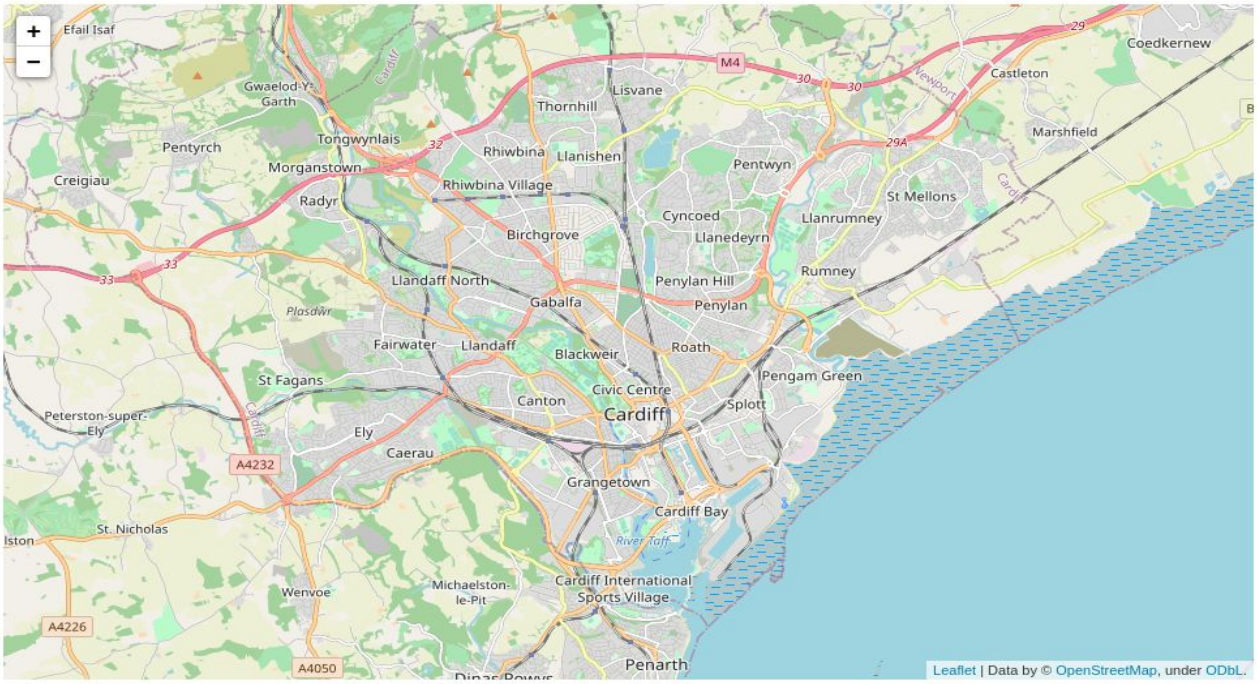
1) INTRODUCTION



1.1) Background

Cardiff, Capital of Wales, attracts crowds of tourists all year long. Along with being a vibrant city, it is full of youthful energy. Hence, Cardiff becomes an optimum choice for a retailer to start its business Wales.

1.2) Problem Statement

A vendor desires to set up a store in Cardiff. However, there are many constraints to this decision like an appropriate neighborhood, less competition in the vicinity. Moreover, the vendor demands to select the admired venue type, as all these factors will collectivity result in expansion and yield benefit to the business.

1.3) Scope

The aim of this project is to discover the most favored venue type around the town and using a clustering algorithm to find a desirable location that fits all the constraints.

2) Data Acquisition

Every JSON file used in this project is acquired from **FourSquare API**, under which lies the Places API which contains all the location data of venues and its attributes all across the world.

In this project we made 3 types of calls to the API namely search, explore, and trending. Among which the data obtained from search and explore calls contribute to our analysis.

3) Methodology

3.1) Phase 1 - Analysing venues around Cardiff

By applying the “explore” call to the API, we get the initial data that has all types of venues in Cardiff within the radius of 10 km.

