Capstone Project

Srikruthi Adiraju

Introduction

The final course of the Data Science Professional Certificate consist of a capstone project where in all the skills and relevant knowledge that one has gathered from this 9 intense courses has to be applied on a final capstone project.

The final problem as well as the analysis is the left for the reader to explore and decide. The idea uses location data with the help of the foursquare api that can be leveraged into coming up with a problem that the foursquare location data
to solve it or just in contrast to compare cities or neighbourhoods of ones own choice

Business Problem

In this ever changing world of technology and reforms the use of AI will dominate and change most of the world and industries as we know so among the two busiest cities in the world which one would a person be willing to start a business in AI. Various factors would be included such as pricing, multiculturism, language barriers and so on would influence this decision.

Data

Paris Dataset

	Place Name	State	County	City	Latitude	Longitude
0	Paris 01 Louvre	Île-de-France	Paris	Paris	48.8592	2.3417
1	Paris 02 Bourse	Île-de-France	Paris	Paris	48.8655	2.3426
2	Paris 03 Temple	Île-de-France	Paris	Paris	48.8637	2.3615
3	Paris 04 Hôtel-de-Ville	Île-de-France	Paris	Paris	48.8601	2.3507
4	Paris 05 Panthéon	Île-de-France	Paris	Paris	48.8448	2.3471

