



IBM Applied Data Science Capstone

**Clustering Five Financial Hub of the World
according to essential venues**

-Nishant Kumar

April 14, 2021

1. Introduction:

1.1 Background:

Financial hub is a location with a concentration of participants in banking, asset management, insurance or financial markets with venues and supporting services for these activities to take place. Participants can include financial intermediaries (such as banks and brokers), institutional investors (such as investment managers, pension funds, insurers, hedge funds), and issuers (such as companies and governments). Trading activity can take place on venues such as exchanges and involve clearing houses, although many transactions take place over-the-counter (OTC), that is directly between participants. Financial centres usually host companies that offer a wide range of financial services, for example relating to mergers and acquisitions, public offerings, or corporate actions; or which participate in other areas of finance, such as private equity and reinsurance. Ancillary financial services include rating agencies, as well as provision of related professional services, particularly legal advice and accounting services.

1.2 Problem:

- Which cities are similar?
- Which city ranks better in terms of essential venues?
- What is the venue is omni present and tops the list?
- Which city have more essential venues, both in terms of number and density?

1.3 Target Audience:

This project aims to help all individuals(bankers, investors, brokers, managers) planning to move to a new city by presenting visualizations and data. Thus helping the one to make informed decision.

1.4 Data Description:

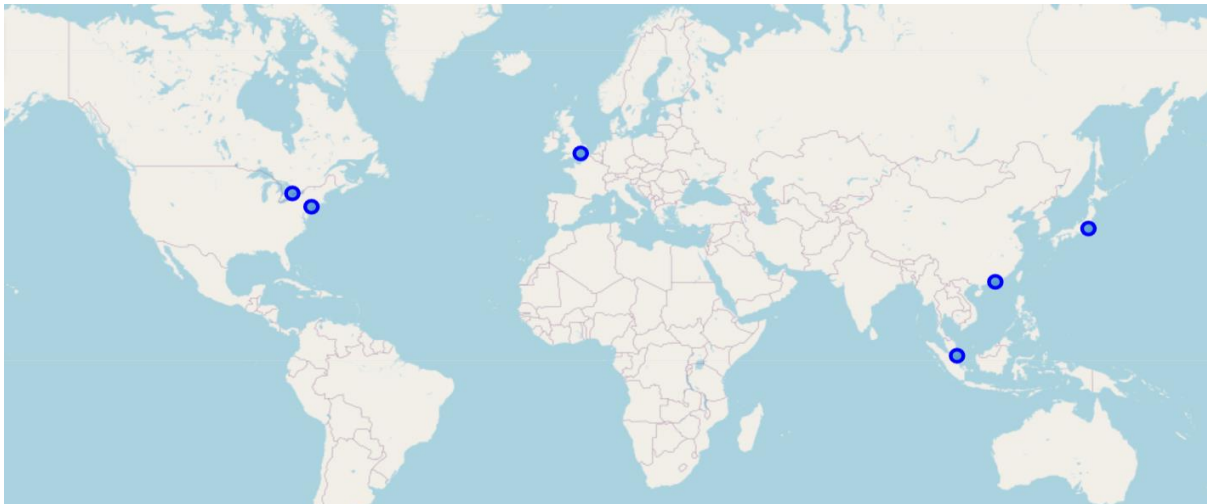
- City Latitude/ Longitude:
 - Data is obtained using opencage library by fetching in city name and country name
 - Website: <https://opencagedata.com>
- City venues:
 - Data is obtained using Foursquare API
 - Website: <https://foursquare.com>
- City, Country, Population:
 - Obtained from google
 - Website: <https://google.com>

