Discovering dining venues and rental accommodation costs in the vicinity of high-tech companies in Dublin, Ireland

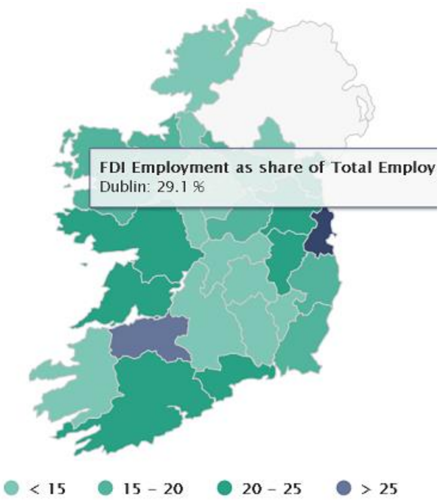
Roger Clarke
July 2021

Dublin City: *Image from Bing.com – licence public domain*

1. Introduction

1.1 Background

Ireland is a European hub to over 1000 foreign direct companies in technology, pharma, social media, and many others. FDI (Foreign Direct Investment) comprises around **20%** of all private sector employment with Dublin playing host to over 29% of that with many big names Such as Microsoft, Google, Amazon, Facebook and LinkedIn being some of the largest employers. FDI companies employ circa 250 thousand people. Dublin, the capital city, is known the world over as vibrant and a focal point of prominent levels of foreign direct investment especially in the tech sector.



Young top talent coming out of universities in many countries are interested securing employment with these firms is coming to the city with no personal transport. The young employees want a short walk to the office of no more than 2.5km and have all the food amenities close to the office so that they can continue enjoying the cafe lifestyle where they can eat out at lunchtime rather than suffer terrible food in a company cafeteria. However, Dublin is also well known for outrageous rental costs and poor-quality accommodation which may be deterrent to taking up employment. According to the Residential Tenancies Board (RTB), the average cost of renting a home in Ireland in 2021 is **€1,745 per** month and that's a 2.1 percent increase year on year.

CSO statistical publication, 15 April 2021, 11am

On-line ISSN: 2009-5236

Table 1.1: Residential Property Price Index

Month	Price Index (2015=100)	Monthly % change	Annual % change
November 2020	136.2	0.7	0.4
December 2020	137.2	0.7	2.3
January 2021	137.8	0.4	2.6
February 2021	138.3	0.3	3.0

The latest three month's RPPI results are provisional and subject to revision.

1.2 Problem

Companies such as Microsoft, Google, Facebook, Amazon, Twitter & LinkedIn, all of which based in Dublin, are always looking for skilled, educated and highly capable people from all over the world. Potential job applicants who are not familiar with Ireland, and who wish to live and socialise close to work, will need to use multiple sources of information to understand which company meets those criteria.

The objective

Simplify selecting a high company to work at in Dublin based on rental cost and dining venues.

2. Data Acquisition, Feature Selection, Methodology and Cleaning

2.1 Data Acquisition

2.1.1 Foreign Direct Employers

There are many foreign direct employers in Dublin, so many that there would be an excess of data points to deal with. Instead, I selected the top 5 employers in Dublin (Twitter, Amazon, Facebook, Google and Microsoft) with one found far from the city centre to demonstrate the usefulness of this project to its intended audience. Many of these companies have several properties in separate locations and again for simplicity, the headquarters of each will be selected. The names, postal codes, address, and geo-coordinates will be manually applied to a dataset.

2.1.2 Rental Availability

There are two primary sources of rental accommodation https://www.Daft.ie and https://www.Myhome.ie. Daft prohibits web scraping, does supply an API but which requires payment for its use. MyHome.ie has no API, and the policy does not prohibit web scraping hence this will the sole source of rental data. Unfortunately, that underrepresents the entire rental market.

2.1.3 Rental Statistics

Daft.ie publishes quarterly statistics on rental price changes across the country including changes in property types. I used the 2021 Q2 **report** and download the nationwide rental statistics for use in the descriptive statistics and visualisation of rental properties around companies.

2.1.4 Postal District(code)
Dublin, in terms of postal addresses, is divided into county and districts. Districts addresses use postal codes where as "County" addresses don't. The postal code becomes important when doing Google Map API searches as it improves accuracy and imputing the cost of properties where the stated cost is missing.

2.1.5 Rental Latitude and Longitude
This distance between each company and rental accommodation is key to fulfilling the requirement that the rental accommodation be within 2km of each company. I used the Google Maps API will supply the latitude and longitude of each company and rental property. The choice of 2km is intentional as its distance quickly walked or cycled.

2.1.6 Foursquare
The Foursquare API will supply all venues within 500m distance of each company selected for evaluation along with their category, latitude, and longitude. The choice of 500m is intentional as its distance you can quickly walk thus allowing more time for dining during the lunch hour.

2.2 Feature Selection
Knowing the features required is key to understand how they will help achieve the goal. I decided on three primary datasets, Companies, Rental and Venues which I later blended into two datasets, Companies and associated Rental and Companies and associated dining venues.

2.2.1 Common Features
Latitude and **Longitude** which, enables me display locations on folium maps and calculate distances between companies, properties, and venues. These are common to companies, rental properties, and venues.

2.2.2 Companies
I selected **company name**, mostly for mapping display purposes, **street address**, **postal code**, and Eircode. Later in the project I discovered I did not need this but the saved scraped data from myhome.ie but kept it.

2.2.3 Rental Properties
This was little more challenging as I first needed to explore myhome.ie with some searches to understand what data would be most useful for this project. In the end I selected property **address**, **type** (apartment, etc), **cost**, number of **bedrooms** and **bathrooms**. There were other features such as energy rating (BER) but decided this was not that relevant in this context.

2.2.3 Foursquare
Foursquare offers a multitude of features but the most useful are **latitude**, **longitude**, **category,** and **venue** (place name).

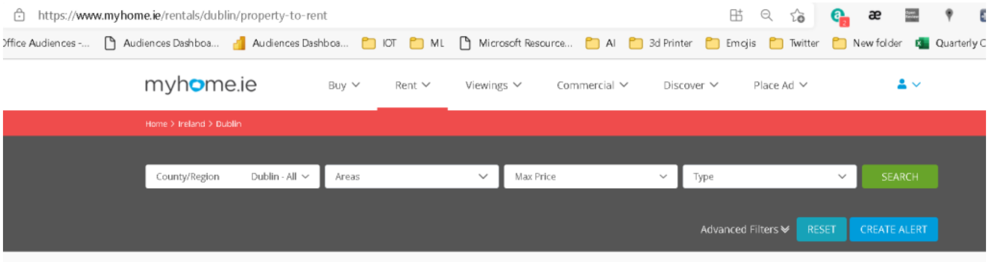
3.1. Methodology

3.1.1 Companies
The companies selected have multiple buildings scattered across the Dublin but the headquarters of each hosts the largest number of employees and job roles. Additionally, including all locations would add map showing all of them, the associated venues and rental accommodation more complex due to the need to create multiple folium maps just to reduce the sheer clutter. I created a companies data frame from a list of dictionaries I populated with a selection of company names and street addresses and then used the Google Maps API to add the latitude and longitude of each company as added features. Using the Google Maps API required to me create a Google account and sign up for the API with a credit card. Fortunately, the usage allowance was generous to avoid any charges. This approach cut the need for post data frame cleansing. The resulting data frame looks like this.

