

# **Predicting to open a Thai restaurant in the town lands in Dublin county of Ireland.**

**Mahesh Matele**

**26th June 2020**

## **1.Introduction**

### **1.1 Background**

Dublin is a city of Ireland with the maximum population. Below is a list of town lands which gives a general idea about the respective town lands in Ireland spread out across the dublin county.

[https://en.wikipedia.org/wiki/Category:Townlands\\_of\\_County\\_Dublin](https://en.wikipedia.org/wiki/Category:Townlands_of_County_Dublin)

These town lands have a lot of entertainment places which generate an immense amount of business based on its success.

One of the major leisure activities for the people of Dublin is visiting an eatery. Out of this exotic dishes are very much in demand.

As part of this research we are concentrating on the below town lands of dublin county as mentioned in the above link namely :

**Athgoe,Ballsbridge,Ballygall,Beggars Bush,Clonskeagh,Merrion Gates, Poppintree, Priorswood, Ringsend, Roebuck, Sandymount.**

The subject to open a thai restaurant was taken on the basis of, after looking at the different kinds of venues available in dublin town lands. We can experiment the same solution with any other venue category. For this analysis stint we are concentrating on thai restaurants.

## **1.2 Problem :**

To open a restaurant we should be having a good idea about the categories of the venues available for food in these town lands.

Opening a thai restaurant in a populous place where there are already other thai restaurant is only going to divide the profit and increase the competition, consequently there will be an addition of monetary investment required to up the competition.

Therefore, we will have to find an appropriate place for opening a Thai restaurant.

## **1.3 Interest :**

The research will be of interest to stakeholders who would like to open any kind of venue in a given category of restaurants.

## **2. Data acquisition and cleaning**

### **2.1 Data sources**

- **Wikimedia** to get the base of town lands used for analysis

Below is the data source used for fetching the information about the town lands of Dublin county.

[https://en.wikipedia.org/wiki/Category:Townlands\\_of\\_County\\_Dublin](https://en.wikipedia.org/wiki/Category:Townlands_of_County_Dublin)

- **Foursquare API** used to :
  - fetch the venues
  - explore the areas to fetch the venues in the surrounding areas
  - get the list of Thai restaurants in the dublin county town lands
- **Geopy** - For getting the co-ordinated of different locations.

### **2.2 Data cleaning**

Data is downloaded or scraped from multiple sources including :  
Geopy ,Foursquare API ,wikimedia.

As the data was scrapped from an API the data which was received was in a json format and not a simple flat file data like a csv file.

This data is updated in a while hence the data received may differ from time and every run of code may generate different results for prediction.

There were some issues with the data that was fetched from wikimedia. Some of the townlands that were received had the county name associated with it, whereas the others did not have it. The pictorial representation of the same is as below.

Out[559]:

townlands	
0	Athgoe
1	Ballsbridge
2	Ballygall
3	Beggars Bush, Dublin
4	Clonskeagh
5	Merrion Gates
6	Poppintree
7	Priorswood
8	Ringsend
9	Roebuck, Dublin
10	Sandymount

This was cleansed to get it back into a standard format where the county name was clipped as below :

Out[560]:

townlands	
0	Athgoe
1	Ballsbridge
2	Ballygall
3	Beggars Bush
4	Clonskeagh
5	Merrion Gates
6	Poppintree
7	Priorswood
8	Ringsend
9	Roebuck
10	Sandymount

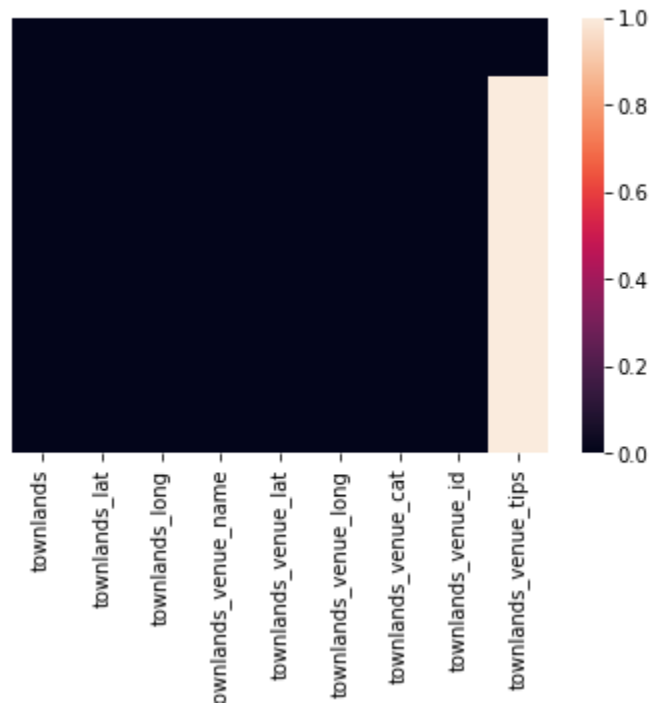
After the data is sourced from the above sources it is enriched every step with co ordinates and other attributes regarding its venues.