2. Data Visualization:

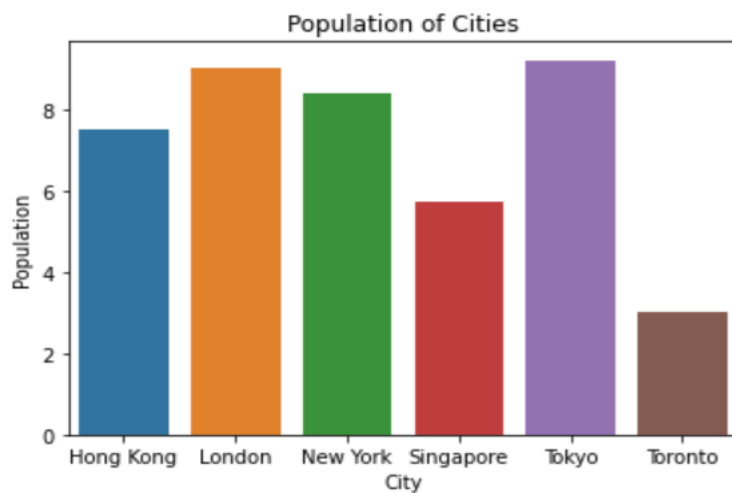
- Initial dataset created

	City	Country	Population
0	Hong Kong	China	7.5
1	London	United Kingdom	9.0
2	New York	USA	8.4
3	Singapore	Singapore	5.7
4	Tokyo	Japan	9.2
5	Toronto	Canada	3.0