	COMPANYNAME	COMPANYADDRESS	POSTALCODE	COMPANYLAT	COMPANYLNG
0	FaceBook	BLOCK J, FACEBOOK DUBLIN BALLSBRIDGE CAMPUS, S...	DUBLIN 4	53.329515	-6.225678
1	Google	ONE, GRAND CANAL PLAZA, GRAND CANAL STREET UPPER	DUBLIN 4	53.340627	-6.239397
2	Amazon	Burlington Plaza, 1 Burlington Rd	DUBLIN 4	53.332500	-6.246058
3	Twitter International	CUMBERLAND PLACE, FENIAN STREET	DUBLIN 2	53.341705	-6.249626
4	LinkedIn	GARDNER HOUSE, WILTON PLAZA, WILTON PLACE	DUBLIN 2	53.334907	-6.249499
5	Microsoft	ONE MICROSOFT PLACE, SOUTH COUNTY BUSINESS PARK	DUBLIN 18	53.268810	-6.196712

3.1.2 Rental Accommodation

I chose to use BeautfulSoup to scrap myhome.ie for rental accommodation. Where a website requires user input to create results for scraping, one would need to use the likes of Mechanize. However, I found a useful work around that reduces the coding complexity and run time. I pre-populated the URL by navigating myhome.ie, selecting the Rent tab and then selecting 'Dublin All, region. This generated a URL (below) that I could use with BeautifulSoup to generate search results.



The search results are divided into pages 

and I noticed that after clicking page 1, the URL changed to https://www.myhome.ie/rentals/dublin/property-to-rent?page=1. By getting page numbers I was able to automate the navigating by passing the number to the URL page parameters and then running BeautifulSoup on each page to parse out the properties. From there I iterated the pages getting the HTML for each page. Each page has a list of a specific HTML class that holds a property from the search result. By getting the length of this list I could then iterate all of pages getting the individual property details. The specific features such as Cost, Rental Type and so on are in a variety of HTML types (a, span and div). Getting the data for the features required getting the text of the proper elements.

- Cost, and payment period needed to be split from a single text value and I removed any other wanted text.
- The number of bedrooms, bathrooms and rental type also needed to be split from a single text value.

Example of how a property's useful features are displayed.

The iteration cycle was quite slow hence I ran the scraping function only once and saved the results to a tab separated file for later loading and processing. All the data cleansing was done from loading the saved results. The resulting data frame looks like this.

	RENTALADDRESS	RENTALTYPE	RENTALLAT	RENTALLNG	RENTALCOST	PAYMENTFREQUENCY	RENTALBEDROOMS	RENTALBATHROOMS	POSTALCODE
356	Apt 2 Monck Place, Phibsboro, Dublin 7, D07 V2R0	Apartment	53.358936	-6.274158	1463	Monthly	2	1	Dublin 7
532	An Teach Deireanach, Cranmer Place, Ballsbrid...	House	53.337182	-6.236755	9625	Monthly	3	3	Dublin 4
543	Sherrard Street Lower, Dublin 1, Dublin	Apartment	53.359325	-6.258375	1938	Monthly	0	0	Dublin 1
566	Sean McDermott Street Apartments, Sean McDerm...	Apartment	53.353859	-6.252517	1938	Monthly	0	0	Dublin 1
658	Kenilworth Road, Rathmines, Dublin 6	Apartment	53.320117	-6.273922	2389	Monthly	0	0	Dublin 6

3.1.3 Rental Statistics Data

The rental statistics were taken from the 2021 Q2 Daft.ie rental report. Page 10 held a table with mean rental values by rental type, number of bedrooms and postal code which I copied into Excel, formatted and then saved to a CSV. I then loaded the CSV into a rental statistics data frame.

	Region	1 Bed Apartment	2 Bed House	3 Bed House	4 Bed House	5 Bed House
0	Dublin 1	€1,500	€1,787	€2,231	€2,710	€3,041
1	Dublin 2	€1,665	€2,038	€2,402	€2,931	€3,170
2	Dublin 3	€1,537	€1,853	€2,103	€2,458	€2,813
3	Dublin 4	€1,817	€2,135	€2,408	€2,796	€3,129
4	Dublin 5	€1,529	€1,732	€1,949	€2,251	€2,462

3.1.4 Companies and Rental data frame

At this stage there are two distinct data frames, one with companies' data and the other rental accommodation. I created a new data frame to hold the joint results by extracting the columns names of both data frames into a list and creating the new data frame using that.

To populate the new data frame, I iterated the companies' data frame and within that loop iterated the rental properties data frame calculating the distance between rows in both. Where the distance between a company and a rental property was <= 2km I created a dictionary with all the column data from both data frames, added the distance a new feature and appended this to a list. When the iterations were complete, I converted the list to a data frame and merged it with the new companies and rental data frame created at the start.

I created a new data frame by blending both based on the distance between the companies and rental properties. To calculate the distance, I used the Haversine formula which takes the latitude and longitude of two locations and calculates the distance in kilometres.

