IBM Data Science Professional Certificate

Applied Data Science Capstone

Recommending Locations for Opening a New Restaurant at Co. Galway in the Republic of Ireland

Jayakarthi Boovendran

jaykarthik11@gmail.com | linkedin.com/in/jayakarthi-boovendran

Table of Contents

Jupyter Notebooks:

- https://github.com/Jay110796
- https://nbviewer.jupyter.org/github/Jay110796

I. Introduction

In the modern business environment, spatial analytics plays a critical role in improving business values by elevating location-based insights to business problems. In this study, we focus primarily on analyzing the geospatial data of the Republic of Ireland and recommend a set of locations that could be fitting for starting a restaurant business. For the same, we employ the Foursquare API to retrieve neighborhood information to analyze and group areas that are similar.

II. Background

The Republic of Ireland has 32 counties and four provinces. Since the businesses are already well-established in the provinces Leinster, Ulster, and Munster, they wanted to expand the company to a county in the Connacht region. For simplification, we chose Co. Galway from Connacht for our study. Co. Galway is ranked 5th in the overall population of Ireland and 1st in the Connacht province. Therefore, setting up the restaurant in Co. Galway will improve the business.

	County	Population	Density	Rank	Province
0	Dublin	12,73,069	1380.8	1	Leinster
1	Antrim	6,18,108	202.9	2	Ulster
2	Down	5,31,665	215.6	3	Ulster
3	Cork	5,19,032	69.0	4	Munster
4	Galway	2,50,541	40.7	5	Connacht
5	Londonderry	2,47,132	119.1	6	Ulster

Figure 1: Counties of Ireland

III. Problem Statement

To analyze the neighborhoods of Co. Galway based on the popular venues and recommend a set of locations that suit well for setting up a new restaurant.

IV. Target Audience:

The target audience may be a franchise like McDonald's looking to start a new restaurant in a new province.

V. Business Requirements

- The neighbourhood should be a busy area with plenty of foot traffic.
- The region should have enough bars and pubs nearby
- Should have relatively less number of restaurants opened in the same spot

VI. Data Understanding

a) Datasets Used

- **Dataset 1**: Ireland Counties Dataset | Available @ list-counties-ireland.csv

	County	Population	Density	Rank	Province
0	Dublin	12,73,069	1380.8	1	Leinster
1	Antrim	6,18,108	202.9	2	Ulster
2	Down	5,31,665	215.6	3	Ulster
3	Cork	5,19,032	69.0	4	Munster
4	Galway	2,50,541	40.7	5	Connacht
5	Londonderry	2,47,132	119.1	6	Ulster

Figure 2: Ireland Counties Data Sample

- **Dataset 2**: Neighborhoods in Co. Galway

  Since no public datasets are available for the neighborhood or town names in Co. Galway, we manually created a data frame with 67 neighborhoods. Given below is a snippet of the data frame.

	Neighborhood
0	Ahascragh
1	Ardrahan
2	Athenry
3	Ballinasloe
4	Ballinderreen
...	...
62	Tully
63	Tully Cross
64	Tuam
65	Turloughmore
66	Woodford

Figure 3: List of Neighborhoods in Co. Galway, Ireland

b) Data Pre-processing and Visualization

Initially, for improved understanding, we plot the 32 counties of Ireland on to a folium map using their latitude and longitude coordinates. Since the dataset has no spatial data, we make use of geopy python library to fetch the latitude and longitude coordinates using the name of the counties.

	County	Population	Density	Rank	Province	Latitude	Longitute
0	Dublin	12,73,069	1380.8	1	Leinster	53.349764	-6.260273
1	Antrim	6,18,108	202.9	2	Ulster	44.994737	-85.113319
2	Down	5,31,665	215.6	3	Ulster	53.597828	-7.351667
3	Cork	5,19,032	69.0	4	Munster	51.897928	-8.470581
4	Galway	2,50,541	40.7	5	Connacht	53.274412	-9.049063
5	Londonderry	2,47,132	119.1	6	Ulster	54.997868	-7.321306