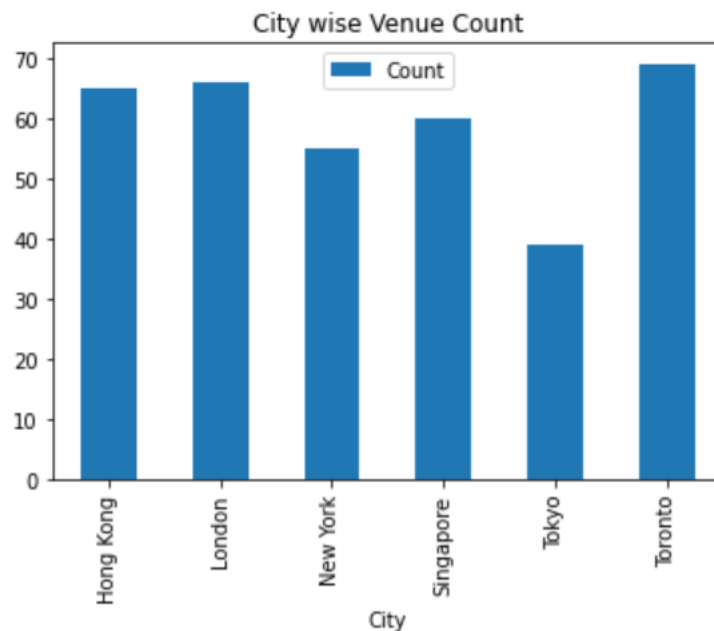
- Cities inside data dataset



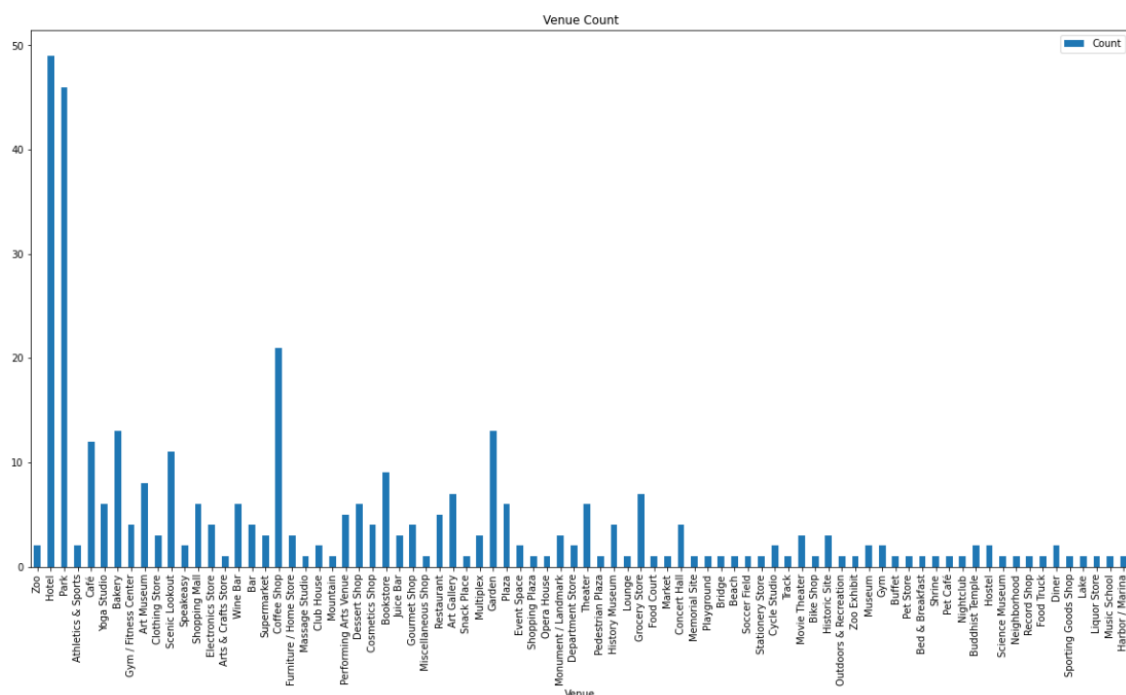
- Population of dataset:



It can be clearly observed that population of Tokyo, New York and London are almost same. But, there is a stark difference in the case of Tokyo and Toronto inside dataset of venues grouped by countries. Which is shown below:



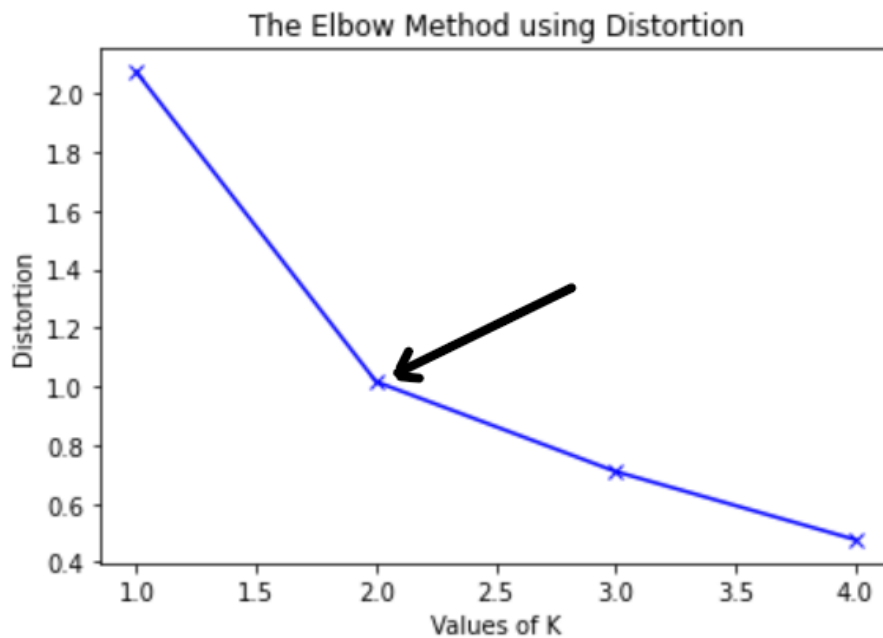
- **Venue count:**



Venue count dataset is topped by Hotel. Which along with parks and coffee shop are ubiquitous.