Data

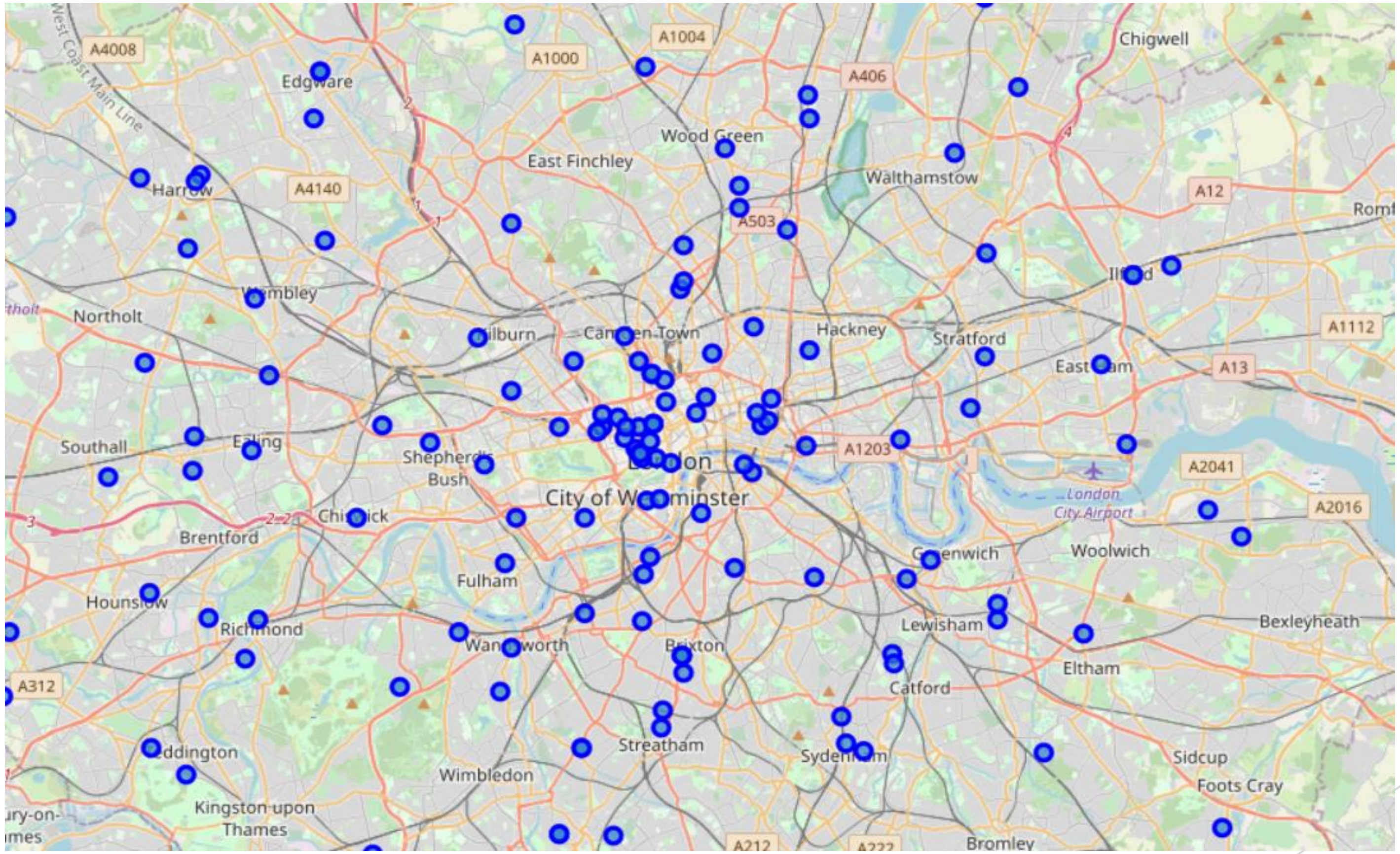
London Dataset

	Postcode	Country	County	District	Latitude	Longitude
0	BR1 1AA	England	Greater London	Bromley	51.401546	0.015415
1	BR1 1AB	England	Greater London	Bromley	51.406333	0.015208
2	BR1 1AD	England	Greater London	Bromley	51.400057	0.016715
3	BR1 1AE	England	Greater London	Bromley	51.404543	0.014195
4	BR1 1AF	England	Greater London	Bromley	51.401392	0.014948

Methodology

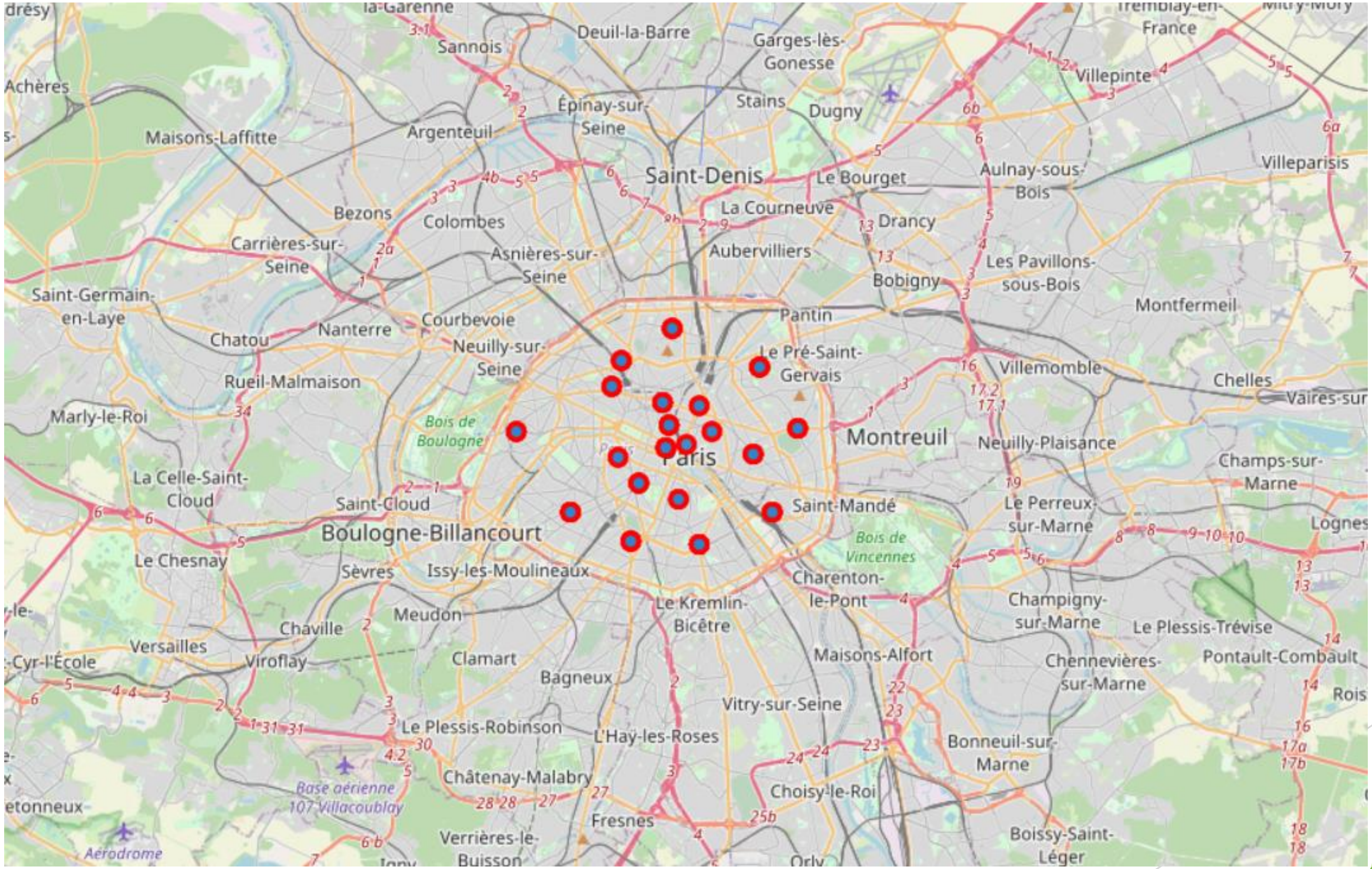
An in-depth research of the dataset has been done and a thorough analysis of the various features and methods have been investigated to ensure the maximum accuracy of the model as possible.
After reduction of the number of features in the data frame by replacing them with more useful data cluster analysis was done to find the best cluster of both Paris and London and then correlation and various other visual graphs were used to compare the two cities.