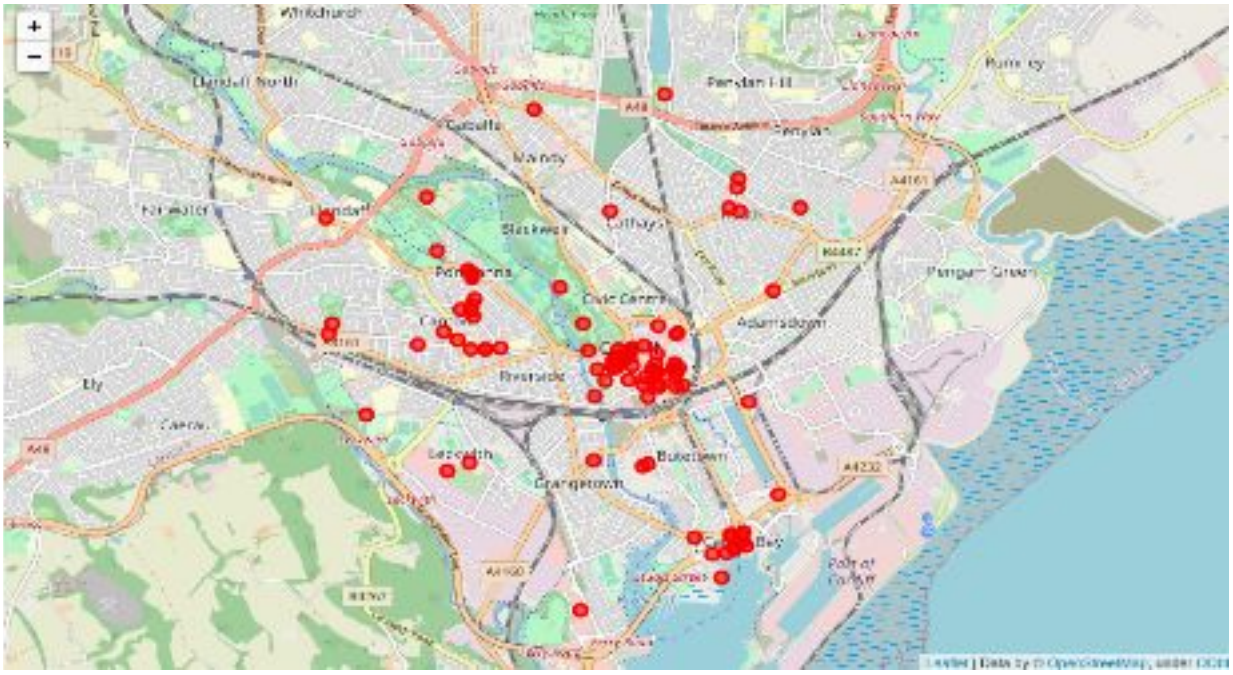


Fig 1: Different venues in Cardiff

Fig 1 depicts the mapping of venues, where we can see most of the stores are concentrated in the city center.