3. Machine Learning:

- Finding optimum value of K using elbow method:

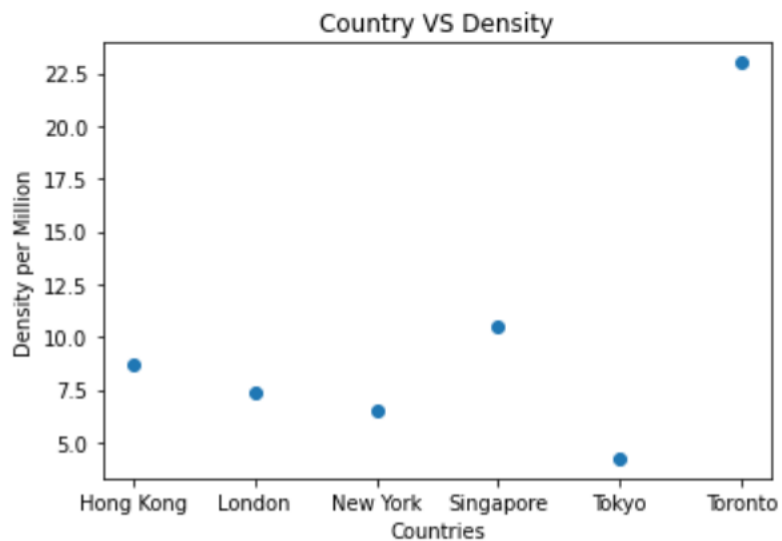


- World Map after performing machine learning:



- Venue density of cities:

	City	Count	Density
0	Hong Kong	65	8.666667
1	London	66	7.333333
2	New York	55	6.547619
3	Singapore	60	10.526316
4	Tokyo	39	4.239130
5	Toronto	69	23.000000



It can be clearly observed that Toronto ranks best in terms of essential services

4. Observation:

- **Cluster 0:**

	City	Zoo	Hotel	Park	Athletics & Sports	Café	Yoga Studio	Bakery	Gym / Fitness Center	Art Museum	...	Food Truck	Diner	Sporting Goods Shop	Lake	Liquor Store	Music School	Harbor / Marina	M I
1	London	0.0	1.111111	1.000000	0.0	0.222222	0.111111	0.111111	0.000000	0.222222	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	New York	0.0	0.238095	1.547619	0.0	0.000000	0.119048	0.476190	0.119048	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Tokyo	0.0	0.652174	0.217391	0.0	0.108696	0.000000	0.000000	0.000000	0.543478	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

3 rows × 86 columns

London , New York and Tokyo which have large population are grouped together. This cluster have lowest 3 cities of the list based on density. This presenting a picture that population had played an important role in resurrection of venues.

- **Cluster 1:**

	City	Zoo	Hotel	Park	Athletics & Sports	Café	Yoga Studio	Bakery	Gym / Fitness Center	Art Museum	...	Food Truck	Diner	Sporting Goods Shop	Lake	Liquor Store	Music School	Harbor Marina
5	Toronto	0.0	1.0	4.0	0.333333	1.666667	0.0	1.666667	0.0	0.0	...	0.333333	0.666667	0.333333	0.333333	0.333333	0.333333	0.333333

1 rows × 86 columns

Cluster 1 contains Toronto, clearly creating the conclusion that this cluster have high density of essential venues.

- **Cluster 2:**

	City	Zoo	Hotel	Park	Athletics & Sports	Café	Yoga Studio	Bakery	Gym / Fitness Center	Art Museum	...	Food Truck	Diner	Sporting Goods Shop	Lake	Liquor Store	Music School	Harbo Marina
0	Hong Kong	0.133333	1.866667	0.533333	0.133333	0.266667	0.400000	0.4	0.4	0.133333	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Singapore	0.175439	2.456140	1.052632	0.000000	0.350877	0.175439	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0

2 rows × 86 columns

Cluster 2 contains Hong Kong and Singapore. This cluster is Asian and have medium density of essential venues.

5. Conclusion:

It can be clearly concluded that cluster 1 which have low population and high density of essential services should be the choice for any individual planning to relocate. Hong Kong and Singapore have similar features and thus are clustered in cluster 2, while London, New York and Tokyo are clustered in to cluster 0.