A Venue-Type-Based City Recommender

XINYUE LUO
MAR 03 2019

A city recommender assists decision-making for vacation destination/residential area choice

The Business Problem

- Looking for a dream city to start a new life or a perfect place to spend a vacation can be energy-draining and time-consuming
- Traditional ways of searching information (e.g. talking to friends, browsing review articles) are low-throughput that hardly cover a large range of data within a short timeframe

Objectives

- build a classification model to generate recommendation of cities based on a user's known favorite city
- further pinpoint an region within a city of interest that matches a user's preferences

Data Collection & Cleaning

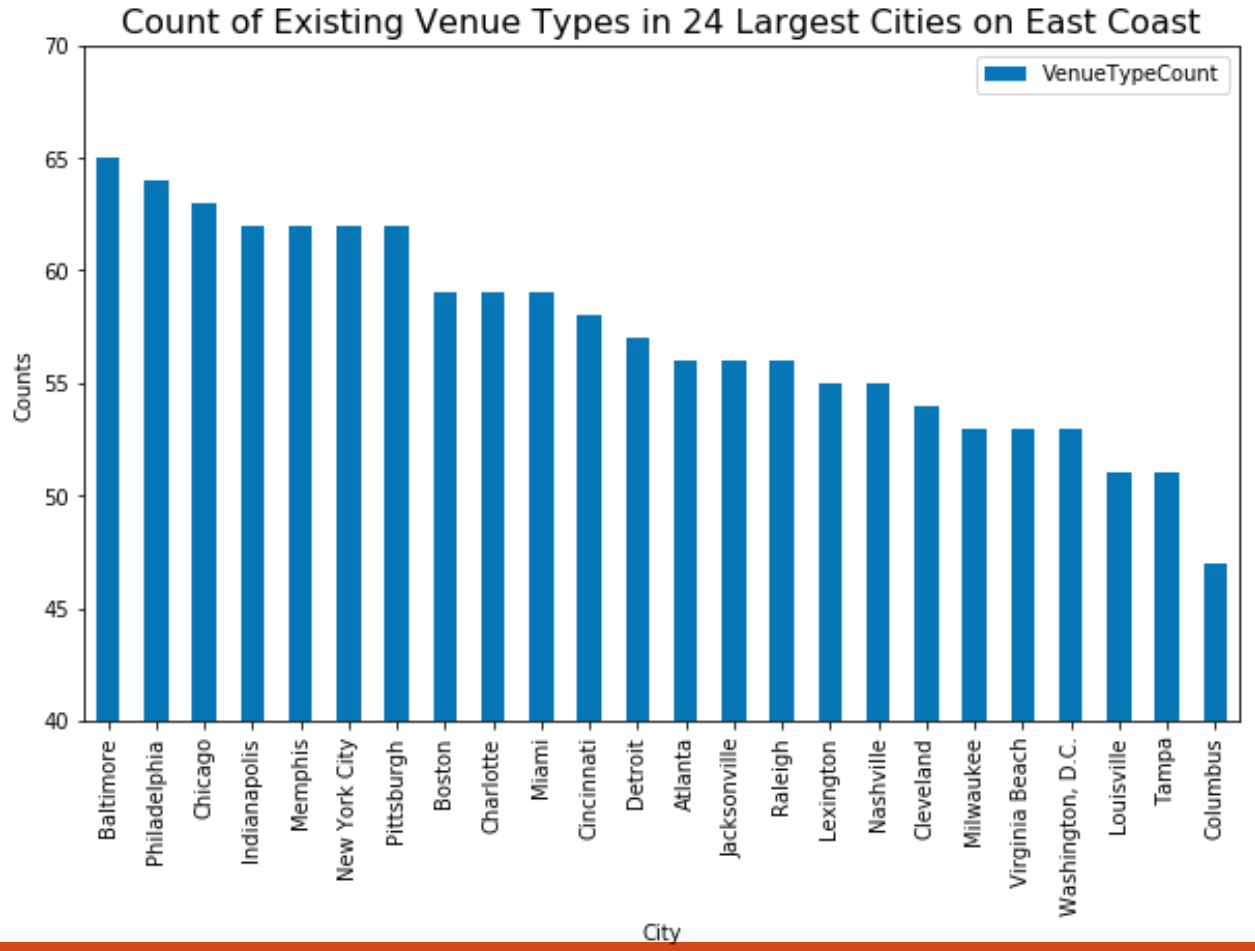
WEB SCRAPING FOR A LIST OF LARGEST CITIES ON EAST COAST FROM WIKIPEDIA PAGES

	City	State	Area (km2)	Population	Radius (m)	Latitude	Longitude
0	New York City	New York	1213.37	8,622,698	19652.675813	40.730862	-73.987156
1	Chicago	Illinois	606.00	2,695,598	13888.692920	41.875562	-87.624421
2	Philadelphia	Pennsylvania	369.62	1,567,827	10846.829036	39.952415	-75.163575
3	Jacksonville	Florida	2265.30	821,784	26852.697912	30.332184	-81.655651
4	Indianapolis	Indiana	953.18	820,445	17418.571047	39.768333	-86.158350
5	Columbus	Ohio	577.85	787,033	13562.277380	39.962260	-83.000707
6	Charlotte	North Carolina	771.00	731,424	15665.788274	35.227087	-80.843127
7	Detroit	Michigan	370.08	713,777	10853.576493	42.331551	-83.046640
8	Washington, D.C.		177.00	703,608	7506.054213	38.895009	-77.036563
9	Boston	Massachusetts	232.14	667,137	8596.072183	42.360253	-71.058291

