Peer-graded Assignment: Capstone Project - The Battle of Neighbourhoods

Subject: Brexit based comparison of London and Dublin

Introduction/Business Problem

The only known fact about Brexit is that it is happening. Nobody can tell all the implications it will have. Currently London city still plays the role of a finance system capital of Europe, but that can change due to Brexit. Financial institutions might prefer to stay within the European Union instead of staying in London. When considering a move of office locations companies need to take into consideration many different variables. One of those is how specific locations are perceived by potential employees they would like to attract. One of the cities being taken into consideration by financial institutions is Irelands Dublin – it's close to London, it's in the EU, it's in an English-speaking country. But will it be attractive enough for potential employees?

Therefore, many of the currently London based financial institutions would find an analysis of similarities between attractiveness of London and Dublin to potential employees not only interesting, but necessary.

Such similarities would include number of restaurants, theatres, cinemas, clubs, sport related places per square kilometre – after all employees need to have somewhere to spend their salaries in. Since number of venues and their ratings usually differ in between different neighbourhoods of each city – more theatres in the city centre and so on – such an analysis would need to be based on comparison not only between whole cities, but also their neighbourhoods.

Financial sector employees won't be interested only in leisure activities in their neighbourhoods. Therefore, financial institutions will also be interested in comparison between whole city parameters such as Safety, Health Care, Cost of Living, Traffic Commute Time and Pollution to get a good basis for a decision about potential move of quarters to Ireland.

Audience**:** London based financial institutions

Data

Both London and Dublin are huge cities with many different neighbourhoods so as a first step I would look at the administrative split of London and Dublin to find the ones that are of interest – city centres, office areas, trendy living areas. So the areas with most venues.

- London data (2014) about boroughs and neighbourhoods is available under the following Wikipedia link https://en.wikipedia.org/wiki/List_of_London_boroughs
- Dublin data about administrative division, based on electoral division, is available in CSV format under the following link https://data.smartdublin.ie/dataset/d80dbc39-7999-4715-a56f-f2ae9c3f5b25/resource/b5d71c3f-3c8c-491c-b3a8-a3d94c42980a/download/fccelectoraldivisionsp20110901-1705.csv

After knowing how to split London and Dublin into smaller neighbourhoods I would base the selection of the ones of interest using Foursquare location data in terms of numbers of restaurants, theatres, cinemas, clubs, sport related places per square kilometre.

After preparing a comparison of neighbourhoods in terms of leisure time I'd move to analysing the comparison of between the whole cities using the data available under the following link

https://www.numbeo.com/quality-of-life/compare_cities.jsp?country1=United+Kingdom&country2=Ireland&city1=London&city2=Dublin
To do that I'd use the BeautifulSoup website scraping library.

Methodology

1. ### Libraries:
    - Pandas – for using dataframes
    - Numpy – for calculations
    - Requests – for importing data from websites
    - BeautifulSoup – for importing data from websites
    - Matplotlib – for generating plots
    - matplotlib.pyplot – for generating plots
    - json_normalize – for tranforming json file into a pandas dataframe library

2. ### Encountered problems and their remediation:
    - London data downloaded from Wikipedia initially caused issues with the encoding. One of the columns had non-visible additional string of characters '\ufeff'. This has been found when using the Latitude values for creating URL to be used in querying Foursquare. The mentioned string of characters was included before the numerical value of Latitude in the URL. Therefore, I had to delete that string of characters from the whole Latitude column of the London geolocation dataframe.
    - After running several parts of the Python code a message "A value is trying to be set on a copy of a slice from a DataFrame." Is being thrown out by Python. I have validated that this does not affect any of the calculations.
    - While using Foursquare and trying to assign relevant part of JSON to venues for London an unexpected error occurred - KeyError: 'venues'. Though the exact same code was earlier used for Dublin and worked fine. While tracing where the problem lies it was found that when sending the GET Request an error appeared 'code': 429: 'quota_exceeded'. That was because of the limitations of the version of Foursquare account. After upgrading the Foursquare account to Personal, the daily limits were increased, and problem stopped occurring.

