CLUSTERING SMART CITIES AROUND THE WORLD BASED ON THEIR VENUES

Bhaskar Rao

PROBLEM STATEMENT

Top 50 smart cities around the world are selected and analysed. A novel way to cluster them together, based on their social signature (social venues), is explored to generate insights. City officials can use the clusters generated to compare how similar or different is their city to a smart city in terms of social life. This will help them evaluate and plan strategies more suitable for their residence, keeping in mind their social signature.

Data Source and Acquisition

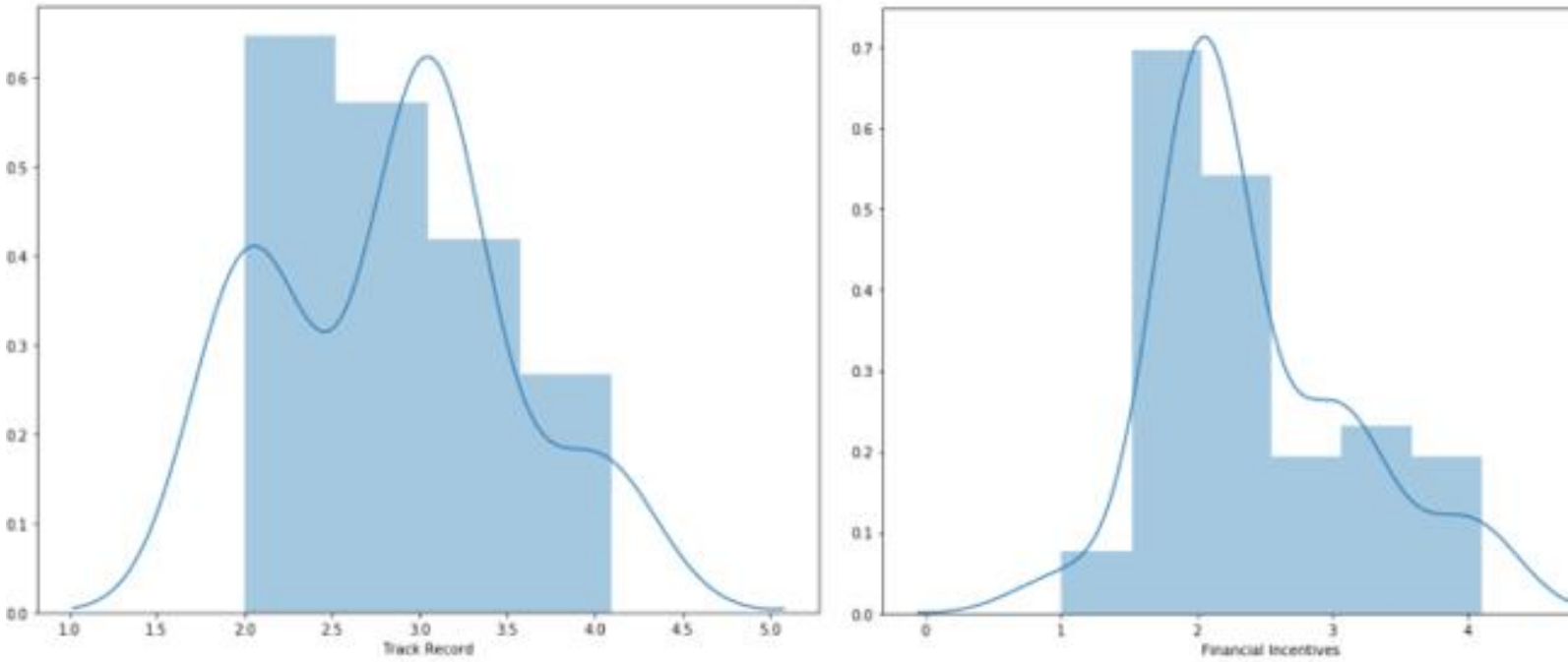
- The list of 50 top global smart cities is extracted by web-scrapping https://www.smartcitygovt.com/
- The latitude and longitude of these city were generated by using the ***geopy*** python library. These city latitude and longitudes are queried against the foursquare data to get the most popular venues around their city centre.
- Foursquare database provided top 10 popular venue around city centre (using ***geopy's*** latitude and longitude) are collected for each of these 50 cities. All these are combined to create the analytical data set, a snapshot of which is provided below