	COMPANYNAME	COMPANYADDRESS	POSTALCODE	COMPANYLAT	COMPANYLNG	RENTALADDRESS	RENTALTYPE	RENTALLAT	RENTALLNG	RENTALCOST	PAYMENTFREQUENCY	RENTALBEDROOMS
0	FaceBook	BLOCK J, FACEBOOK DUBLIN BALLSBRIDGE CAMPUS, S...	Dublin 4	53.329515	-6.225678	The Locks, Charlotte Quay Dock, Grand Canal D...	Apartment	53.342225	-6.236034	3200	Monthly	3
1	FaceBook	BLOCK J, FACEBOOK DUBLIN BALLSBRIDGE CAMPUS, S...	Dublin 2	53.329515	-6.225678	Hanover Riverside, Grand Canal Dk, Dublin 2	Penthouse	53.345558	-6.237870	3250	Monthly	2
2	FaceBook	BLOCK J, FACEBOOK DUBLIN BALLSBRIDGE CAMPUS, S...	Dublin 2	53.329515	-6.225678	Hanover Riverside, Grand Canal Dk, Dublin 2	Penthouse	53.345558	-6.237870	3250	Monthly	2
3	FaceBook	BLOCK J, FACEBOOK DUBLIN BALLSBRIDGE CAMPUS, S...	Dublin 2	53.329515	-6.225678	41 The Waterfront, Grand Canal Dk, Dublin 2	Apartment	53.344003	-6.234158	2400	Monthly	2
4	FaceBook	BLOCK J, FACEBOOK DUBLIN BALLSBRIDGE CAMPUS, S...	Dublin 4	53.329515	-6.225678	The Waterside, Charlotte Quay Dock, Grand Can...	Apartment	53.342225	-6.236034	1570	Monthly	1

3.1.5 Preparing the companies and rental data frame for visualisation

I wanted to offer the map user the opportunity to toggle on/off property types with a legend to guide navigation via a collapsed tile. The toggle functionality was provided by creating a feature group for each rental property type and adding properties only of that that. Each property marker consisted of the rental cost and a colour coding to show if the cost is above or below the average rental cost for that property type.

I chose three colours to use, red, green, and orange and to decide on which to apply I did the following:

- When creating the folium map, I iterated the joint companies and rental data to add properties to their respective feature groups. During the iteration:
    - Using a property type, number of bedrooms and postal address, I searched the rental statistics data set for a matching mean rental value and set a meanrental variable.
    - The statistics dataset has no data for 2-n bedroom apartments and 6-n bedroom houses. In these cases, I didn't use the rental statistics but instead searched the joint companies and rental properties dataset for all properties matching the same criteria used earlier and calculate the mean of their cost and set the meanrental variable.
    - The cost of current property in the iteration was compared to the meanrental and if it was less the marker colour was set to green, more red and otherwise orange. I then created a map legend with the colour coding and cost ranges.

3.1.6 Venues

I used the Foursquare API to get venue types, names, latitude, longitude of each company. The first step was to create an account and get the API client id and secret. I used a function that was already defined as part of this course, to get the venues within a 500m radius of each company which results in a data frame that looks like this:

	COMPANYNAME	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	FaceBook	53.329515	-6.225678	The Good Food Store	53.330478	-6.224752	Food & Drink Shop
1	FaceBook	53.329515	-6.225678	Base Wood Fired Pizza	53.328869	-6.230202	Pizza Place
2	FaceBook	53.329515	-6.225678	Baan Thai Restaurant	53.328832	-6.230290	Thai Restaurant
3	FaceBook	53.329515	-6.225678	InterContinental Dublin	53.326608	-6.226079	Hotel
4	FaceBook	53.329515	-6.225678	Pembroke Wanderers Hockey Club	53.329011	-6.223840	Hockey Field

Any query to Foursquare, unless the URL is specifically tailored, will return an extensive range of venue types from Pubs, Parks, ATMs and Canals. In the context of this project, many of those are not needed and, if left in the dataset, would cause serious clutter on a folium map. In total, 62 venue types were returned. After examining all the unique types, I dropped all rows that did not relate to dining. The final choice of venues was narrow down to the below. This would ensure the greatest range of culinary experiences available in the vicinity of each company.

Pub, Coffee Shop, Café, Gastropub, Pizza Place, Sandwich Place, Burrito Place, Restaurant, Thai Restaurant, Indian Restaurant, French Restaurant, Italian Restaurant, Japanese Restaurant, Argentinian Restaurant, Asian Restaurant, Modern European Restaurant, Bakery, Bistro, Steak house, Mexican Restaurant, Seafood Restaurant, Chinese Restaurant, Middle Eastern Restaurant

Finally, I filtered the venues data frame with include only the above categories and created a new data frame with the companies and results of the above venue filtering.

At the end of the preparation phase I had two data frames:

- df_company_venues (companies and venues within a 500m radius)
- df_companiesrental_merged (companies and rental properties within a 500m radius)

2.4. Cleaning

2.4.1 Rental Properties

Property agents manually enter properties into the site with the usual issues of missing data related to human error. This becomes problematic for rental costs and types where there is either no value or a non-numerical value. Post web scraping, there was a considerable amount of data cleansing needed which, if not doing would prevent statistical analysis and negatively affect all the visualisations.

- Rental type and address where both object types so I converted these to strings.
- The repayment period fluctuates between monthly weekly and daily. To simplify, I summed all costs to monthly where applicable in the data frame.
- 2.09 percent of all properties had no defined property type. This is not something one can easily infer from the address hence I chose to remove these from the data frame.
- Rental costs were more challenging to deal with. 1.84 percent of all properties had their cost defined as "POA" which makes running statistical functions impossible. Additionally, the RENTALCOST feature was defined as object while an INT is best type for number whole numbers. Even though 1.84 percent is small, on a point of principle I decided to impute the costs using the mean rental costs of equivalent properties. To deal with this I used the following approach.
    - Extract all rentals with POA as their cost into a new data frame and removed the POA rows from the properties data frame.
    - Change the rental type in the properties data frame to INT.
    - I selected rental type and postal code as the best feature combination for finding comparable properties in the non POA data frame.
    - Created another data frame from the POA set and remove duplicate rental type and postal code combinations.
    - I iterated that set and got the mean rental cost of matching properties (by rental type and postal code) and then update the non-de-duplicated non-POA data frame using the same criteria.
    - Finally, I converted the rental cost to an INT and appended this revised non-POA set to the properties data frame. This resulted in a properties data frame with all costs as INTs and no missing properties due to cost issues in the data.

2.4.2 Rental Statistics

The rental statistics data frame needed to removal of euro symbols and commas from the cost columns and a conversion of those columns to numeric. The rental cost column in the properties data frame is numeric and later I would be doing calculations on the costs in both data frames hence the need to convert to numeric.

2.4.3 Venues

There was no cleaning required per-se. As mentioned in the venues method, I reduced the venues data frame to selected categories whilst creating the companies and venues data frame.

3. Exploratory Data Analysis

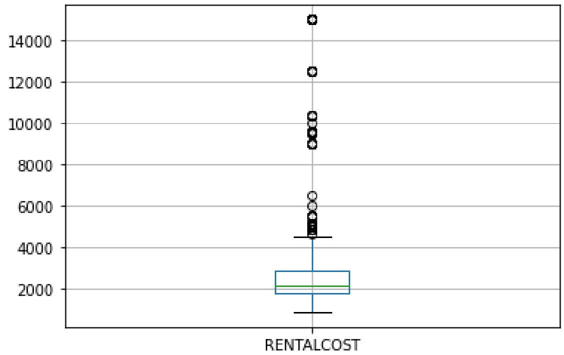
3.1. Companies and Rental Accommodation

The average rental cost for properties around companies was €2, 753 which is actually higher than the publicly reported average. Draft.ie publishes a quarterly report on rental price and availability changes and the 2021 Q2 report shows average rental costs for the south city at €2,111 and South County at €2,221. Given both companies reside with these two parts of the city, the mean rent of €2, 753 would appear to conflict.