Map London



Chigwell
Edgware
Wood Green
East Finchley
Walthamstow
Harrow
Northolt
Hackney
Stratford
Southall
Ealing
Bush
City of Westminster
Chiswick
Brentford
Greenwich
Woolwich
Fulham
Hounslow
Richmond
Lewisham
Bexleyheath
Brixton
Eltham
Catford
Streatham
Wimbledon
Sidcup
Foots Cray
Kingston upon
Thames
Bromley
London
City Airport
A4008
A1000
A1004
A406
A4140
A503
A12
A1112
A13
A1203
A2041
A2016
A312
A212
A222

Map Paris



Sannois
Deuil-la-Barre
Garges-lès-Gonesse
Achères
Maisons-Laffitte
Argenteuil
Épinay-sur-Seine
Stains
Dugny
Villepinte
Saint-Denis
Le Bourget
Aulnay-sous-Bois
Villeparisis
Bezons
Colombes
La Courneuve
Drancy
Carrières-sur-Seine
Asnières-sur-Seine
Aubervilliers
Bobigny
Les Pavillons-sous-Bois
Saint-Germain-en-Laye
Montfermeil
Chatou
Nanterre
Courbevoie
Pantin
Neuilly-sur-Seine
Le Pré-Saint-Gervais
Villemomble
Chelles
Rueil-Malmaison
Marly-le-Roi
Bois de Boulogne
Paris
Montreuil
Neuilly-Plaisance
Champs-sur-Marne
La Celle-Saint-Cloud
Saint-Cloud
Saint-Mandé
Le Perreux-sur-Marne
Boulogne-Billancourt
Bois de Vincennes
Lognes
Le Chesnay
Sèvres
Issy-les-Moulineaux
Charenton-le-Pont
Le Kremlin-Bicêtre
Champigny-sur-Marne
Meudon
Chaville
Le Plessis-Trévise
Versailles
Viroflay
Clamart
Maisons-Alfort
Chennevières-sur-Marne
Pontault-Combault
Bagneux
Vitry-sur-Seine
Le Plessis-Robinson
L'Haÿ-les-Roses
Bonneuil-sur-Marne
Châtenay-Malabry
Base aérienne 107 Villacoublay
Choisy-le-Roi
Fresnes
Verrières-le-Buisson
Boissy-Saint-Léger
Aérodrome
Orly