Figure 4: Counties of Ireland with Latitude & Longitude coordinates

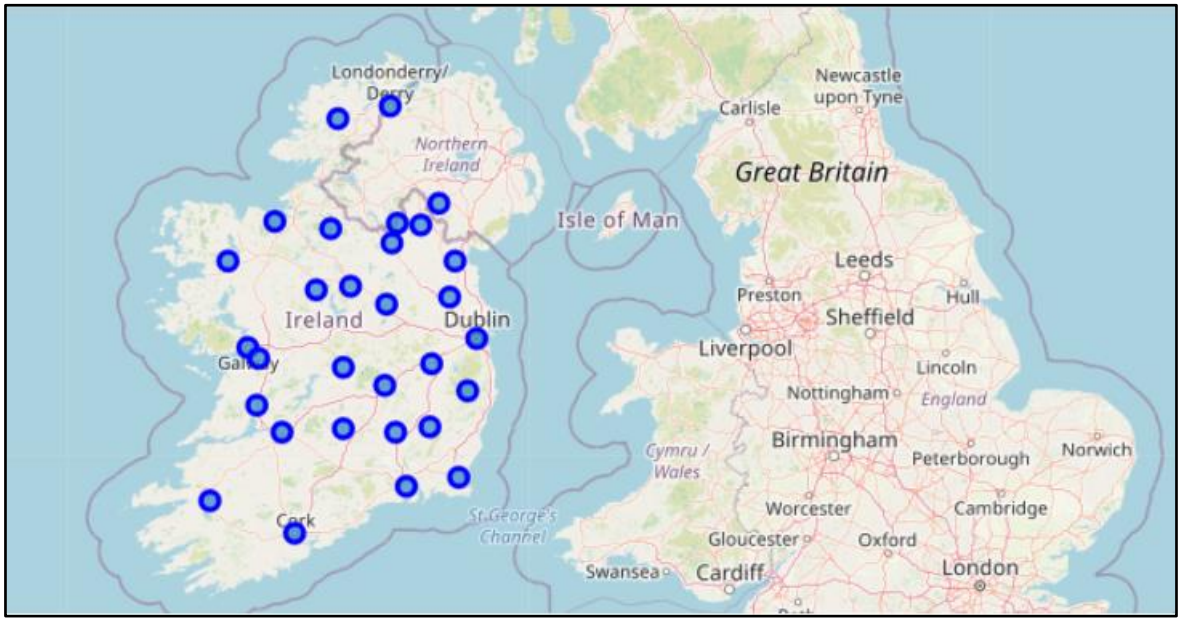

Figure 5: Counties of Ireland plotted on world map

Plotting Co. Galway

The primary objective of the project is to identify locations or neighborhoods in Co. Galway that might interest the stakeholders in starting a restaurant. Therefore, we filter only the coordinated of Co. Galway and plot it onto a folium map.

```
galway_lat=df_irish_counties.loc[df_irish_counties['County'] == 'Galway', 'Latitude'].iloc[0]
galway_lon=df_irish_counties.loc[df_irish_counties['County'] == 'Galway', 'Longitute'].iloc[0]
print('The geograpical coordinates of County Galway are {} and {}'.format(galway_lat,galway_lon))
```

```
The geograpical coordinates of County Galway are 53.2744122 and -9.0490632
```

Figure 6: Fetching only the coordinates of Co. Galway

Figure 7: Location of Co. Galway in the Republic of Ireland

Fetching Neighborhoods in Co. Galway

As the next step, we fetch the latitude and longitude coordinates of the cities in Co. Galway using the geopy library. It is to be noted that the neighborhoods or towns in Co. Galway is scraped manually from the web. Finally, we plot the regions onto a world map.

	Neighborhood	Latitude	Longitude
0	Ahascragh	53.397315	-8.332594
1	Ardrahan	53.156750	-8.805991
2	Athenry	53.299190	-8.746830
3	Ballinasloe	53.330272	-8.221376
4	Ballinderreen	53.186252	-8.908921
...	...	...	...
62	Tully	53.595796	-9.975448
63	Tully Cross	53.588201	-9.962930
64	Tuam	53.514792	-8.851287
65	Turloughmore	53.380699	-8.856270
66	Woodford	53.052181	-8.399454