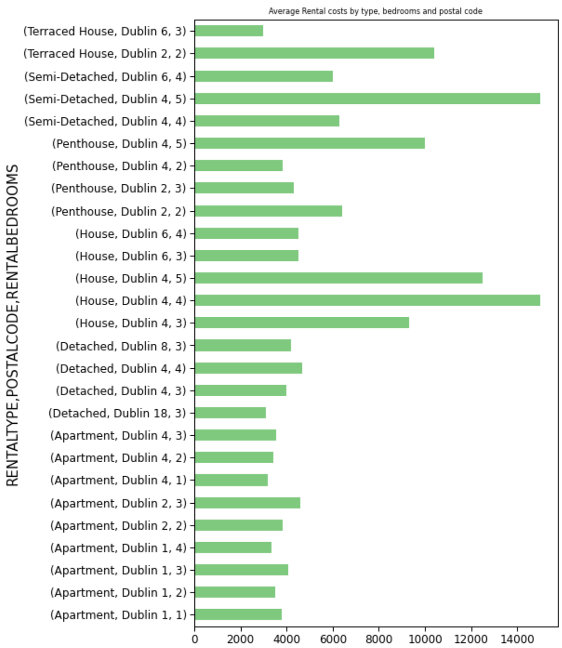
Merged companies and rental properties statistics

	COMPANYLAT	COMPANYLNG	RENTALLAT	RENTALLNG	RENTALCOST	RENTALBEDROOMS	RENTALBATHROOMS	RENTALDISTANCE
count	983.000000	983.000000	983.000000	983.000000	983.000000	983.000000	983.000000	983.000000
mean	53.335337	-6.242811	53.338209	-6.242947	2753.237030	1.858596	1.491353	1.228800
std	0.009657	0.009922	0.012651	0.014101	2137.392175	0.819409	0.830722	0.473615
min	53.268810	-6.249626	53.257847	-6.278848	900.000000	0.000000	0.000000	0.100000
25%	53.332500	-6.249499	53.331969	-6.251457	1800.000000	1.000000	1.000000	0.890000
50%	53.334907	-6.246058	53.340836	-6.238491	2150.000000	2.000000	1.000000	1.250000
75%	53.340627	-6.239397	53.345148	-6.233438	2900.000000	2.000000	2.000000	1.620000
max	53.341705	-6.196712	53.358351	-6.180413	15000.000000	5.000000	7.000000	2.000000

A boxplot of the dataset shows a considerable number of outliers in the 3k euro range which is skew the data.

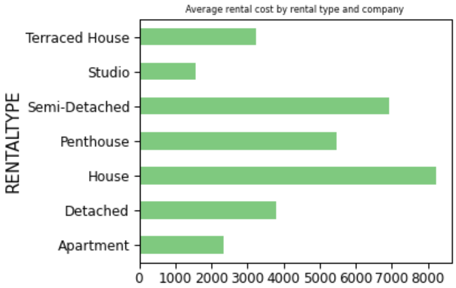


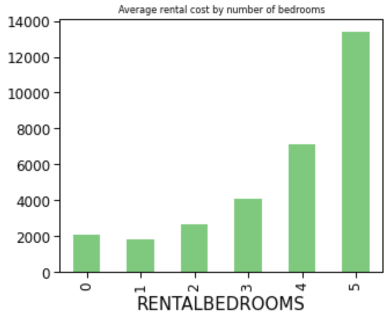
I wanted to understand what property types, number of bedrooms and postal districts these outliers belong to. To do that, I used the 75 rental cost percentile from the above descriptive statistics and filtered all properties whose rental cost is great than that value. Grouping the results by type, postal code and no. of bedrooms, it quickly became clear that Dublin 4, the most affluent part of the city, indeed entire country, to rent property was driving up the with average with houses in the 4- and 5-bedrooms range.



3.1.1 Relationship between rental cost and accommodation type

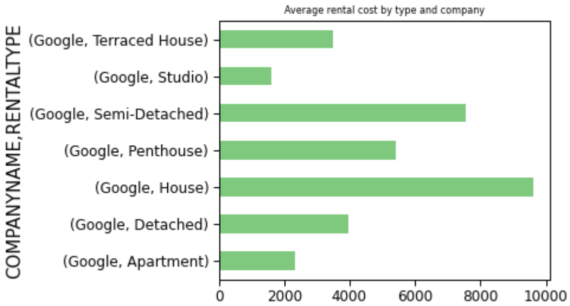
As expected, there is a strong relationship between rental type, number of bedrooms and rental cost.

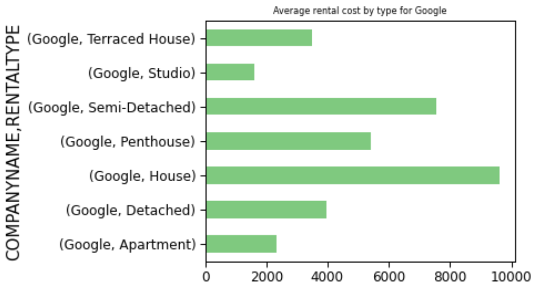




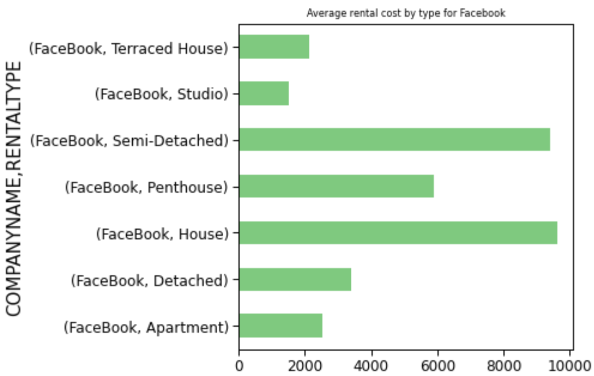
3.1.2 Relationship between average rental cost, type, and company

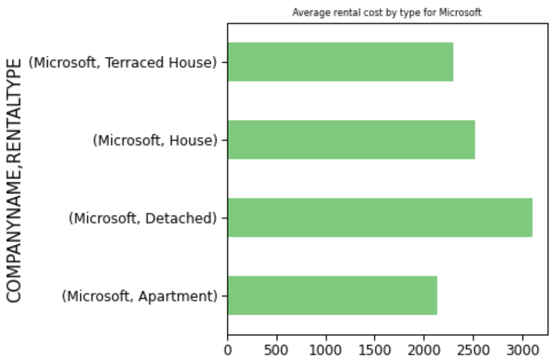
Apartments and studios are the cheapest to rent near Google with houses being the most expensive. It makes sense since Google is located by the Liffey River where there are few houses available but a growing demand for them, especially since working from home became the norm.