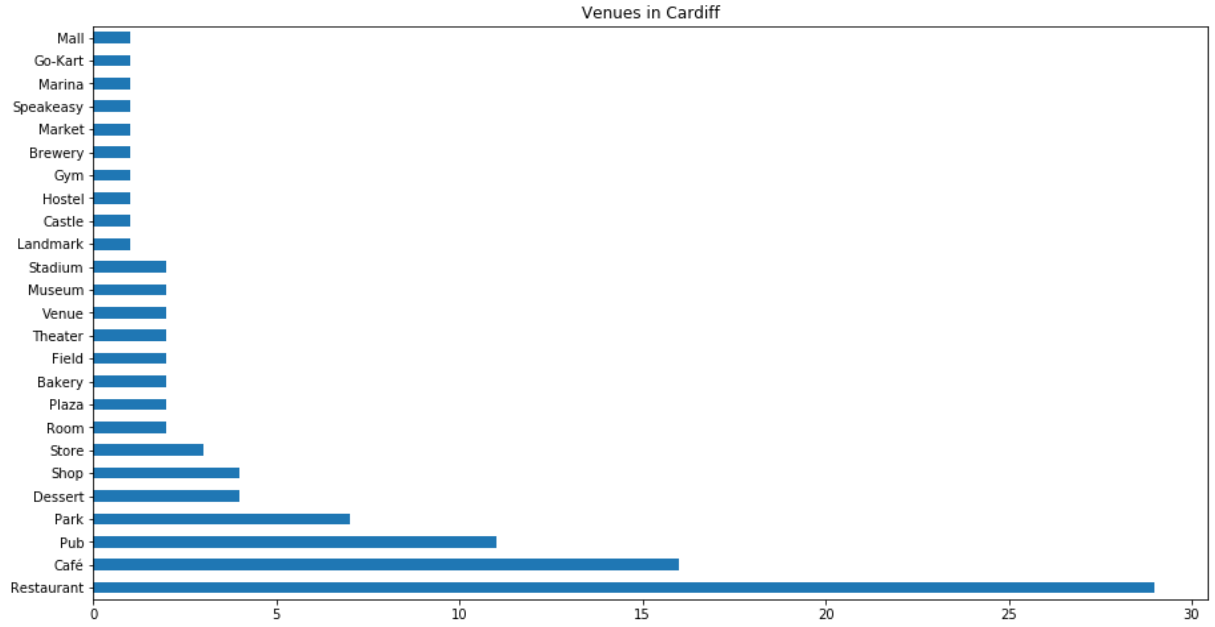


Fig 2: Frequencies of different venue category

By visualizing the category field in the data frame we see that the restaurant's rate is the highest. Therefore narrowing the study towards restaurants, further, plotting the restaurants on a Folium Map *(Fig 3)* which will visually guide us to a specific area to explore.

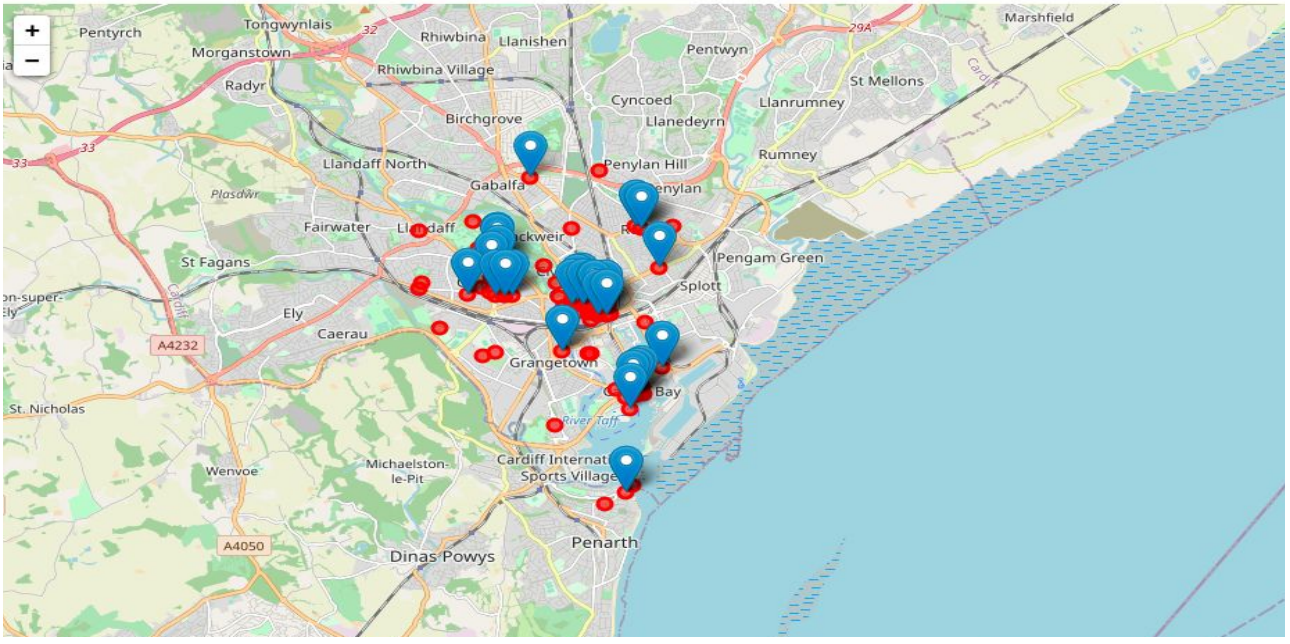


Fig 3: Map with venues (Red markers) & Restaurants (Blue Markers)

3.2) Phase 2 - Extracting the location data to work with

After analyzing the neighborhood of Cardiff, the conclusion is that the in-demand venue is a food joint or Cafe. In this phase, the goal is to get the basic neighborhood data. Since there was no Cardiff neighborhood data available, we will create our own by calling the search function into the Foursquare API and passing restaurants as a query *(Fig 4)*.