Figure 8: Latitude & Longitude coordinates of Neighborhoods of Co. Galway

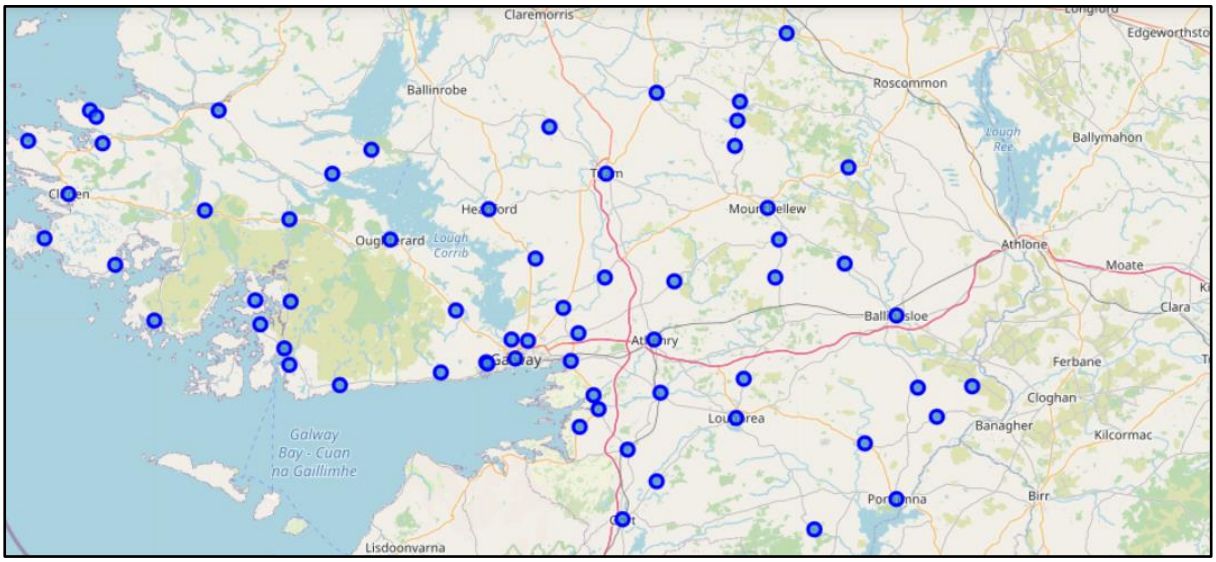

Figure 9: Neighborhoods of Co. Galway

Fetching Venues using Foursquare API

Using Foursquare API, we identify the venues at each of the neighborhood in Co. Galway. Further, we create data frame with information on the venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ahascragh	53.397315	-8.332594	Cahill's	53.397293	-8.335465	Irish Pub
1	Ardrahan	53.156750	-8.805991	Ardrahan Bus Stop	53.156631	-8.806193	Bus Stop
2	Ardrahan	53.156750	-8.805991	Joyces	53.157442	-8.806357	Gastropub
3	Ardrahan	53.156750	-8.805991	Tarpeys	53.160487	-8.806054	Bar
4	Athenry	53.299190	-8.746830	Athenry Castle	53.299940	-8.745710	Castle

Figure 10: Venues fetched using Foursquare API

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ahascragh	1	1	1	1	1	1
1	Ardrahan	3	3	3	3	3	3
2	Athenry	4	4	4	4	4	4
3	Ballinasloe	5	5	5	5	5	5
4	Ballinderreen	2	2	2	2	2	2
5	Ballyconneely	2	2	2	2	2	2
6	Ballygar	3	3	3	3	3	3
7	Barna	4	4	4	4	4	4
8	Carna	2	2	2	2	2	2
9	Casla	3	3	3	3	3	3
10	Castlegar	1	1	1	1	1	1

Figure 11: Venue Info grouped by Neighborhood

Assumptions:

- Venue Category (**Figure 11**) is directly proportional to Foot traffic
- **Positive Venues Categories**: Bars and Pubs (desired venues)
- **Negative Venues Categories**: Restaurant, Fast Food Restaurant, Seafood Restaurant, Vegan Restaurant, Middle Eastern Restaurant, Chinese Restaurant, Mexican Restaurant, Indian Restaurant, Asian Restaurant, French Restaurant, Japanese Restaurant and Italian Restaurant (the location should not have many restaurants)

Explanation

In the following sections, we will cluster neighborhoods based on the count of Venue Category (foot traffic), number of positive venues like bar and pubs, also on the number of negative venues like other restaurants. Finally, we compute an overall score based on the cluster values and categorize locations into low-value, mid-value and high-value neighborhoods.

In simple terms, we need a location that is busy, and has venues like bar and pubs. However, the number of restaurants in the same point should be less.

