Coursera Capstone

IBM Applied Data Science Capstone

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

By: Roger Clarke

May 2021



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

Introduction

Ireland is a European hub to over 1000 foreign direct companies in technology, pharma, social media, and many others. FDI (Foreign Direct Investment) in Ireland comprises around [**20%**](https://enterprise.gov.ie/en/What-We-Do/Trade-Investment/Foreign-Direct-Investment-FDI-/) 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](https://www.siliconrepublic.com/jobs/employment-fdi-multinationals-ireland-ida) thousand people. Dublin, the capital city of the Republic of Ireland, is known the world over as vibrant and a focal point of prominent levels of foreign direct investment especially in the tech sector.

Map

Description automatically generated

Young top talent coming out of universities in many countries are interested securing employment with these firms. However, Dublin is also well known for outrageous rental costs which may be deterrent to taking up employment. According to the Residential Tenacies Board ([RTB](https://www.rtb.ie)), the state body that regulates the industry, the average cost of renting a home in Ireland in 2021 is [**€1,745 per**](https://www.rtb.ie/research/average-rent-dataset) month with a 2.1% increase year on year.

Table

Description automatically generated

Problem

Companies such as Microsoft, Google, Facebook, Amazon, Twitter & LinkedIn which all based in Dublin, are always looking for skilled, educated and highly capable people from all over the world. This project is aimed at potential hires researching the cost of living in Dublin before they apply for advertised role.

The objective is to simplify the discovery of rental accommodation availability, type and cost and desirable social venues within a 2-kilometre radius of the potential employer's location.

Data Required

To cluster rental accommodation and social venues around specific employers in Dublin, the following data is needed:

**1 Companies’ names and address**

For this project I have selected Microsoft, LinkedIn, Amazon, Google, Twitter & Facebook all of which are based in Dublin and will use their business addresses to obtain their latitude and longitude.

**2 Rental Accommodation**

There are two primary sources of information for accommodation both rent and sale: <https://www.daft.ie> and <https://www.myhome.ie>. Daft.ie prohibits web scraping and does supply and API which requires payment for use. MyHome has no API, web scraping is permissible hence. This sole source of information on rental accommodation unfortunately represents only a portion of the entire market.

**3 EIRCODE**

An Eircode is Ireland's equivalent of a modern postal code. It’s a very precise unit represent the exact location of an address. Accommodation addresses from myhome.ie don’t always have an Eircode. For each instance, I will use [Eircode](https://www.eircode.ie) to get the code associated with the addresses which can then be used to get the exact latitude/longitude of the address in question.

**5 Foursquare**

Foursquare will supply the selected venue types, names and latitude/longitude within a pre-defined radius of each employer location.

# **Methodology**

**Data Retrieval, Cleaning, Preparation, and Feature Engineering**

**Employer’s address and location**

For this project, the five largest technology employers in Dublin have been selected (). I will use the Google maps API to get their full street address and latitude/longitude storing these details in a dataframe.

**Rental Accommodation**

Retrieving and preparing this data here the most complex part of the data retrieval phase. I will use BeautifulSoup from BS4 to scrap myhome.ie for rental accommodation in Dublin. To use myhome.ie, the user selects from a set of tabs (Buy, Rent, etc) and then selects the county/areas/price/etc from a set of query dropdowns. This could be automated using mechanise, but will be simplified by using myhome.ie to directly create the URL needed for Beautiful Soup:

*https://www.myhome.ie/rentals/dublin/property-to-rent*

The results are contained in a set of nested divs and lists grouped into units of 10. If you look on the site, only 10 properties are presented at a time. The div classes of interest are:

Each group is a set of divs with the class name PropertyListingCard\_\_Content. This div class has sub classes having all the features needed.

**Price**

Price is taken from the inner html text of 'div.PropertyListingCard\_\_Price' and provides the price and payment period which can be monthly/weekly. In general rent is paid monthly hence where the stated period is weekly, I will calculate the monthly repayment. The text contains also contains additional data such as “/ month”, /” week”, “/ ft2”, price increase/decrease “€150 on 5th Mar 18” we can be ignored. I will use regular expressions to each the data from text string.

**Address**

Address is taken from the inner html text of 'div. PropertyListingCard\_\_Address'. Working with addresses in Ireland is a little complex because it’s a mixture of postal codes, eircodes and neither. Postal codes are unique to the core of the city and don’t apply to specific areas outside. Eircodes were designed to find the precise location of each property yet there are many properties that lack them, for example Google’s head quarters in Dublin. I will need the exact latitude/longitude of each address and Eircode are helpful in dropping duplicate results especially for apartment in large groups of blocks. To detect a missing Eircode I will use a regular expression and, where the Eircode is missing, I will use the [www.eircode.ie](http://www.eircode.ie) site to search the address and retrieve the Eircode. This will be automated using mechanise and BeautifulSoup. Once the address is a precise as I can get it, I will use gmaps.geocode to query the latitude and longitude for the given address.

