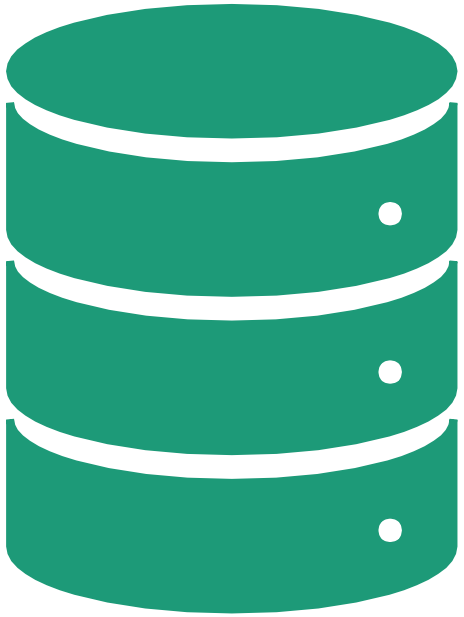
Schools and Venues in Northern Ireland and Belfast

Applied Data Science Capstone

Part 2

Introduction

- In this Capstone project I will be identifying areas in Northern Ireland and Belfast City that have schools and what other venues are in these locations. The project may support young couples or families in Northern Ireland that are looking to invest in their first home or move while considering local schools if they hope to raise or are raising children in this new home, while also considering what other venues are within the vicinity.
- This project will allow me to leverage the Foursquare location data to solve and execute the results, which will identify the venues around the schools. The school locations were obtained from Open Data NI.
- This audience of young couples and families would care about this problem because they may be making a large investment on a new home and would like to know if the location is suitable for their needs and suit raising a family in the future based on where their children can go to school and which town has suitable venues that suit their needs.



Data

- The data I am using is from a csv file which is publicly available on Open Data NI from the following link;
- https://www.opendatani.gov.uk/dataset/locate-a-school/resource/d0947faf-5d84-4ce4-80dd-ce4fa0e1c0d5
- I included the whole csv file in my notebook, however I found the following columns particularly useful;

Institution_Name	Town_Name	County_Name	Postcode	Latitude	Longitude
174 Trust Playgroup	BELFAST	ANTRIM	BT146BP	54.61253	-5.93645
Abbey Christian Brothers Grammar School	NEWRY	DOWN	BT342QN	54.19349	-6.32846
Abbey Community College	NEWTOWNABBEY	ANTRIM	BT370EA	54.69015	-5.91848
Abbey Primary School	NEWTOWNARDS	DOWN	BT238RQ	54.59601	-5.66723
Abbots Cross Primary School	NEWTOWNABBEY	ANTRIM	BT379QW	54.66801	-5.91408