Venues

London

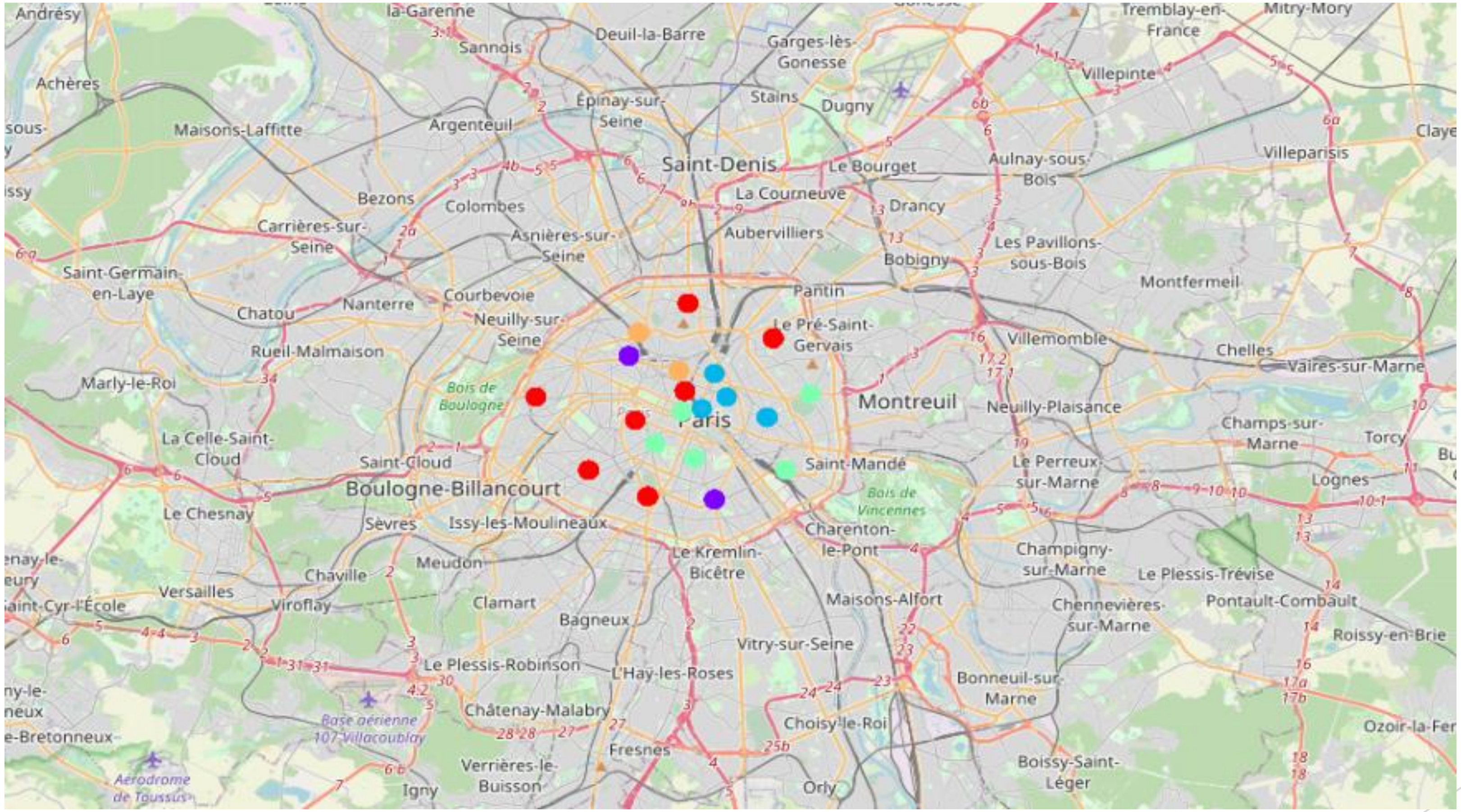
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Barnet	Turkish Restaurant	Italian Restaurant	Sushi Restaurant	Grocery Store	Indian Restaurant	Bakery	Deli / Bodega	Portuguese Restaurant	Coffee Shop	Gym / Fitness Center
1	Brent	Pub	Coffee Shop	Park	Platform	Indian Restaurant	Eastern European Restaurant	Supermarket	Food Truck	Japanese Restaurant	Deli / Bodega
2	Bromley	Pizza Place	Supermarket	Coffee Shop	Grocery Store	Pub	Stationery Store	Indian Restaurant	Fish & Chips Shop	Pharmacy	Café
3	Camden	Japanese Restaurant	Pizza Place	Coffee Shop	Beer Bar	Italian Restaurant	Tapas Restaurant	Malay Restaurant	Market	Hotel	Mexican Restaurant
4	City of London	Boxing Gym	Hotel	Burrito Place	Steakhouse	Department Store	Pizza Place	Indie Movie Theater	Event Space	French Restaurant	Botanical Garden

