

SEPTEMBER 15, 2019

INTRODUCTION

Which metropolitan areas are most similar?

A business planning to expand into new countries or regions may want to know which metropolitan areas are similar to the areas in which they already have existing branches.

There are different ways to classify similar. In this project, I will classify similar based on the types of venues foursquare returns for top picks in the metro area.

DATA

DATA FROM WIKIPEDIA FOR THE 100 MOST POPULATED METROPOLITAN AREAS

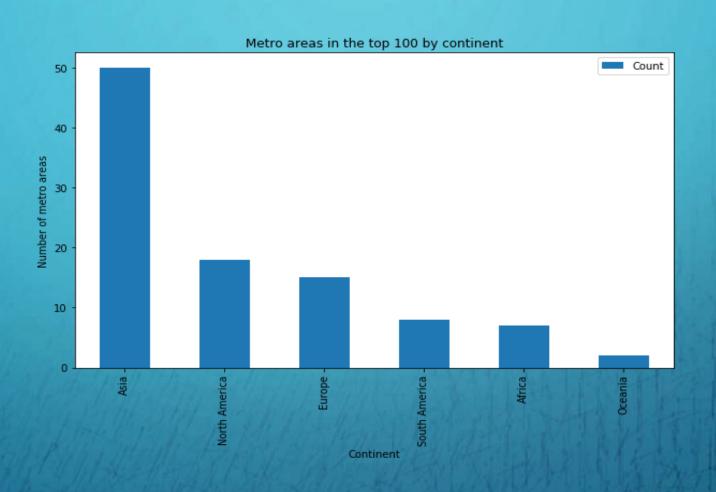
	Metropolitan	Country	Continent
0	Tokyo	Japan	Asia
1	Delhi	India	Asia
2	Shanghai	China	Asia
3	Jakarta	Indonesia	Asia
4	Seoul	South Korea	Asia