	ID	Name	Category	Address	Latitude	Longitude	Distance	Detailed Address
0	5079a1d0498ec6be8d9d88d9	Bill's Restaurant	English Restaurant	27-39 Wyndham Arcade	51.477874	-3.175822	481	[27-39 Wyndham Arcade (Mill Ln), Cardiff, CF10...
1	4bc6fffc2f94d13a8a10117f	Cardiff University Trevithick Restaurant	College Cafeteria	5 The Parade	51.484519	-3.170129	704	[5 The Parade (Queen's Buildings), Cardiff, CF...
2	4bb4b156613fb7130e0a94e6	Restaurant Minuet	Italian Restaurant	42 Castle Arcade	51.480996	-3.181087	150	[42 Castle Arcade, Cardiff, CF10 1BW, United K...
3	4c23a61ef1272d7fe9c481c5	Riverside Cantonese Restaurant	Chinese Restaurant	Leckwith Road	51.477772	-3.201222	1587	[Leckwith Road, Cardiff, CF11 6AH, United King...
4	5b54976383e380002c83f883	The Restaurant & Bar at Clayton Hotel Cardiff	Hotel Bar	St Mary St	51.476196	-3.176718	631	[St Mary St, Cardiff, CF10 1GD, United Kingdom]

Fig 4: A snippet of the location data

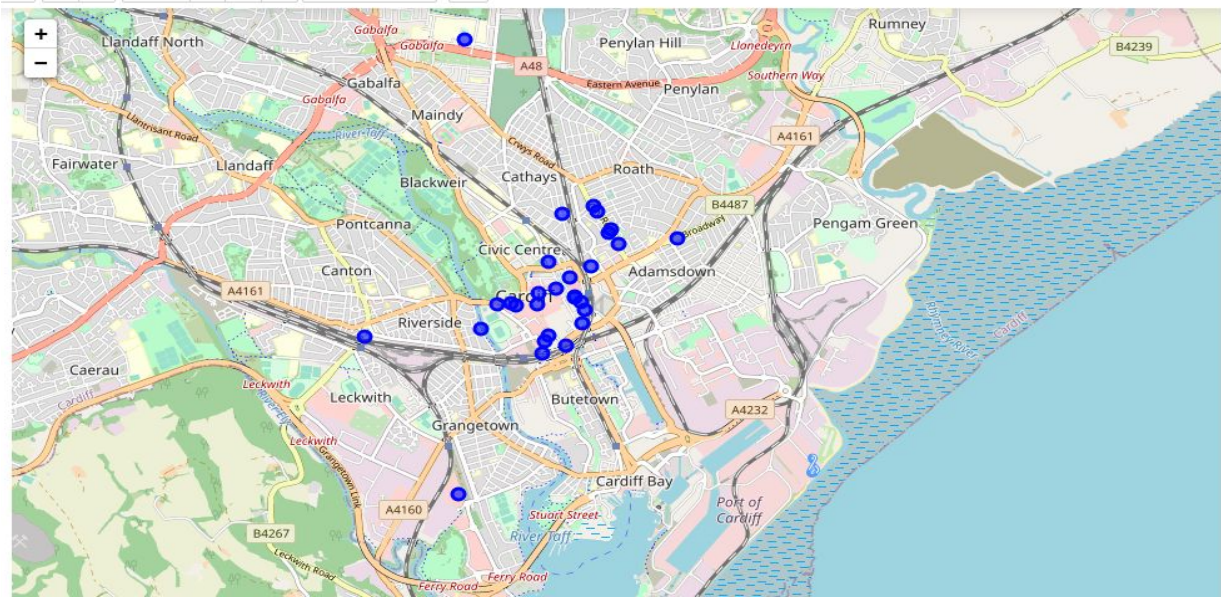


Fig 5: Plotting all the locations from the location data

3.3) Phase 3 - Getting nearby venues and choosing a venue category

Applying search calls on each venue from the location data to get nearby venues.

	R_name	R_Latitude	R_Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bill's Restaurant	51.477874	-3.175822	Uncommon Ground Coffee Co.	51.478297	-3.177394	Coffee Shop
1	Bill's Restaurant	51.477874	-3.175822	The Plan	51.478930	-3.177093	Café
2	Bill's Restaurant	51.477874	-3.175822	Wahaca	51.478059	-3.175173	Mexican Restaurant
3	Bill's Restaurant	51.477874	-3.175822	Miller & Carter	51.479687	-3.176286	Steakhouse
4	Bill's Restaurant	51.477874	-3.175822	St David's Dewi Sant	51.479429	-3.175565	Shopping Mall

Fig 6: Dataframe head with nearby venues around each location

Fig 6 is a glimpse of the data frame acquired where the venues nearby a particular location fall under "venue" and "Venue Category" is the type of venue. *Fig 7* depicts the visualization of category type which clearly states "Coffee shop" is prominent in Cardiff. Hence, it will be the user's preference.

