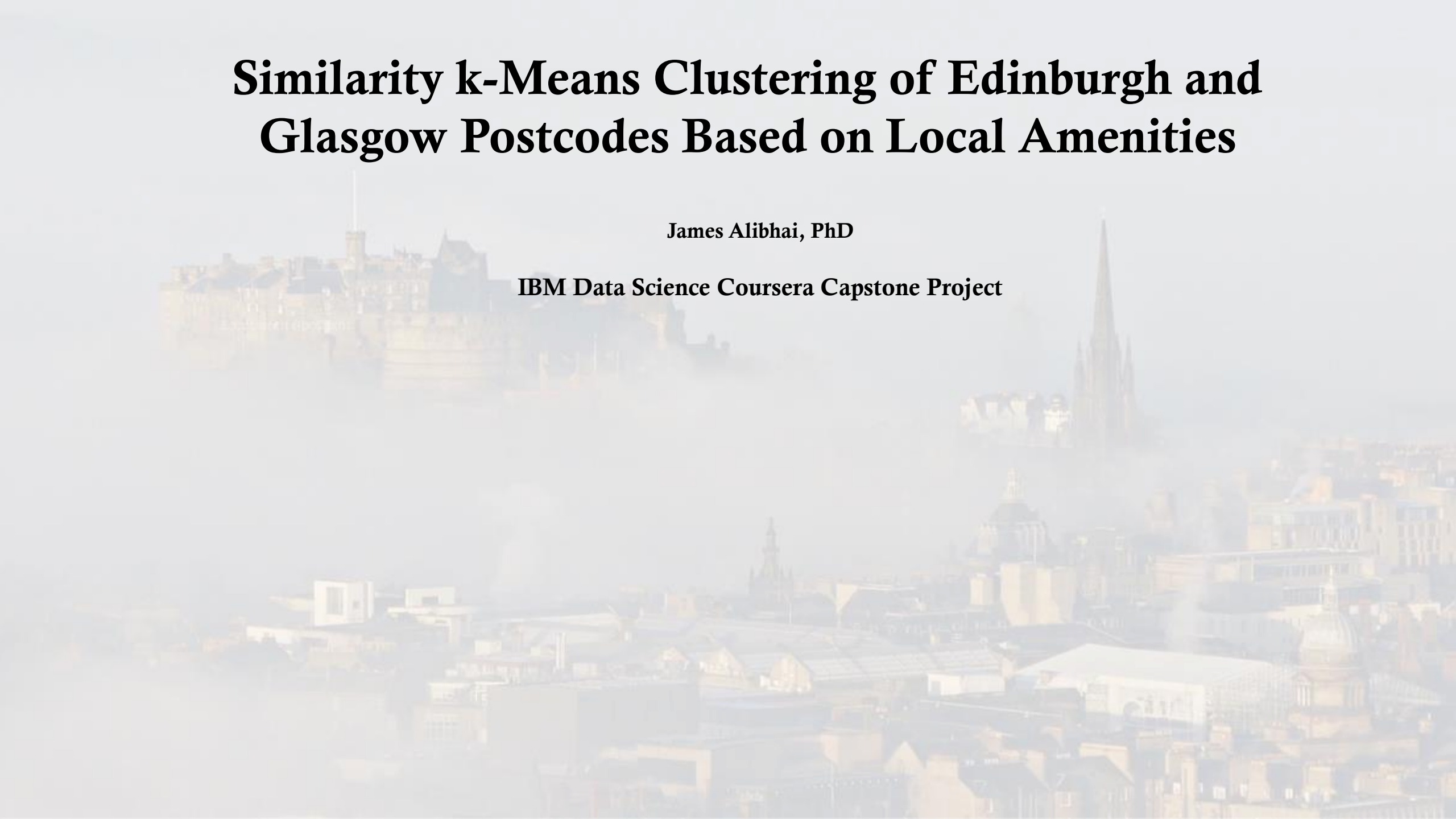
Similarity k-Means Clustering of Edinburgh and Glasgow Postcodes Based on Local Amenities

James Alibhai, PhD

IBM Data Science Coursera Capstone Project



Business background

Moving home is a challenging time for anyone, especially if that person is moving to a location that is unfamiliar.

Image Source: Property Reporter

The traditional role of an estate agent is to provide specialist advice about local areas

Online property searching become the normal working practice with online property search engines, such as RightMove, the first port of call for a property searcher.

