The battle of the Neighbourhoods – Coventry 2021 – Final Report

Aziz Meknassi

Contents

Applied Data Science Capstone Project – The Restaurant Battle of Neighbourhoods in Coventry UK

Assignment Requirements

Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighbourhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve. If you cannot think of an idea or a problem, here are some ideas to get you started:

In Module 3, we explored New York City and the city of Toronto and segmented and clustered their neighbourhoods. Both cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighbourhoods of the two cities and determine how similar or dissimilar they are. Is New York City more like Toronto or Paris or some other multicultural city? I will leave it to you to refine this idea.

In a city of your choice, if someone is looking to open a restaurant, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office?

These are just a couple of many ideas and problems that can be solved using location data in addition to other datasets. No matter what you decide to do, make sure to provide sufficient justification of why you think what you want to do or solve is important and why would a client or a group of people be interested in your project.

I have decided to go with, for tourists vising Coventry, what's there to eat in Coventry United Kingdom

Prologue:

It has been just over a year since the first lockdown was announced in England due to the corona virus pandemic, from there on 2 more lockdowns until now April 2021.

With nowhere to go I decided to use the time and learn about data science. In January 2021 I had completed Data Science Fundamentals with Python and SQL certificate that has 4 courses but I wanted more, so I enrolled in IBM Data Science Professional Certificate that consists of 10 online courses that promised to provide students with the latest job-ready tools and skills, including open source tools and libraries, Python, databases, SQL, data visualization, data analysis, statistical analysis, predictive modelling, and machine learning algorithms. It was also promised that students will learn data science through hands-on practice in the IBM Cloud using real data science tools and real-world data sets.

Upon successfully completing these courses, students will have built a portfolio of data science projects to provide them with the confidence to plunge into an exciting profession in data science. More info here if the reader would like to know more, I highly recommend this course to anyone interested in data science https://www.coursera.org/professional-certificates/ibm-data-science

The final assignment is the capstone project course which gives students a taste of what data scientists go through in real life when working with data.

After learning about location data and different location data providers, such as Foursquare. I learned how to make RESTful API calls to the Foursquare API to retrieve data about venues in different neighbourhoods around the world. I also learned how to be creative in situations where data are not readily available by scraping web data and parsing HTML code. I utilised Python and its pandas library to manipulate data, exploring and analysing data.

Finally, as students we are required to use the Folium library to create maps of geospatial data and to communicate our results and findings.

I chose the city of Coventry UK as a case in the this final assignment where the requirements are to create a Jupyter notebook with python as the programming language and add all necessary comments, finally publish the results as either a blog or a presentation and a final report. This is the final report, enjoy.

1. A brief introduction to Coventry

1.1 Business problem

For many people when they hear Coventry, their minds immediately serve them Lady Godiva's story, if you are not one of them and you are not familiar with the city of Coventry or lady Godiva, stick around, as today, we are going to learn a bit about the 9th largest city in England and 11th largest in the United Kingdom as per 2011 census when its population reached 316,915.

Coventry is located in the West Midlands, it is also the most central city in England, in other words, it is far from the sea in every direction. Coventry is well known for the legendary 11th century exploits of Lady Godiva who rode through the city naked on horseback in protest at high taxes being levied on the city folk by her husband Leofric, Earl of Mercia. The residents of the city were commanded to look away as she rode, but one man did not and was allegedly struck blind. He became known as Peeping Tom thus originating a new idiom, or metonym, in English. There is a Grade II listed statue of her in the city centre.

On several occasions Coventry was briefly the capital of England. One of those occasions was in 1404, when Henry IV summoned a parliament in Coventry as he needed money to fight rebellion, which wealthy cities such as Coventry lent to him, while both Henry V and VI frequently sought loans from the city to meet the expense of the war with France.

City of Coventry, where the author lives, attracts a large number of tourists, not least due to its famous cathedral that was destroyed in the world war II, it is also as a home to the following 13 museums and galleries,

- Charterhouse Coventry
- Coventry Transport Museum
- Coventry Watch Museum
- Herbert Art Gallery and Museum
- Lunt Roman Fort
- Midland Air Museum
- Coventry Cathedral
- Coventry Police Museum
- Revival Cars
- 2-Tone Village Coventry
- St Mary's Guildhall
- The Coventry Music Museum
- The Weaver's House

In 2017, the city of Coventry won the title of UK City of Culture 2021, which means more tourists will visit.

Finding the right place to eat can be a challenge, due to different tastes, health, religious, cultural, and moral reasons. hence, the purpose of this exercise is to give a simple recommendation to tourists in Coventry: for example, where to each thai or middle eastern food, where to get fast food? Etc

1.2 Target audience:

Tourists visiting Coventry who would want to know what type of restaurants are available and where to eat

A business entrepreneur that wants open a new restaurant in Coventry.

Business Analyst or Data Scientists, who wish to analyse the neighbourhoods of Coventry using python, Jupyter notebook and some machine learning techniques.

Someone curious about data that want to have an idea, how beneficial it is to open a restaurant and what are the pros and cons of this business.

2. Data section

It is requested by the assignment to use Foursquare API, in this report I will use foursquare data to find restaurants in Coventry.

- To start and in Wikipedia, I found the list of postcodes for the City of Coventry and its surrounding areas, this will help segment the area. https://en.wikipedia.org/wiki/CV_postcode_area
- Many areas around the city use the postcode starting CV, however, in this study I will focus on the areas (I called boroughs for lack of better terms) that are under the Coventry Council authority

Coverage [edit]

The approximate coverage of the postcode districts:

Postcode district	Post town	Coverage	Local authority area(s)
CV1	COVENTRY	Coventry C (Coventry City Centre, Gosford Green, Hillfields, Spon End, Coventry University)	Coventry
CV2	COVENTRY	Coventry NE (Walsgrave, Wyken, Stoke, Bell Green, Wood End, Potters Green, Aldermans Green, Clifford Park, Woodway Park)	Coventry
CV3	COVENTRY	Coventry SE (Binley, Whitley, Willenhall, Cheylesmore, Styvechale, Finham, Fenside, Stoke Aldermoor, Green Lane, Ernesford Grange, Binley Woods)	Coventry, Rugby
CV4	COVENTRY	Coventry SW (Tile Hill, Canley, Cannon Park, Lime Tree Park, Gibbet Hill, Westwood Heath, University of Warwick)	Coventry
		Coventry NW (Allesley, Allesley Park, Allesley Green,	

As can been seen above, the areas (neighbourhoods) are segregated by postcodes, this will limit our project and may affect our results, because we need a bit of detailed per neighbourhood data.