**Number of Bedrooms/Bathrooms/Type/Energy Rating**

The property class “PropertyListingCard\_\_Address” contains a list of divs with the class “PropertyInfoStrip\_\_Detail PropertyInfoStrip\_\_Detail…”. There four in the four containing the number of bathrooms/bedrooms/propertype and energy rating. I will extract the innerhtml text of each to get the values.

**Navigating the myhome.ie query results.**

The query results have a class called DesktopPagination. Within this class is a list of page numbers which are displayed on the site as boxes. I will get the pages using soup.find\_all( class\_ = "DesktopPagination" ) which give me a list of pages. I can then select the first page and get a list of numbers by using pages[0].select('li.small-screen'). The numbers are in a string separated by a ‘/’. Using the python split function, I can then create a range of page numbers using plist[0].text.replace(' ', '').split('/'). Using this range, I can extend the myhome URL query to add a query parameter “page=” and append the range value as appropriate. For example:

*https://www.myhome.ie/rentals/dublin/share/shared-accommodation-in-dublin-north?page=2*

**Feature Selection**

Employer name

Employer Address

Employer Latitude

Employer Longitude

Property Cost

Property Address

Property Type

Property No Bedrooms

Property No Bathrooms

Property Energy Rating

Property Latitude

Property Longitude

Property Distance from Employer

Venue Name

Venue Type

Venue Distance from Employer

[Exploring Rental Rates in Canada — Hands-on Project making use of Foursquare API with Python | by Muhammad Al-Fairuz | Medium](https://medium.com/@backupfairuz/exploring-rental-rates-in-canada-hands-on-project-making-use-of-foursquare-api-with-python-413793d2dfca)

NaN values

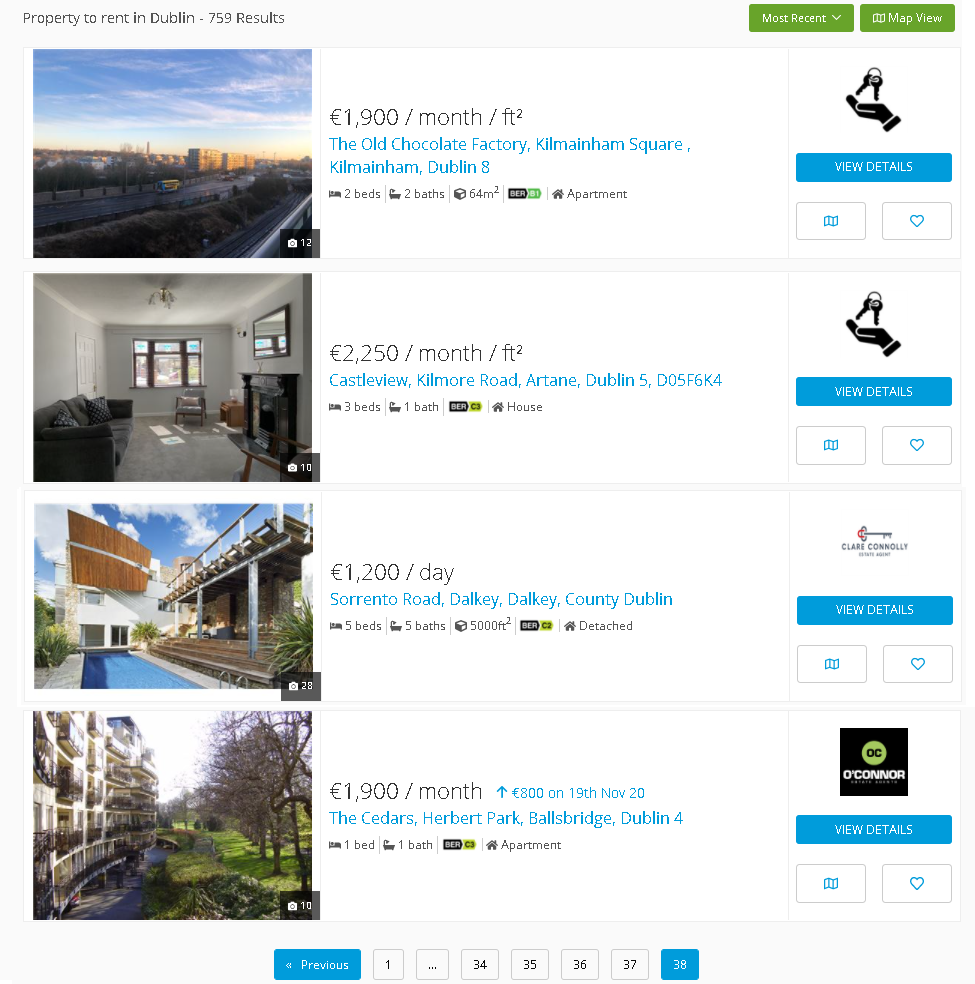
Preview folium employers

Calculate distance

Visualise data

Getting the vendues of each rental accommodation using Foursquare API

<image





Companies’ names and address

Eircode

Mechanise

#### Feature Selection

Rental Type, Cost, Eircode, Latitude, Longitude

Exploratory Data Analysis

Conclusions