Choosing the best neighborhood for a potential pub of a Pub chain

I. Introduction

1. Background

   Pin’t is a pub chain based in London. the chain plans on expanding and opening a new first pub in Dublin. The chain has several pubs in different neighborhoods in London with different ROI (Revenue On Investment). The variables between the existing pubs are location and ROI. The chain identified some potential neighbourhoods in Dublin. But they need to refine their choice based on the two variables.

2. Business Problem

   The problem is identifying a neighborhood, based on its location and the chain’s history and experience in other neighbourhoods, based on the available data:

   - Geolocation of existing pubs and potential neighbourhoods for new pub
   - ROI of existing pubs

3. Approach:

   My approach is to use Foursquare location data and a clustering algorithm in order to cluster the potential neighbourhoods and the neighbourhoods where the chain has already established a pub, according to the category of venues close to each neighbourhood. And then choose the neighbourhood that belongs to the cluster with the most success rate. If there is a conflict: two potential neighbourhoods belong to most successful cluster I’ll be using other metrics: choose the neighbourhood with less pubs (for example)

4. Interest:

The study concerns the Pub chain Pin't but can be reused by the chain for their next openings. And help other brands to make decisions based on their existing locations and profit from those locations.

II. Data

1. Data description:
    - Categories of the venues in each neighbourhood venues (Foursquare API): It will help us group similar neighbourhoods based on the venues nearby.
    - Existing pubs data: It will help us identify the most successful cluster based on existing pubs ROI. https://docs.google.com/spreadsheets/d/1UrTzTew7otS3l2AZS3aDeT2LqeLolHafxf1-3wlig08/edit?usp=sharing
    - Potential neighbourhoods in Dublin: https://docs.google.com/spreadsheets/d/1QNCQQIdCHGsi6Re8IfioJaHm8ab5C11RQTGTXm-756M/edit?usp=sharing
2. Data acquisition:

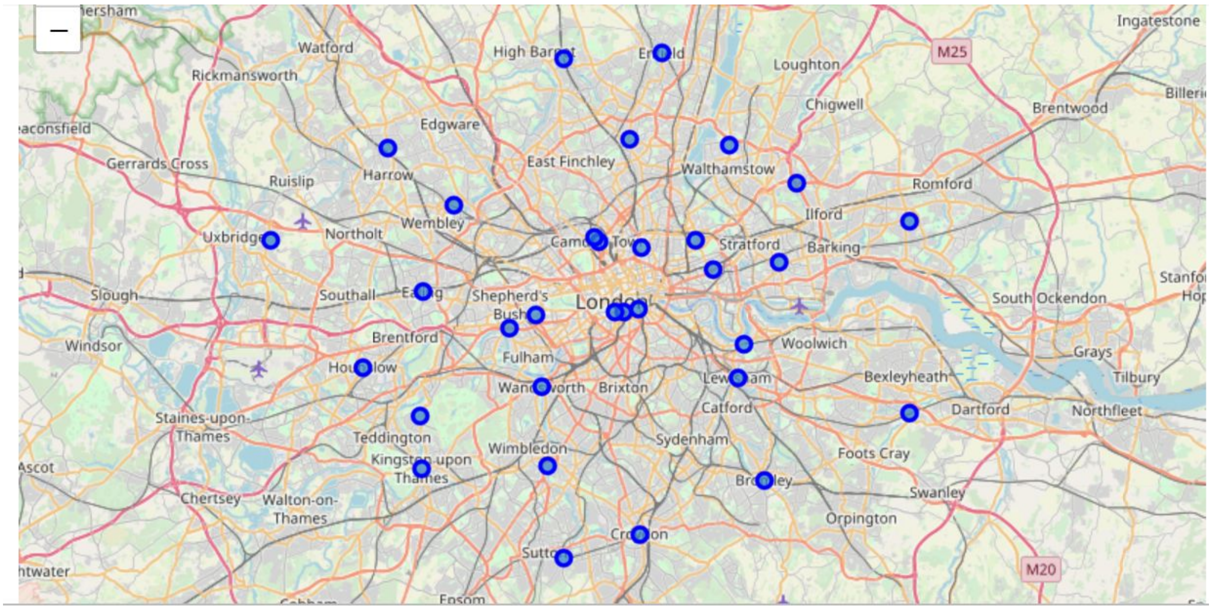
In order to conduct the study I needed to get the exact coordinates of each neighbourhood, that I used to get venues nearby.