Using Beautiful Soup library, I pulled the data into Jupyeter notebooks,

Out[53]:

	postcode	Post town	Coverage	Local authority area(s)
0	CV1	COVENTRY	Coventry C (Coventry City Centre, Gosford Gree...	Coventry
1	CV2	COVENTRY	Coventry NE (Walsgrave, Wyken, Stoke, Bell Gre...	Coventry
2	CV3	COVENTRY	Coventry SE (Binley, Whitley, Willenhall, Chey...	Coventry, Rugby
3	CV4	COVENTRY	Coventry SW (Tile Hill, Canley, Cannon Park, L...	Coventry
4	CV5	COVENTRY	Coventry NW (Allesley, Allesley Park, Allesley...	Coventry
5	CV6	COVENTRY	Coventry N (Holbrooks, Coundon, Radford, Longf...	Coventry, Nuneaton and Bedworth
6	CV7	COVENTRY	Exhall, Ash Green, Keresley, Meriden, Balsall ...	Nuneaton and Bedworth, North Warwickshire, Cov...
7	CV8	COVENTRY	Wolston, Ryton-on-Dunsmore	Warwick, Rugby
8	CV8	KENILWORTH	Kenilworth, Baginton, Bubbenhall, Burton Green...	Warwick, Rugby
9	CV9	ATHERSTONE	Atherstone, Mancetter, Grendon, Baddesley Enso...	North Warwickshire, Hinckley and Bosworth
10	CV10	NUNEATON	Nuneaton N & W (Weddington, Stockingford, Camp...	Nuneaton and Bedworth, North Warwickshire
11	CV11	NUNEATON	Nuneaton C & E (town centre, Abbey Green, St N...	Nuneaton and Bedworth, Rugby, Hinckley and Bos...
12	CV12	BEDWORTH	Bedworth (except Exhall and Ash Green), Bulkin...	Nuneaton and Bedworth
13	CV13	NUNEATON	Barlestone, Barton in the Beans, Bilstone, Cad...	Hinckley and Bosworth
14	CV21	RUGBY	Rugby (north), Brownsover	Rugby
15	CV22	RUGBY	Rugby (south), Bilton, Cawston, Dunchurch	Rugby
16	CV23	RUGBY	Thurlaston, Princethorpe, Stretton-on-Dunsmore...	Rugby, West Northamptonshire
17	CV31	LEAMINGTON SPA	Leamington Spa (south), Sydenham, Whitnash, Ra...	Warwick
18	CV32	LEAMINGTON SPA	Leamington Spa (north), Cubbington, Lillington	Warwick
19	CV33	LEAMINGTON SPA	Harbury and surrounding villages	Warwick, Stratford-on-Avon
20	CV34	WARWICK	Warwick	Warwick

After few steps in wrangling this data, it seems that it will not yield what we need, as it clustered the neighbourhoods by a postcode, which will cause us two issues:

1) We will have to do more work in separating the coverage into neighbourhoods
2) We will have to find a post code for each area within the areas provided

Since this data was lacking, I started looking elsewhere, luckily I found UK-GeoJSON API and website that will not only give me the breakdown of Coventry areas but it also has the latitude and longitude as well as geometry, but the postcodes are missing.

In [328]:
```
#load the file from https://martinjc.github.io/UK-GeoJSON/
coventry_neighborhood_geodf = gpd.read_file('coventry_boundaries.geojson')
coventry_neighborhood_geodf.head()
```

Out[328]:

	id	WD13CD	WD13CDO	WD13NM	WD13NMW	geometry
0	E05001218	E05001218	00CQFU	Bablake	None	POLYGON ((-1.52929 52.41871, -1.53465 52.42134...
1	E05001219	E05001219	00CQFW	Binley and Willenhall	None	POLYGON ((-1.46293 52.37639, -1.46539 52.37802...
2	E05001220	E05001220	00CQFX	Cheylesmore	None	POLYGON ((-1.50450 52.37779, -1.50448 52.37812...
3	E05001221	E05001221	00CQFY	Earlsdon	None	POLYGON ((-1.50938 52.37876, -1.51609 52.37982...
4	E05001222	E05001222	00CQFZ	Foleshill	None	POLYGON ((-1.50921 52.41305, -1.50951 52.41563...

After cleaning the above data, we are left with 3 columns, neighbourhoods, latitude and longitude.

	Neighborhood	Latitude	Longitude
0	Bablake	52.439341	-1.565273
1	Binley And Willenhall	52.391733	-1.457056
2	Cheylesmore	52.387161	-1.492103
3	Earlsdon	52.393886	-1.531970
4	Foleshill	52.428311	-1.497469
5	Henley	52.433886	-1.449003
6	Holbrook	52.443502	-1.513496
7	Longford	52.446222	-1.478762
8	Lower Stoke	52.404403	-1.473717
9	Radford	52.424336	-1.517428
10	St Michael'S	52.407433	-1.501943

My next step, I used an API call MapIt UK and I created a function to loop through the neighbourhood ids and provide the postcodes, once that the data was ready I merged it with the latest dataframe.