Online searching risks losing specialist advice from local property experts, with many location summaries largely become predictable and too generic for individual clients.

Business case

An online tool that can be used to search similarity of unfamiliar locations using data from known locations.

If a client is moving from Edinburgh to Glasgow and is unfamiliar with Glasgow and its suburbs, it is in best interests of the client to have a tool available that can allow a search of Glasgow neighbourhoods.

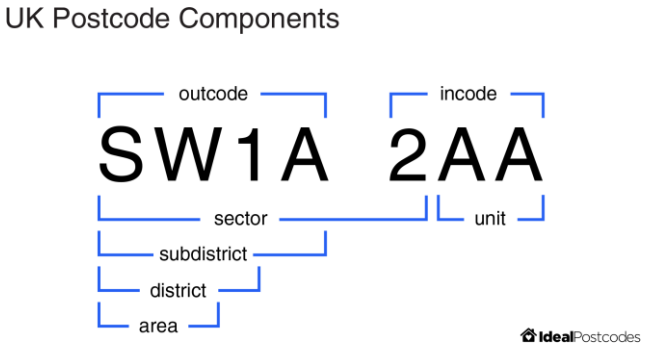
Generally a client will relate their desired area to a known area where they live and will want to know what locations best match their known desired location – this tool will provide the client with exactly that!

Data acquisition and cleaning

(i) UK postcode, latitude and longitude data is sourced from doogle.co.uk.

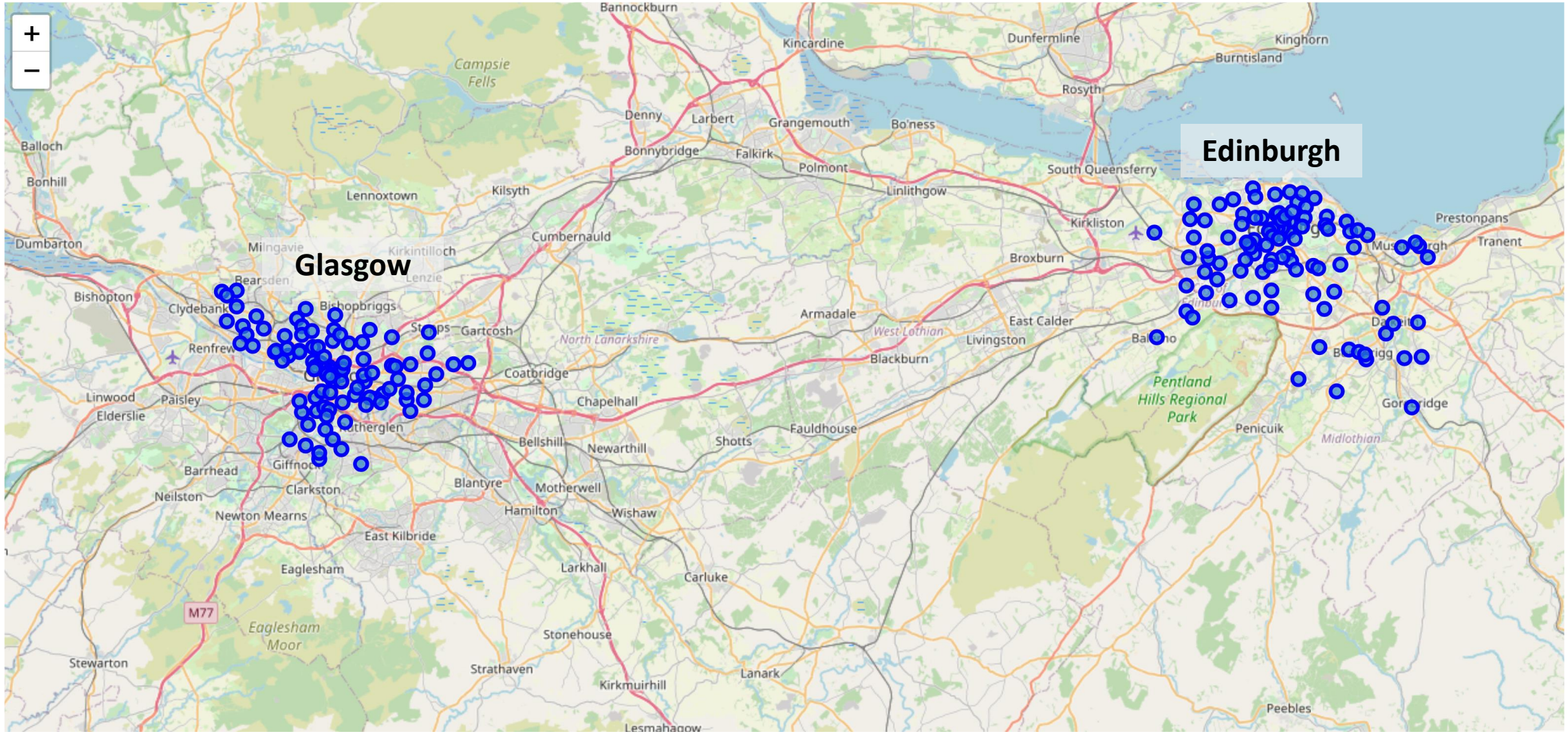
(ii) Local amenitites data is sourced from www.foursquare.com