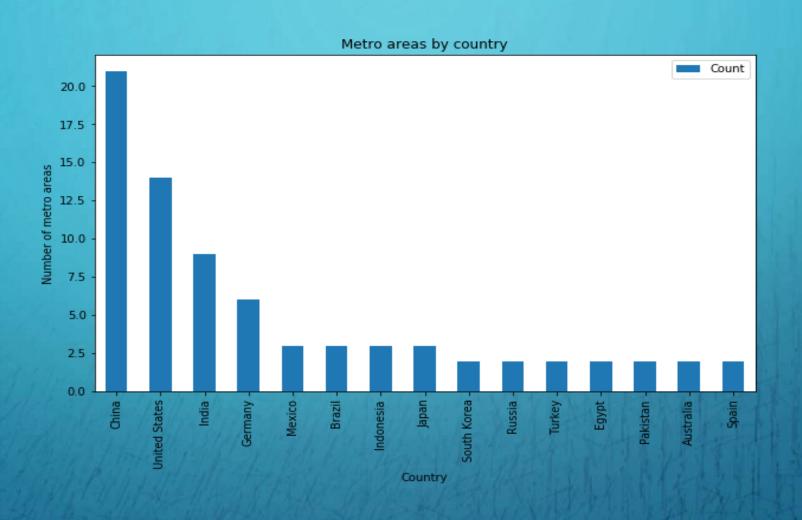
GEOPY GEOCODERS NOMINATIM LATITUDE AND LONGITUDE FOR EACH METRO AREA

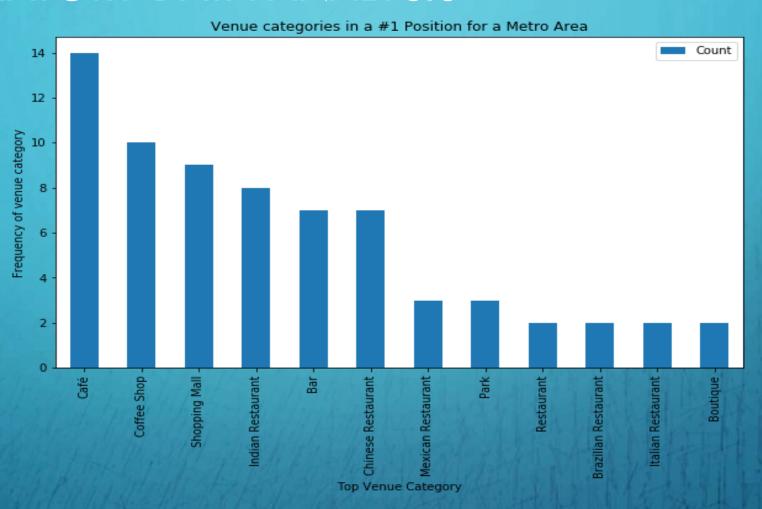
	Metropolitan	Country	Continent	Latitude	Longitude
0	Tokyo	Japan	Asia	35.6828	139.759
1	Delhi	India	Asia	28.6517	77.2219
2	Shanghai	China	Asia	31.2323	121.469
3	Jakarta	Indonesia	Asia	-6.17539	106.827
4	Seoul	South Korea	Asia	37.5667	126.978

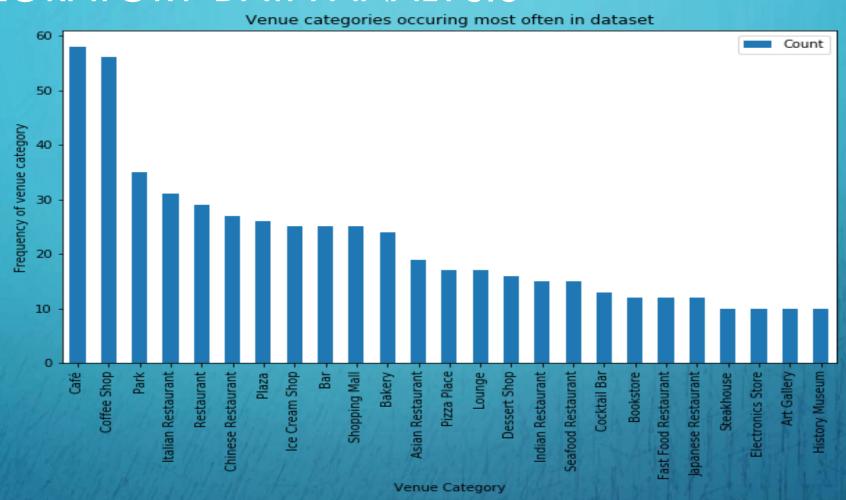
THE FOURSQUARE API 'TOP PICKS' VENUE CATEGORIES

непороная	Country	Continent	Laftude	Longitude	Calegory 1	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7	Category 8
Tokyo	Japan	Asia	35.682839	139.759455	Caté	Japanese Restaurant	talian Restaurant	Sushi Restaurant	Lounge	Garden	Flench Restaurant	Dessert Shop
Delhi	India	Asia	28.651718	77.221939	Indian Restaurant	Bar	Café	Flea Market	Lounge	Asian Restaurant	ice Cream Shop	South Indian Restaurant
Merico City	Nexico	North America	19.432601	-99 133342	Merican Restaurant	Art Museum	lluseum	Arts & Crafts Store	Bar	loe Cream Shop	Boutique	Bakery
São Paulo	Brazil	South America	-23.550651	-4 6 633382	Brazilian Restaurant	Japanese Restaurant	Café	Bakery	Snack Place	Bookstore	Cosmetics Shop	Asian Restaurant
Lagos	Năgeria	Africa	6.455057	3.394179	Lounge	African Restaurant	Bar	Caté	Shopping Mail	Pizza Place	Hotel	Art Gallery
	Delhi Mexico City São Paulo	Menico City Menico São Paulo Brazil	Delhi India Asia Merico City Merico North America São Paulo Brazil South America	Dehi India Asia 28.651718 Merico City Merico Morth 19.432601 São Paulo Brazil America 23.550651	Dehi India Asia 28.651718 77.221939 Merico City Merico North 19.432601 -99.133342 São Paulo Brazil Stuth -23.550851 -46.633382	Dehi India Asia 28.651718 77.221939 Indian Restaurant Merico City Merico Mortin 19.432601 99.133342 Restaurant São Paulo Bradi America -23.550651 -46.633382 Restaurant	Delhi India Asia 28.651718 77.221938 Indian Bar Mexico City Mexico North 19.432601 -99.133342 Restaurant Museum São Paulo Brazil South -23.550651 -49.633382 Restaurant Restaurant Restaurant	Delhi India Asia 28.651718 77.221938 Indian Bar Café Mexico City Nexico Mortin 19.432601 -99.133342 Mexican Art Nuseum São Paulo Brazil South -23.550851 -46.633382 Restaurant Restaurant Café	Delhi India Asia 28.651718 77.221939 Indian Bar Café Narket Merico City Nerico North 19.432601 99.133342 Restaurant Museum Huseum Store São Paulo Brazil South -23.550651 -46.633382 Restaurant Restaurant Café Bakery	Delhi India Asia 28.651718 77.221939 Indian Bar Café Rea Lounge Merico City Merico North 19.432601 99.133342 Restaurant Museum Inseum Café Bar São Paulo Brazil South -23.550651 45.633382 Brazilian Japanese Café Bakery Place	Delhi India Asia 28.651718 77.221939 Indian Bar Cafe Market Lounge Asian Mexico City Mexico North 19.432601 -99.133342 Mexican Art Museum Crafts Bar Cream Shop São Paulo Brazil South -23.550651 -48.633382 Brazilian Japanese Restaurant Restaurant Cafe Bakery Place Bookstore	Delhi India Asia 28.651718 77.221939 Indian Bar Café Market Lounge Asian Ice Cream Shop Merico City Mexico North 19.432601 99.133342 Restaurant Museum Museum Crafts Store Bar Shop Boulique São Paulo Brazil South -23.550651 -46.633382 Brazilian Japanese Café Bakery Flace Bookstore Cosmetics Shop