I loaded the file in a dataframe and merged it with the first dataframe,

	Neighbourhoods	Latitude	Longitude	Neighbourhoods ids	postcodes
0	Bablake	52.439341	-1.565273	8907	CV6 2AG
1	Cheylesmore	52.387161	-1.492103	8910	CV3 4AH
2	Earlsdon	52.393886	-1.531970	8909	CV3 5AZ
3	Foleshill	52.428311	-1.497469	8914	CV1 4NZ
4	Henley	52.433886	-1.449003	8905	CV2 1EN
5	Holbrook	52.443502	-1.513496	8922	CV6 6HX
6	Longford	52.446222	-1.478762	8921	CV2 1NW
7	Lower Stoke	52.404403	-1.473717	8918	CV2 5FR
8	Radford	52.424336	-1.517428	8917	CV6 1DT
9	Sherbourne	52.416080	-1.535556	8916	CV6 1GH
10	Upper Stoke	52.420333	-1.477083	8920	CV2 3EH
11	Wainbody	52.379894	-1.542019	8908	CV3 6DS
12	Westwood	52.396166	-1.579312	8906	CV4 8BP
13	Whoberley	52.410833	-1.549512	8913	CV5 8GD
14	Woodlands	52.412161	-1.583586	8912	CV4 9TN
15	Wyken	52.412569	-1.448599	8919	CV2 2NB

We now have a dataframe with areas names, latitude, longitude ad postcodes and ready to use the Foursquare API

3. Methodology

3.1 Business understanding

The aim of this exercise is to help tourists find where to eat and what restaurants are available in the City of Coventry

3.2 Analytical Approach

The total number of neighbourhoods under Coventry Local authority are 16, hence we need to find a way to cluster them based on their similarities.
After few steps in Data Cleaning and Data Exploration, I will be using a K-Means algorithm to extract the clusters and produce a map and finally conclude based on the final results.

3.3 Data Exploration

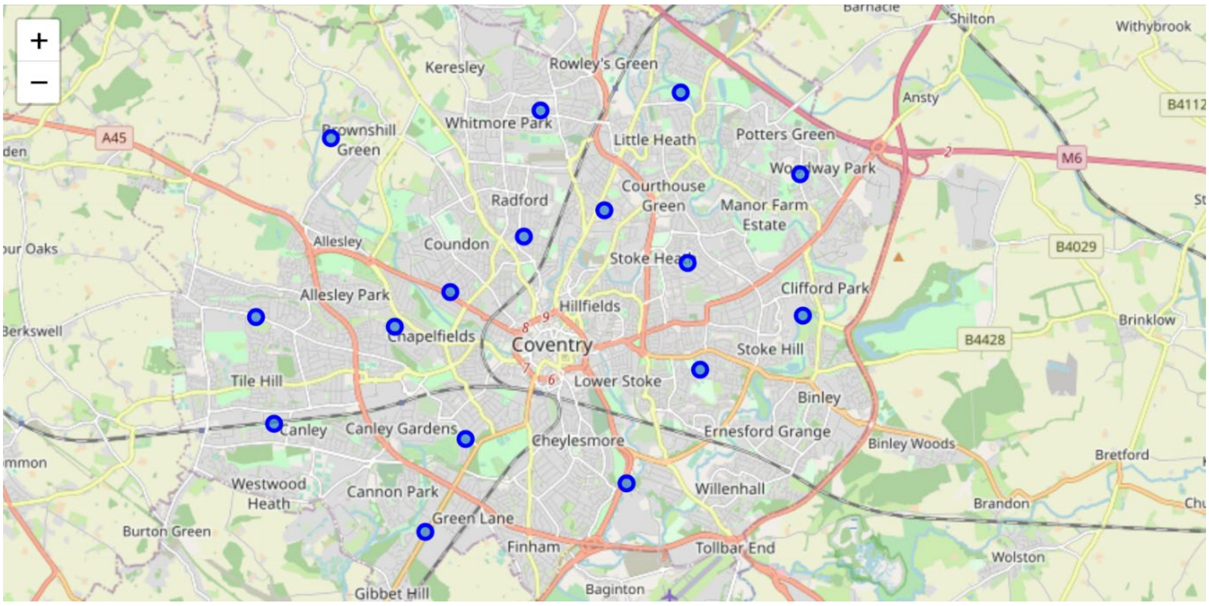
Since we have all the data with geometry, latitude and longitude, I used Folium, a python library that helps create interactive leaflet maps using coordinate data, to create centroids for every neighbourhood in Coventry.
Sample code :

```
[361]: # create map of Coventryusing latitude and longitude values
map_coventry = folium.Map(location=[latitude,longitude], zoom_start=10)

# add markers to map
for lat, lng,postcodes,  neighborhood in zip(coventry_data_all['Latitude'],
                                             coventry_data_all['Longitude'],
                                             coventry_data_all['postcodes'],
                                             coventry_data_all['Neighbourhoods']):
    label = '{}'.format(neighborhood)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_coventry)

map_coventry
```

And the resulting map is as follows:



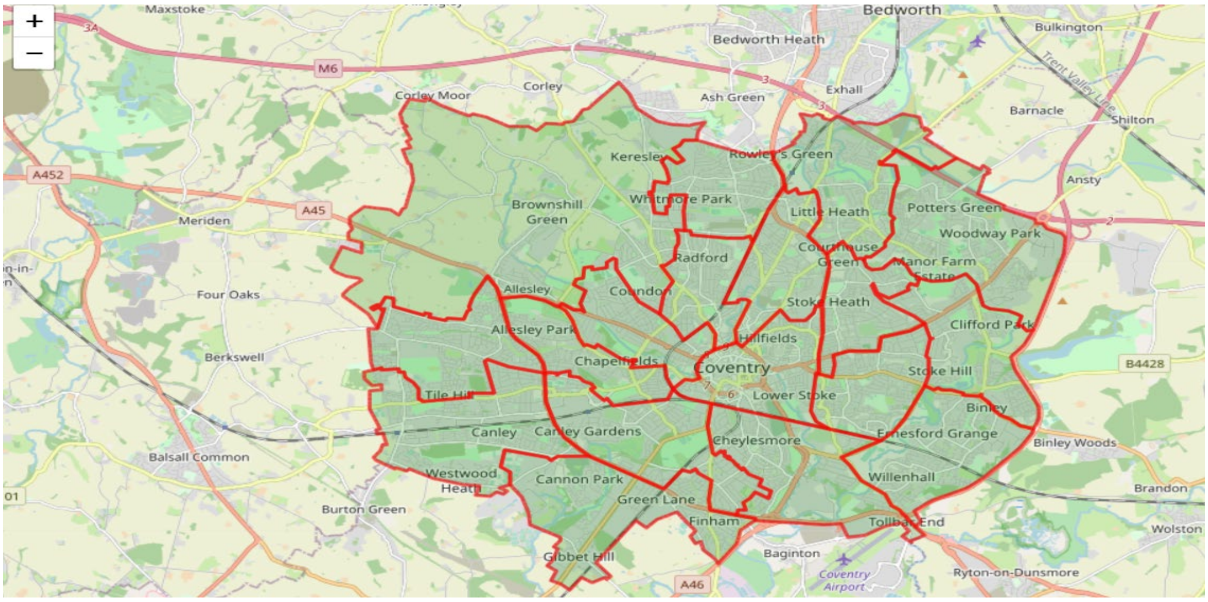
Since I also have the geometry data I created a GeoJSON map for Coventry with its areas highlighted on the map:
Code:

```
In [365]: m = folium.Map([latitude, longitude], zoom_start=12)

          folium.GeoJson(coventry_neighborhood_geodf,
              style_function=lambda x: {
                  'color' : 'red',
                  'opacity': 0.6,
                  'fillColor' : 'green',
                  }).add_to(m)

          m
```