2.1. Neighbourhoods' coordinates:

I used python's geopy library to get the coordinates of each neighbourhood (existing and potential)

	Neighbourhood	ROI	latitude	longitude
0	Barking and Dagenham	0.20	51.554117	0.150504
1	Barnet	0.76	51.648784	-0.172913
2	Bexley	0.38	51.441679	0.150488
3	Brent	0.74	51.563826	-0.275760
4	Bromley	0.29	51.402805	0.014814

London Pubs coordinates

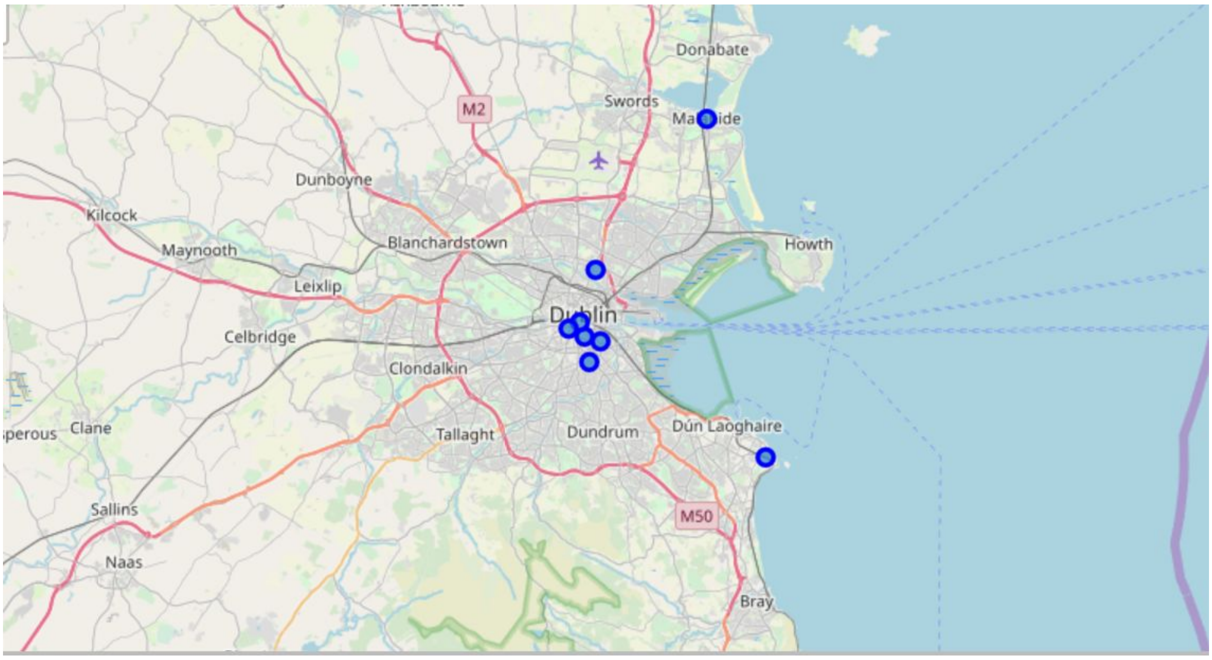


London pubs visualized

2]:

	Neighbourhood	latitude	longitude
0	St. Stephen's Green	53.337990	-6.259073
1	Temple Bar	53.345496	-6.263114
2	Christchurch	53.342689	-6.272784
3	Ranelagh and Rathmines	53.325218	-6.255050
4	Ballsbridge and Donnybrook	53.335711	-6.245229
5	Drumcondra	53.372525	-6.249515
6	Malahide	53.450840	-6.153670
7	Dalkey	53.275607	-6.103188

Potential pubs coordinates



Dublin neighbourhoods visualized

2.2. Neighbourhoods' nearby venues: I used the Foursquare API to get maximum of 30 venues nearby each neighbourhood.

14]:

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking and Dagenham	51.554117	0.150504	Central Park	51.559560	0.161981	Park
1	Barking and Dagenham	51.554117	0.150504	Harrow Lodge Park	51.555648	0.197926	Park
2	Barking and Dagenham	51.554117	0.150504	Capital Karts	51.531792	0.118739	Go Kart Track
3	Barking and Dagenham	51.554117	0.150504	The Eva Hart (Wetherspoon)	51.570460	0.130342	Pub
4	Barking and Dagenham	51.554117	0.150504	Hylands Park	51.572074	0.191155	Park

Venues nearby existing pubs

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	St. Stephen's Green	53.33799	-6.259073	St Stephen's Green	53.338151	-6.259160	Park
1	St. Stephen's Green	53.33799	-6.259073	Hatch & Sons	53.339515	-6.258460	Café
2	St. Stephen's Green	53.33799	-6.259073	Dolce Sicily	53.340942	-6.258772	Café
3	St. Stephen's Green	53.33799	-6.259073	Peruke & Periwig	53.340086	-6.258542	Cocktail Bar
4	St. Stephen's Green	53.33799	-6.259073	Iveagh Gardens	53.335680	-6.261059	Park

Venues nearby potential neighbourhoods

3. Data cleaning:
   - Converting variables

   After getting the venues nearby each neighbourhood I convert the categorical variable “Venue Category” column into indicator variables, thanks to .get_dummies.

