IBM

Applied Data Science Capstone

The Battle of Neighborhoods

Introduction:

The city of Edinburgh, a capital of Scotland is one of the most beautiful cities in the world, and on top of the most targeted places by tourists and pubs.

Fig.1. City of Edinburgh, Scotland.

Source:

https://fr.wikipedia.org/wiki/%C3%89dimbourg#/media/Fichier:EdinburghMontage.png

"The steps bar" available in Glasgow, Scotland, is about to open a second business in the capital, the beautiful city of Edinburgh. Meanwhile, their bar in Glasgow has always been a success in term of income due to its perfect location. So, the owners want the same success story within their second adventure, and would like to keep the same
strategy by opening the bar in a crowded area surrounded by coffee shops, restaurants, other bars and all that attracts people daily. So according to their needs, the problem is to find the best location for a pub in the city of Edinburgh by collecting and analysing Geo location data about places and the venues they contain.

Target audience

The targets for this project are “the steps bar” owners or any other bar, pub or restaurant’s owners who are planning to move to Edinburgh.

Data Section:

To solve the problem discussed above, we will first start by collecting data about the place’s names in Edinburgh with their corresponding Postal codes (for future needs) from this Wikipedia page (https://en.wikipedia.org/wiki/EH_postcode_area). One we get this first data set; we use the places postal codes to get the venues and business present in each place along with their corresponding Latitude and Longitude coordinates from the Foursquare API. To finally regroup them and extract clusters from this last. And by analysing the clusters with their locations we will be able to determine which place contains the most of pubs and restaurants and would suits the best our "Steps bar".

Edinburgh EH1 to EH17 & EH28 to EH30 [edit]

Postcode district	Post town	Coverage	Local authority area
EH1	EDINBURGH	Mostly consists of Edinburgh's Old Town. Also hosts the old GPO building (at EH1 1AA) and the areas immediately to the north of this are also included, that is St. James Centre and the areas down Leith Street and Broughton Street.	
EH2	EDINBURGH	The New Town and central commercial area of Edinburgh which includes Princes Street.	
EH3	EDINBURGH	An odd shaped area surrounding EH1 and EH2 to the north, west and south. It can be divided into three distinct areas: • The New Town area, north of Queen Street and northwards to Stockbridge. • The West End, including some of Edinburgh's financial district • The Tollcross and Fountainbridge areas.	
EH4	EDINBURGH	Radiates from the older and more central areas out to the suburban areas added to them, as the city grew outwards during the 20th century. EH4, then, branches out from Dean Village and Cornely Bank along a corridor centred on the A90 as far as Barnton and Cramond. Central to this area is the Sainsbury's in Craigleith Retail Park adjacent to the A90.	
EH5	EDINBURGH	Based on a village formerly separate from the city; in this case, Granton on the Firth of Forth. EH5 extends inwards from the coast to Ferry Road.	
EH6	EDINBURGH	Covers Leith, as well as Newhaven bordering it on the west.	
EH7	EDINBURGH	The inner city area between central Edinburgh and Leith, and radiates out to Restalrig and Craigentinny.	
EH8	EDINBURGH	The inner city Southside, Newington and Canongate areas, in the west of the postcode area. It spreads eastwards around Holyrood Park to take in areas like Abbeyhill and Mountcastle; being in east Edinburgh these areas have no obvious connection with the Southside. However the postcode effectively takes in the area surrounding Holyrood Park.	
EH9	EDINBURGH	The inner city, Marchmont and Grange, Blackford and around Minto Street, including Causewayside.	
EH10	EDINBURGH	A corridor along the A702 from Bruntsfield, through Morningside as far as Fairmilehead.	City of Edinburgh, Midlothian
EH11	EDINBURGH	A corridor (rather thin in shape) along the A71 from Haymarket Station, through Gorgie and Stenhouse to Sighthill and the Calders.	
EH12	EDINBURGH	A corridor along the A8 from Haymarket through Murrayfield and Corstorphine as far as the Gyle.	
EH13	EDINBURGH	Based on the previously separate village of Colinton, and including Oxgangs.	
EH14	BALERNO, CURRIE, EDINBURGH, JUNIPER GREEN	A corridor in south-west Edinburgh starting at Slateford, through Longstone, Wester Hailes, Juniper Green, Currie and on to Balerno.	
EH15	EDINBURGH	Based on Portobello and Duddingston, formerly separate settlements to Edinburgh.	
EH16	EDINBURGH	Based on the formerly separate village of Liberton. It extends north and north-east to Cameron Toll, Craigmillar and Niddrie.	
EH17	EDINBURGH	Based on the formerly separate village of Gilmerton, taking in Moredun and extending westwards as far as Mortonhall.	
EH28	NEWBRIDGE	Newbridge, Ratho	City of Edinburgh Council
EH29	KIRKLISTON	Kirkliston	City of Edinburgh Council
EH30	SOUTH QUEENSFERRY	South Queensferry	
EH77	EDINBURGH	Special postcode used by Census 2021	
EH91	EDINBURGH	Special postcode used by *Jobcentre Plus*	
EH95	EDINBURGH	Special postcode used by Scottish Gas.	
EH99	EDINBURGH	Special postcode used by the Scottish Parliament.	