The patterns change the further away from the city you are. The Facebook building, I selected, is in an affluent residential area as is Microsoft headquarters both with a shortage of apartments.

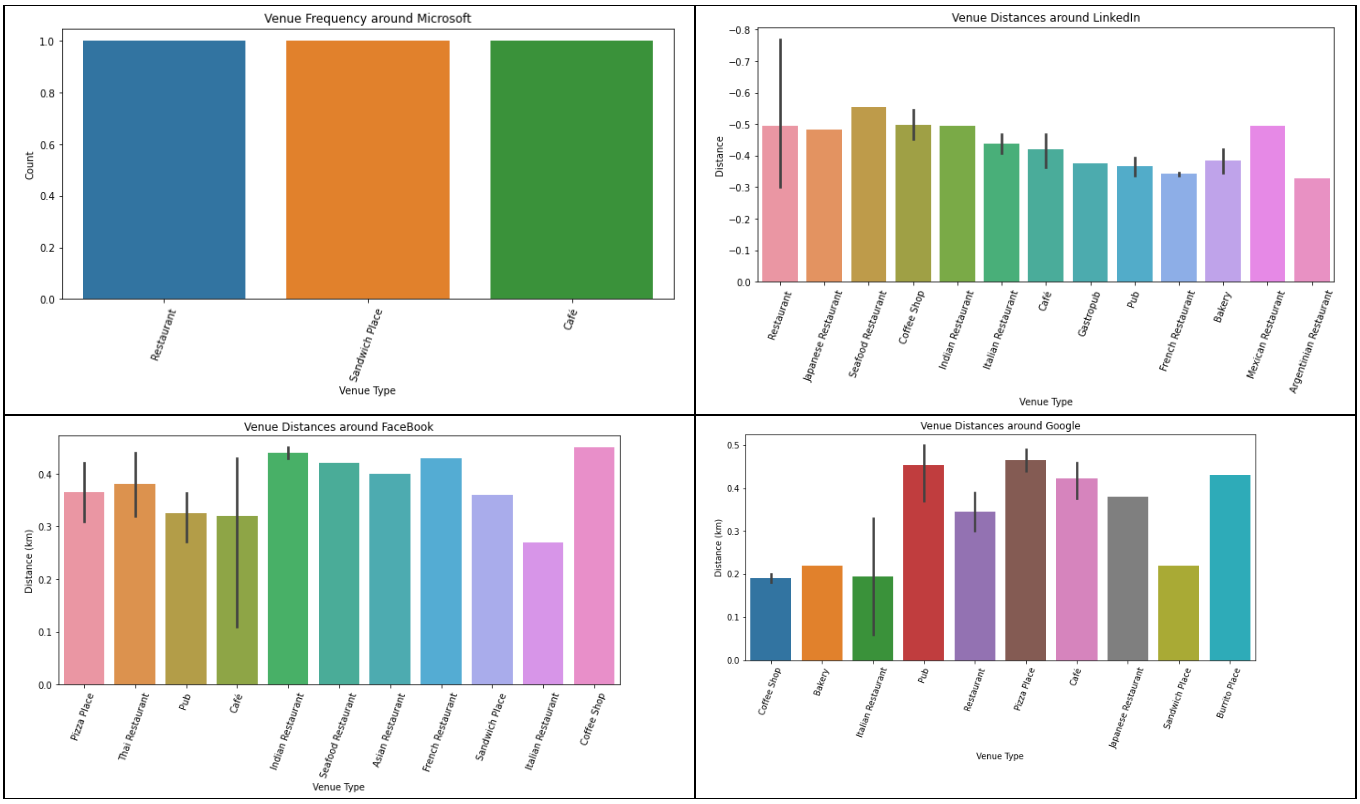




3.2 Companies and Venues

In total there were 68 unique venues but the target user for this project is interested in dining and not in Canals or ATMs. To reduce the choice to dining options, I created a list and added all the cherry-picked types related to eating out. I then used this list to drop non-related rows using the pandas ‘isin’ function.

I’ve taken a sample of companies in the city centre and suburban residential areas to show case both the distance to each venue from a given company and the variety of options. The Restaurant by Microsoft is, in fact it’s cafeteria.



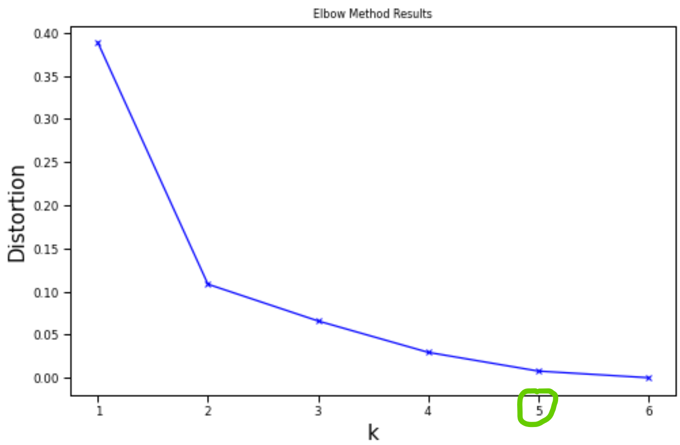
4. Modelling

I used the supervised learning algorithm k-means to cluster the venues arounds the companies.

- Got the frequency of venue types around each company and to do this using a one hot encoding.
- Grouped the results by company and get the average frequency occurrence.