And the resulting map:



FOURSQUARE API

Now it is time to use the above api, as per the course, I took coventry city centre as a sample, to check the quality of the data sent from foursquare APi.

Then, applied the same logic on the rest of the areas, bringing us 917 venues:

In [374]:
```
coventry_venues = getNearbyVenues(names=coventry_data['Neighbourhoods'],
                                   latitudes=coventry_data['Latitude'],
                                   longitudes=coventry_data['Longitude']
                                  )
```

```
Bablake
Cheylesmore
Earlsdon
Foleshill
Henley
Holbrook
Longford
Lower Stoke
Radford
Sherbourne
Upper Stoke
Wainbody
Westwood
Whoberley
Woodlands
Wyken
```

Let's check the size of the resulting dataframe

In [375]:
```
print(coventry_venues.shape)
coventry_venues.head()
```

```
(917, 7)
```

Then checked how many venues are there per neighbourhood, pandas head only brings the first 5 results by default

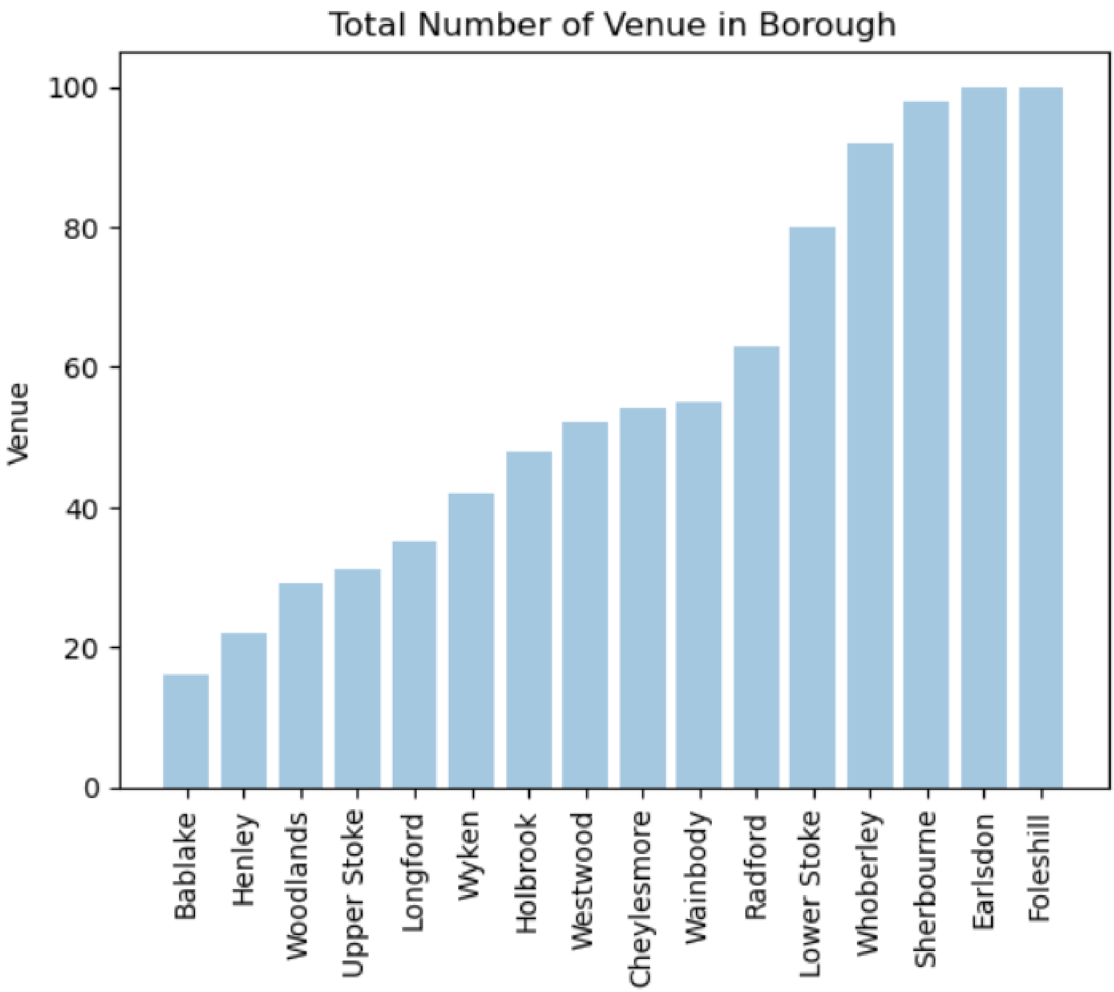
Let's check how many venues were returned for each area/neighbourhood and sort them in count

In [376]:
```
summary = coventry_venues.groupby('Borough').count().reset_index()
summary['Count'] = summary['Venue']
summary = summary.drop(['Borough Latitude', 'Borough Longitude', 'Venue', 'Venue Latitude', 'Venue Longitude','Venue Category'], ax:
summary = summary.sort_values('Count').reset_index(drop=True)
summary.head()
```

Out[376]:

	Borough	Count
0	Bablake	16
1	Henley	22
2	Woodlands	29
3	Upper Stoke	31
4	Longford	35

For the visual audience I have created graphic for the total number of venues per area