Out[561]:

	townlands	townlands_lat	townlands_long
0	Athgoe	53.280190	-6.520800
1	Ballsbridge	53.331060	-6.234010
2	Ballygall	53.386643	-6.291895
3	Beggars Bush	53.336496	-6.235518
4	Clonskeagh	53.305630	-6.236670
5	Merrion Gates	53.315579	-6.204262
6	Poppintree	53.399555	-6.276376
7	Priorswood	53.399357	-6.196738
8	Ringsend	53.340410	-6.232040
9	Roebuck	53.299291	-6.219382
10	Sandymount	53.328690	-6.208870

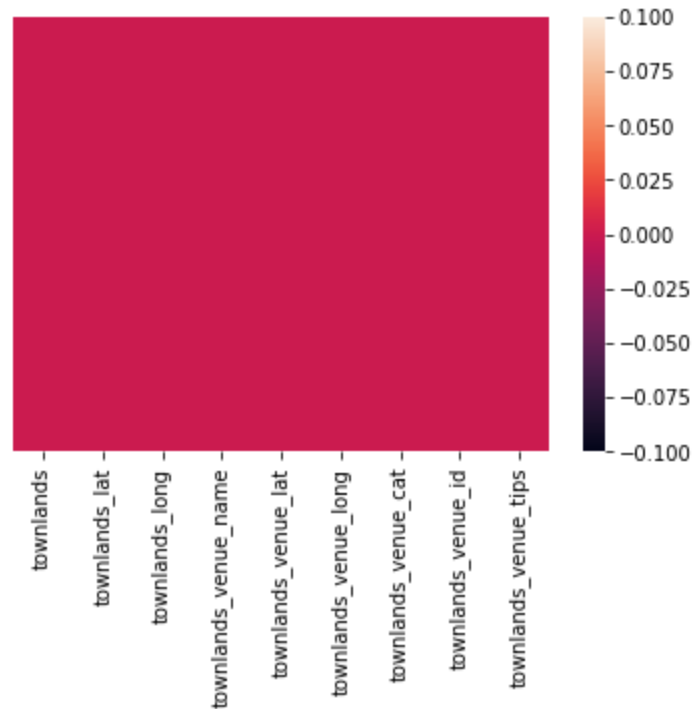
It is checked for data quality using the null check and the same is replaced with mean of the columns, so that there no major standard deviation.

```
Out[581]: <matplotlib.axes._subplots.AxesSubplot at 0x1baad128ec8>
```



The null values are removed by replacing them with the mean of that column for reducing the standard deviation.

```
Out[585]: <matplotlib.axes._subplots.AxesSubplot at 0x1baab3fc3c8>
```



From the many ways that the data is sliced for verification , all the categories are replaced with onehot functionality to convert them into numbers as below :

```
Out[589]:
```

	onehot_townlands	Bar	Burger Joint	Café	Coffee Shop	Convenience Store	Fast Food Restaurant	Gastropub	Grocery Store	Gym	Hotel	Indian Restaurant	Italian Restaurant	Park	Pizza Place	Plaza
0	Athgoe	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Ballsbridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Ballsbridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Ballsbridge	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
4	Ballsbridge	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
5	Ballsbridge	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
6	Ballsbridge	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
7	Ballsbridge	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
8	Ballsbridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Ballsbridge	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

