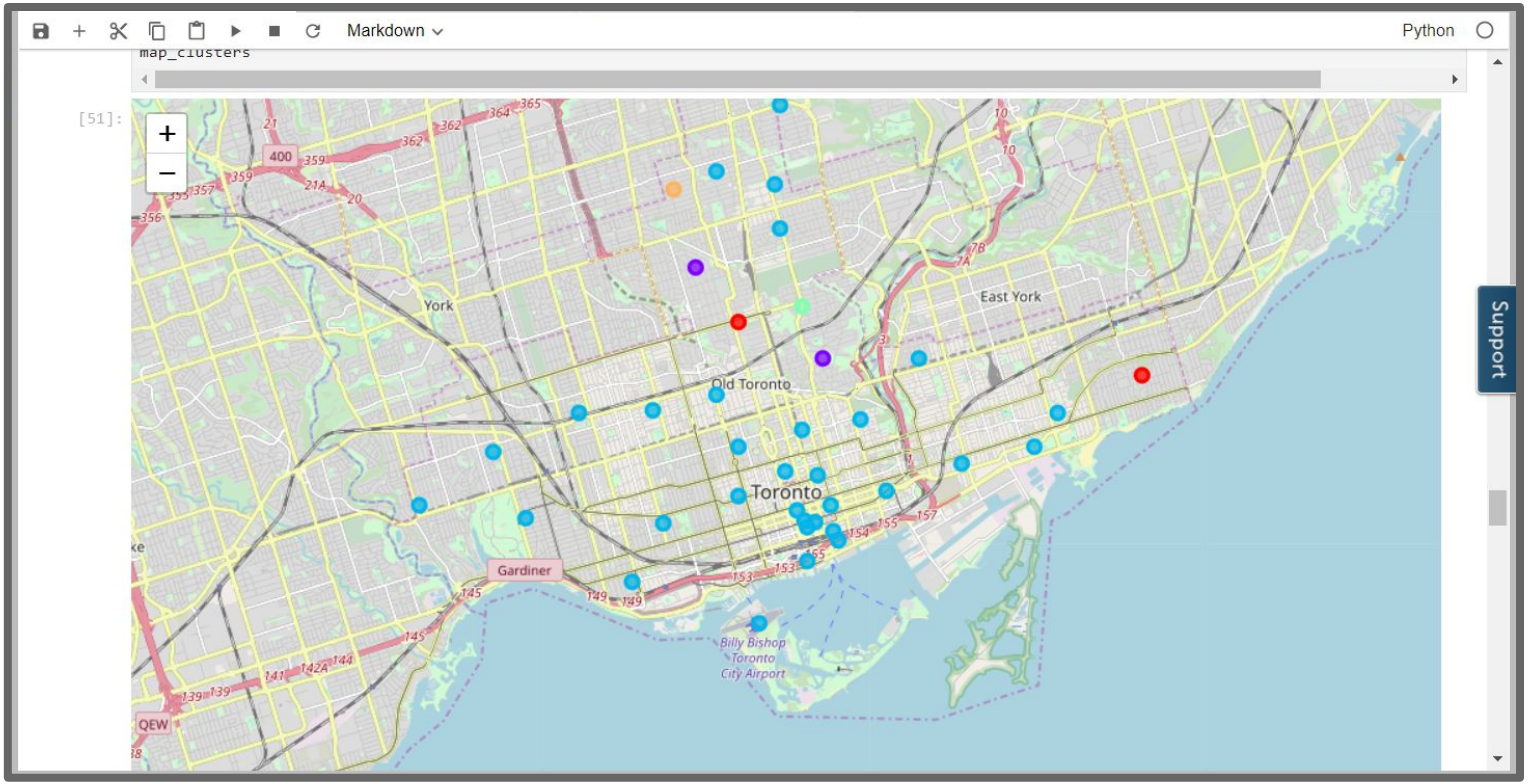
In [37]:
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toron
to_merged.shape[1]))]]
```

Out[37]:

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
41	The Danforth West,Riverdale	2	Greek Restaurant	Coffee Shop	Ice Cream Shop	Italian Restaurant	Furniture / Home Store	Bookstore	Brewery
	The Beaches West,India		Pizza						Sandwich

- Finally, these places are grouped and clustered based on their user ratings and nearby venues
- The number of clusters is determined by the total travel days the user has at hand
- On the left, is a list of places segregated cluster wise along with nearby venues

Plotting Clusters on map



Conclusions and future scope

Built useful models to predict and plan out the places of visit.

- Accuracy of the models has room for improvement
- Capture more of places' individual traits along with user ratings
- Ideas include:
    - Physical data (best time to visit, etc.)
    - Financial data (amount of pay, etc.)
    - Team interaction data (age groups, entry fees, interest levels, etc.)