Venues

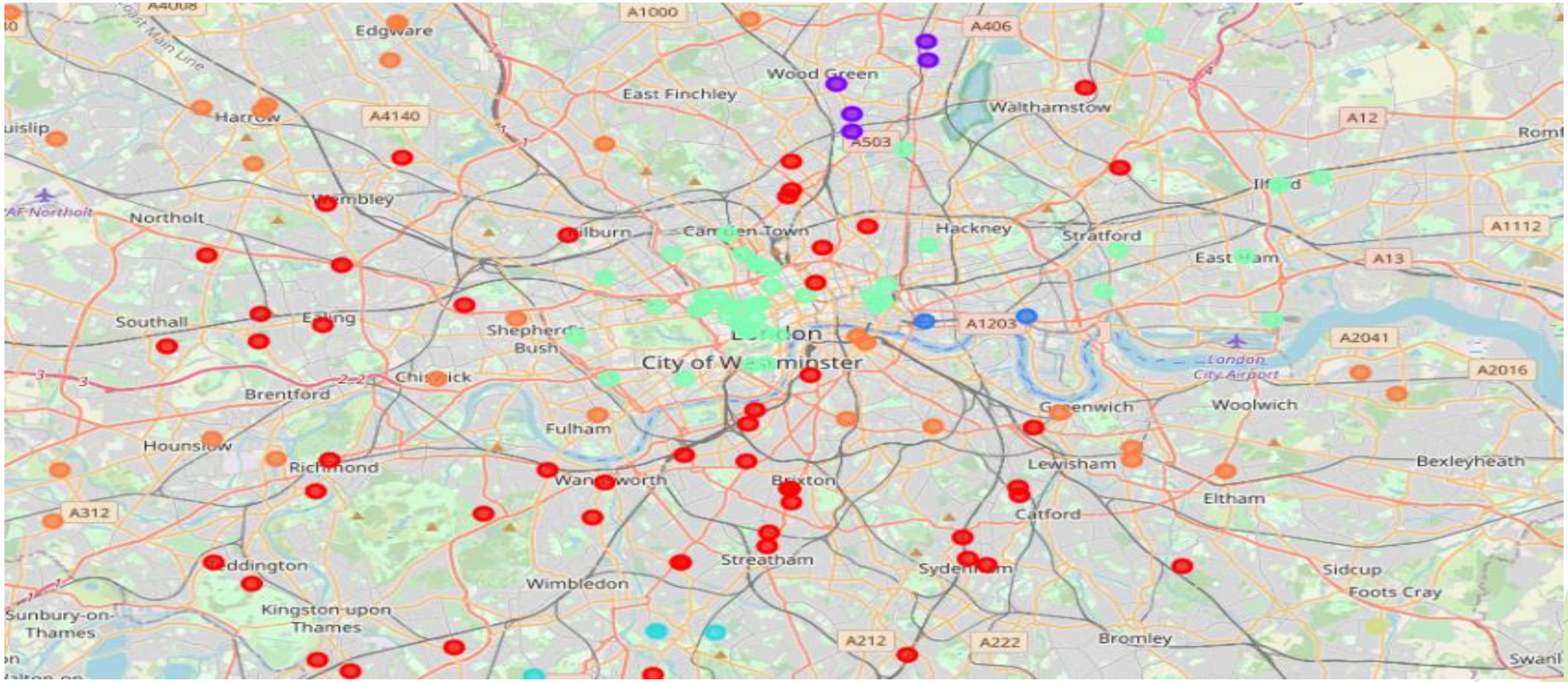
Paris

	Cluster Labels	Neighborhood	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	3	Paris 01 Louvre	Plaza	French Restaurant	Cocktail Bar	Church	Pedestrian Plaza	Chinese Restaurant	Park	Coffee Shop	Art Gallery	Garden
1	0	Paris 02 Bourse	French Restaurant	Plaza	Bakery	Ramen Restaurant	Restaurant	Souvlaki Shop	Perfume Shop	Bookstore	Farmers Market	Coffee Shop
2	2	Paris 03 Temple	Sandwich Place	Wine Bar	Park	Tea Room	Burger Joint	Restaurant	Cocktail Bar	Seafood Restaurant	Farmers Market	Wine Shop
3	2	Paris 04 Hôtel-de-Ville	Ice Cream Shop	Souvenir Shop	Art Gallery	Art Museum	Cocktail Bar	Fountain	Gourmet Shop	Lebanese Restaurant	Pub	Alsatian Restaurant
4	3	Paris 05 Panthéon	Plaza	French Restaurant	Bar	Korean Restaurant	Monument / Landmark	Science Museum	Ice Cream Shop	Bakery	Creperie	Grocery Store

