Coursera Presentation

Introduction

- In this Capstone project I will be identifying planning permission applications in Cork City from 2008 onwards. This data is in csv format and has been obtained from the City Council's open data website https://data.corkcity.ie.
- I will be looking at what neighbourhoods these applications take place in. Here the Foursquare data will help us define these neighbourhoods by the amenities available. I will also be using the supplemental data in this source to provide further useful comparisons between the segmented neighbourhoods once the clustering has been mapped.
- This information would be of significant interest to many parties. From building contractors and developers to county planners. It would help local businesses identify changing land use patterns and help banks identify which areas are going to experience significant growth.

Data

- The data I am using is a csv file I obtained from https://data.corkcity.ie.
- It is reasonably large measuring 4,975 rows by 34 columns. However of these the majority are unimportant to our investigation. As part of data preparation the following columns were kept and the rest dropped:
- Application Number
  - As Primary Key
- Latitude
- Longitude
- OneOffHouseFlag
  - This signals if the permission was for a one off house
- NumberOfResidentialUnits
  - This signifies how many homes are being built
- AreaOfSite
  - The total area of the property in km2
- Decision
  - Whether the application was accepted for rejected

Methodology

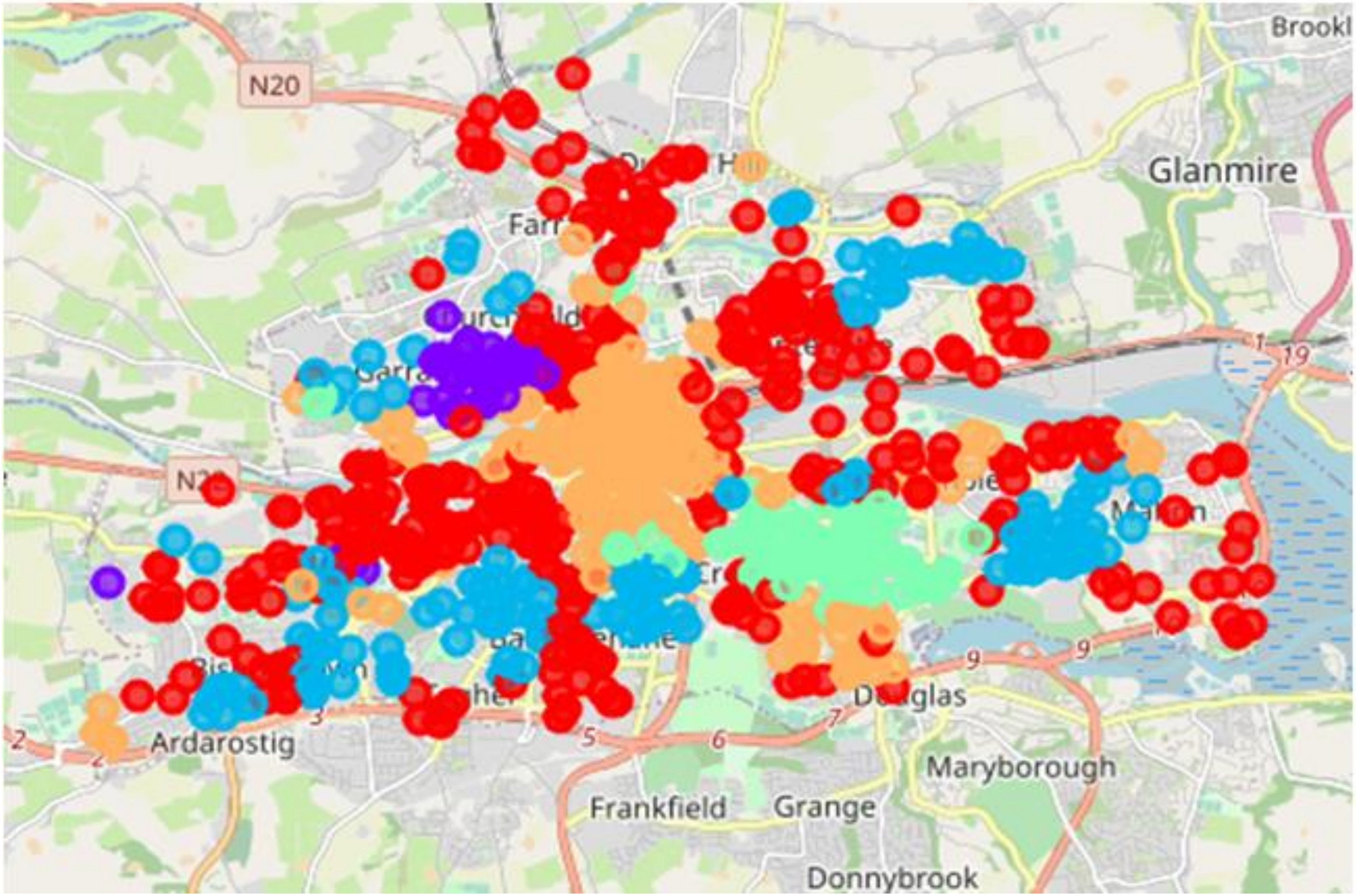
- The first step was to run the various applications through the Foursquare API to assemble a list of most common nearby venues. The radius was set to 500m as this is appropriately sized for a city of Cork's size. The limit was reduced to 20 for ease of computation.
- The initial API results (with one return per venue type) were first aggregated to application level with the most common venue for each application type being listed. Then these individual results were transposed onto a larger table with the application numbers serving as primary key:

Methodology

- From here we could use k means clustering to identify which neighbourhoods were the most similar to each other. We aimed for four clusters in this method and it returned the following table:

	_id	ApplicationNumber	Latitude	Longitude	OneOffHouseKPI	NumResidentialUnits	AreaOfSite	Decision	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
0	3	832737	51.875175	-8.496067	No	0	0.043	Refused	0.0	Convenience Store	Bar
1	4	832738	51.898020	-8.414907	No	0	0.074	Granted (Conditional)	0.0	Trail	Market
2	5	832739	51.883356	-8.482564	No	0	0.020	Granted (Conditional)	2.0	Plaza	Chinese Restaurant
3	6	832740	51.887688	-8.473576	Yes	1	0.042	Refused	3.0	Soccer Stadium	Park
4	7	832741	51.917226	-8.469818	No	0	0.850	Granted (Conditional)	0.0	Coffee Shop	Soba Restaurant

Map


Brookl
N20
Glanmire
N20
Ardarostig
Frankfield
Grange
Maryborough
Donnybrook
Douglas

Discussions

- The results show the following clusters:

1. City Centre
2. NW Inner City
3. SE Suburbia
4. City Centre Periphery
5. General City Area (poor cluster fit)

Conclusion

- This was clearly a success