3. ### Explanation of each analysis step:
    A. Quality of Life data

    A.1. The project assignment was started with importing the Quality of Life data from Numbeo website. That data included the specific ratios for both London and Dublin that were supposed to be analysed.

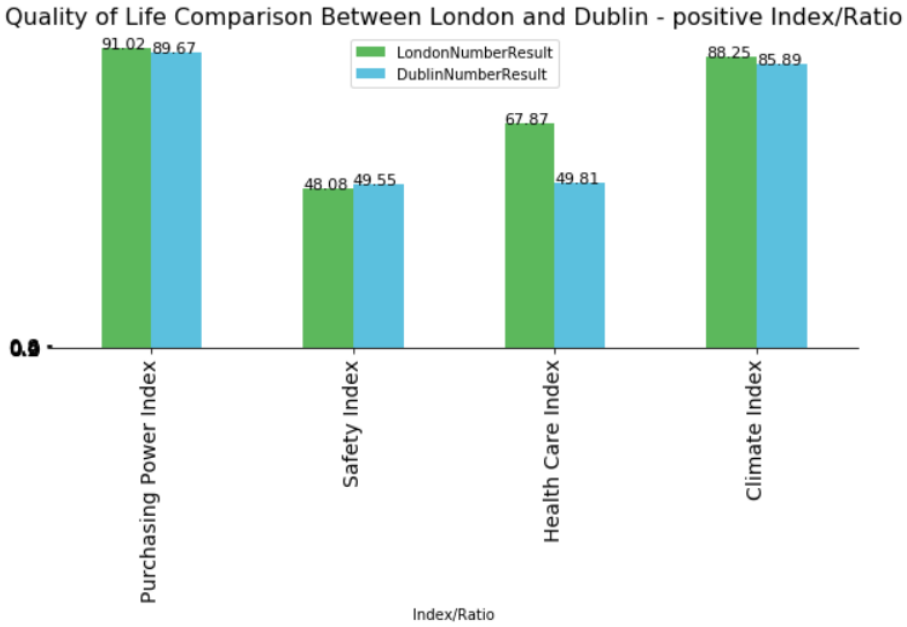
    A.2. Some of the ratios and indexes among this data are negative – the higher the value the worse for inhabitants. Therefore, I have separated the negative indexes/ratios into a separate table ang given them minus values. This way when presenting the data on charts it's clear that those are negative.

    The Negative ones are:

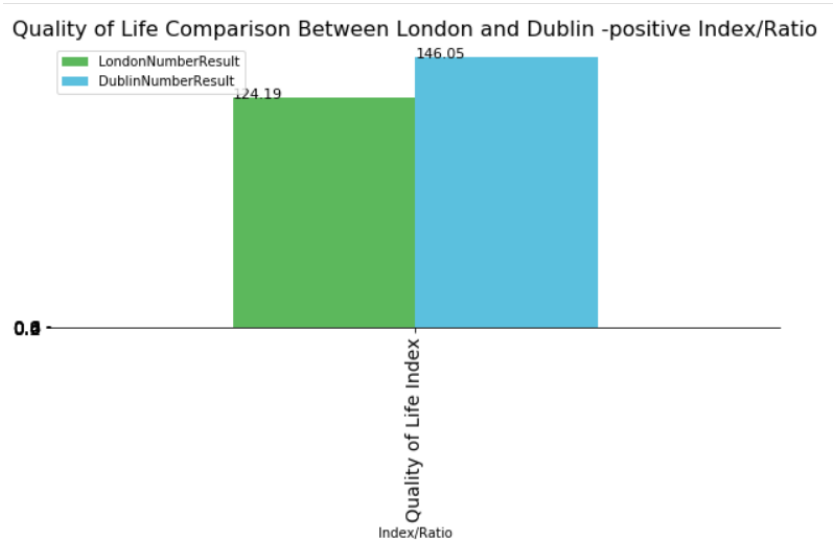
    - Cost of Living Index
    - Property Price to Income Ratio
    - Traffic Commute Time Index
    - Pollution Index