	COMPANYNAME	Argentinian Restaurant	Asian Restaurant	Bakery	Bistro	Burrito Place	Café	Coffee Shop	French Restaurant	Gastropub	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	Pizza Place	Pub	Restaurant	Sandwich Place	Seafo Restaur
0	Amazon	0.032258	0.000000	0.000000	0.064516	0.032258	0.096774	0.129032	0.064516	0.032258	0.032258	0.032258	0.032258	0.000000	0.064516	0.225806	0.096774	0.000000	0.0000
1	FaceBook	0.000000	0.047619	0.000000	0.000000	0.000000	0.142857	0.047619	0.047619	0.000000	0.095238	0.095238	0.000000	0.000000	0.095238	0.238095	0.000000	0.047619	0.047
2	Google	0.000000	0.000000	0.043478	0.000000	0.000000	0.304348	0.130435	0.000000	0.086957	0.000000	0.086957	0.000000	0.000000	0.043478	0.217391	0.043478	0.043478	0.0000
3	LinkedIn	0.033333	0.000000	0.066667	0.000000	0.000000	0.133333	0.166667	0.066667	0.033333	0.033333	0.066667	0.033333	0.033333	0.000000	0.200000	0.100000	0.000000	0.033
4	Microsoft	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.400000	0.400000	0.0000
5	Twitter International	0.000000	0.000000	0.107143	0.000000	0.000000	0.142857	0.178571	0.107143	0.000000	0.000000	0.142857	0.035714	0.000000	0.000000	0.178571	0.107143	0.000000	0.0000

- Calculated the optimum k-means K value using the Elbow technique to determine where on the slope of the tangent line is almost flat and the distortion value was non-zero.



- Sorted the venues in descending order and select on the top 10 for each company.

	COMPANYNAME	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
0	Amazon	Pub	Coffee Shop	Restaurant	Café	Thai Restaurant	Bistro	French Restaurant	Pizza Place	Italian Restaurant	Japanese Restaurant
1	FaceBook	Pub	Café	Indian Restaurant	Pizza Place	Italian Restaurant	Thai Restaurant	Coffee Shop	French Restaurant	Asian Restaurant	Sandwich Place
2	Google	Café	Pub	Coffee Shop	Gastropub	Italian Restaurant	Bakery	Sandwich Place	Restaurant	Pizza Place	Argentinian Restaurant
3	LinkedIn	Pub	Coffee Shop	Café	Restaurant	Bakery	French Restaurant	Italian Restaurant	Argentinian Restaurant	Japanese Restaurant	Seafood Restaurant
4	Microsoft	Sandwich Place	Restaurant	Café	Argentinian Restaurant	Italian Restaurant	Seafood Restaurant	Pub	Pizza Place	Mexican Restaurant	Japanese Restaurant

- Ran k-means clustering on the grouped results with K = 5 and merged the clustered results with the top 10 companies and venues.

	COMPANYNAME	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	VENUEDISTANCE	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	9th Most Common Venue	10th Mo Comm Ven
44	Amazon	53.3325	-6.246058	Zakura Noodle & Sushi	53.333827	-6.244926	Japanese Restaurant	0.17	3	Pub	Coffee Shop	Restaurant	Café	Thai Restaurant	Bistro	French Restaurant	Pizza Place	Italian Restaurant	Japane Restau
45	Amazon	53.3325	-6.246058	Kerala Kitchen	53.333855	-6.244952	Indian Restaurant	0.17	3	Pub	Coffee Shop	Restaurant	Café	Thai Restaurant	Bistro	French Restaurant	Pizza Place	Italian Restaurant	Japane Restau
46	Amazon	53.3325	-6.246058	Searson's	53.333217	-6.243007	Pub	0.22	3	Pub	Coffee Shop	Restaurant	Café	Thai Restaurant	Bistro	French Restaurant	Pizza Place	Italian Restaurant	Japane Restau
47	Amazon	53.3325	-6.246058	Eathos	53.333699	-6.244699	Café	0.16	3	Pub	Coffee Shop	Restaurant	Café	Thai Restaurant	Bistro	French Restaurant	Pizza Place	Italian Restaurant	Japane Restau
48	Amazon	53.3325	-6.246058	Coffee 2 Go	53.333993	-6.245597	Coffee Shop	0.17	3	Pub	Coffee Shop	Restaurant	Café	Thai Restaurant	Bistro	French Restaurant	Pizza Place	Italian Restaurant	Japane Restau

4.1 Examining the Clusters

To decide how both companies and venues cluster, I created a dataset from the result of the k-means clustering merged with the 10 top venues, filtered on cluster number and grouped by venue popularity position. This was repeated for all five clusters.

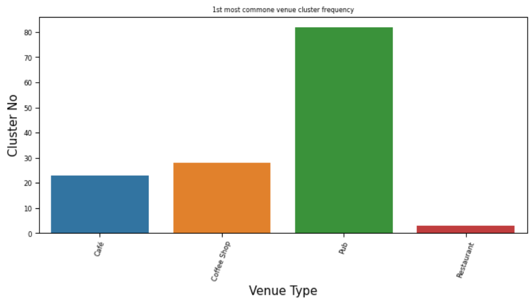
Summarising the clusters compared to the number of venue types, Cluster 1 offers the most variety with cluster 0 having the least.

```
Cluster 0 size: (5, 15)
Cluster 1 size: (58, 15)
Cluster 2 size: (21, 15)
Cluster 3 size: (31, 15)
Cluster 4 size: (23, 15)
```

4.1.1 Most common venues by company

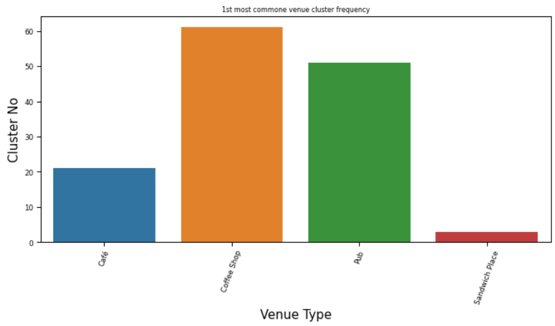
1st most common venue

1st Most Common Venue	COMPANYNAME	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	VENUEDISTANCE	Cluster Labels
Café	Google	23	23	23	23	23	23	23	23
Coffee Shop	Twitter International	28	28	28	28	28	28	28	28
Pub	Amazon	31	31	31	31	31	31	31	31
	FaceBook	21	21	21	21	21	21	21	21
	LinkedIn	30	30	30	30	30	30	30	30
Restaurant	Microsoft	3	3	3	3	3	3	3	3



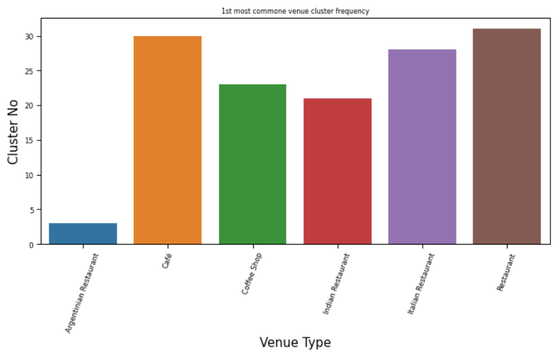
2nd most common venue

2nd Most Common Venue	COMPANYNAME	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	VENUEDISTANCE	Cluster Labels
Café	FaceBook	21	21	21	21	21	21	21	21
Coffee Shop	Amazon	31	31	31	31	31	31	31	31
	LinkedIn	30	30	30	30	30	30	30	30
Pub	Google	23	23	23	23	23	23	23	23
	Twitter International	28	28	28	28	28	28	28	28
Sandwich Place	Microsoft	3	3	3	3	3	3	3	3