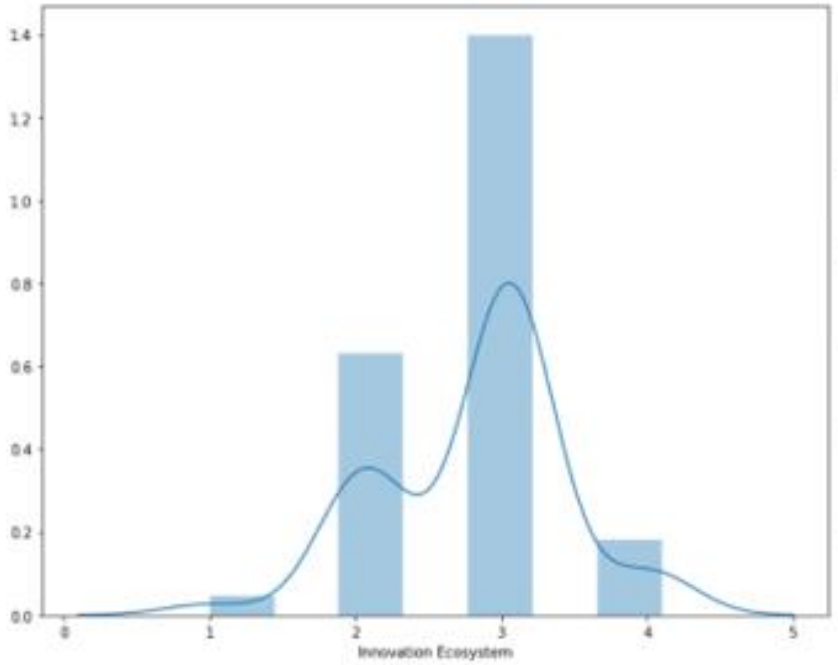
Master Data View

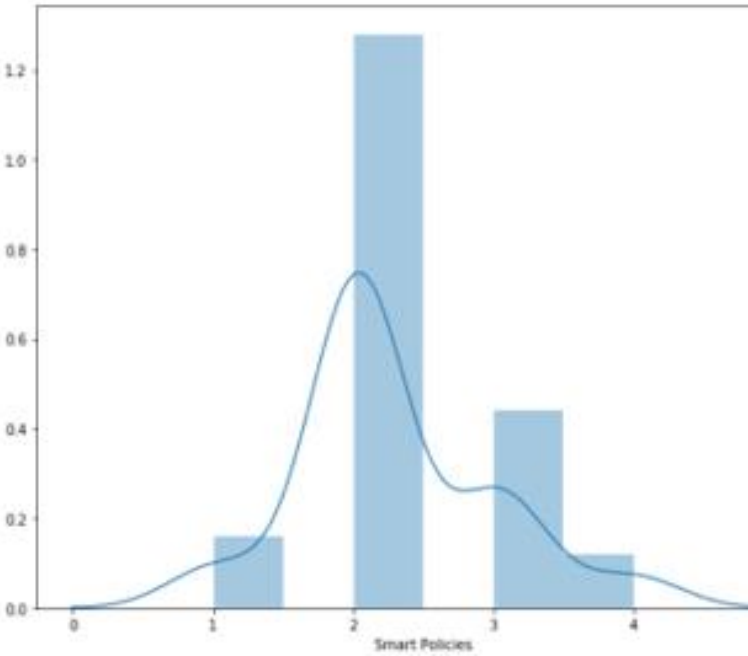
City	Total Score	Vision	Leadership	Budget	Financial Incentives	Support Programmes	Talent-Readiness	People Centricity	Innovation Ecosystem	Smart Policies	Track Record	Latitude	Longitude
London	33.5	3.1	4.0	3.0	4.0	3.0	3.1	3.0	4.1	3.1	3.1	51.507322	-0.127647
Singapore	32.3	3.0	4.0	3.0	4.1	3.0	3.1	2.0	3.1	4.0	3.0	1.357107	103.819499
Seoul	31.4	3.1	3.0	3.0	2.2	3.0	3.0	4.1	3.0	3.0	4.0	37.566679	126.978291
New York	31.3	3.0	3.0	3.0	3.1	3.0	3.1	3.0	4.0	2.0	4.1	40.712728	-74.006015
Helsinki	31.2	3.0	2.0	4.0	3.1	3.0	4.0	3.0	3.1	2.0	4.0	60.167410	24.942577

Exploratory Data Analysis

- Each of the parameters which have been used to score the cities have been analysed by looking at their distribution using histograms.
- Most of these cities were scored below average for track record and financial incentives whereas for innovative ecosystem, they received above average score. They score just about average for smart policies.