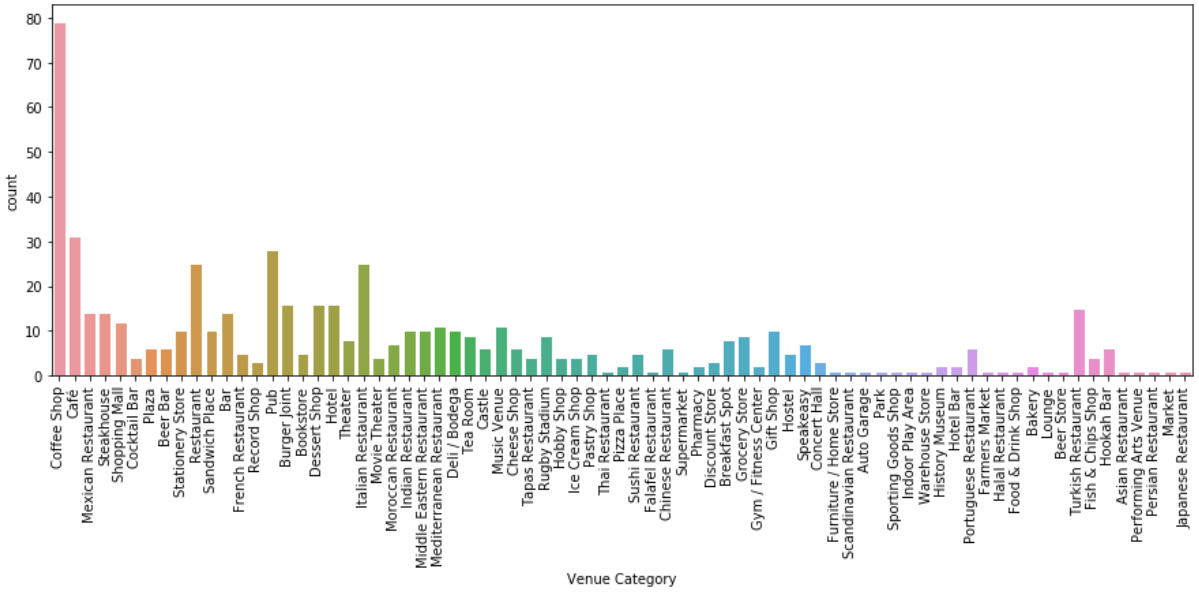


Fig 7: Frequency of Venue Category

3.4) Phase 4 - Algorithmic Modeling

3.4.1) Preparing the data

Before feeding the data we need to regroup it and adjust its values to a standard value. First, we need to change the Categorical data to Numerical Data which the model would understand. Here, we have used *GET_DUMMIES* function for the conversion of a particular field in the data frame of *Fig 6*.

	R_name	Asian Restaurant	Auto Garage	Bakery	Bar	Beer Bar	Beer Store	Bookstore	Breakfast Spot	Burger Joint	...	Stationery Store	Steakhouse	Supermarket	Sushi Restaurant	Taj Restaur
0	Bill's Restaurant	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
1	Bill's Restaurant	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
2	Bill's Restaurant	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
3	Bill's Restaurant	0	0	0	0	0	0	0	0	0	...	0	1	0	0	
4	Bill's Restaurant	0	0	0	0	0	0	0	0	0	...	0	0	0	0	

5 rows × 76 columns

Fig 9: Conversion of Categorical to Numerical on Venue Category

The Dataframe in *Fig 9* has the count of each venue near a restaurant. Since the data is disarranged, we group it by the column "R_name" and apply the mean function to get the normalized data *(Fig 10)*

	R_name	Asian Restaurant	Auto Garage	Bakery	Bar	Beer Bar	Beer Store	Bookstore	Breakfast Spot	Burger Joint	...	Stationery Store	Steakhouse	Supermarket	Sushi Restaurant	Ta Restau
0	Al-Rayan Restaurant	0.0	0.0	0.0	0.00	0.00	0.0	0.00	0.00	0.00	...	0.00	0.05	0.0	0.0	(
1	Alwaly Omani Restaurant	0.0	0.0	0.0	0.00	0.00	0.0	0.00	0.05	0.00	...	0.00	0.00	0.0	0.0	(
2	Bill's Restaurant	0.0	0.0	0.0	0.05	0.05	0.0	0.05	0.00	0.05	...	0.05	0.05	0.0	0.0	(
3	Cardiff University Trevithick Restaurant	0.0	0.0	0.0	0.05	0.00	0.0	0.00	0.00	0.00	...	0.00	0.00	0.0	0.0	(
4	Castell's Restaurant	0.0	0.0	0.0	0.00	0.00	0.0	0.00	0.00	0.00	...	0.00	0.00	0.0	0.0	(

5 rows × 76 columns

Fig 10: Regrouped & Normalised data

3.4.2) Feeding the Data into the model

We have used the K-means Clustering algorithm, this will cluster the locations on the basis of venue count and the number of clusters is kept at 5. After fitting the data, we have to evaluate each cluster and find which fits our constraints.