ACQUISITION OF LOCAL VENUE INFORMATION FROM FOURSQUARE API

	City	Afghan Restaurant	African Restaurant	American Restaurant		Yoga Studio	Zoo	Zoo Exhibit
0	Atlanta	0.00	0.00	0.05		0.00	0.00	0.00
1	Baltimore	0.01	0.01	0.04		0.02	0.00	0.00
2	Boston	0.00	0.00	0.02		0.01	0.00	0.00
3	Charlotte	0.00	0.00	0.04		0.02	0.00	0.00
4	Chicago	0.00	0.00	0.01	•••	0.02	0.00	0.00
5	Cincinnati	0.00	0.00	0.04		0.00	0.00	0.00
6	Cleveland	0.00	0.00	0.05		0.00	0.00	0.00
7	Columbus	0.00	0.00	0.07		0.00	0.00	0.00
8	Detroit	0.00	0.00	0.05		0.00	0.00	0.00
9	Indianapolis	0.00	0.00	0.02		0.00	0.01	0.01

Exploratory analysis:
Counts of venue types in 24 largest cities on east coast



Exploratory analysis:
Top venue types

TOP 5 VENUE TYPES IN EACH CITY (SHOWN FIRST 5 CITIES BY ALPHABET)

	City	Top 1	Top 2	Top 3	Top 4	Top 5
0	Atlanta	Trail	American Restaurant	Park	Southern / Soul Food Restaurant	Mexican Restaurant
1	Baltimore	Italian Restaurant	Coffee Shop	Seafood Restaurant	American Restaurant	Lounge
2	Boston	Park	Bakery	Hotel	Italian Restaurant	Seafood Restaurant
3	Charlotte	Brewery	BBQ Joint	American Restaurant	Italian Restaurant	Deli / Bodega
4	Chicago	Hotel	Park	New American Restaurant	Theater	Italian Restaurant
5	Cincinnati	Bar	Park	Italian Restaurant	American Restaurant	Coffee Shop

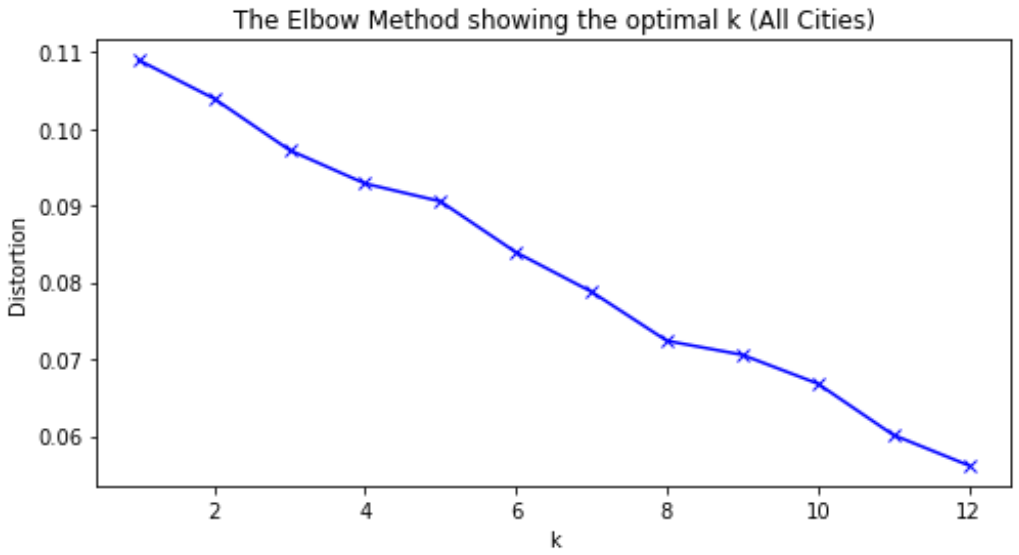
TOP 10 MOST POPULAR VENUE TYPES ACROSS 24 CITIES