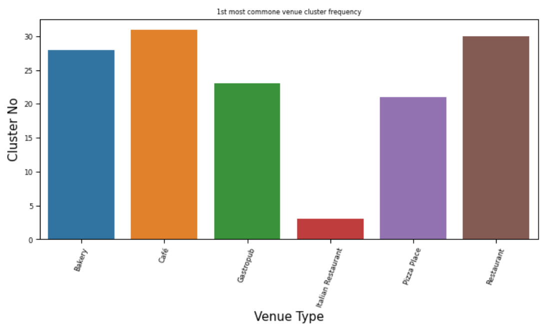
3rd most common venue

3rd Most Common Venue	COMPANYNAME	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	VENUEDISTANCE	Cluster Labels
Argentinian Restaurant	Microsoft	3	3	3	3	3	3	3	3
Café	LinkedIn	30	30	30	30	30	30	30	30
Coffee Shop	Google	23	23	23	23	23	23	23	23
Indian Restaurant	FaceBook	21	21	21	21	21	21	21	21
Italian Restaurant	Twitter International	28	28	28	28	28	28	28	28
Restaurant	Amazon	31	31	31	31	31	31	31	31



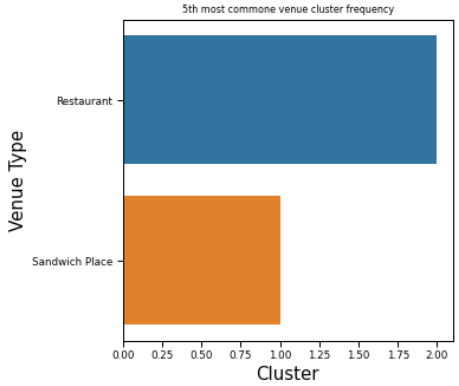
4th most common venue

4th Most Common Venue	COMPANYNAME	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	VENUEDISTANCE	Cluster Labels
Bakery	Twitter International	28	28	28	28	28	28	28	28
Café	Amazon	31	31	31	31	31	31	31	31
Gastropub	Google	23	23	23	23	23	23	23	23
Italian Restaurant	Microsoft	3	3	3	3	3	3	3	3
Pizza Place	FaceBook	21	21	21	21	21	21	21	21
Restaurant	LinkedIn	30	30	30	30	30	30	30	30

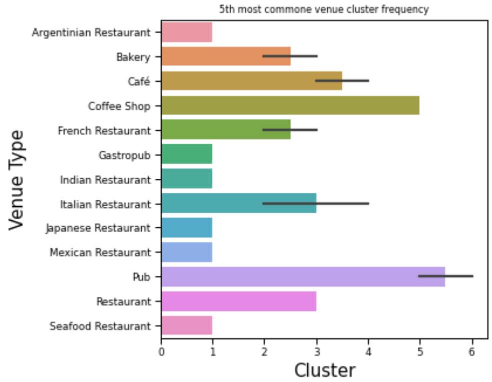


4.1.2 Venues and companies by cluster

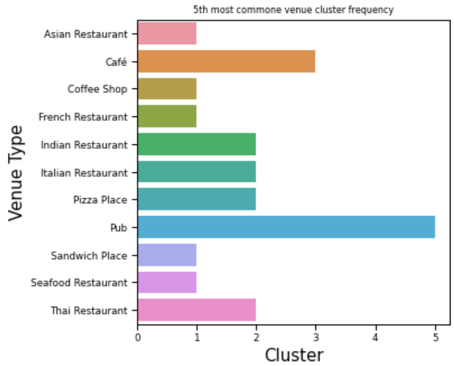
Cluster 0 (Microsoft)



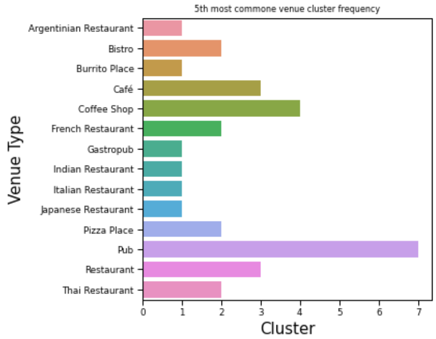
Cluster 1 (LinkedIn & Twitter)



Cluster 2 (Facebook)

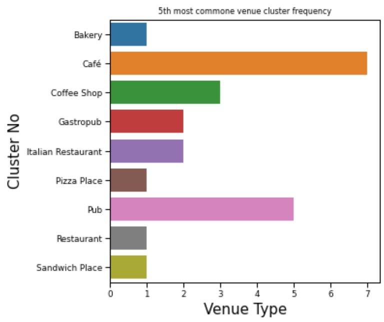


Cluster 3 (Amazon



)

Cluster 4 (Google)

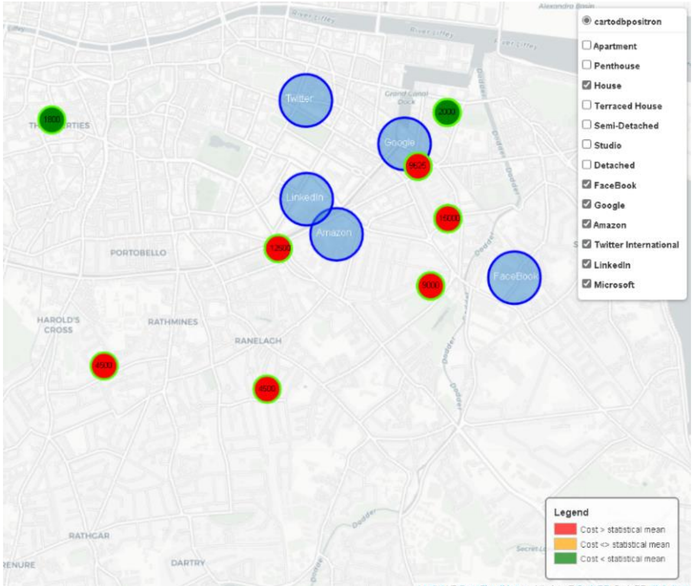


5. Mapping Visualisation

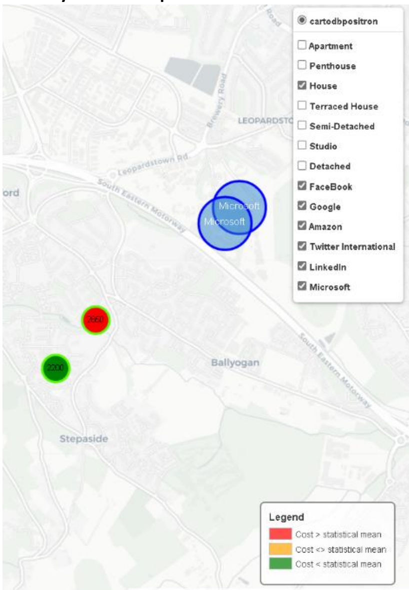
5.1 Companies and Rental Accommodation

I used folium to create maps with a combination of markers, circles tiles with rental type categories and a legend to explain the cost colour coding.