The Positive ones are:

- Purchasing Power Index
- Safety Index
- Health Care Index
- Climate Index



A.3. That data also included an overall Quality of Life Index



A.4. Analysis of the data shown that though in case of the positive indexes and ratios London has better values than Dublin – especially on the Health Care Index – it has much worse values when it comes to the negative indexes and ratios. Therefore, the overall Quality of Life index clearly shows that Dublin is a better place to live than London – of course when taking those indexes and ratios into consideration. It was mainly the Property Price to Income Ratio, Traffic Commute Time Index and Pollution Index that seem to cause that. One possible explanation would be that London is a much bigger city – but that will be explained in the next part of this paper.

A.5. Main takeaway – Dublin is a better place to live than London

B. Administrative split data

B.1. The data about London administrative split of London was downloaded from Wikipedia and populated into a pandas dataframe. The data had to be cleaned up and prepared for analysis. The area column had to be recalculated from square miles into square kilometres to be comparable to the later used data about Dublin.

B.1.1. When finalizing this part of the London data exercise the dataframe had following columns: Borough

- Area_KM2
- Population (2013 est)
- Latitude
- Longitude

B.2. The data about Dublin administrative split of London was downloaded from data.smartdublin.ie website and populated into a pandas dataframe. The data had to be cleaned up and prepared for analysis.

B.2.1. When finalizing this part of the exercise the dataframe had following columns: Borough

- Name
- Latitude
- Longitude
- Area_KM2

C. Foursquare data

C.1. Foursquare data regarding number of venues - restaurants, bars, theatres, clubs, sport related places - was needed for each area of both London and Dublin.

Therefore, it was necessary to download it using FOR loops. The Foursquare data was then added to the London dataframe and to Dublin dataframe respectively (from point B above). Also, a sum of venues column was added – that is because while for some potential employees' number of restaurants might be the most important for others it might be theatres or sport centres. Therefore, it was decided that it's best to look at the sum of venues. Afterwards another column with the sum of venues per square kilometre was added to take into consideration the large difference between the sizes of London and Dublin areas – as also mentioned in section A.4. above.

Top five areas in London:

	Borough	Area_KM2	Population (2013 est)	Latitude	Longitude	Nr of restaurants	Nr of bars	Nr of theatres	Nr of clubs	Nr of sport	Sum of venues	Sum of venues per KM2
18	Kensington and Chelsea	12.121144	155594	51.5020	-0.1947	30.0	30.0	10.0	30.0	10.0	100.0	8.250046
17	Islington	14.866532	215667	51.5416	-0.1022	30.0	30.0	7.0	16.0	2.0	83.0	5.583010
5	Camden	21.755900	229719	51.5290	-0.1255	30.0	30.0	30.0	30.0	21.0	120.0	5.515745
11	Hammersmith and Fulham [note 4]	16.394625	178685	51.4927	-0.2339	25.0	30.0	6.0	28.0	7.0	89.0	5.428609
31	Westminster	21.471001	226841	51.4973	-0.1372	30.0	30.0	13.0	30.0	13.0	103.0	4.797168

Top five areas in Dublin:

	Name	Latitude	Longitude	Area_KM2	Nr of restaurants	Nr of bars	Nr of theatres	Nr of clubs	Nr of sport	Sum of venues	Sum of venues per KM2
13	BLANCHARDSTOWN-ROSELAWN	53.385	-6.38	0.47	2.0	14.0	1.0	1.0	5.0	23.0	48.936170
39	SWORDS VILLAGE	53.460	-6.22	0.92	5.0	23.0	NaN	6.0	6.0	40.0	43.478261
41	TURNAPIN	53.400	-6.24	0.94	6.0	7.0	NaN	3.0	7.0	23.0	24.468085
11	BLANCHARDSTOWN-DELWOOD	53.380	-6.39	1.01	1.0	10.0	NaN	NaN	1.0	12.0	11.881188
9	BLANCHARDSTOWN-COOLMINE	53.390	-6.40	2.75	1.0	19.0	1.0	1.0	7.0	29.0	10.545455