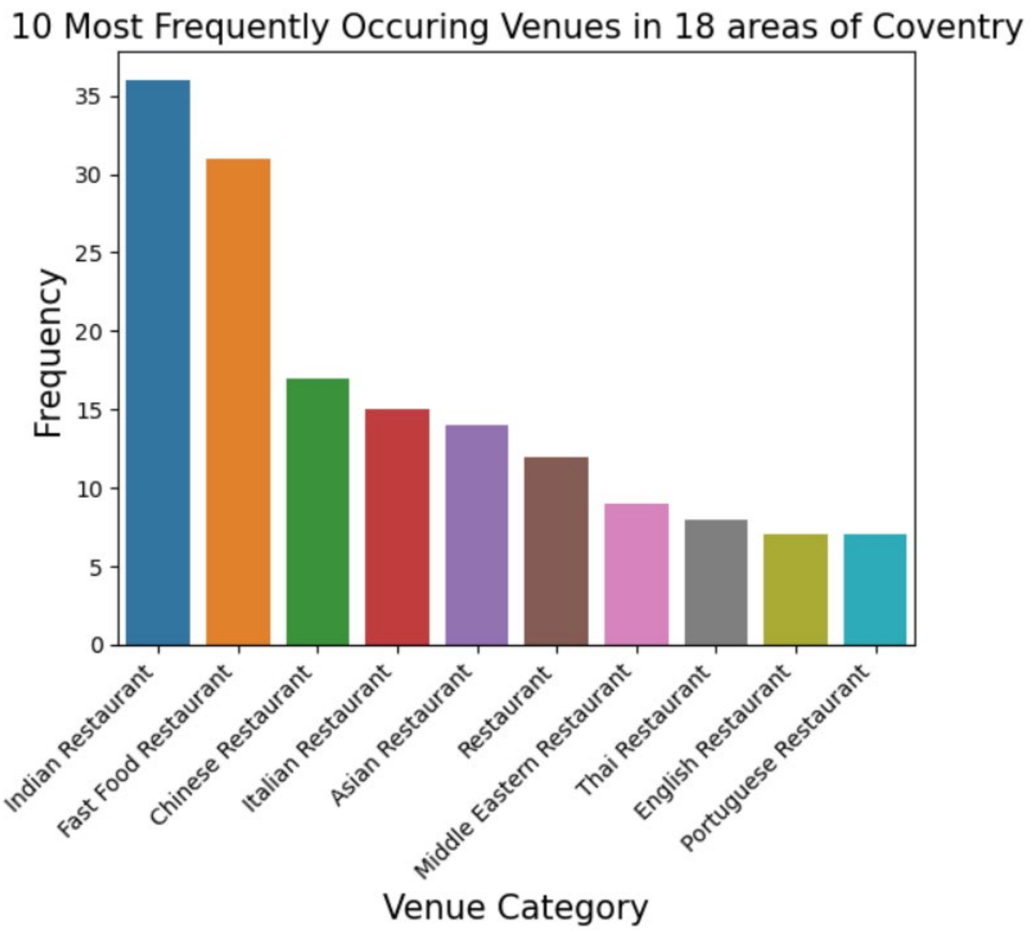


Since our focus is on the restaurants, I filtered the results for only restaurants, giving the following

```
In [380]: print (coventry_Venues_only_restaurant['Venue Category'].value_counts())
```

```
Indian Restaurant             36
Fast Food Restaurant          31
Chinese Restaurant            17
Italian Restaurant            15
Asian Restaurant              14
Restaurant                    12
Middle Eastern Restaurant      9
Thai Restaurant                8
English Restaurant             7
Portuguese Restaurant          7
Turkish Restaurant             7
Lebanese Restaurant            6
Latin American Restaurant      5
Japanese Restaurant            4
American Restaurant            4
Greek Restaurant               3
Persian Restaurant             3
Mexican Restaurant             3
Korean Restaurant              2
Pakistani Restaurant           1
Name: Venue Category, dtype: int64
```

A Graphic was also created for the 10 most frequently occurring Venus in Coventry from Foursquare API



3.4 Clustering

To find clusters of restaurant types in the different city areas, I first transformed the data frame with the restaurant venues, associated to city areas, by one-hot encoding (0/1), as seen in the picture below

[390]:

	Borough	American Restaurant	Asian Restaurant	Chinese Restaurant	English Restaurant	Fast Food Restaurant	Greek Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Korean Restaurant	Latin American Restaurant	Leb Resta
1	Bablake	0	0	0	1	0	0	0	0	0	0	0	
2	Cheylesmore	0	0	0	0	0	0	0	0	0	0	0	
3	Cheylesmore	0	0	1	0	0	0	0	0	0	0	0	
4	Cheylesmore	0	0	0	0	0	0	0	0	0	0	0	
5	Cheylesmore	0	0	0	0	0	0	0	0	0	0	0	

using grouping to show the frequency of each category of restaurants in each city area.