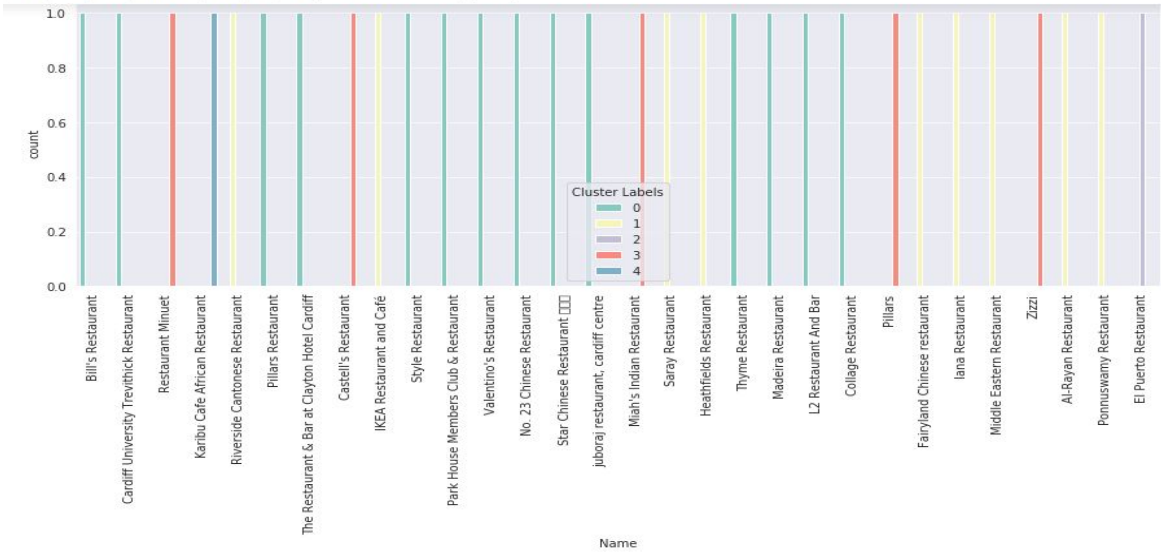


Fig 11: Locations belonging to different clusters

3.5) Analysing the clusters

3.5.1) Cluster 1

Cluster 1 consists of majorly coffee shops as the most common venue near a particular location and has 14 entries. Therefore, Cluster1 won't be of much use to us.

	Name	Latitude	Longitude	Distance	Detailed Address	Cluster Labels	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
0	Bill's Restaurant	51.477874	-3.175822	481	[27-39 Wyndham Arcade (Mill Ln), Cardiff, CF10...	0	Coffee Shop	Café	Plaza	Pub	Restaurant	Sandwich Place	Burger Joint
1	Cardiff University Trevithick Restaurant	51.484519	-3.170129	704	[5 The Parade (Queen's Buildings), Cardiff, CF...	0	Coffee Shop	Hotel	Pub	Italian Restaurant	Dessert Shop	Moroccan Restaurant	Theater
5	Pillars Restaurant	51.480870	-3.177420	150	[United Kingdom]	0	Coffee Shop	Restaurant	Café	Deli / Bodega	Shopping Mall	Cheese Shop	Gift Shop
6	The Restaurant & Bar at Clayton Hotel Cardiff	51.476196	-3.176718	631	[St Mary St, Cardiff, CF10 1GD, United Kingdom]	0	Coffee Shop	Plaza	Shopping Mall	Cocktail Bar	Mexican Restaurant	Café	Pub
9	Style Restaurant	51.482369	-3.174849	311	[United Kingdom]	0	Coffee Shop	Gift Shop	Italian Restaurant	Deli / Bodega	Concert Hall	Café	Restaurant
10	Park House Members Club & Restaurant	51.484933	-3.175978	427	[20 Park Place, Cardiff, CF10 3DQ, United King...	0	Coffee Shop	Hotel	Pub	Dessert Shop	Theater	Hotel Bar	History Museum
11	Valentino's Restaurant	51.483365	-3.173055	466	[5 Windsor Place, Cardiff, CF10 3BD, United Ki...	0	Coffee Shop	Hotel	Dessert Shop	Italian Restaurant	Café	Burger Joint	Pub
12	No. 23 Chinese Restaurant	51.481632	-3.172262	480	[23 Churchill Way, Cardiff, C F10, United King...	0	Coffee Shop	Bar	Italian Restaurant	Gift Shop	Movie Theater	Portuguese Restaurant	Mediterranean Restaurant
	Star				[23 Churchill								

Fig 12: Cluster 1

3.5.2) Cluster 2

This cluster has potential locations although it would need some cleaning for optimality.

