Recommendation system Based on Similarities Between Venues

MUSA ŞAHIN – IBM DATA SCIENCE COURSE CAPSTONE PROJECT

Introduction/Business Problem

- People want to see similar cities when they move.
- Real estate companies can recommend similar places to their customers.

Data

Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	...	Vegetarian / Vegan Restaurant	Veterinarian	Video Game Store
Battery Park City	0.0	0.0	0.0	0.0	0.033333	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0
Carnegie Hill	0.0	0.0	0.0	0.0	0.033333	0.0	0.0	0.000000	0.0	0.0	...	0.033333	0.0	0.0
Central Harlem	0.0	0.0	0.0	0.1	0.066667	0.0	0.0	0.033333	0.0	0.0	...	0.000000	0.0	0.0
Chelsea	0.0	0.0	0.0	0.0	0.033333	0.0	0.0	0.000000	0.0	0.0	...	0.033333	0.0	0.0
Chinatown	0.0	0.0	0.0	0.0	0.033333	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0

Data includes frequency of each venue for the neighborhood. One for Manhattan and one for Toronto

Foursquare API parameters:
Radius: 500 meters
Limit: 30 venues
Location: Neighborhood location

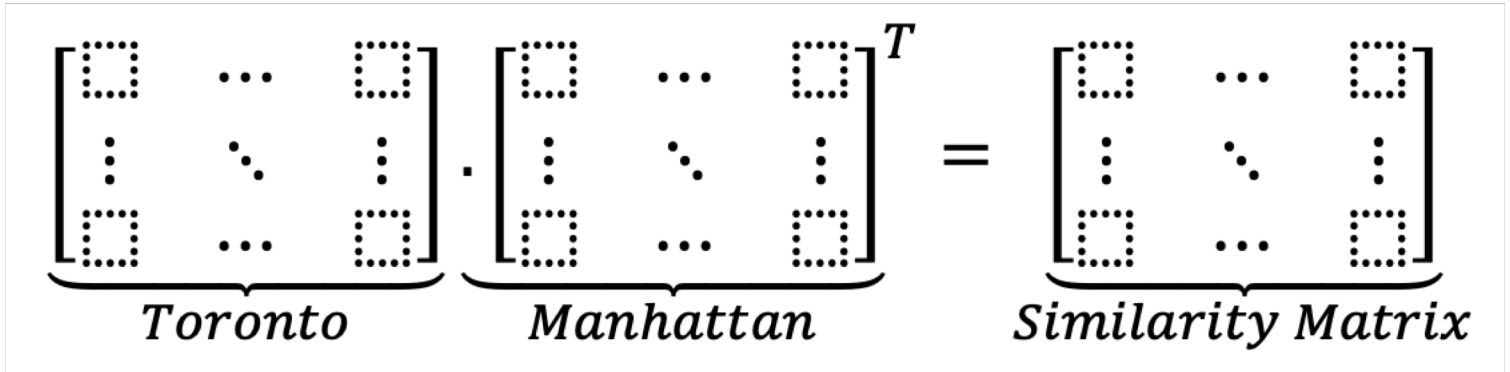
Data Preprocessing

```
print("Venues categories that are not included at Toronto: " + str(len(set(venueListMan) - set(venueListTor))))
print("Venues categories that are not included at Manhattan: " + str(len(set(venueListTor) - set(venueListMan))))
print("Venues categories that are included at both city: " + str(len(set(venueListMan) & set(venueListTor))))
```

```
Venues categories that are not included at Toronto: 99
Venues categories that are not included at Manhattan: 57
Venues categories that are included at both city: 135
```

Toronto and Manhattan data frames do not have same venue categories. 135 mutual venues must be considered, different ones are dropped from data frames.

Method



Similarity score = sum of products of same venue category frequencies
Similarity matrix consist of similarity scores.

Result Matrix

Toronto→ / Manhattan↓	Adelaide, King, Richmond	Berczy Park	Brockton, Exhibition Place, Parkdale Village	Business Reply Mail Processing Centre 969 Eastern	CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	Cabbagetown, St. James Town	Central Bay Street	Chinatown, Grange Park, Kensington Market	Christie	Church and Wellesley	...
Battery Park City	0.010345	0.008889	0.008642	0.010526	0.006250	0.011111	0.022222	0.002222	0.020833	0.007778	...
Carnegie Hill	0.018391	0.011111	0.019753	0.012281	0.006250	0.015556	0.030000	0.012222	0.016667	0.013333	...
Central Harlem	0.009195	0.008889	0.006173	0.007018	0.004167	0.004444	0.008889	0.008889	0.010417	0.005556	...
Chelsea	0.019540	0.012222	0.009877	0.000000	0.004167	0.011111	0.014444	0.012222	0.014583	0.005556	...
Chinatown	0.009195	0.006667	0.001235	0.007018	0.000000	0.003333	0.014444	0.005556	0.000000	0.006667	...
Civic Center	0.010345	0.010000	0.016049	0.010526	0.004167	0.012222	0.022222	0.007778	0.008333	0.006667	...
Clinton	0.014943	0.003333	0.007407	0.007018	0.000000	0.002222	0.003333	0.004444	0.006250	0.001111	...
East Harlem	0.008046	0.018889	0.012346	0.001754	0.002083	0.016667	0.011111	0.022222	0.012500	0.010000	...
East Village	0.013793	0.005556	0.009877	0.005263	0.008333	0.012222	0.022222	0.013333	0.008333	0.011111	...
Financial District	0.026437	0.013333	0.013580	0.010526	0.004167	0.014444	0.022222	0.006667	0.012500	0.005556	...
Flatiron	0.008046	0.004444	0.017284	0.008772	0.000000	0.006667	0.007778	0.007778	0.006250	0.005556	...
Gramercy	0.016092	0.008889	0.012346	0.015789	0.004167	0.015556	0.027778	0.010000	0.016667	0.012222	...

For example, similarity score between Christie and Central Harlem is 0.010417

Result evaluation

- First three recommendation pair: 'Stuyvesant Town' - 'Rosedale', 'Battery Park City' - 'Rosedale' and 'Stuyvesant Town' - 'Moore Park, Summerhill East'. All neighborhoods have parks substantially.
- Coffee shops are most frequent venue in the Central Bay Street. Recommendations for this neighborhood: Hamilton Heights, Carnegie Hill, Morningside Heights and Murray Hill. They have many coffee shops as well.
- Recommendations are generally successful.

Discussion

- Lack of variety of venues prevents to make recommendations. For example, Roselawn has all zeros for every similarity scores.
- In addition to venue similarity, real estate companies must also consider other parameters to recommend a place for their customers.
- For better recommendations a new metric can be created rather than similarity score which can utilize dissimilar places.