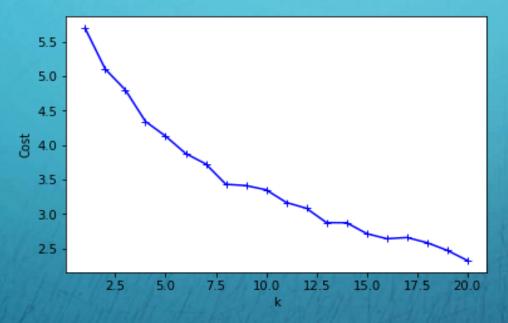






METHODS

I used KMeans to cluster the metro areas into groups. There was not obvious best k (number of clusters) to use, so based on cluster size, I chose k=16.



RESULTS



CONCLUSIONS WHICH METROPOLITAN AREAS ARE MOST SIMILAR?

- Using Foursquare venue category data, I grouped the 100 most populated metropolitan areas into 16 clusters
- 10 clusters with two or more metro areas
- 6 clusters with a single metro area, the outliers

CONCLUSIONS WHICH METROPOLITAN AREAS ARE MOST SIMILAR?

- Most of the outlier metro areas had fewer than ten venue categories returned by Foursquare.
- Of the remaining 10 clusters, four contained only metro areas in Asia.
- Clusters 1, 3, 9, 13, and 14 contained a mixture of metro areas from different continents.
- Geographic areas tended to contain only a few clusters with the exception of China. China had the most metro areas in the list.

NOTEBOOKS

- Part 1:
 - https://nbviewer.jupyter.org/github/KathrynDH/IBMCapstoneFinalProject/blob/master/Final%20Project%20Get%20Data.ipynb
- Part 2: https://nbviewer.jupyter.org/github/KathrynDH/IBMCapstoneFinalProject/blob/master/Final%20Project%20-%20Get%20FourSquare%20Data.ipynb
- Part 3:
 https://nbviewer.jupyter.org/github/KathrynDH/IBMCapstoneFinalProject/blob/master/Explore%20Location%20Data.ipynb
- Part 4:
 https://nbviewer.jupyter.org/github/KathrynDH/IBMCapstoneFinalProject/blob/master/Cluster%20Metro%20Areas.ipynb
- Part 5:
 https://nbviewer.jupyter.org/github/KathrynDH/IBMCapstoneFinalProject/blob/master/Explore%20the%20Clusters.ipynb

REFERENCES

Course Reference:

https://www.coursera.org/learn/applied-data-science-capstone

Data sources:

https://en.wikipedia.org/wiki/List of metropolitan areas by population

https://developer.foursquare.com/