Fig.2. Data set 1 scrapped from Wikipedia page.

Source: https://en.wikipedia.org/wiki/EH_postcode_area

Methodology:

The first step consists of scrapping the data from the Wikipedia web page shown above using the BeautifulSoup package. These data do not contain the latitude and longitude coordinates of the places so we had to extract them by requesting another package called Geocoder. The picture below shows the geo data plotted on the map of Edinburgh.

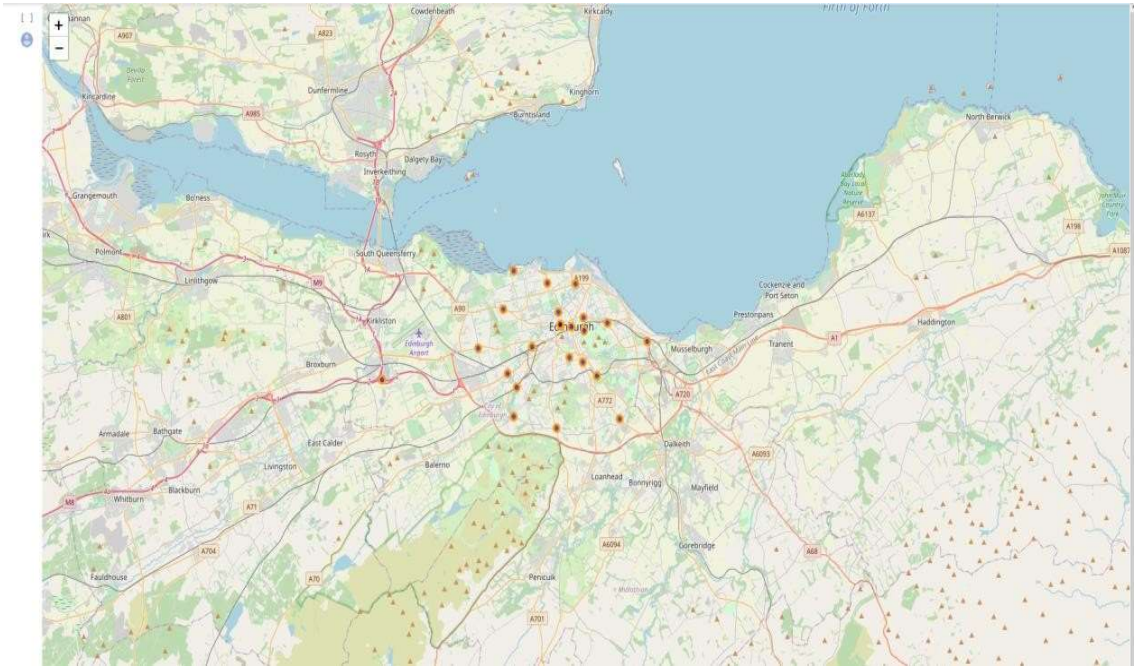

Fig.3. Geo data plotted on Edinburgh map.

Now we have got the places along with their coordinates but this data is still useless for our problem. The interesting part is to use it by requesting the Foursquare API to get a maximum of 500 venues within a radium of 1000 meter for each place. We get then a total of 1109 venues including bars, restaurants, parks, hotels … along with their corresponding Latitude and Longitude plus the Venue Category. Picture below shows the 5 first rows from our data set.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mostly consists of Edinburgh's Old Town. Also ...	55.95309	-3.18899	Dishoom	55.953726	-3.192540	Indian Restaurant
1	Mostly consists of Edinburgh's Old Town. Also ...	55.95309	-3.18899	The Balmoral Hotel	55.953113	-3.189509	Hotel
2	Mostly consists of Edinburgh's Old Town. Also ...	55.95309	-3.18899	SCOTCH Whisky Bar	55.953062	-3.190095	Whisky Bar
3	Mostly consists of Edinburgh's Old Town. Also ...	55.95309	-3.18899	Apple Princes Street	55.953354	-3.189947	Electronics Store
4	Mostly consists of Edinburgh's Old Town. Also ...	55.95309	-3.18899	The Milkman	55.950650	-3.191010	Coffee Shop

Fig.4. List of some venues extracted using Foursquare API.