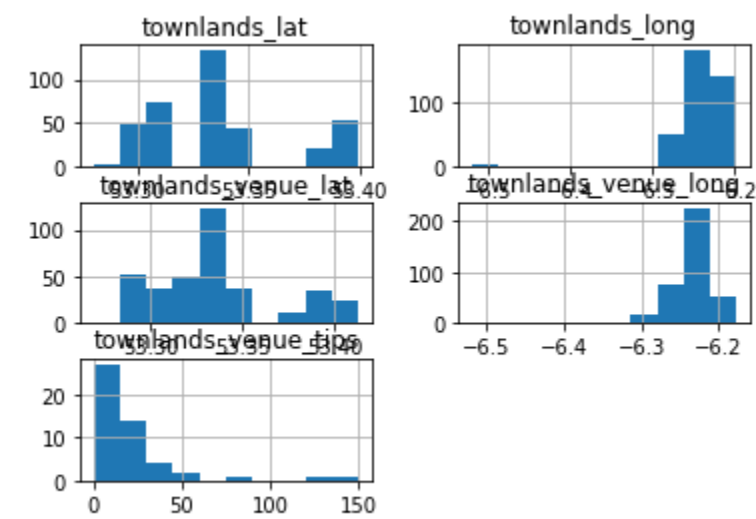
### 3.1 Feature selection

After the data cleansing we found that there were 370 samples and limited to 35 attributes. The surrounding venues were limited to just 30 categories.

The samples thus generated were limited the town lands in dublin county and the categories of venues.

As we are clear with the objective of this exercise we chose the townhall where we can have a business of thai restaurant as the target variable.

Below is the histogram of all the numerical attributes , once the data was available for processing. There was no possible linear/polynomial relationship between the attributes.



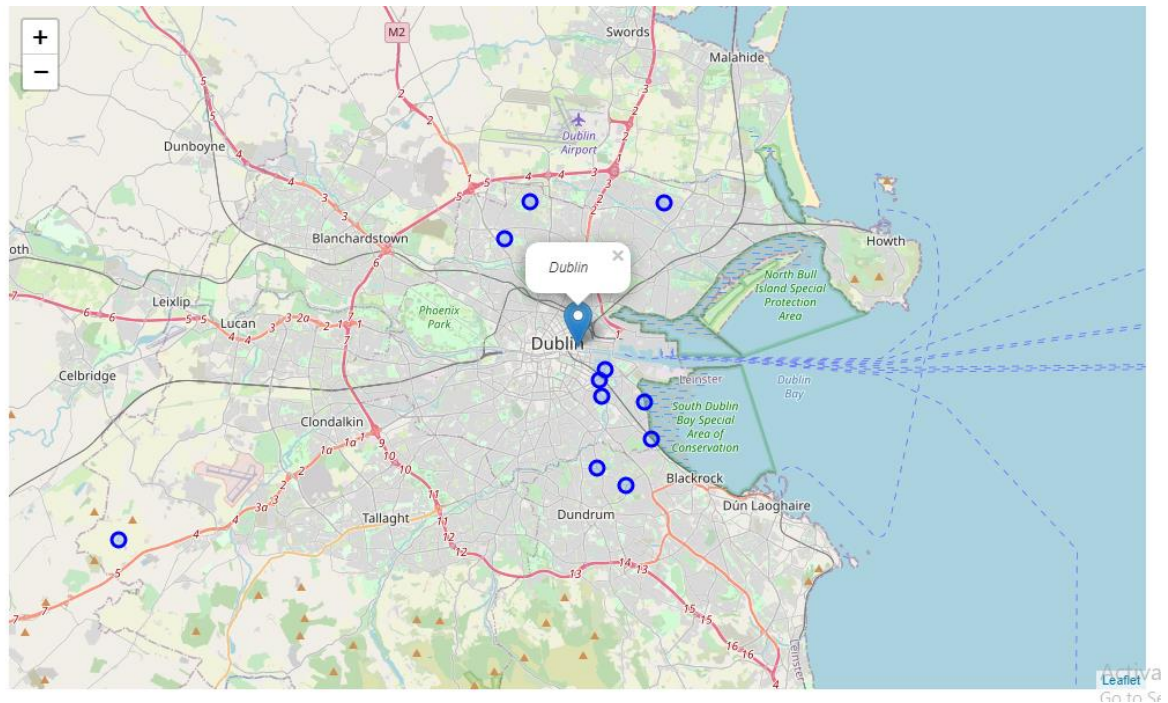
The data which was made available to us from the web scraping is used to create a word cloud which depicts the top venue categories in the town lands as below :

strn_townlands_venue_cat	strn_townlands_venue_cat_count
0	Pub45
1	Café45
2	Coffee Shop39
3	Park32
4	Hotel23
5	Restaurant23
6	Supermarket23
7	Bar16
8	Pizza Place15
9	Grocery Store13
10	Fast Food Restaurant12





Out[562]:



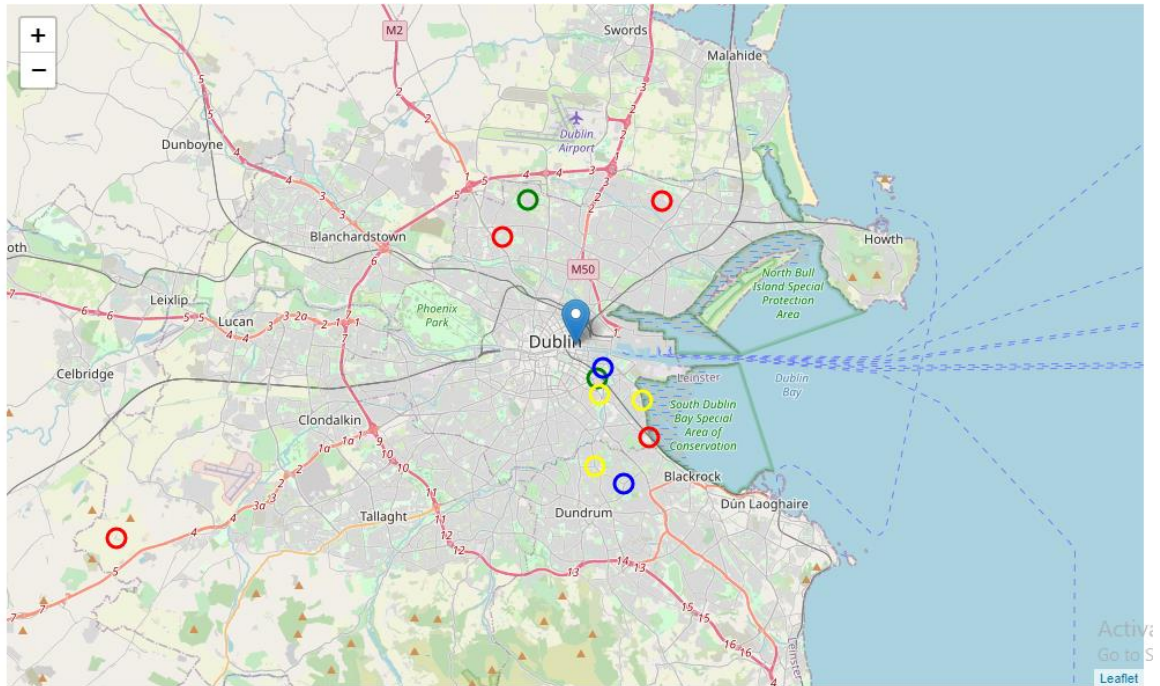
Based on the density of a thai restaurant in the 30 surrounding venues of a town land the mean value is calculated and again mapped on the folium map to check its intensity factor in the area, as below :

Out[609]:

	onehot_townlands	Thai Restaurant
0	Athgoe	0.000000
1	Ballsbridge	0.038462
2	Ballygall	0.000000
3	Beggars Bush	0.032258
4	Clonskeagh	0.037736
5	Merrion Gates	0.000000
6	Poppintree	0.033333
7	Priorswood	0.000000
8	Ringsend	0.022727
9	Roebuck	0.021277
10	Sandymount	0.040000

Here first of all the a random K of 4 is taken to cluster the similar town lands as below :

Out[620]:



From above clustering it is seen that cluster-1 (in red color rings) has no existing Thai Restaurants and we expect the prediction be a subset of this list of venues.

With the highest ranking of thai restaurants present is cluster-4 with yellow rings

With moderate ranking of are thai restaurants is cluster-2 and cluster-3 with green and blue rings located in the not so outer skirts of Dublin county.

#### **4. Predictive Modeling**

Predictive modeling is done using K-nearest neighbors (KNN) algorithm which is a type of supervised ML algorithm which can be used for both classification as well as regression predictive problems. However, it is mainly used for classification predictive problems in industry.

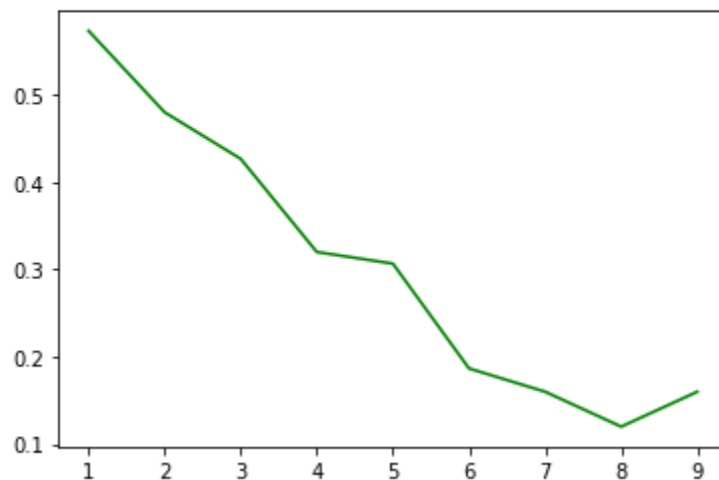
K-nearest neighbors (KNN) algorithm uses 'feature similarity' to predict the values of new data points which further means that the new data point will be assigned a value based on how closely it matches the points in the training set.