VII. Methodology

a) Clustering based on Foot Traffic

A neighborhood is assumed to have higher foot traffic if there are a higher number of venues located in that area. Therefore, we cluster neighborhoods based on the count of places under each city.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ahascragh	1	1	1	1	1	1
1	Ardrahan	3	3	3	3	3	3
2	Athenry	4	4	4	4	4	4
3	Ballinasloe	5	5	5	5	5	5
4	Ballinderreen	2	2	2	2	2	2
5	Ballyconneely	2	2	2	2	2	2

Figure 12: Count of Venues under each Neighborhood

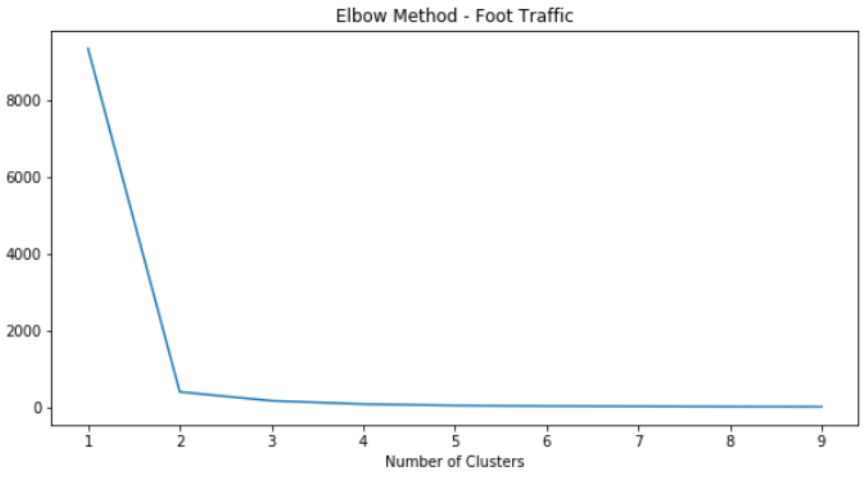

Figure 13: Elbow method – Optimal value for K is 4

Venue Traffic Cluster	count	mean	std	min	25%	50%	75%	max
0	17.0	1.882353	0.781213	1.0	1.00	2.0	2.00	3.0
1	22.0	5.227273	1.377777	4.0	4.00	5.0	6.00	9.0
2	2.0	14.500000	3.535534	12.0	13.25	14.5	15.75	17.0
3	1.0	100.000000	NaN	100.0	100.00	100.0	100.00	100.0

Figure 14: Cluster Description

From the elbow method, we identified the optimal value for k as 4. Performing the k-means clustering algorithm clusters the data into four groups. These clusters are then arranged based on their mean values. From **Figure 14** it is evident that the **Cluster 3** is the most desirable as neighborhoods under this cluster has an average of 100 venues, whereas, Cluster 1 is the least desirable cluster with locations having almost two venues. The Final output is given in Figure 15.

	Neighborhood	Latitude	Longitude	Venue Traffic Cluster
0	Ahascragh	53.397315	-8.332594	0
1	Ardrahan	53.156750	-8.805991	0
2	Athenry	53.299190	-8.746830	1
3	Ballinasloe	53.330272	-8.221376	1
4	Ballinderreen	53.186252	-8.908921	0
5	Ballyconneely	53.431491	-10.074751	0
6	Ballygar	53.522323	-8.326138	0
7	Barna	53.270434	-9.112482	1
8	Carna	53.324842	-9.835653	0
9	Casla	53.288519	-9.552989	0