16]:

	Neighbourhood	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	Bakery	Bar	Beer Bar	Beer Garden
0	Barking and Dagenham	0	0	0	0	0	0	0	0	0	0	0	0
1	Barking and Dagenham	0	0	0	0	0	0	0	0	0	0	0	0
2	Barking and Dagenham	0	0	0	0	0	0	0	0	0	0	0	0

4. Feature selection:

   The clustering of neighbourhoods was based on the categories of the venues nearby each neighbourhood. After grouping the venues categories by neighborhoods I calculated the frequency of each category:

9]:

	Neighbourhood	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	Bakery	Bar	Beer Bar	Beer Garden	B St
0	Barking and Dagenham	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000	0.0000
1	Barnet	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.033333	0.000000	0.000000	0.0000
2	Bexley	0.033333	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
3	Brent	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
4	Bromley	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.0000
5	Camden	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
6	Croydon	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
7	Ealing	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.033333	0.000000	0.000000	0.0000
8	Enfield	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
9	Greenwich	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
10	Hackney	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000	0.0000
11	Hammersmith and Fulham	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.033333	0.000000	0.0000
12	Haringey	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.033333	0.000000	0.000000	0.0000
13	Harrow	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0333
14	Havering	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000	0.0000

III. Methodology

In this part I worked on my neighbourhoods data in order to cluster my neighbourhoods into similar groups based on the venues nearby. The potential neighbourhood that belongs to the most successful cluster (cluster with highest mean of ROI or cluster where the neighbourhood with highest ROI belongs) will be the location of my next pub in Dublin

1. Data exploration:
   - Relationship between distance from city center and ROI

Before clustering all neighborhoods I wanted to see if there is a correlation between existing pubs ROI and their distance from Central London. That I can mirror by distance from Center of Dublin in potential pubs. So I calculated the distances:

[36]:

	Neighbourhood	ROI	latitude	longitude	distance from central London
10	Hackney	0.78	51.543240	-0.049362	12.936147
29	Waltham Forest	0.77	51.598169	-0.017837	14.649178
1	Barnet	0.76	51.648784	-0.172913	26.349405
21	Lewisham	0.75	51.462432	-0.010133	10.257794
19	Kingston upon Thames	0.74	51.409627	-0.306262	31.534463
3	Brent	0.74	51.563826	-0.275760	28.578105
23	Newham	0.72	51.530000	0.029318	7.356572
17	Islington	0.67	51.538429	-0.099905	16.095262
15	Hillingdon	0.64	51.542519	-0.448335	39.924256
27	Sutton	0.64	51.357511	-0.173640	26.104506

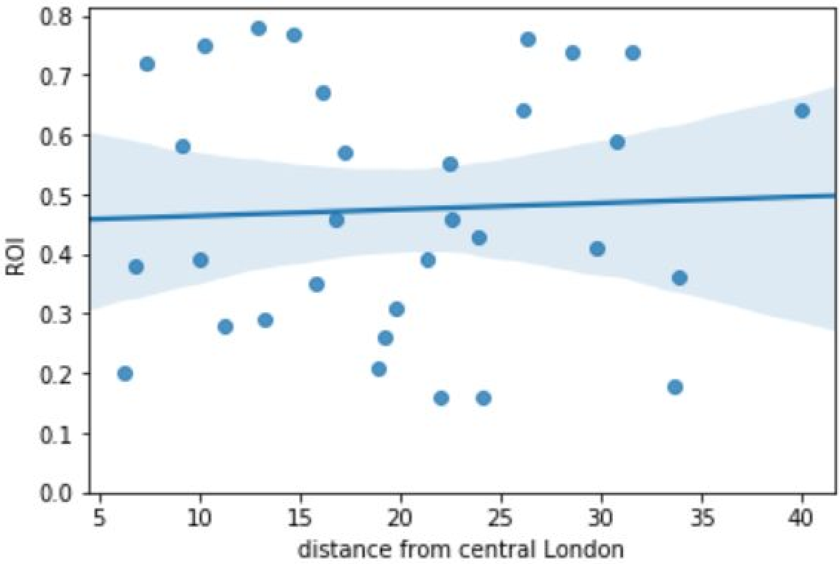
then calculated the correlation between the two variables by using the pandas method: corr().

