Where to go in smaller cities when there are no-online guides?

An exercise in using Data Science to guide us!

By Brian Naylor, July 2020

Table of contents

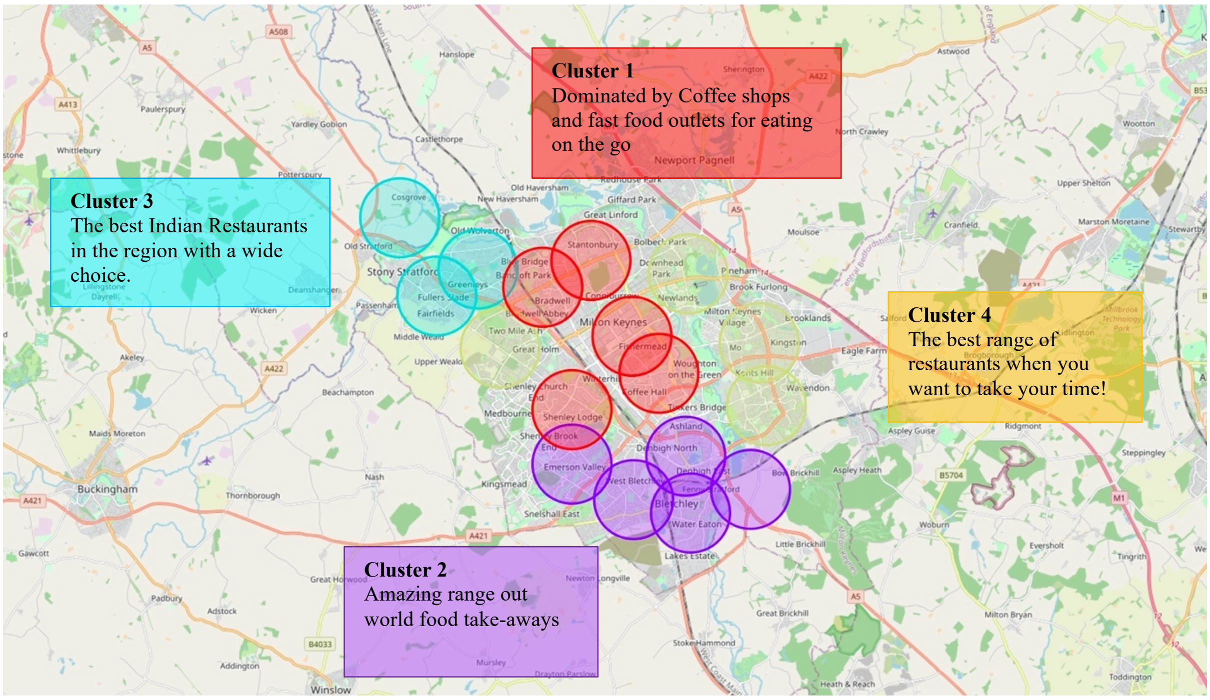
Introduction

The final part of the **IBM Data Science professional certificate**, a series of nine courses, requires the preparation of a report and working Python Notebook, documenting a business problem, how it was addressed, what data was used, and what conclusions were drawn using Data Science techniques. This report details those results, and is aimed at fellow students on the course, to both review what has been done, and potentially inspire others to follow.

Report Summary

For those of you who don't want to read the details of all of the analysis and how I came to the conclusions within this report, the 50,000ft summary is as follows.

By combining a variety of publicly available data sets, and using a series of Data Science and machine learning techniques, we were able to analyse the data available to make a series of recommendations on where to go in Milton Keynes, to find the areas with similar characteristics. Namely, if you want to find the best choice of Indian Restaurants, head for the north of Milton Keynes to the areas of Stony Stratford and Old Wolverton.



We did this to answer the problem "**Where to go in smaller cities when there are no-online guides?**" So by using a variety of data science techniques, we have shown that even for smaller cities with generally less data available, we can use machine learning to derive insightful outcomes.

If you are interested in how we came to these conclusions, then read on into the following sections, where we describe in detail the approach and outcomes.

Problem Definition

Background

When you live in a big city like London or New York, its relatively easy to figure out where to go for some amazing restaurants, or banging street food in the cuisine of your choice, or what to do when you have a few hours to kill? But when you live in a much smaller city (as most of us do!), the choice is less and often the decisions are harder. There aren't many guides or websites for smaller cities. When you are visiting a place for a first time, wouldn't it be great to figure out the happening neighbourhoods or boroughs to hang out, or where to go for the best restaurants?