	Name	Latitude	Longitude	Distance	Detailed Address	Cluster Labels	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 M Comm Ver
4	Riverside Cantonese Restaurant	51.477772	-3.201222	1587	[Leckwith Road, Cardiff, CF11 6AH, United King...	1	Chinese Restaurant	Coffee Shop	Indian Restaurant	Pharmacy	Pizza Place	Falafel Restaurant	F
8	IKEA Restaurant and Café	51.462658	-3.188317	2207	[IKEA (Cardiff Bay Retail Park), Cardiff, CF11...	1	Discount Store	Warehouse Store	Boat or Ferry	Park	Scandinavian Restaurant	Furniture / Home Store	Sport Goods Sh
16	Saray Restaurant	51.490210	-3.169716	1157	[United Kingdom]	1	Turkish Restaurant	Middle Eastern Restaurant	Fish & Chips Shop	Coffee Shop	Chinese Restaurant	Mexican Restaurant	Mediterrane Restaur
17	Heathfields Restaurant	51.506220	-3.187372	2792	[University Hospital Wales (Heath Park), Cardi...	1	Grocery Store	Bakery	Food & Drink Shop	Lounge	Beer Store	Stationery Store	Gym / Fitn Cer
23	Fairyland Chinese restaurant	51.489556	-3.174021	949	[116 Salisbury Road, Cardiff, CF24 4AE, United...	1	Turkish Restaurant	Pub	Asian Restaurant	Café	Indian Restaurant	Mediterranean Restaurant	Mexi Restaur
24	Iana Restaurant	51.487663	-3.167638	1043	[United Kingdom]	1	Italian Restaurant	Coffee Shop	Turkish Restaurant	Sandwich Place	Mexican Restaurant	Mediterranean Restaurant	Moroc Restaur
25	Middle Eastern Restaurant	51.488002	-3.167371	1082	[United Kingdom]	1	Italian Restaurant	Turkish Restaurant	Café	Mexican Restaurant	Sandwich Place	Mediterranean Restaurant	F
					[42 City								

Fig 13: Cluster 2

3.5.3) Cluster 3

Although cluster 3 has only one entry, this particular cluster is the most optimum from the rest 4 as the first four common venues do not consist of coffee shop

	Name	Latitude	Longitude	Distance	Detailed Address	Cluster Labels	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
29	El Puerto Restaurant	51.445283	-3.16818	4120	[Cardiff Bay Barrage, Penarth, Vale of Glamorg...	2	Seafood Restaurant	Harbor / Marina	Trail	Athletics & Sports	Coffee Shop	Pub	Park	Steakhouse

Fig 14: Cluster 3

3.5.4) Cluster 4

Cluster 4 locations are surrounded by Deli, Pubs, Coffee shop. Although the frequency of coffee shops are not much in most common venues, but the constraint is no Cafes nearby.

	Name	Latitude	Longitude	Distance	Detailed Address	Cluster Labels	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
2	Restaurant Minuet	51.480996	-3.181087	150	[42 Castle Arcade, Cardiff, CF10 1BW, United K...	3	Deli / Bodega	Coffee Shop	Tea Room	Pub	Rugby Stadium	Ice Cream Shop	Café	Music Venue
7	Castell's Restaurant	51.480875	-3.182924	272	[Castle Street, Cardiff, CF10 1SZ, United King...	3	Pub	Tea Room	Rugby Stadium	Ice Cream Shop	Music Venue	Deli / Bodega	Coffee Shop	Hobby Shop
15	Miah's Indian Restaurant	51.478537	-3.185263	545	[Swansea, City and County of Swansea, United K...	3	Hostel	Pub	Rugby Stadium	Castle	Halal Restaurant	Italian Restaurant	Farmers Market	Music Venue
22	Pillars	51.481915	-3.177360	130	[United Kingdom]	3	Coffee Shop	Pub	Deli / Bodega	Gift Shop	Italian Restaurant	Castle	Café	Bookstore
26	Zizzi	51.480723	-3.180434	134	[27 High Street, Cardiff, CF10 1PU, United Kin...	3	Deli / Bodega	Coffee Shop	Pub	Rugby Stadium	Ice Cream Shop	Music Venue	Hobby Shop	Cheese Shop

Fig 15: Cluster 4

3.5.5) Cluster 5

The data inside might make this cluster as a potential cluster but the location itself is a cafe. Hence, making it not a viable option.