2]:

	distance from central London	ROI
distance from central London	1.000000	0.045609
ROI	0.045609	1.000000

The correlation was equal to 0.045 which indicated a weak correlation: there was no need to include it in clustering method.

[44]: (0, 0.8117919339817881)



2. K-means Clustering

2.1. Clustering the neighbourhoods

```
2]: kclusters = 5

neighbourhoods_clustering = all_neighbourhoods.drop('Neighbourhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(neighbourhoods_clustering)
```

Then I grouped the clusters by ROI mean

```
[218]: all_venues.groupby('Cluster Labels').mean()
```

[218]:

Cluster Labels	ROI
0	0.415000
1	0.475385
2	0.440000
3	0.610000
4	0.460000

Notice that the cluster number 3 (4th cluster) is the most successful cluster.

2.2. Exploring the clusters

- Cluster 1: it contains only london neighbourhoods

```
[221]: print(all_venues.loc[all_venues['Cluster Labels'] == 0,['ROI','City']].mean())
       all_venues.loc[all_venues['Cluster Labels'] == 0,['ROI','City','Neighbourhood']]
```

```
ROI      0.415
dtype: float64
```

[221]:

	ROI	City	Neighbourhood
0	0.20	London	Barking and Dagenham
16	0.36	London	Hounslow
27	0.64	London	Sutton
30	0.46	London	Wandsworth

- Cluster 2: Contains only one Dublin Neighbourhood

```
ROI    0.475385
dtype: float64
```

[222]:

	ROI	City	Neighbourhood
2	0.38	London	Bexley
4	0.29	London	Bromley
6	0.39	London	Croydon
7	0.41	London	Ealing
8	0.16	London	Enfield
9	0.58	London	Greenwich
15	0.64	London	Hillingdon
19	0.74	London	Kingston upon Thames
22	0.43	London	Merton
23	0.72	London	Newham
24	0.39	London	Redbridge
28	0.28	London	Tower Hamlets
29	0.77	London	Waltham Forest
39	NaN	Dublin	Temple Bar

- Cluster 3:

```
ROI    0.44
dtype: float64
```

3]:

	ROI	City	Neighbourhood
3	0.74	London	Brent
5	0.21	London	Camden
11	0.16	London	Hammersmith and Fulham
13	0.18	London	Harrow
14	0.26	London	Havering
17	0.67	London	Islington
18	0.55	London	Kensington and Chelsea
21	0.75	London	Lewisham

- Cluster 4 (most successful):

```
ROI    0.61
dtype: float64
```

24]:

	ROI	City	Neighbourhood
1	0.76	London	Barnet
10	0.78	London	Hackney
12	0.31	London	Haringey
25	0.59	London	Richmond upon Thames
32	NaN	Dublin	Ballsbridge and Donnybrook
33	NaN	Dublin	Christchurch
34	NaN	Dublin	Dalkey
35	NaN	Dublin	Drumcondra
36	NaN	Dublin	Malahide
37	NaN	Dublin	Ranelagh and Rathmines
38	NaN	Dublin	St. Stephen's Green

```
ROI    0.61
dtype: float64
```

4]:

	ROI	City	Neighbourhood
1	0.76	London	Barnet
10	0.78	London	Hackney
12	0.31	London	Haringey
25	0.59	London	Richmond upon Thames
32	NaN	Dublin	Ballsbridge and Donnybrook
33	NaN	Dublin	Christchurch
34	NaN	Dublin	Dalkey
35	NaN	Dublin	Drumcondra
36	NaN	Dublin	Malahide
37	NaN	Dublin	Ranelagh and Rathmines
38	NaN	Dublin	St. Stephen's Green

- Cluster 5: Cluster 5 is empty

2.3. Decision making

The most successful cluster has more than 1 Dublin Neighbourhood.

In order to make a decision I needed to find the neighbourhood that is most similar to the most successful neighbourhood in London:Hackney

I got the most common venues categories in each neighbourhood from the 4th cluster, by adding cluster labels to each neighbourhood:

]:

	ROI	Neighbourhood	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
10	0.78	Hackney	Coffee Shop	Café	Pub	Park	Indie Movie Theater	Market	Canal Lock	Butcher	Flea Market
1	0.76	Barnet	Café	Park	Coffee Shop	Pub	Supermarket	Turkish Restaurant	Dessert Shop	Theater	Fish & Chips Shop
25	0.59	Richmond upon Thames	Park	Café	Garden	Coffee Shop	Hotel	Italian Restaurant	Bakery	Scenic Lookout	Pub
12	0.31	Haringey	Café	Mediterranean Restaurant	Turkish Restaurant	Park	Coffee Shop	Pizza Place	Gourmet Shop	Garden Center	Indie Movie Theater
32	NaN	Ballsbridge and Donnybrook	Café	Coffee Shop	Park	Hotel	Lounge	Concert Hall	Cocktail Bar	Outdoor Sculpture	Pizza Place
33	NaN	Christchurch	Café	Pub	Coffee Shop	Music Venue	Park	Cocktail Bar	Restaurant	Ice Cream Shop	Irish Pub
34	NaN	Dalkey	Beach	Coffee Shop	Scenic Lookout	Restaurant	Park	Pub	Café	Seafood Restaurant	Italian Restaurant
35	NaN	Drumcondra	Coffee Shop	Café	Pub	Clothing Store	Restaurant	Hotel	Donut Shop	Discount Store	Italian Restaurant
36	NaN	Malahide	Café	Italian Restaurant	Beach	American Restaurant	Pub	Gourmet Shop	Garden	Gym	Hotel
37	NaN	Ranelagh and Rathmines	Café	Park	Coffee Shop	Restaurant	Hotel	Burger Joint	Chinese Restaurant	Falafel Restaurant	Pub
38	NaN	St. Stephen's Green	Coffee Shop	Hotel	Park	Café	Burger Joint	Ice Cream Shop	Pub	Cheese Shop	Lounge

]:

	ROI	Neighbourhood	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
10	0.78	Hackney	Coffee Shop	Café	Pub	Park	Indie Movie Theater	Market	Canal Lock	Butcher	Flea Market
1	0.76	Barnet	Café	Park	Coffee Shop	Pub	Supermarket	Turkish Restaurant	Dessert Shop	Theater	Fish & Chips Shop
25	0.59	Richmond upon Thames	Park	Café	Garden	Coffee Shop	Hotel	Italian Restaurant	Bakery	Scenic Lookout	Pub
12	0.31	Haringey	Café	Mediterranean Restaurant	Turkish Restaurant	Park	Coffee Shop	Pizza Place	Gourmet Shop	Garden Center	Indie Movie Theater
32	NaN	Ballsbridge and Donnybrook	Café	Coffee Shop	Park	Hotel	Lounge	Concert Hall	Cocktail Bar	Outdoor Sculpture	Pizza Place
33	NaN	Christchurch	Café	Pub	Coffee Shop	Music Venue	Park	Cocktail Bar	Restaurant	Ice Cream Shop	Irish Pub
34	NaN	Dalkey	Beach	Coffee Shop	Scenic Lookout	Restaurant	Park	Pub	Café	Seafood Restaurant	Italian Restaurant
35	NaN	Drumcondra	Coffee Shop	Café	Pub	Clothing Store	Restaurant	Hotel	Donut Shop	Discount Store	Italian Restaurant
36	NaN	Malahide	Café	Italian Restaurant	Beach	American Restaurant	Pub	Gourmet Shop	Garden	Gym	Hotel
37	NaN	Ranelagh and Rathmines	Café	Park	Coffee Shop	Restaurant	Hotel	Burger Joint	Chinese Restaurant	Falafel Restaurant	Pub
38	NaN	St. Stephen's Green	Coffee Shop	Hotel	Park	Café	Burger Joint	Ice Cream Shop	Pub	Cheese Shop	Lounge

I notice that the 3 most common venues categories in the most successful pubs in London are: Coffee Shops, cafés ,pubs and parks

=> In order to make a decision I need to choose the neighbourhood which 3 most common venues categories are in the list

The potential neighbourhoods that check the criteria are:

- Ballsbridge and Donnybrook

- Christchurch
- Drumcondra
- Ranelagh and Rathmines

To refine my choices I noticed that **Ranelagh and Rathmines** is the most similar neighbourhood to Hackney where the most successful pub is established: with the same first 3 most common venues categories.

IV. Conclusions

In this study I studied the similarity between neighbourhoods in two different cities based on the venues nearby each neighbourhoods. Thanks to the Foursquare API. I also studied the correlation between the ROI of existing pubs in London and their distance from the center, and I discovered that it was weak. I used k-means clustering to group similar neighbourhoods (in London and Dublin) in order to find the best potential neighbourhood in Dublin. I found out that the most successful pubs in London (in Hackney and Barnet) belong to the same cluster (which emphasize the impact of the venues nearby). And I made a decision by choosing the neighbourhood that resembles the most to the Pub with the highest ROI. Which was **Ranelagh and Rathmines.**