Now actually this is an easier problem when you have lots of data, and bigger cities also have much more wrangled data available....but with smaller cities, the data acquisition and wrangling is much harder, and there is less data to be statistically relevant.

Problem

Where to go in smaller cities when there are no-online guides?

In the 1960s, the UK Government decided that a further generation of new towns in the South East of England was needed to relieve housing congestion in London. This new town (in planning documents, "new city"), Milton Keynes, was to be the biggest yet, with a target population of 250,000 and a "designated area" of about 22,000 acres (9,000 ha). At designation, its area incorporated the existing towns of Bletchley, Wolverton, and Stony Stratford, along with another fifteen villages and farmland in between.

I live in this amazing city and wondered whether data science could teach me a thing a or two about the place I have lived in for the last 25 years? Also, being a new town just 50 years old, there isn't much history to rely upon hundreds of years of established businesses.

But more specifically, if you were a visitor to my home town;

- What would be the areas within the city you would go to, for a range of different activities?
- Where would you go if you wanted a range of fast food outlets?
- Where do you go for the best Supermarkets and food shops?
- Which parts of the city have the best museums?

So could data science shed light on where to go in a city of 250,000 people?

Just in case you have never heard of Milton Keynes, its here, in the middle of England, half way between London and Birmingham.

Data Understanding

In a city of 250,000 people covering 100 km squared, we're going to need to chunk up the city into small enough areas to be explored yet big enough to be of interest. In the UK we use postcode (similar to ZIP codes) to segment each area. Typically postcodes are 6 or 7 characters (eg. MK7 8AA) which go down to a level of 10-15 houses. That will be too granular for our analysis, so we will use a higher level of the postcodes going to just 3 or 4 characters (eg. MK7 or MK78) to give us a postal region cover a few square kilometres each.

There are THREE main sources of data that we will need for our analysis:

- Neighbourhood or Borough data, of all of the regions within Milton Keynes;
- Location coordinate data within the Boroughs so that we can search for venues around a locale;
- Venue information, its location and classification;

Neighbourhood Data

For the **Neighbourhood data**, the internet is our friend, and we should be able to get this from Wikipedia. Actually our good friends at Wikipedia have posted a list of all of the major postal code towns, their major post codes, and neighbourhoods within. Since its in Wikipedia, we can screen scrape it, and transform it into a dataframe!

Coverage [edit]

The approximate coverage of the postcode districts:

Postcode district	Post town	Coverage	Local authority area(s)
MK1	MILTON KEYNES	Denbigh, Mount Farm	Milton Keynes
MK2	MILTON KEYNES	Brickfields, Central Bletchley, Fenny Stratford, Water Eaton	Milton Keynes
MK3	MILTON KEYNES	Church Green, Far Bletchley, Old Bletchley, Newton Leys, West Bletchley	Milton Keynes, Aylesbury Vale
MK4	MILTON KEYNES	Emerson Valley, Furzton, Kingsmead, Oxley Park, Shenley Brook End, Snelshall West, Tattenhoe, Tattenhoe Park, Westcroft, Whaddon, Woodhill	Milton Keynes, Aylesbury Vale
MK5	MILTON KEYNES	Crownhill, Elfield Park, Grange Farm, Oakhill, Knowlhill, Loughton, Medbourne, Shenley Brook End, Shenley Church End, Shenley Lodge, Shenley Wood	Milton Keynes
MK6	MILTON KEYNES	Ashland, Beanhill, Bleak Hall, Coffee Hall, Eaglestone, Fishermead, Leadenhall, Netherfield, Oldbrook, Peartree Bridge, Redmoor, Springfield, Tinkers Bridge, Woughton on the Green, Woughton Park, Simpson, Winterhill	Milton Keynes
MK7	MILTON KEYNES	Browns Wood, Caldecotte, Kents Hill, Tilbrook, Walton, Walton Hall, Wavendon, Wavendon Gate, Old Farm Park	Milton Keynes
MK8	MILTON KEYNES	Crownhill, Grange Farm, Great Holm, Hazeley, Loughton, Loughton Lodge, Two Mile Ash, Whitehouse, Wymbush	Milton Keynes
MK9	MILTON KEYNES	Central Milton Keynes, Campbell Park	Milton Keynes
MK10	MILTON KEYNES	Brinklow, Broughton, Kingston, Middleton, Monkston, Oakgrove	Milton Keynes

https://en.wikipedia.org/wiki/MK_postcode_area.