Figure 15: Cluster Output

b) Clustering Based on Positive and Negative Venues

As the next step, we categorize neighborhoods based on their locations. From the Business Requirements, it is clear that the neighborhoods should have pubs and bars and less number of restaurants. Therefore, we consider pubs and bars as Positive Venues and Restaurant categories as Negative Venues. Given in Figure 17 is the count of Positive and Negative venues of each neighborhood.

```
array(['Irish Pub', 'Bus Stop', 'Gastropub', 'Bar', 'Castle', 'Café',
       'Train Station', 'Restaurant', 'Hotel', 'Fast Food Restaurant',
       'Pub', 'Auto Workshop', 'Beach', 'Pharmacy', 'Grocery Store',
       'Gym', 'Home Service', 'Business Service', 'Food Service',
       'Waste Facility', 'Convenience Store',
       'Financial or Legal Service', 'Shopping Mall',
       'Gym / Fitness Center', 'Tea Room', 'Steakhouse',
       'Construction & Landscaping', 'Boat or Ferry', 'Bakery',
       'Supermarket', 'Breakfast Spot', 'Harbor / Marina',
       'Bed & Breakfast', 'Trail', 'History Museum', 'Rest Area',
       'RV Park', 'Other Great Outdoors', 'Ice Cream Shop', 'Pizza Place',
       'Department Store', 'Plaza', 'Clothing Store', 'Food & Drink Shop',
       'Asian Restaurant', 'French Restaurant', 'Japanese Restaurant',
       'Bistro', 'Hostel', 'Coffee Shop', 'Italian Restaurant',
       'Bookstore', 'Sculpture Garden', 'Pie Shop', 'Spanish Restaurant',
       'Flea Market', 'Cheese Shop', 'Theater', 'Outdoor Sculpture',
       'Mexican Restaurant', 'Indian Restaurant',
       'Middle Eastern Restaurant', 'Chinese Restaurant', 'Nightclub',
       'Molecular Gastronomy Restaurant', 'Hotel Bar', 'Soccer Field',
       'Vegetarian / Vegan Restaurant', 'Pet Store', 'Diner',
       'Souvenir Shop', 'National Park', 'Gift Shop',
       'Tourist Information Center', 'Playground', 'Boxing Gym', 'Park',
       'Health & Beauty Service', 'Jewelry Store', 'Seafood Restaurant',
       'Farmers Market', 'Sandwich Place', 'Gas Station'], dtype=object)
```

Figure 16: Unique Venue Categories in Co. Galway

	Neighborhood	Latitude	Longitude	Negative Venues	Positive Venues
0	Ahascragh	53.397315	-8.332594	0	0
1	Ardrahan	53.156750	-8.805991	0	1
2	Athenry	53.299190	-8.746830	0	1
3	Ballinasloe	53.330272	-8.221376	2	0
4	Ballinderreen	53.186252	-8.908921	0	0
5	Ballyconneely	53.431491	-10.074751	0	1
6	Ballygar	53.522323	-8.326138	1	0
7	Barna	53.270434	-9.112482	0	0
8	Carna	53.324842	-9.835653	0	0
9	Casla	53.288519	-9.552989	0	0
10	Castlegar	53.298620	-9.022817	0	0

Figure 17: Number of Positive and Negative Venues in each location

Negative Venues Cluster	count	mean	std	min	25%	50%	75%	max
0	1.0	22.000000	NaN	22.0	22.0	22.0	22.0	22.0
1	7.0	2.428571	0.534522	2.0	2.0	2.0	3.0	3.0
2	5.0	1.000000	0.000000	1.0	1.0	1.0	1.0	1.0
3	29.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0

Figure 18: Negative Venue Clusters

From the clustering output, it is evident that **Cluster 3** is most desirable as it comprises of locations that have no restaurants (mean=0), whereas, Cluster 1 is not profitable as it includes sites that have nearly 22 restaurants on an average. Therefore, our target neighborhoods are in Cluster 3. Given in Figure 19 is the output after clustering.

	Neighborhood	Latitude	Longitude	Venue Traffic Cluster	Negative Venues Cluster
0	Ahascragh	53.397315	-8.332594	0	3
1	Ardrahan	53.156750	-8.805991	0	3
2	Athenry	53.299190	-8.746830	1	3
3	Ballinasloe	53.330272	-8.221376	1	3
4	Ballinderreen	53.186252	-8.908921	0	3
5	Ballyconneely	53.431491	-10.074751	0	3
6	Ballygar	53.522323	-8.326138	0	3
7	Barna	53.270434	-9.112482	1	3
8	Carna	53.324842	-9.835653	0	3
9	Casla	53.288519	-9.552989	0	3

Figure 19: Clustering Output

Positive Venues Cluster	count	mean	std	min	25%	50%	75%	max
0	31.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	7.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0
2	2.0	2.0	0.0	2.0	2.0	2.0	2.0	2.0
3	2.0	3.0	0.0	3.0	3.0	3.0	3.0	3.0

Figure 20: Positive Venue Clusters

Similarly, we cluster neighborhoods based on the number of favorable venues they have. Also, from Figure 20 it is evident that **Cluster 3** is most desirable as it comprises of locations that have the maximum number of favorable venues like pubs and bars, whereas, Cluster 0 has locations with no promising sites (mean=0)