(iii) Each full UK postcode data is binned into the sector postcode, e.g. EH1 1 or G2 1, resulting in 203 unique postcode sectors



(iv) Latitudes and longitudes of every EH (Edinburgh) and G (Glasgow) postcode that falls under the sector postcode is averaged by the mean.

Map of all Edinburgh and Glasgow postcode areas



IBM Data Science Coursera Capstone Project

James Alibhai, PhD

Venue information merged with Edinburgh and Glasgow dataframe

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Old Town	55.9509	-3.1891	The Milkman	55.950650	-3.191010	Coffee Shop
1	Old Town	55.9509	-3.1891	The Royal Mile	55.950029	-3.188567	Road
2	Old Town	55.9509	-3.1891	The Devil's Advocate	55.950309	-3.191643	Cocktail Bar
3	Old Town	55.9509	-3.1891	Whiski Bar & Restaurant	55.950318	-3.186471	Bar
4	Old Town	55.9509	-3.1891	Civerino's	55.949738	-3.188043	Pizza Place
5	Old Town	55.9509	-3.1891	St. Giles' Cathedral	55.949531	-3.191716	Church
6	Old Town	55.9509	-3.1891	The Edinburgh Larder	55.950080	-3.186088	Café
7	Old Town	55.9509	-3.1891	Angels With Bagpipes	55.949782	-3.191055	Restaurant
8	Old Town	55.9509	-3.1891	The Balmoral Hotel	55.953113	-3.189509	Hotel
9	Old Town	55.9509	-3.1891	Old Town Chambers	55.950165	-3.191188	Hotel

A total of 265 unique venue categories have been called from the FourSquare API for all postcode areas (‘neighborhood’)

Venue information merged with Edinburgh and Glasgow dataframe

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Alderman Road	3	3	3	3	3	3
Alexandra Parade	7	7	7	7	7	7
Allan Park	14	14	14	14	14	14
Andersonian Library	16	16	16	16	16	16
Angle Park Terrace	19	19	19	19	19	19
...	...	...	...	...	...	...
Western General Hospital	16	16	16	16	16	16
Willowbrae Road	4	4	4	4	4	4
Wyndford	3	3	3	3	3	3
Wyndford Road	5	5	5	5	5	5
York Place	52	52	52	52	52	52

(i) The unique venue categories are summarised in all postcode areas (‘Neighborhood’)