C.2. The analysis showed that though the number of different venues is much higher in London than in Dublin, they are not as widely spread around the areas and therefore can be less accessible. Taking into relation point A.2. above (the Traffic Commute Time Index which is much worse for London) a conclusion comes that the many venues in London might be too far away for people to visit them.

C.3. Main takeaway – Dublin seems to be a better place to live than London

D. Conclusions about use of further analysis methods

D.1. Initial assumption was that London and Dublin both being capital cities of large EU countries would have more similarities between their neighbourhoods. While comparing the size of specific neighbourhoods and the number of their venues it came to light that such an assumption was wrong. Therefore, the initial plan to use Machine Learning/Clustering for finding clusters of similar neighbourhoods in London and Dublin had to be abandoned. Such an analysis could lead to faulty conclusions by clustering London suburban areas together with Dublin centre areas.

D.2. Python toolset also allows great ways of visualizing geolocation data on maps using folium library. For the same reason of scale as explained in section D.1. such visualization would only create confusion since visualizing Dublin and London on maps of the same scale would make Dublin blurry with venues very close to each other, while London would seem to have a little amount of venues due to the wider area, while the truth is London has much more of those. Therefore such visualization was abandoned as well.

D.3. Therefore it was decided that a graph analysis of the ratios/indexes of Quality of life together with a comparison of number of venues per square kilometre give the best overview needed for drawing conclusions

Results

The main takeaway of the assignment is that Dublin is a better place to live than London and therefore should attract potential financial sector employees.

Though London is a much larger city than Dublin and has much more venues those are less available to city inhabitants – details in sections A.4. and C.1.

As presented in section A it is easier to purchase property in Dublin than it is in London, which would make it easier for potential employees to move to that city from other places for a job in the financial sector.

Besides property prices also the overall cost of living is significantly smaller in Dublin than in London. That indicates that the same expenses spent by financial institutions on salaries would have a more positive perception for employees working in Dublin than in London.

Section A also indicates that employees would find that working in Dublin makes them waste much less time on commuting to work leaving them more free time for enjoying life which would also be healthier due to cleaner air – the effect of a healthier life should outweigh the worse healthcare services in Dublin.

The main deficiency of Dublin in comparison to London is its size. Dublin is much smaller than London. This might indicate that there is a limitation to how many businesses and people can still move into Dublin.

Discussion

Based on the analysis results it is advised that Europe wide (or wider) Financial Institutions move their offices from London to Dublin to remain in the European Union. It is also advised for Financial Institutions to consider their move to Dublin as soon as possible due to the limitations of space a much smaller Dublin might have. There is of course a potential for Dublin to grow larger, but before that happens the prices of property might increase.

Conclusion

The undertaken analysis has proven that Dublin is a place worth considering a move of offices to from London for Financial Institutions. The analysis has been performed mainly from a perspective of attracting employees. That is because currently the European market is clearly an employee one – financial sector included. The financial industry requires skilled employees and those are not easy to find and therefore every financial institution must consider attracting them.

There are of course also other perspectives that financial institutions should be taking into consideration such as the political situation of Ireland and European Union as such. Breixit is a huge unknown and even during the course of preparing this analysis the situation has been changing. Perhaps despite Brexit London will remain under enough EU regulations for Financial Institutions not to move their offices elsewhere. That remains an unknown and experience of many recessions in the past show that uncertainty is not good for the financial industry. Dublin seems to be a good potential safe harbour for the financial industry.