In our example we have data points which are the 30 surrounding venues of the town lands. Based on these data points the similarity of the town lands will be achieved by the algorithm.

We have determined the K value by using the elbow graphical representation as below :

```
In [680]: #Plotting the accuracy for different values of K  
plt.plot(range(1,ks),mean_acc,'g')
```

```
Out[680]: [<matplotlib.lines.Line2D at 0x1baaf721448>]
```



So, following the K as 8 I created the training and test data.

As all the values should be numerical the categories were assigned a numerical value as shown below :

```
[34 1 4 4 7 4 7 8 12 14 7 1 4 4 1 7 23 21 14 1 34 1 7 7
1 4 1 29 7 8 34 7 14 5 36 5 34 1 21 1 1 1 5 1 34 1 34 5
34 5 29 14 1 23 29 35 26 14 1 26 5 23 7 7 1 5 31 1 50 31 1 1
49 8 7 7 23 23 34 34 4 4 5 5 12 5 26 36 36 34 34 23 23 5 1 1
7 36 1 23 49 1 5 49 50 1 1 5 1 29 29 5 8 14 29 8 1 7 1 29
8 49 7 23 8 1 1 4 29 1 8 23 23 29 7 34 8 1 7 14 1 35 8 35
8 1 7 1 7 8 7 8 12 21 29 7 50 34 1 5 7 23 5 52 5 12 36 1
14 14 1 23 23 23 35 34 35 7 1 1 5 1 49 35 50 5 1 49 35 49 50 1
1 50 50 31 1 23 5 31 29 1 1 23 34 7 7 4 18 7 1 50 8 23 23 50
1 1 50 1 23 23 23 23 31 34 5 23 23 49 8 7 7 21 34 1 34 1 1 23
5 7 7 1 35 34 5 7 1 5 14 23 5 35 35 34 1 5 34 7 4 1 23 7
34 4 1 23 4 34 14 7 1 4 5 5 29 8 49 23 49 14 23 7 5 5 5 52
29 29 29 35 1 35 23 23 23 23 35 1 1 35 7 35 8 23 34 8 5 35 7 31
35 35 1 5 8 5 35 1 52 36 5 36 12 4 1 4 14 1 4 1 1 29 7 8
14 14 7 36 5 7 34 1 23 1 5 1 7 29 1 36 4 4 1 7 8 1 1 14
35 35 1 1 4 4 8 7 29 1 1]
```

Ultimately, the prediction is realized for every town land on which business is beneficial for the particular town land , including the thai restaurant ,as shown below :

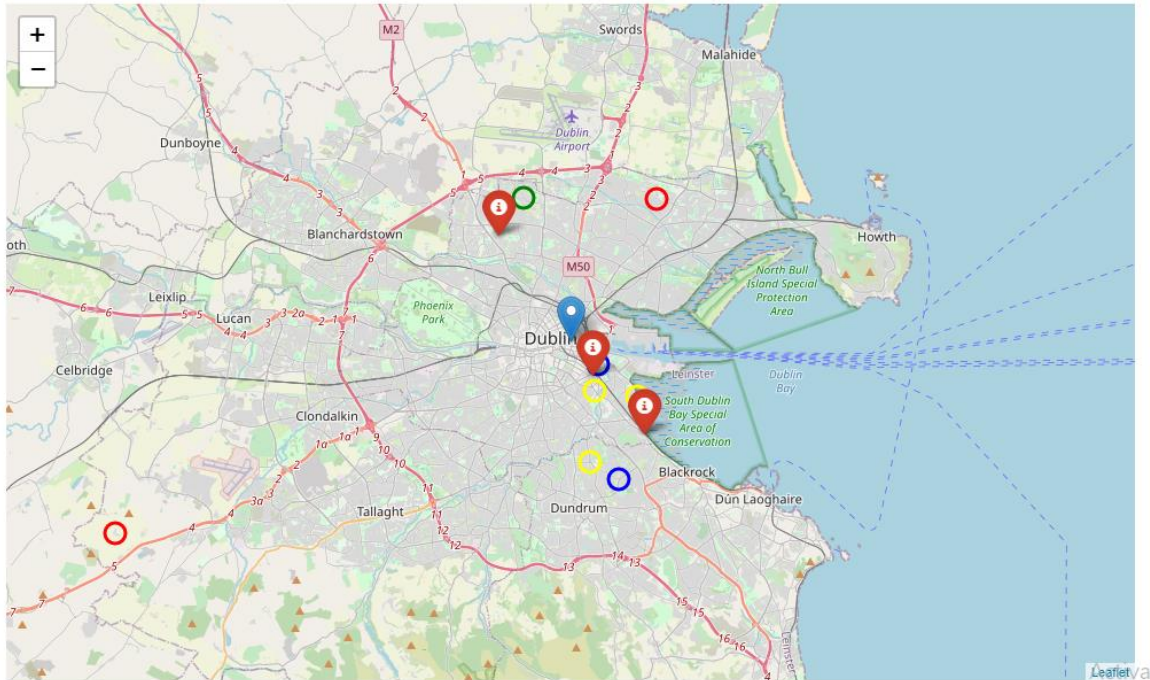
Out[689]:

	Prediction	townlands	surr_ven1	surr_ven2	surr_ven3	surr_ven4	surr_ven5	surr_ven6	surr_ven7	surr_ven8	surr_ven9	surr_ven10	su
0	Plaza	Athgoe	Moving Target	Pub	Construction & Landscaping	Stables	NaN	NaN	NaN	NaN	NaN	NaN	
1	Pub	Ballsbridge	Pub	Thai Restaurant	Park	Wine Bar	Café	Pizza Place	Soccer Stadium	Café	Pizza Place	Museum	I
2	Thai Restaurant	Ballygall	Pet Store	Park	Farmers Market	Restaurant	Pub	Supermarket	Coffee Shop	Supermarket	Fast Food Restaurant	Supermarket	E
3	Thai Restaurant	Beggars Bush	Pizza Place	Café	Gourmet Shop	Café	Coffee Shop	Soccer Stadium	Gym / Fitness Center	Plaza	Plaza	Theater	
4	Café	Clonskeagh	Fast Food Restaurant	Pub	Pub	Park	Café	Café	Restaurant	Park	Indian Restaurant	Pub	
5	Thai Restaurant	Merrion Gates	Waterfront	Italian Restaurant	Beach	Coffee Shop	Gym	Café	Supermarket	Train Station	Hotel	Pub	
6	Café	Poppintree	Gym / Fitness Center	Hotel	Pizza Place	Furniture / Home Store	Diner	Park	Theater	Park	Supermarket	Fish Market	
7	Pizza Place	Priorswood	Supermarket	Supermarket	Coffee Shop	Coffee Shop	Coffee Shop	Sports Club	Supermarket	Chocolate Shop	Diner	Coffee Shop	Su
8	Stadium	Ringsend	Gourmet Shop	Pizza Place	Café	Gym / Fitness Center	Café	Gym	Plaza	Pool	Pub	Plaza	
9	Hotel	Roebuck	Park	Farm	Hotel	Bistro	Restaurant	Indian Restaurant	Grocery Store	Fast Food Restaurant	Pub	Restaurant	ativat
10	Café	Sandymount	Beach	Burger Joint	Gym	Garden	Beach	Waterfront	Food & Drink Shop	Bar	Stadium	Hockey Field	Go to Seti

Below is the prediction of thai restaurant mapped on the folium map :



Out[694]:



## 5 Conclusion

From above analysis we can infer that cluster-1 (in red color rings) has no existing Thai Restaurants and we expect the prediction be a subset of this list of venues.

With the highest ranking of thai restaurants present is cluster-4 with yellow rings.

With moderate ranking of are thai restaurants is cluster-2 and cluster-3 with green and blue rings located in the not so outer skirts of Dublin county.

This analysis presents a great opportunity to entrepreneurs to tap into the unutilized potential of the pockets of the Dublin county by opening thai restaurants.

It is also evident that cluster-3 has a very high competition when it comes to opening a thai restaurant, hence investment in this area should be avoided by investors.

Investors with unique selling propositions that can stand out from the moderate competition in cluste-3 and cluster-2 and can take moderate risk and attract the customers already visiting the locality of this cluster because of the existing thai restaurant.