(ii) Onehot encoded and averaged by the mean

(iii) Finally fit to a k-Means cluster machine learning algorithm

Venue information merged with Edinburgh and Glasgow dataframe

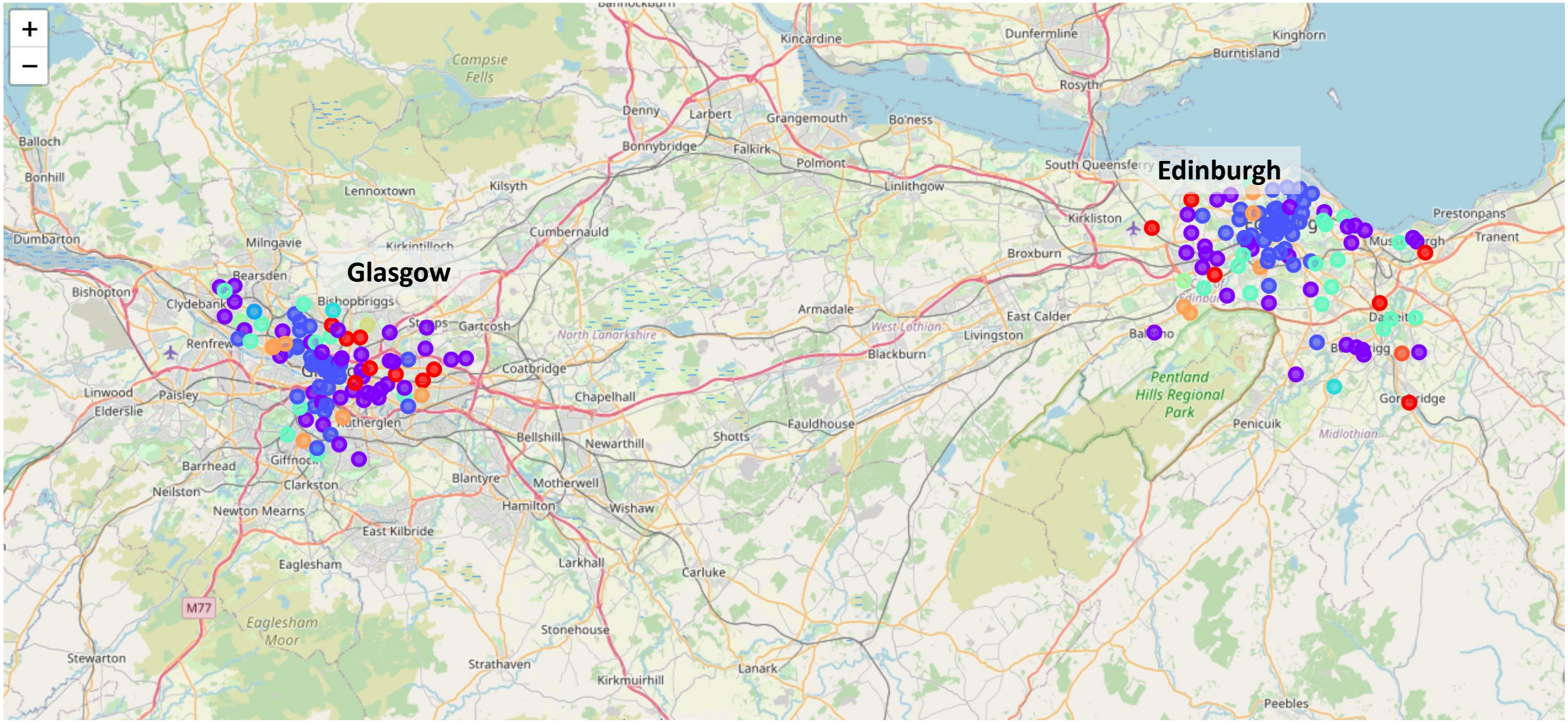
	Postcode	Latitude	Longitude	Area	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	EH1 1	55.9509	-3.18910	Old Town	2	Hotel	Pub	Bar
1	EH1 2	55.9492	-3.19825	Edinburgh Castle	2	Café	Pub	Historic Site
2	EH1 3	55.9568	-3.18752	York Place	2	Hotel	Bar	Pub
3	EH1 9	55.9432	-3.23293	Haymarket	2	Hotel	Grocery Store	Trail
4	EH2 1	55.9550	-3.19491	Queen Street	2	Bar	Café	Hotel
5	EH2 2	55.9532	-3.19631	George Street	2	Café	Bar	Coffee Shop
6	EH2 3	55.9533	-3.20072	Frederick Street	2	Bar	Café	Seafood Restaurant
7	EH2 4	55.9521	-3.20676	Charlotte Square	2	Bar	Hotel	Italian Restaurant
8	EH2 9	55.9451	-3.22670	West Coates	2	Hotel	Grocery Store	Italian Restaurant
9	EH3 0	55.9613	-3.20204	Brandon Street	2	Pub	Coffee Shop	Hotel

(i) The unique venue categories are summarised in all postcode areas (‘Neighborhood’)

(ii) Onehot encoded and averaged by the mean

(iii) Finally fit to a k-Means cluster machine learning algorithm

Cluster analysis of Edinburgh and Glasgow postcodes



IBM Data Science Coursera Capstone Project

James Alibhai, PhD

Conclusions and future directions

Fit Edinburgh and Glasgow postcode location data to a k-Means cluster machine learning algorithm

Shows similar locations between both cities based on local amenitities

Has potential for scale up by broadening database range to UK wide

Has potential for additional data sources to be added that would be desirable for someone looking to move home, such as crime safety data or school performance tables.