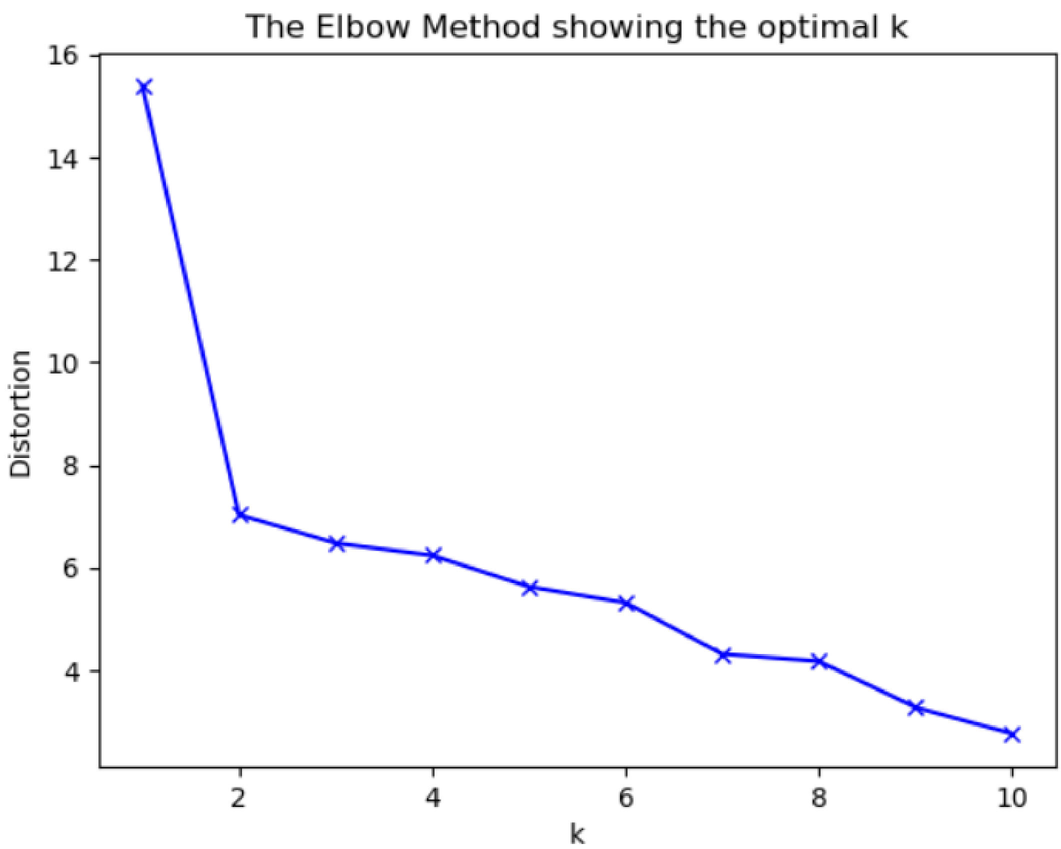
	Borough	American Restaurant	Asian Restaurant	Chinese Restaurant	English Restaurant	Fast Food Restaurant	Greek Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Korean Restaurant	Latin American Restaurant	Le Res
0	Bablake	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
1	Cheylesmore	0.000000	0.000000	0.222222	0.000000	0.222222	0.000000	0.111111	0.000000	0.000000	0.000000	0.000000	C
2	Earlsdon	0.000000	0.080000	0.000000	0.000000	0.040000	0.040000	0.320000	0.120000	0.040000	0.000000	0.040000	0.
3	Foleshill	0.043478	0.086957	0.086957	0.000000	0.173913	0.043478	0.000000	0.130435	0.043478	0.043478	0.043478	0.
4	Henley	0.000000	0.000000	0.000000	0.000000	0.400000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
5	Holbrook	0.200000	0.000000	0.000000	0.000000	0.400000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.
6	Longford	0.200000	0.000000	0.000000	0.000000	0.400000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.
7	Lower Stoke	0.000000	0.090909	0.136364	0.045455	0.227273	0.000000	0.090909	0.000000	0.045455	0.045455	0.045455	0.
8	Radford	0.000000	0.125000	0.062500	0.000000	0.000000	0.000000	0.250000	0.125000	0.000000	0.000000	0.062500	0.
9	Sherbourne	0.000000	0.115385	0.038462	0.038462	0.076923	0.038462	0.230769	0.115385	0.038462	0.000000	0.038462	0.
10	Upper Stoke	0.000000	0.111111	0.222222	0.000000	0.111111	0.000000	0.111111	0.000000	0.000000	0.000000	0.000000	0.
11	Wainbody	0.000000	0.111111	0.000000	0.000000	0.111111	0.000000	0.444444	0.111111	0.000000	0.000000	0.000000	0.
12	Westwood	0.000000	0.000000	0.500000	0.000000	0.250000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
13	Whoberley	0.000000	0.045455	0.090909	0.090909	0.090909	0.000000	0.318182	0.136364	0.000000	0.000000	0.000000	0.
14	Woodlands	0.000000	0.000000	0.500000	0.250000	0.250000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
15	Wyken	0.111111	0.000000	0.000000	0.111111	0.555556	0.000000	0.111111	0.000000	0.000000	0.000000	0.000000	0.

Then put all the information in a new dataframe to show the 10 most common venues in Coventry by area

)ut[395]:

	Borough	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	Bablake	English Restaurant	American Restaurant	Lebanese Restaurant	Thai Restaurant	Restaurant	Portuguese Restaurant	Persian Restaurant	Pakistani Restaurant	Middle Eastern Restaurant	Mexican Restaurant
1	Cheylesmore	Chinese Restaurant	Fast Food Restaurant	Turkish Restaurant	Thai Restaurant	Indian Restaurant	Middle Eastern Restaurant	Lebanese Restaurant	Mexican Restaurant	Restaurant	Portuguese Restaurant
2	Earlsdon	Indian Restaurant	Italian Restaurant	Thai Restaurant	Restaurant	Asian Restaurant	Latin American Restaurant	Fast Food Restaurant	Greek Restaurant	Japanese Restaurant	Portuguese Restaurant
3	Foleshill	Fast Food Restaurant	Italian Restaurant	Turkish Restaurant	Asian Restaurant	Chinese Restaurant	Restaurant	Portuguese Restaurant	Middle Eastern Restaurant	Mexican Restaurant	Lebanese Restaurant
4	Henley	Restaurant	Fast Food Restaurant	Portuguese Restaurant	American Restaurant	Lebanese Restaurant	Thai Restaurant	Persian Restaurant	Pakistani Restaurant	Middle Eastern Restaurant	Mexican Restaurant

Now, with all this data, I could finally run an unsupervised machine learning algorithm, more specifically, a k-means clustering algorithm from the scikit-learn package. One could use the elbow method to systematically define the k value, but I simply chose k to be 6.