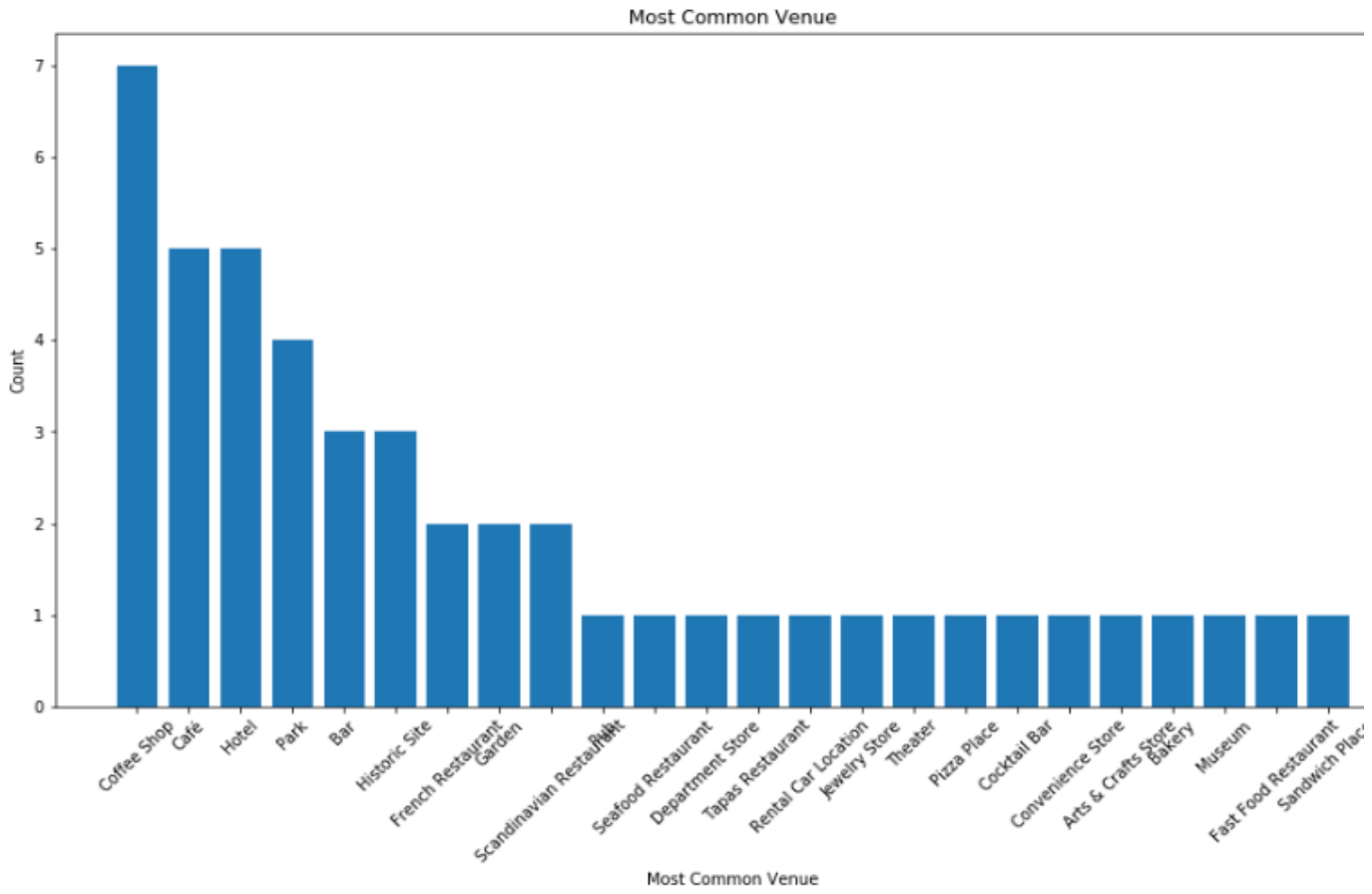






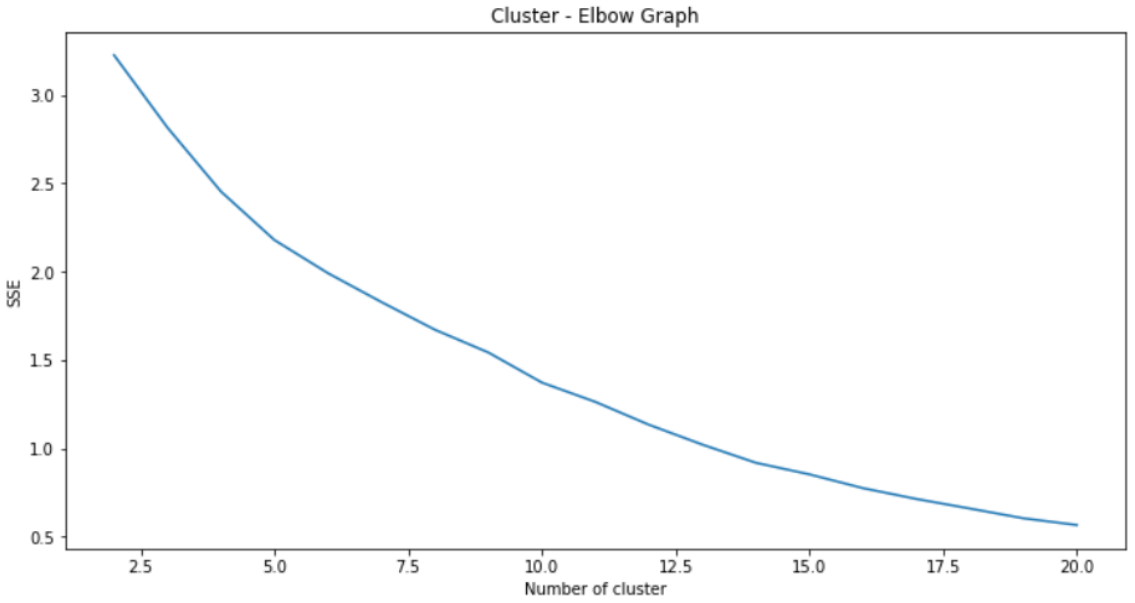
Most popular venue type across all 50 cities

Coffee shops, cafés and hotels are the topmost popular spots around these smart cities.



Clustering Smart Cities

- K-means clustering algorithm is selected for this unsupervised learning project, as the aim is to discover relations/insights from these cities (in terms of venues).
- K-means clustering is done for the 50 smart cities, using their venues as features. An elbow plot is generated to determine the optimal number of clusters, based on their *SSE*.



Results

The 50 cities were grouped into 7 clusters, they are provided below, based on the similarities of their popular venues

Clusters						
0	1	2	3	4	5	6
Copenhagen	Ahmedabad	Pune	Kigali	Barcelona	Atlanta	Bhubaneswar
Dubai	Amsterdam	Shenzhen	Montreal	Dublin	Beijing	New Delhi
Jakarta	Berlin	Sydney		Los Angeles	Rio de Janeiro	Shanghai
Melbourne	Boston			Lyon	Singapore	
Stockholm	Charlotte			Seoul		
Tel Aviv	Chicago			Taipei		
Wellington	Columbus			Tokyo		
	Helsinki			Toronto		
	Hong Kong			Vancouver		
	Jaipur			Vienna		
	Kansas City					
	London					
	New York					
	Paris					
	Philadelphia					
	Reykjavik					
	San Francisco					
	Seattle					
	Washington, DC					

Cluster Plot on World Map

Discussion

- From the clustering results, it can be seen that majority of cities fall under cluster 1.
- A city's leadership can analyse their popular venues and can compare their city with one of the clusters. This will inform them how similar or different is their social signature is with one of these clusters.
- This comparison can be used to take inspirations about how smart cities similar to their cities are tackling problems and take learnings from them to develop their own cities.

Conclusion

- Smart cities were clustered based on their ‘social signature’, which is defined by their popular venues.
- Tools and techniques like K-means clustering, foursquare API, geopy and leaflet are used for this project.
- A new way of using the social factor of a city to compare and contrast with other cities is explored and implemented.