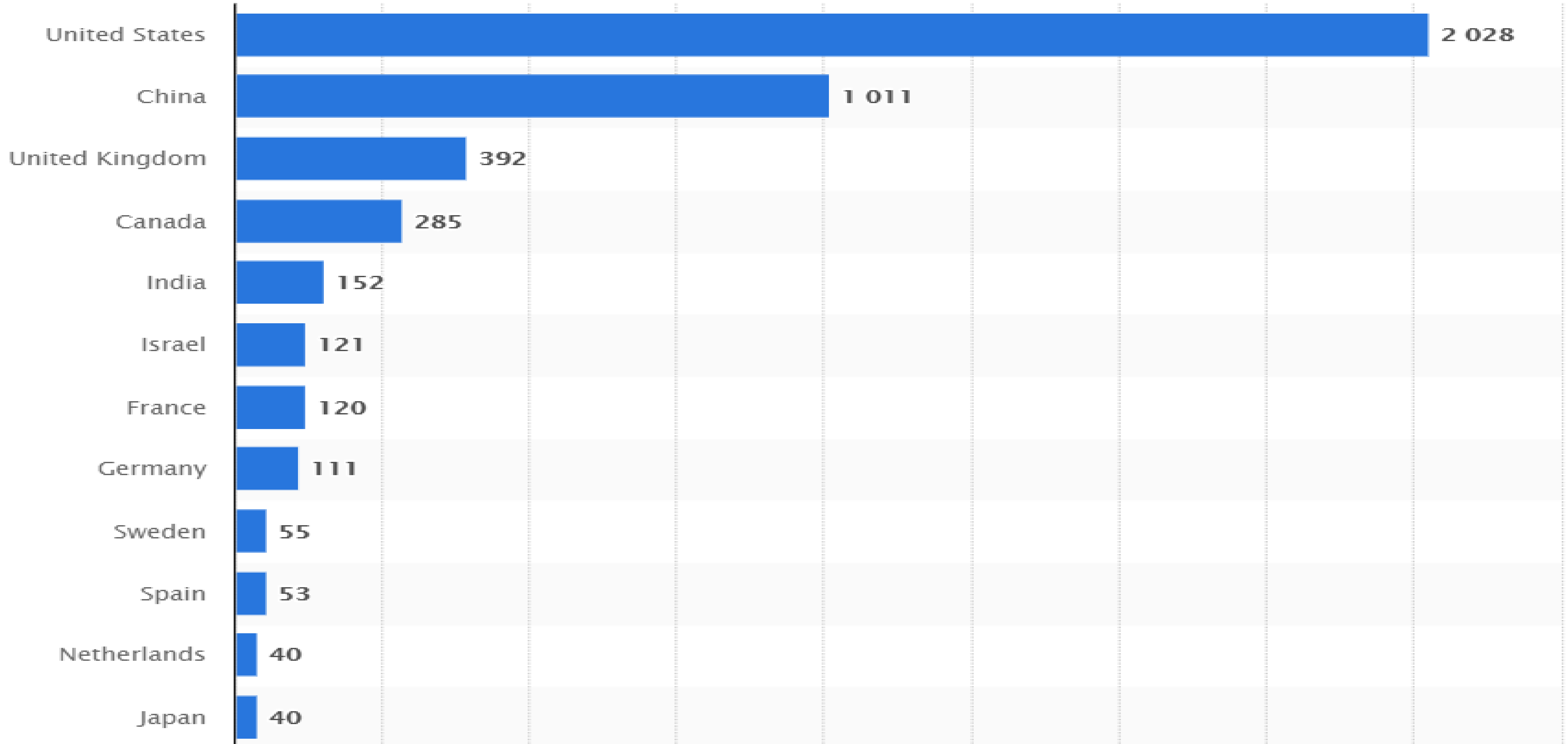
K Means Clustering Map - Paris



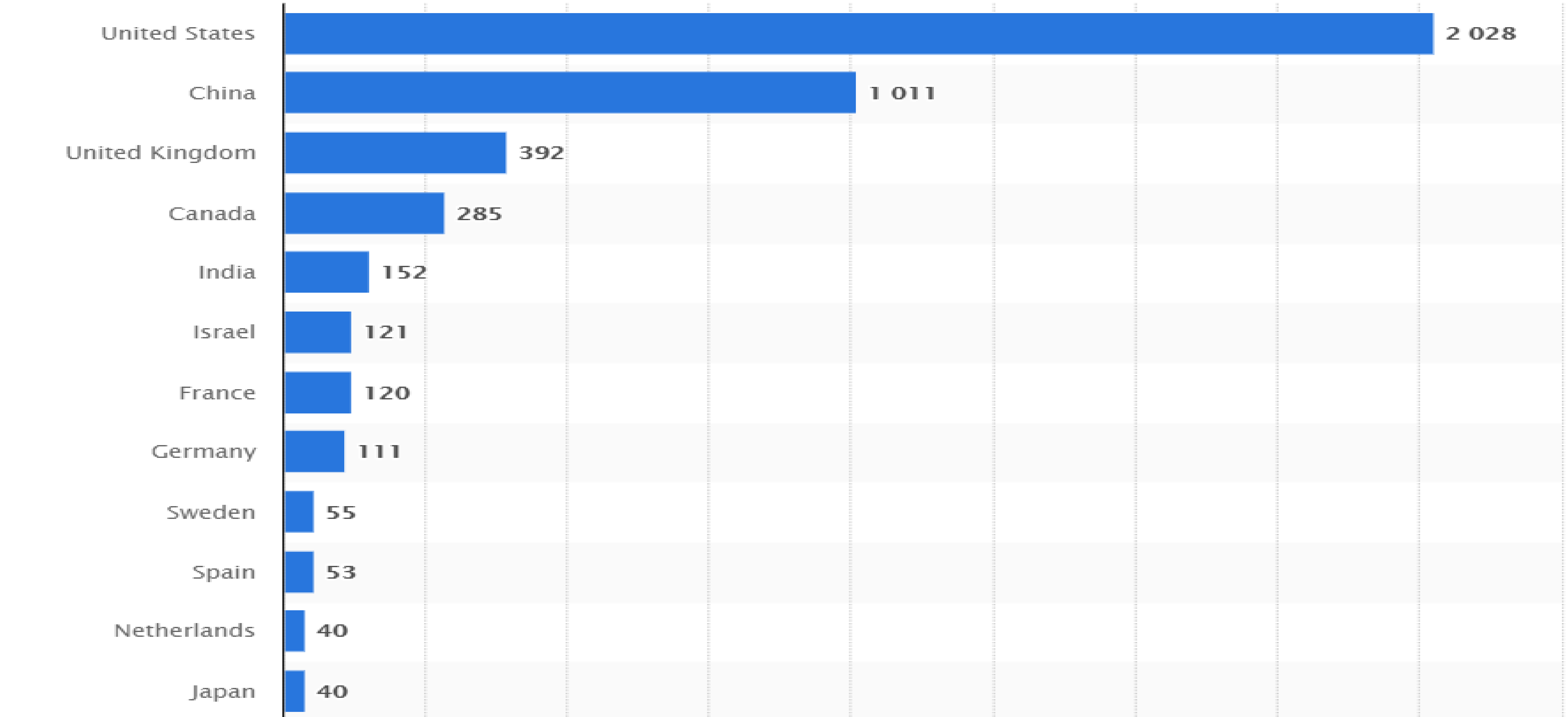
K Means Clustering Map - London



Artificial Intelligence



Artificial Intelligence



Results and Discussion

Similarities:
Both cities are multicultural and diverse in their own ways and share a rich history of their own.
Most of the famous neighbourhoods have a restaurant as its top most Common Venue.

Differences:
While looking at the maps one can observe that Paris is more compact and one can walk around much more freely without the use of transport
In terms of population density Paris definitely outweighs London by a ratio of 4:1.

Results and Discussion

Artificial Intelligence

Tech hub							
	2013	2014	2015	2016	2017	2018	Total
San Francisco	£418.08m	£1.83bn	£2.07bn	£4.46bn	£806.03m	£1.84bn	**£11.44bn**
Beijing	£11.66m	£53.75m	£197.32m	£599.56m	£1.63bn	£1.07bn	**£3.57bn**
New York	£79.43m	£165.62m	£318.28m	£667.51m	£593.85m	£1.2bn	**£3.05bn**
Shanghai	—	£1.28m	£400.93m	£16.10m	£1.6bn	£453.61	**£2.47bn**
London	£9.85m	£41.16m	£67.04m	£166.04m	£228.97m	£326.90m	**£839.96m**
Paris	£1.92m	£2.83m	£23.49m	£61.49m	£99.45m	£132.40m	**£321.48m**
Singapore	£13.76m	£13.89m	£70.92m	£55.59m	£106.52m	£30.81m	**£291.49m**
Tel Aviv	£14.80m	£17.12m	£5.49m	£39.04m	£112.25m	£89.01m	**£277.71m**
Berlin	£7.09m	£0.79m	£23.60m	£17.41m	£17.67m	£21.06m	**£87.62m**
Bangalore	£1.31m	£32.29m	£45.75m	£1.96m	£36.71m	£18.65m	**£136.67m**

Results and Discussion

Artificial Intelligence

The dataset for the Artificial Intelligence wasn't readily available and so had to be scrapped from multiple sources which often leads to inconsistency happening as well as errors.
The districts have too complex geometry which would bring an error in our analysis if the venues are too close to each other.
The data obtained through the API calls would return different results each time its called. Multiple trials and error runs are required to get the desired result.

Conclusion

Artificial Intelligence is a booming topic and recently more people have started investing into it as well as companies automating their processes
Both cities offer a wide range of opportunities for anyone starting to invest in Artificial Intelligence or even start a company and various factors were shown.
Finally a better model could be made by various other methods and much stronger Machine Learning Algorithms like KD Tree which have a much faster run time algorithm.
Furthermore, clustering however did help us to highlight the most optimal venues and areas.
Finally correlation does not imply causation and so any result here is subject to change on various other trends and opinions and datasets.