We have got 173 different categories of venues in our data set.

Next step in to analyse this data to extract some useful information to get to solve our problem. We proceed then by regrouping the data by category of venues and computing for each place (or Neighborhood) the frequency of having each category of venue. This analyse gives us a first look on the places that have more venues than others.
Then we sort the most common venues for each place (as show in picture 5) to apply some machine learning on it.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	A corridor (rather thin in shape) along the A7...	Supermarket	Grocery Store	Fish & Chips Shop	Gas Station	Coffee Shop
1	A corridor along the A702 from Bruntsfield, th...	Convenience Store	Pub	Outdoor Supply Store	Mediterranean Restaurant	Pizza Place
2	A corridor along the A8 from Haymarket through...	Grocery Store	Pub	Gym / Fitness Center	Supermarket	Chinese Restaurant
3	A corridor in south-west Edinburgh starting at...	Supermarket	Coffee Shop	Grocery Store	Gas Station	Tennis Court
4	An odd shaped area surrounding EH1 and EH2 to ...	Café	Bar	Pub	Coffee Shop	Bakery

Fig.5. Most common venues per neighbourhood.

To answer the main question of this project which is “what is the best place to open a bar in Edinburgh” and using the data set we have built, the ideal would be to cluster our venues grouped by their category and plot the result on the map of the city to visualize where should “The steps bar” open in Edinburgh. Unsupervised machine learning is used with the Kmeans algorithm to get 8 clusters from our data.

Results:

After pre-processing the data and applying a Kmeans clustering we have got the following results along with the map of the clusters:

- Cluster n°1 has 14 group of neighbourhoods.
- Cluster n°2 has 2 group of neighbourhoods.
- Cluster n°3 has 1 group of neighbourhoods.
- Cluster n°4 has 1 group of neighbourhoods.
- Cluster n°5 has 1 group of neighbourhoods.
- Cluster n°6 has 3 group of neighbourhoods.
- Cluster n°7 has 1 group of neighbourhoods.
- Cluster n°8 has 1 group of neighbourhoods.

The cluster 1 is the one that interests us because it contains mainly pubs, bars, restaurants and all kind of businesses that creates a crowded atmosphere and attracts a lot of people, and according to the map (picture 6 below) we can see that the cluster 1 (red color) is situated in the city center of Edinburgh, which is obvious.

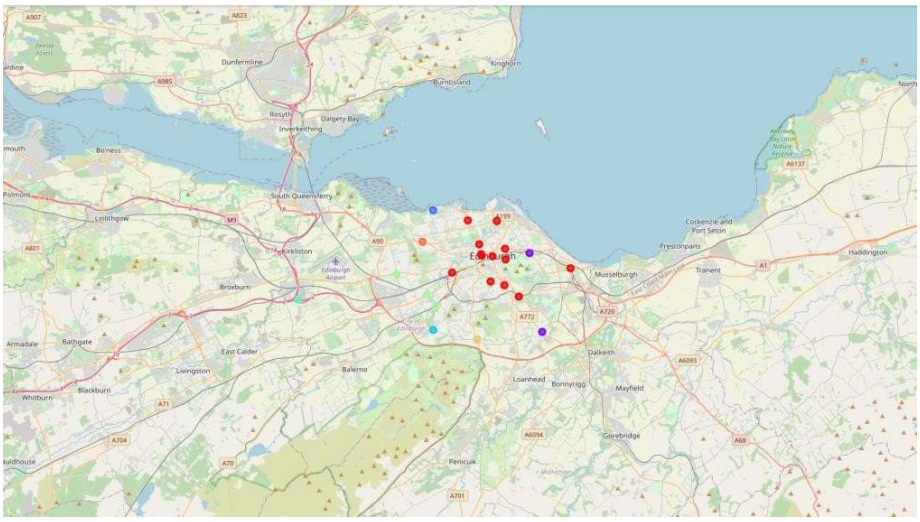

Fig.6. Clusters location in Edinburgh.

Conclusion:

This project allowed us to determine which area is the best for a specific business (a bar in our case), it can for sure be used for any other category of venue in any other city. We have seen how by using data science we can answer some questions to solve complex problems and provide a deep study. A possible further work for this project would be comparing the frequencies of pubs and restaurants in different cities in the UK and give some recommendations to the bars owners where it would be beneficial for them to open next.