Top 1s

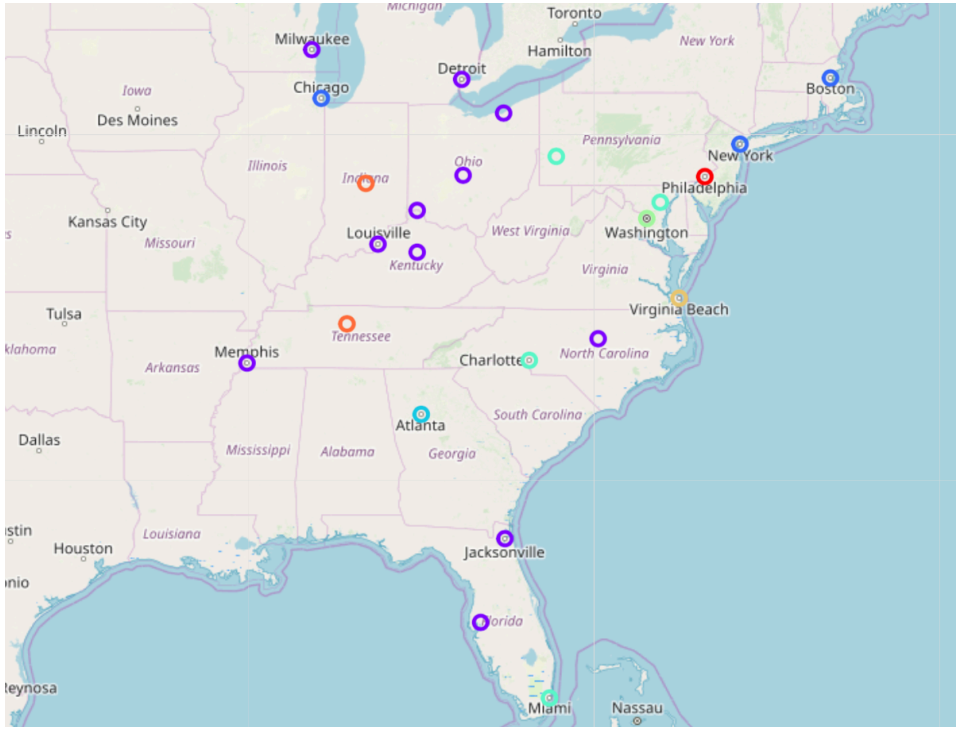
Venue type	Count
American Restaurant	19
Park	19
Coffee Shop	16
Hotel	16
Pizza Place	12
Italian Restaurant	12
Bar	12
Brewery	10
Restaurant	8
Seafood Restaurant	7

Classification with K-Means for 24 cities

ELBOW METHOD TO SELECT THE OPTIMAL K



MAP VISUALIZATION OF 8 CLUSTERS



Result:
Example of city recommendation for Fremont, CA

GENERATE A CITY RECOMMENDATION LIST

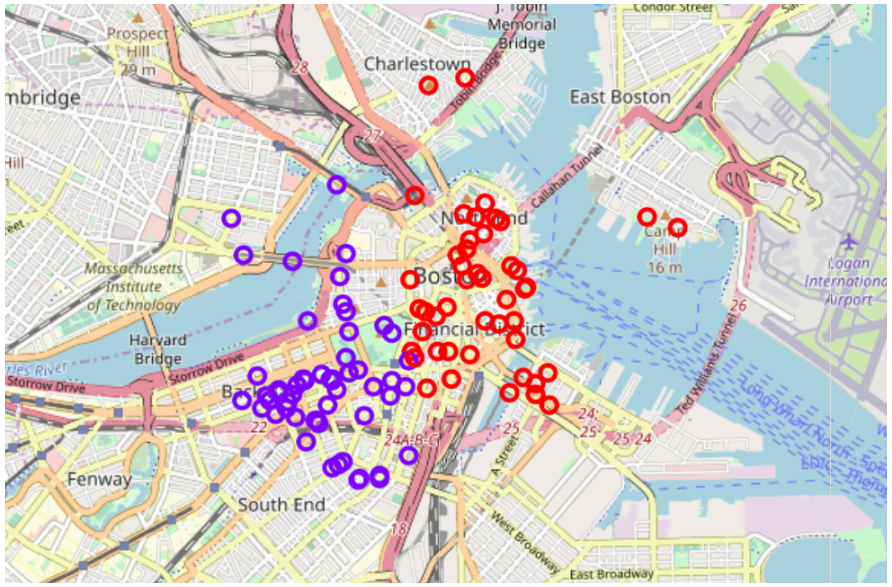
```
Welcome to city recommender! Type your favorate city here: fremont

Type the state for the above city here: ca
```

```
Welcome to city recommender! Type your favorate city here:fremont
Type the state for the above city here:ca
Cluster Label for fremont, ca is  2
A list of cities on the east coast you may like:  ['Boston', 'Chicago', 'New York City'] . Corresponding states:  ['
Massachusetts', 'Illinois', 'New York']
```

Recommendation list of cities

FURTHER ANALYSIS OF ONE RECOMMENDED CITY: BOSTON




```
Cluster Label for your favorate city is  2
You might also be interested in Cluster 0 in Boston
You might also be interested in Cluster 1 in Boston
```

Conclusion & Future Direction

Conclusion

- A tool with K-Means algorithm as core, used to generate city recommendations for a user based on the user's input of a known favorite city
- An extension to analyze whether each venue cluster within a city matches the user's preference

Future Directions

- Expand list of venues involved in model building
- Expand feature types involved in model building (demographic, etc.)
- Expand pool of candidate cities beyond 24
- Experiment with other classification models