Comparing Koreatown with a Korean City

Koreatowns compared to a Korean City

How similar are western Koreatowns to the Korean capital Seoul?

What are the most common venues in these areas?

The cities compared with Seoul are Los Angeles, Toronto and New York

	City	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
91	NewYork	40.741997	-73.985496	Equinox Gramercy	40.740749	-73.985771	Gym
92	NewYork	40.741997	-73.985496	CAVA	40.740842	-73.985510	Mediterranean Restaurant
93	NewYork	40.741997	-73.985496	Upland	40.741891	-73.984640	New American Restaurant
94	NewYork	40.741997	-73.985496	Madison Square Park	40.742262	-73.988006	Park
95	NewYork	40.741997	-73.985496	Barry's Bootcamp	40.742532	-73.984152	Gym / Fitness Center

	City	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
286	Seoul	37.5665	126.978	무교동북어국집	37.567852	126.979753	Korean Restaurant
287	Seoul	37.5665	126.978	Läderach chocolatier suisse (레더라)	37.568153	126.978265	Chocolate Shop
288	Seoul	37.5665	126.978	철철복집	37.567393	126.981310	Seafood Restaurant
289	Seoul	37.5665	126.978	The Plaza Hotel (더 플라자)	37.564621	126.978060	Hotel
290	Seoul	37.5665	126.978	Seoul Plaza (서울광장)	37.565475	126.977937	Pedestrian Plaza

Foursquare API to get venues categories in the areas

Data

	City	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bar	Beer Bar	Bike Shop	
0	LA	0.000000	0.010989	0.00	0.00	0.000000	0.021978	0.021978	0.00	0.032967	0.021978	0.000000	0.000000	(
1	NewYork	0.060000	0.000000	0.02	0.00	0.000000	0.000000	0.000000	0.00	0.010000	0.010000	0.010000	0.000000	(
2	Seoul	0.000000	0.000000	0.01	0.03	0.000000	0.000000	0.010000	0.01	0.000000	0.000000	0.000000	0.000000	(
3	Toronto	0.010526	0.000000	0.00	0.00	0.010526	0.000000	0.010526	0.00	0.042105	0.031579	0.021053	0.010526	(

Methodology

- The data was first transformed to frequency of each venue category for the Koreatowns and Seoul
- Then both a correlation matrix was calculated as well as a clustering algorithm with two clusters

	City	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 Most Common Venue	10th Most Common Venue
0	LA	Korean Restaurant	Ice Cream Shop	Coffee Shop	Japanese Restaurant	Café	Hotel	Bakery	Bubble Tea Shop	Restaurant	Asian Restaurant
1	NewYork	American Restaurant	New American Restaurant	Italian Restaurant	Hotel	Korean Restaurant	Indian Restaurant	Gym / Fitness Center	Dessert Shop	Japanese Restaurant	Pizza Place
2	Seoul	Hotel	Korean Restaurant	Café	Coffee Shop	Historic Site	Sushi Restaurant	History Museum	Japanese Restaurant	Palace	Art Museum
3	Toronto	Korean Restaurant	Grocery Store	Park	Bakery	Coffee Shop	Pizza Place	Ice Cream Shop	Café	Bar	Karaoke Bar

Results

The top ten venues for each of the Koreatowns and for the city Seoul

	Cluster Labels	City
0	0	LA
1	1	NewYork
2	1	Seoul
3	0	Toronto

	Seoul	NewYork	Toronto	LA
Seoul	1.000000	0.258378	0.275085	0.503148
NewYork	0.258378	1.000000	0.334066	0.333328
Toronto	0.275085	0.334066	1.000000	0.787754
LA	0.503148	0.333328	0.787754	1.000000

Results

- Clustering resulted in New York being assigned the same cluster as Seoul
- The correlational matrix showed that Los Angeles had the highest correlation with Seoul
- Further Los Angeles and Toronto had a high correlational value

Discussion

- First the observation that Hotel is the most common venue in Seoul is noted. This is might be different than from Koreatowns in other cities as the Koreatowns formed from immigration. It is reasonable to believe that hotels are more prevalent and used in the tourism and travel areas. These might not be the main function of a Koreatown in a city as the hotels in that city might be contained in the center of the city. This is beyond the question asked in the similarities but still worth noting.
- Performing a clustering algorithm with only four examples is not a particular insightful method however it did not assign Seoul to one cluster and the Koreatowns in another cluster. Instead, it assigned Seoul and New York Koreatown to the same cluster and the two other Koreatown to the second cluster. This is interesting but since there only are four examples it is to be taken very lightly.
- From the correlational matrix between the cities, it can be concluded that LA has the most similar frequencies of venues to Seoul. This contradicts the clustering but as concluded the clustering results should be taken with a grain of salt and this result is more reliable.
- The correlation matrix also showed a correlation between the values of Los Angeles and Toronto suggesting that the Koreatown in these cities could look similar which is interesting.

Conclusion

- From these results it can be concluded that although this is correlation and exactly similarity, we can see that all the Koreatowns seems to have some features of Seoul, due to a notable correlational value. Keeping in mind that it is not actually fair to compare venues of an entire city compared to a community or part of a town.
-
- Future work could explore other areas outside of Koreatown to show if there is a large difference between the correlation of these and the correlation of Koreatowns.