	Neighborhood	Latitude	Longitude	Venue Traffic Cluster	Negative Venues Cluster	Positive Venues Cluster
0	Ahascragh	53.397315	-8.332594	0	3	0
1	Ardrahan	53.156750	-8.805991	0	3	0
2	Athenry	53.299190	-8.746830	1	3	0
3	Ballinasloe	53.330272	-8.221376	1	3	0
4	Ballinderreen	53.186252	-8.908921	0	3	0
5	Ballyconneely	53.431491	-10.074751	0	3	0
6	Ballygar	53.522323	-8.326138	0	3	0
7	Barna	53.270434	-9.112482	1	3	0
8	Carna	53.324842	-9.835653	0	3	0
9	Casla	53.288519	-9.552989	0	3	0
10	Castlegar	53.298620	-9.022817	0	3	0

Figure 21: Clustering Output

c) Overall Scoring

In the above sections, we have arranged the clusters in such a way that the most desirable group has the largest cluster label. In our case, it is **Cluster 3**.

As a next step, we compute an overall score by adding the cluster labels of Venue Traffic Cluster, Positive Venues Cluster and Negative Venues Cluster. We then segment them into low-value, mid-value and high-value neighborhoods. For instance, if a neighborhood is in Cluster 3 for all the clustering, the overall score will be 3+3+3 = 9. Therefore, the location is most suited for starting a restaurant. In our case, we obtained a maximum overall score of 6 (Figure 22).

Overall Score	Venue Traffic Cluster	Positive Venues Cluster	Negative Venues Cluster
2	0.500000	0.500000	1.000000
3	0.176471	0.176471	2.647059
4	1.062500	0.250000	2.687500
5	1.000000	1.666667	2.333333
6	2.000000	1.500000	2.500000

Figure 22: Overall Score Info

	Neighborhood	Latitude	Longitude	Venue Traffic Cluster	Negative Venues Cluster	Positive Venues Cluster	Overall Score
0	Ahascragh	53.397315	-8.332594	0	3	0	3
1	Ardrahan	53.156750	-8.805991	0	3	0	3
2	Athenry	53.299190	-8.746830	1	3	0	4
3	Ballinasloe	53.330272	-8.221376	1	3	0	4
4	Ballinderreen	53.186252	-8.908921	0	3	0	3
5	Ballyconneely	53.431491	-10.074751	0	3	0	3
6	Ballygar	53.522323	-8.326138	0	3	0	3
7	Barna	53.270434	-9.112482	1	3	0	4
8	Carna	53.324842	-9.835653	0	3	0	3
9	Casla	53.288519	-9.552989	0	3	0	3

Figure 23: Output after computing Overall Scores

(Overall score for record highlighted in Orange 1+3+0 = 4)

d) Segmentation

Based on the Overall Score, we segment neighborhoods into Low-value, Mid-value and High-value locations. Finally, we plot the locations based on their segments.

```
cluster_data['Segment'] = 'Low-Value'
cluster_data.loc[cluster_data['Overall Score']>3,'Segment'] = 'Mid-Value'
cluster_data.loc[cluster_data['Overall Score']>4,'Segment'] = 'High-Value'
cluster_data
```

	Neighborhood	Latitude	Longitude	Venue Traffic Cluster	Negative Venues Cluster	Positive Venues Cluster	Overall Score	Segment
0	Ahascragh	53.397315	-8.332594	0	3	0	3	Low-Value
1	Ardrahan	53.156750	-8.805991	0	3	0	3	Low-Value
2	Athenry	53.299190	-8.746830	1	3	0	4	Mid-Value
3	Ballinasloe	53.330272	-8.221376	1	3	0	4	Mid-Value
4	Ballinderreen	53.186252	-8.908921	0	3	0	3	Low-Value
5	Ballyconneely	53.431491	-10.074751	0	3	0	3	Low-Value
6	Ballygar	53.522323	-8.326138	0	3	0	3	Low-Value
7	Barna	53.270434	-9.112482	1	3	0	4	Mid-Value
8	Carna	53.324842	-9.835653	0	3	0	3	Low-Value
9	Casla	53.288519	-9.552989	0	3	0	3	Low-Value
10	Castlegar	53.298620	-9.022817	0	3	0	3	Low-Value
11	Claregalway	53.340482	-8.946239	1	3	0	4	Mid-Value
12	Clarinbridge	53.228065	-8.880166	1	3	0	4	Mid-Value
13	Cleggan	53.556459	-10.110690	1	3	0	4	Mid-Value
14	Clifden	53.488492	-10.021067	2	3	0	5	High-Value
15	Clonbur	53.544594	-9.363610	1	3	0	4	Mid-Value