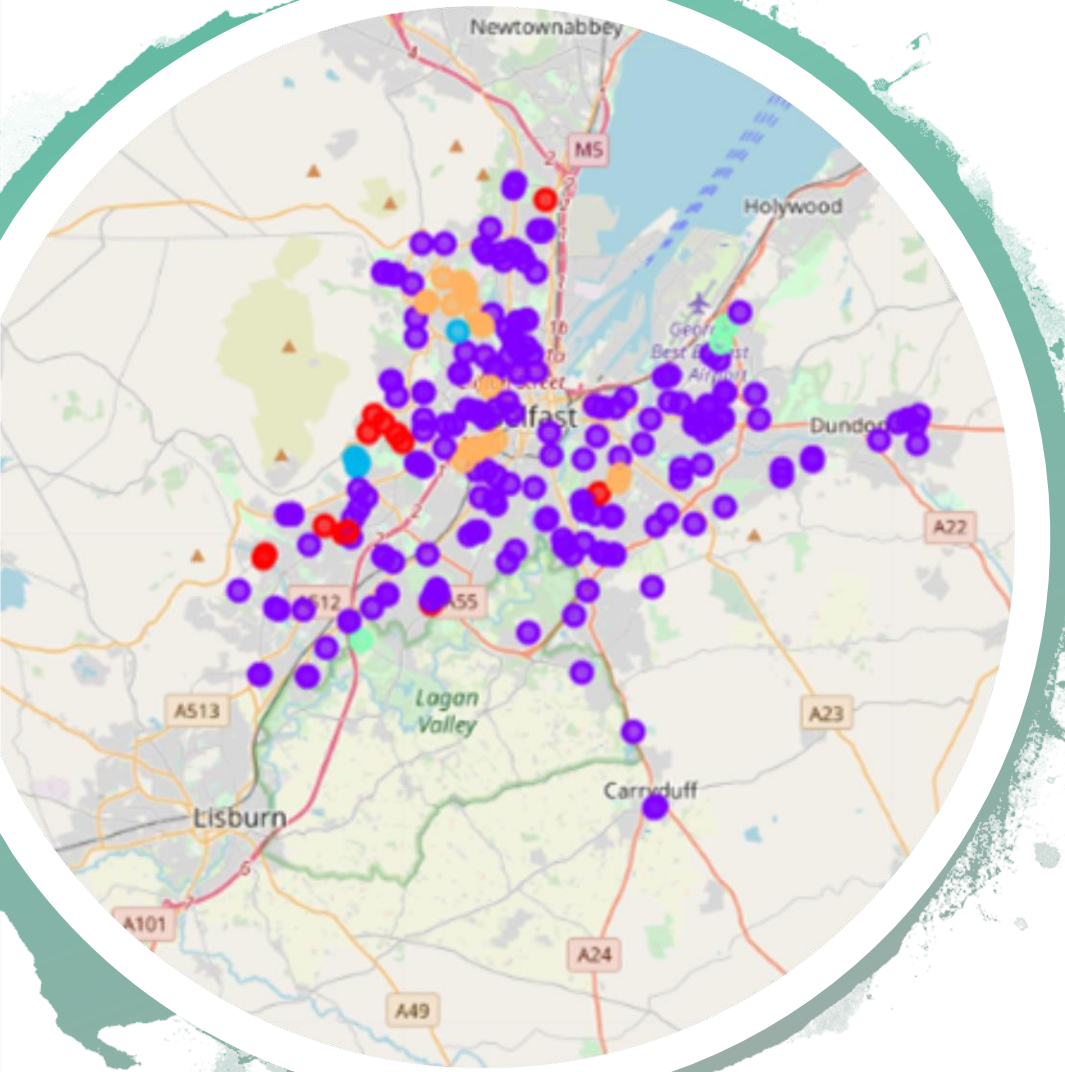
Data

- I used the csv file in combination with the Foursquare location data which was used to explore the schools and segment them. I experienced and resolved issues with the API call volume which has taught me to be conscious while using Foursquare APIs in the future.
- The foursquare data identified venues such as Supermarket, Park and Bus Stop which may be venues our audience require near their new home. The data frame from Foursquare looked like this;

	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	BT100DR	Sandwich Place	Health & Beauty Service	Fast Food Restaurant	Coffee Shop	Grocery Store	Supermarket	Flower Shop	Construction & Landscaping	Weight Loss Center	Forest
1	BT100HT	Warehouse Store	Hotel Bar	Fast Food Restaurant	Supermarket	Weight Loss Center	Flower Shop	French Restaurant	Forest	Food Court	Food & Drink Shop
2	BT100JB	Chinese Restaurant	Train Station	Auto Workshop	Pub	Weight Loss Center	Flower Shop	French Restaurant	Forest	Food Court	Food & Drink Shop
3	BT100LE	Park	Coffee Shop	Bus Stop	Weight Loss Center	Food	Fried Chicken Joint	French Restaurant	Forest	Food Court	Food & Drink Shop
4	BT100LF	Athletics & Sports	Weight Loss Center	Flower Shop	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Forest	Food Court	Food & Drink Shop	Food

Methodology

- Having identified the venues I analysed each school using one hot encoding where I converted the categorical variables. I identified the top 5 venues at each school which would help distinguish differences for the audience.
- This resulted in a data frame size of (1358, 147) and once grouped by postcode the new data frame size was (184, 147). I then investigated each school's top five venues.
- Following this I used k-means to cluster the schools with similar venue nearby into 5 clusters. The new data frame which included the clusters as well as the top 10 venues for each school



Results

- Once empty rows were dropped and data types amended the five clusters were plotted on the cluster map was as follows for Belfast City using Folium.
- As you can see the clusters are not very even in size, they appear to overlap. I examined further to determine the differentiating venue categories that distinguish each cluster.

Discussion

Given the audience of this project may be young couples or families trying to find a home based on schools I have made a suggestion on each what may suit certain individuals reviewing each cluster as follows

Cluster 1

Construction & Landscaping is the 1st most common venue, this may suit someone who has a career in this industry or perhaps suggests new developments in the area.

Cluster 2

This is the largest cluster. The most common venues would appear to suit active individuals with Athletics and Sports, Parks, Sports Clubs and Golf course as the common venues here.

Cluster 3

This cluster appears to suit a foodie however weight loss centres are ironically the 2nd most common venue.

Cluster 4

Food venues are very popular in this cluster, with Forests as the 7th most common venue this distinguishes Cluster 4 from the others. This may suit individuals who enjoy nature and the outdoors rather than the Belfast City lifestyle.

Cluster 5

This is another large cluster. What stood out here was that Grocery Stores appeared to be the 1st most common venue which would be convenient to have locally for a busy family lifestyle.

Conclusion

- In conclusion reviewing the common areas has told me more rather than just looking at the cluster on the map. The clustering appears to have worked. This may assist the audience to purchase a home in a cluster that suits their needs.