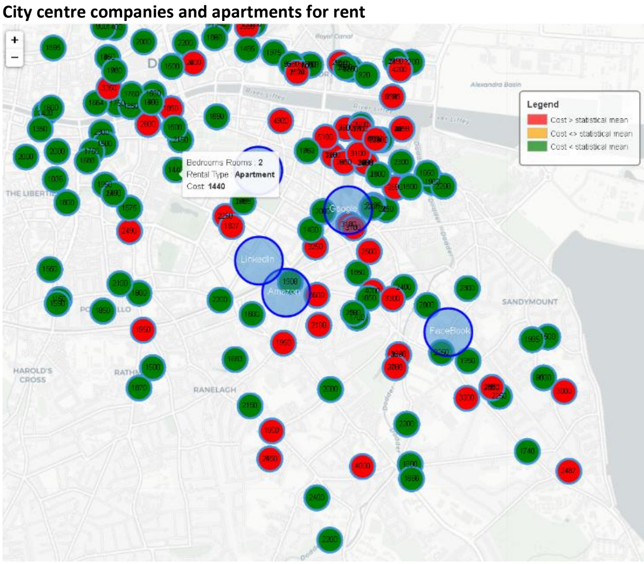
City centre companies and houses for rent



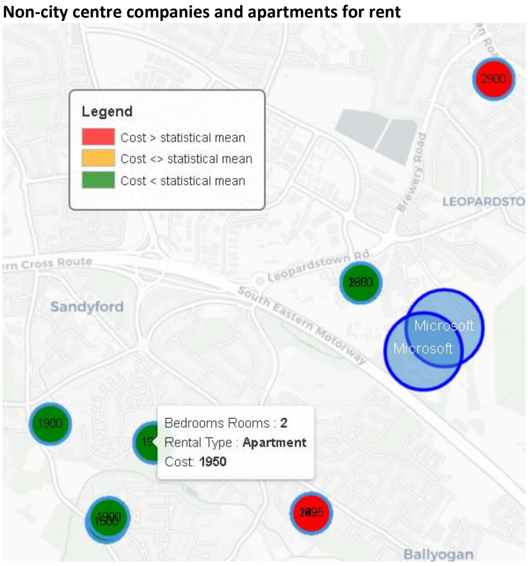
Non-city centre companies and houses for rent



City centre companies and apartments for rent

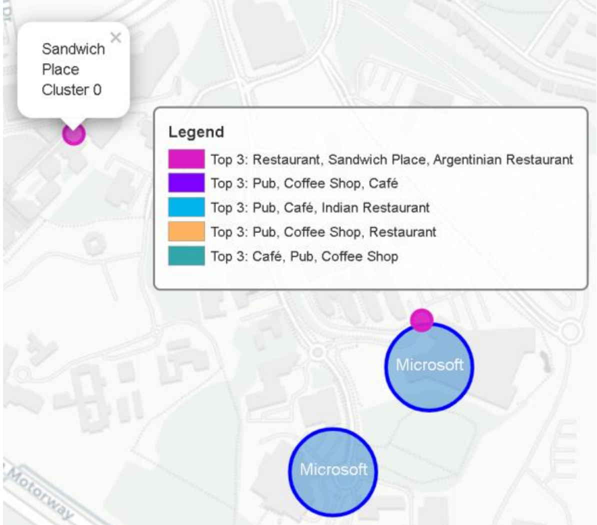


Non-city centre companies and apartments for rent

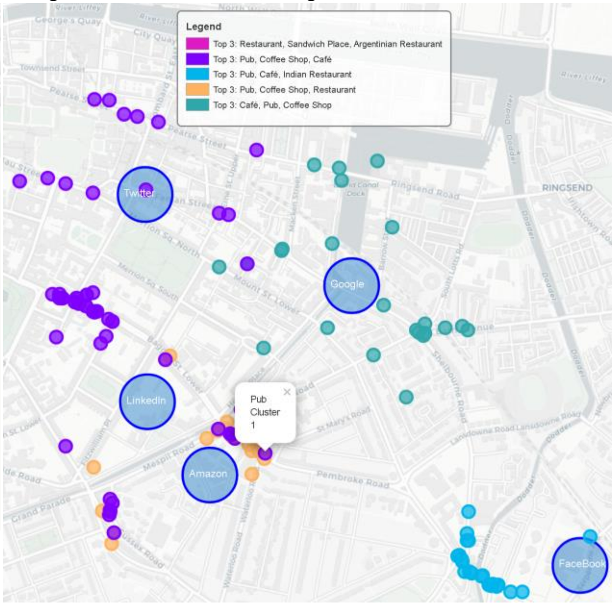


5.2 Venue Clusters around companies

Dining venues around Microsoft



Dining venues around Twitter, Google, Amazon, LinkedIn & Facebook



6. Conclusions

- Microsoft, being located well outside the city centre, offers the least dining varieties, rental options and highest rental costs. The results are unsurprising given its location at the edge of a business park close to the motorway and in an old dense urban area where building homes is underdeveloped. Whilst overall there are sufficient volumes of rental properties in the myhome.ie dataset, we may see more options around Microsoft if daft.ie data was available.
- Facebook fares a little better being closer to the city centre but again the choice of accommodation is limited by its location in an old urban area.
- Twitter, LinkedIn. Amazon and Google would be ideal choice of any employee looking for a wide availability of rental accommodation with prices being lower than suburban areas.
- The city centre offers the most variety of dining experiences with the most frequent venues being the pub followed by café and coffee shop. Not surprising for Dublin.
- LinkedIn, Twitter, and Amazon between them offer up to 59 dining venues with pubs and coffee shops making up the 1st most common venue. As you look at the 2nd to 10th most common venues, this cluster offers the highest cuisine variety ranging from Italian, French, Indian, Mexican, Japanese, Middle Eastern, Argentinian, Seafood, Pizza and Thai. In total, there are 16 distinct venue types.