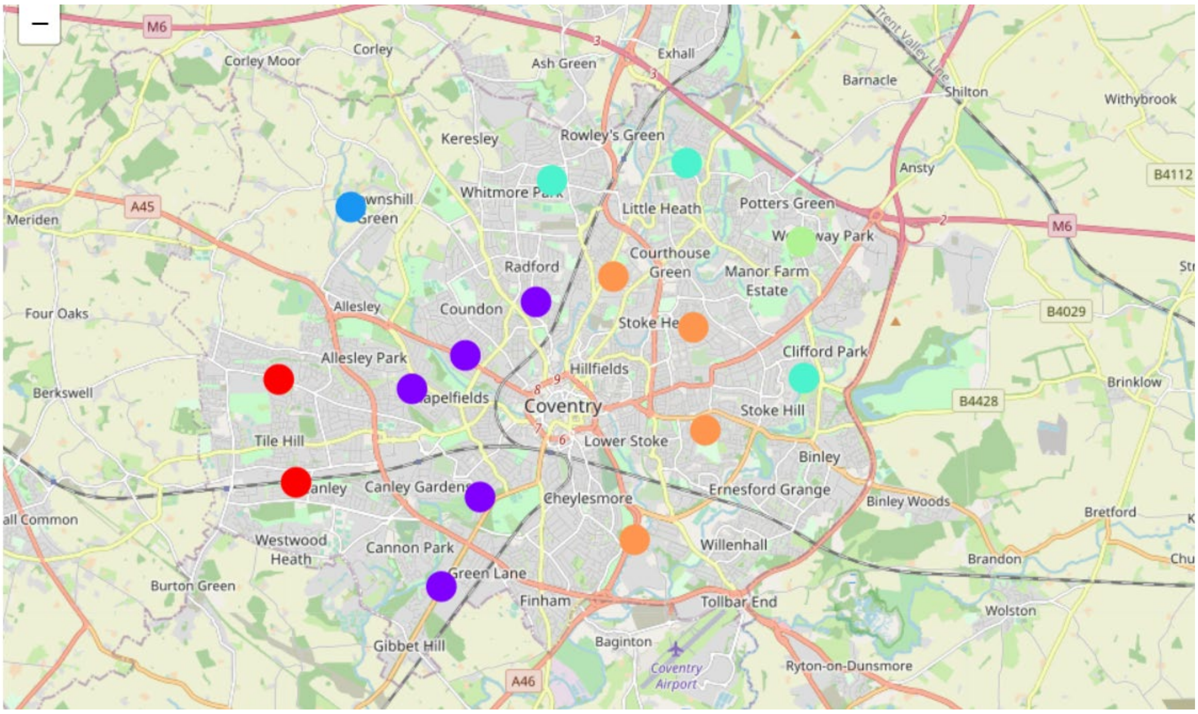


4. Results

Finally, the results are here:

	Neighbourhoods	Latitude	Longitude	Neighbourhoods ids	postcodes	Cluster Labels	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Co
0	Bablake	52.439341	-1.565273	8907	CV6 2AG	2	English Restaurant	American Restaurant	Lebanese Restaurant	Thai Restaurant	Restaurant	Portuguese Restaurant	F Resi
1	Cheylesmore	52.387161	-1.492103	8910	CV3 4AH	5	Chinese Restaurant	Fast Food Restaurant	Turkish Restaurant	Thai Restaurant	Indian Restaurant	Middle Eastern Restaurant	Let Resi
2	Earlsdon	52.393886	-1.531970	8909	CV3 5AZ	1	Indian Restaurant	Italian Restaurant	Thai Restaurant	Restaurant	Asian Restaurant	Latin American Restaurant	Fas Resi
3	Foleshill	52.428311	-1.497469	8914	CV1 4NZ	5	Fast Food Restaurant	Italian Restaurant	Turkish Restaurant	Asian Restaurant	Chinese Restaurant	Restaurant	Portu Resi
4	Henley	52.433886	-1.449003	8905	CV2 1EN	4	Restaurant	Fast Food Restaurant	Portuguese Restaurant	American Restaurant	Lebanese Restaurant	Thai Restaurant	F Resi
5	Holbrook	52.443502	-1.513496	8922	CV6 6HX	3	Fast Food Restaurant	American Restaurant	Indian Restaurant	Mexican Restaurant	Lebanese Restaurant	Thai Restaurant	Resi
6	Longford	52.446222	-1.478762	8921	CV2 1NW	3	Fast Food Restaurant	American Restaurant	Indian Restaurant	Mexican Restaurant	Lebanese Restaurant	Thai Restaurant	Resi
7	Lower Stoke	52.404403	-1.473717	8918	CV2 5FR	5	Fast Food Restaurant	Chinese Restaurant	Turkish Restaurant	Asian Restaurant	Indian Restaurant	Thai Restaurant	Portu Resi
8	Radford	52.424336	-1.517428	8917	CV6 1DT	1	Indian Restaurant	Middle Eastern Restaurant	Italian Restaurant	Asian Restaurant	Latin American Restaurant	Chinese Restaurant	Resi
9	Sherbourne	52.416080	-1.535556	8916	CV6 1GH	1	Indian Restaurant	Italian Restaurant	Asian Restaurant	Fast Food Restaurant	Restaurant	Latin American Restaurant	F Resi

The above is a table of the city areas and their most common venues, and they now have been assigned six different cluster labels from 0 to 5. We can now use the cluster labels to show the city areas marked with a cluster-specific colour on a map, again using folium:



What's visible on the map are 16 bubbles representing 16 city areas in 6 different colours.

Using the above we can show the 6 clusters of restaurant type concentration for the city of Coventry.

So let's see these clusters what they hold

A. First are the two red bubbles for the Chinese restaurants

Cluster 1

In [412]:
```
coventry_merged.loc[coventry_merged['Cluster Labels'] == 0, coventry_merged.columns[[1] + list(range(5, coventry_merged.shape[1]))]]
```

Out[412]:

	Latitude	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 Most Common Venue
12	52.396166	0	Chinese Restaurant	Fast Food Restaurant	Restaurant	American Restaurant	Lebanese Restaurant	Thai Restaurant	Portuguese Restaurant	Persian Restaurant	Pakistani Restaurant	Middle Eastern Restaurant
14	52.412161	0	Chinese Restaurant	English Restaurant	Fast Food Restaurant	American Restaurant	Mexican Restaurant	Thai Restaurant	Restaurant	Portuguese Restaurant	Persian Restaurant	Pakistani Restaurant

B. The Indian restaurants represented with purple colour

Cluster 2

In [413]:
```
coventry_merged.loc[coventry_merged['Cluster Labels'] == 1, coventry_merged.columns[[1] + list(range(5, coventry_merged.shape[1]))]]
```

Out[413]:

	Latitude	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 Most Common Venue
2	52.393886	1	Indian Restaurant	Italian Restaurant	Thai Restaurant	Restaurant	Asian Restaurant	Latin American Restaurant	Fast Food Restaurant	Greek Restaurant	Japanese Restaurant	Portuguese Restaurant
8	52.424336	1	Indian Restaurant	Middle Eastern Restaurant	Italian Restaurant	Asian Restaurant	Latin American Restaurant	Chinese Restaurant	Thai Restaurant	Restaurant	Portuguese Restaurant	Lebanese Restaurant
9	52.416080	1	Indian Restaurant	Italian Restaurant	Asian Restaurant	Fast Food Restaurant	Restaurant	Latin American Restaurant	Persian Restaurant	Chinese Restaurant	English Restaurant	Greek Restaurant
11	52.379894	1	Indian Restaurant	Restaurant	Fast Food Restaurant	Italian Restaurant	Asian Restaurant	American Restaurant	Mexican Restaurant	Thai Restaurant	Portuguese Restaurant	Persian Restaurant
13	52.410833	1	Indian Restaurant	Italian Restaurant	Chinese Restaurant	English Restaurant	Fast Food Restaurant	Middle Eastern Restaurant	Thai Restaurant	Restaurant	Asian Restaurant	Persian Restaurant

C. The blue bubble for the English restaurant

Cluster 3

```
In [414]: coventry_merged.loc[coventry_merged['Cluster Labels'] == 2, coventry_merged.columns[[1] + list(range(5, coventry_merged.shape[1]))]]
```

Out[414]:

	Latitude	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 Most Common Venue
0	52.439341	2	English Restaurant	American Restaurant	Lebanese Restaurant	Thai Restaurant	Restaurant	Portuguese Restaurant	Persian Restaurant	Pakistani Restaurant	Middle Eastern Restaurant	Mexican Restaurant

D. The Fast Food Restaurants in cyan colour

Cluster 4

```
In [415]: coventry_merged.loc[coventry_merged['Cluster Labels'] == 3, coventry_merged.columns[[1] + list(range(5, coventry_merged.shape[1]))]]
```

Out[415]:

	Latitude	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 Most Common Venue
5	52.443502	3	Fast Food Restaurant	American Restaurant	Indian Restaurant	Mexican Restaurant	Lebanese Restaurant	Thai Restaurant	Restaurant	Portuguese Restaurant	Persian Restaurant	Pakistani Restaurant
6	52.446222	3	Fast Food Restaurant	American Restaurant	Indian Restaurant	Mexican Restaurant	Lebanese Restaurant	Thai Restaurant	Restaurant	Portuguese Restaurant	Persian Restaurant	Pakistani Restaurant
15	52.412569	3	Fast Food Restaurant	American Restaurant	Indian Restaurant	Turkish Restaurant	English Restaurant	Greek Restaurant	Chinese Restaurant	Italian Restaurant	Japanese Restaurant	Korean Restaurant

E. A none type restaurant in light green (this one is hiding like finding Wally, tricky to see,but it's by the Woodway park)

Cluster 5

```
In [416]: coventry_merged.loc[coventry_merged['Cluster Labels'] == 4, coventry_merged.columns[[1] + list(range(5, coventry_merged.shape[1]))]]
```

Out[416]:

	Latitude	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 Most Common Venue
4	52.433886	4	Restaurant	Fast Food Restaurant	Portuguese Restaurant	American Restaurant	Lebanese Restaurant	Thai Restaurant	Persian Restaurant	Pakistani Restaurant	Middle Eastern Restaurant	Mexican Restaurant

F. Last cluster with orange is for 1 Chinese restaurant and 3 fast food restaurants

Cluster 6

```
In [417]: coventry_merged.loc[coventry_merged['Cluster Labels'] == 5, coventry_merged.columns[[1] + list(range(5, coventry_merged.shape[1]))]]
```

Out[417]:

	Latitude	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 Most Common Venue
1	52.387161	5	Chinese Restaurant	Fast Food Restaurant	Turkish Restaurant	Thai Restaurant	Indian Restaurant	Middle Eastern Restaurant	Lebanese Restaurant	Mexican Restaurant	Restaurant	Portuguese Restaurant
3	52.428311	5	Fast Food Restaurant	Italian Restaurant	Turkish Restaurant	Asian Restaurant	Chinese Restaurant	Restaurant	Portuguese Restaurant	Middle Eastern Restaurant	Mexican Restaurant	Lebanese Restaurant
7	52.404403	5	Fast Food Restaurant	Chinese Restaurant	Turkish Restaurant	Asian Restaurant	Indian Restaurant	Thai Restaurant	Portuguese Restaurant	Middle Eastern Restaurant	Lebanese Restaurant	Latin American Restaurant
10	52.420333	5	Chinese Restaurant	Turkish Restaurant	Thai Restaurant	Fast Food Restaurant	Indian Restaurant	Portuguese Restaurant	Asian Restaurant	Middle Eastern Restaurant	Mexican Restaurant	Restaurant

Conclusion:

As the analysis is performed on small set of data, we can achieve better results by increasing the neighbourhood information.

Coventry is an international city with many different types of restaurant businesses to offer, i believe we have gone through the process of identifying the business problem, specifying the data required, clean the datasets, performing a machine learning algorithm using k-means clustering and providing some useful tips to our stakeholder.

We achieved the goal presented at the outset of this report: tourists can see in the results which city areas best match their food desires. This is just one example of fantastic data science uses cases one can realize applying technology which is available for free today!.

I ended the study by visualising the data and clustering information on the Coventry map. In future studies, web or telephone applications can be carried out to direct investors.

Discussion and Recommendation

Next steps I recommend would be:

● Use a different Venue API with more data. There seem to be some data lacking in Foursquare as it is not very popular here. Mostly users prefer Google Maps or Facebook.

● Find and use updated demographics data about Coventry's Neighbourhood.

● Try a Neighbourhood-Based Clustering.

Acknowledgement & sources

I would like to thank Coursera and IBM for this successful partnership to bring courses to everyone with an interest and internet connection, the tutors and my fellow students were very helpful and instrumental in bringing this final project to light.

On the wider community, I would like to thank my fellow developers in Stackoverlow and github and youtubers and data scientists who all don't mind share their knowledge, this is what makes open source great.