Figure 24: Output after Segmentation

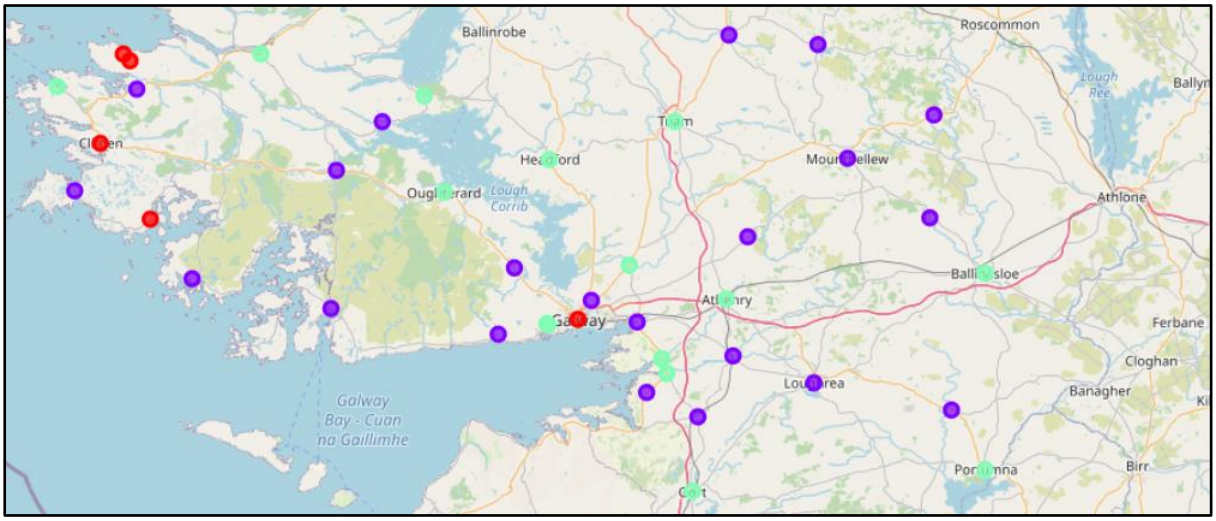

Figure 25: Cluster Output (**Red**: High-value, **Green**: Mid-value, **Purple**: Low-value)

VIII. Results

Finally, we obtained the potential locations for setting up a new restaurant (marked as red in Figure 25). Further, we examine only the high-value neighborhoods to get a clear idea of the top venues.

The Potential Neighborhoods for setting up a Restaurant are:

1. Clifden
2. Galway City
3. Roundstone
4. Tully
5. Tully Cross

	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
14	Clifden	Bar	Pub	Restaurant	Café	Breakfast Spot	Fast Food Restaurant	Bakery	Harbor / Marina	Bed & Breakfast	Supermarket
21	Galway	Pub	Coffee Shop	Hotel	Café	Restaurant	Hostel	Italian Restaurant	Pizza Place	Bakery	Bookstore
38	Roundstone	Pub	Seafood Restaurant	Waste Facility	Construction & Landscaping	Department Store	Diner	Farmers Market	Fast Food Restaurant	Financial or Legal Service	Flea Market
40	Tully	Farmers Market	Grocery Store	Waste Facility	French Restaurant	Diner	Fast Food Restaurant	Financial or Legal Service	Flea Market	Food & Drink Shop	Food Service
41	Tully Cross	Hotel	Pub	Bar	Bed & Breakfast	Gas Station	Farmers Market	Fast Food Restaurant	Financial or Legal Service	Flea Market	Food & Drink Shop

Figure 26: Top Venues in the Target Locations

IX. Discussion

In addition to the above methodology, the Foursquare API can be employed to further understand the top locations of the neighborhoods mentioned above and to drill down to one final location from the filtered set 5 neighborhoods.

X. Conclusion

In real-time, the Foursquare API can be used to leverage the applications of location data for improving business or for promoting new opportunities like the one in this study.