	Name	Latitude	Longitude	Distance	Detailed Address	Cluster Labels	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	9 C
3	Karibu Cafe African Restaurant	51.487146	-3.158276	1573	[12 Broadway, Cardiff, CF24 1NF, United Kingdom]	4	Pub	Persian Restaurant	Bakery	Bed & Breakfast	Grocery Store	Middle Eastern Restaurant	Warehouse Store	Gift Shop	

Fig 16: Cluster 5

4) Observations

- Cardiff is filled with Restaurants followed by coffee shops and pubs
- City Center is the hub of stores and also attracts locals as well as tourists
- Cluster no. 1 *(Fig 12)* which has the most locations is primarily clustered with respect to coffee shops

5) Result

The aim was to recommend a location and venue type for the user. After expanding the location data, we observed that the recurrent category type is a coffee shop. However, if we combine all the restaurant types, the count supersedes the tally of the coffee house, but the vendor seeks for a particular type. Therefore, the venue category chosen is a **Coffee Shop.**

For the location, after evaluating the clusters, the options narrowed down clusters 2&3 *(Fig 13&14)* respectively. After cleaning cluster 2 as it had some undesired value, the locations are visualized in *Fig. 17* and the locations are **Cathays, Heath Park, Grangetown, Penarth Marina.**

Cathays and Heath Park are closeby to Cardiff University and city center, also they have plenty of residential buildings with student residents, therefore making it an optimum location.

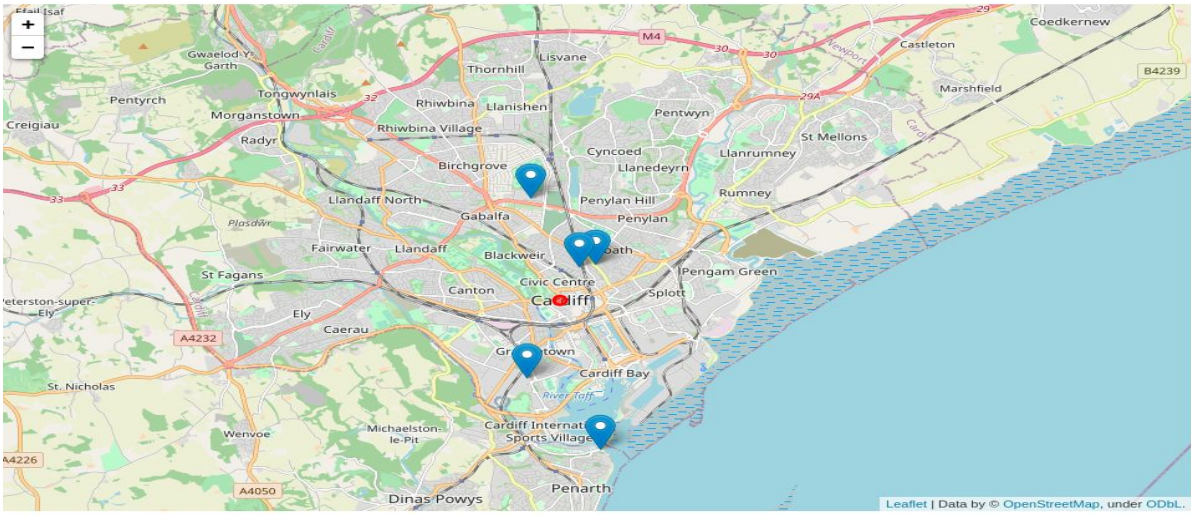


Fig 17: Desirable locations for coffee shop

	Name	Latitude	Longitude	Detailed Address
0	IKEA Restaurant and Café	51.462658	-3.188317	[IKEA (Cardiff Bay Retail Park), Cardiff, CF11...
1	Saray Restaurant	51.490210	-3.169716	[United Kingdom]
2	Heathfields Restaurant	51.506220	-3.187372	[University Hospital Wales (Heath Park), Cardi...
3	Fairyland Chinese restaurant	51.489556	-3.174021	[116 Salisbury Road, Cardiff, CF24 4AE, United...
4	El Puerto Restaurant	51.445283	-3.168180	[Cardiff Bay Barrage, Penarth, Vale of Glamorg...

Fig 18: Dataframe for the user