Location Coordinates

For **the Location Coordinates**, the UK Government's Office for National Statistics (https://www.ons.gov.uk/) publishes much data for each postal code, such as the income levels, the OS grid references, the parliamentary constituency, and for our purposes, the longitude and latitude of each postcode. Now a 6/7-digit postcode is going to be too granular

NSPL_MAY_2020_UK_MK

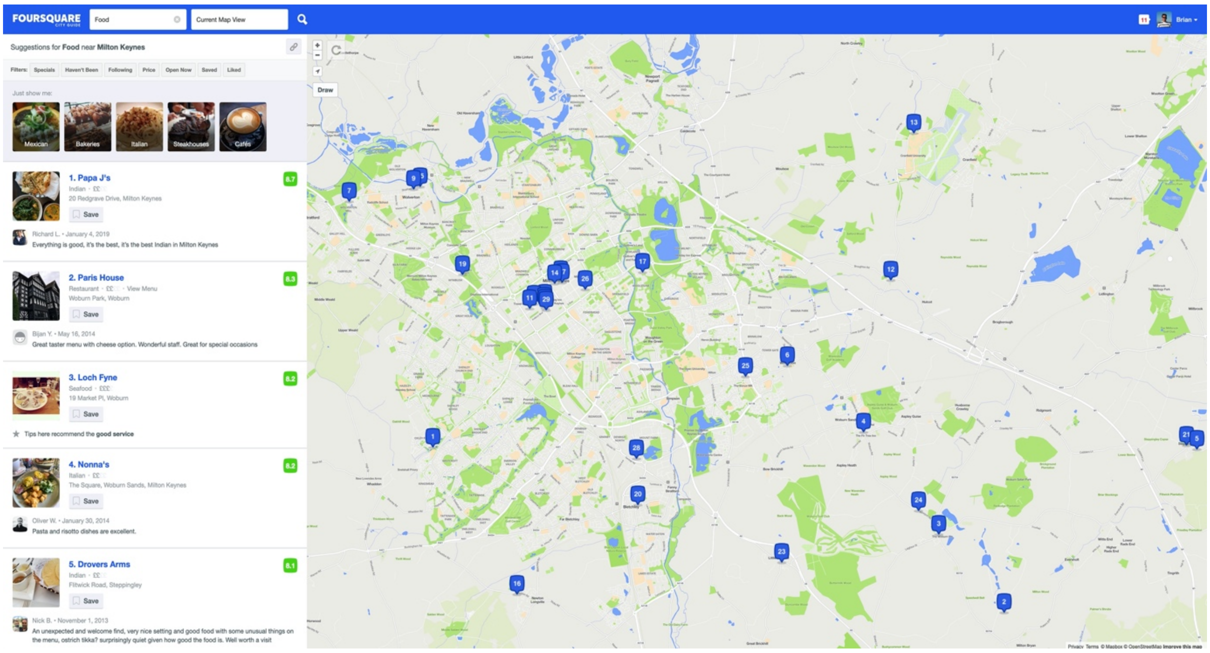
pcd	pcd2	pcds	dointr	doterm	usertype	oseast1m	osnrth1m	osgrdind	oa11	cty
MK1 1AA	MK1 1AA	MK1 1AA	199201	201108	1	487550	235487	1	E00175386	E99999999
MK1 1AB	MK1 1AB	MK1 1AB	199501	199812	1	487334	234906	6	E00084328	E99999999
MK1 1AD	MK1 1AD	MK1 1AD	199507	201012	1	487550	235487	1	E00175386	E99999999
MK1 1AE	MK1 1AE	MK1 1AE	199501	199612	1	487900	234700	8	E00084328	E99999999
MK1 1AF	MK1 1AF	MK1 1AF	199501	199806	1	487334	234906	6	E00084328	E99999999
MK1 1AG	MK1 1AG	MK1 1AG	200706	200707	1	487550	235487	1	E00175386	E99999999
MK1 1AH	MK1 1AH	MK1 1AH	199512	201108	1	487758	235493	1	E00175386	E99999999

pcd	ced	laua	ward	hlthau	nhser	ctry	rgn	pcon	eer	teclec	ttv
MK1 1AA	E99999999	E06000042	E05009406	E18000009	E40000007	E92000001	E12000008	E14000822	E15000008	E24000026	E3
MK1 1AB	E99999999	E06000042	E05009407	E18000009	E40000007	E92000001	E12000008	E14000822	E15000008	E24000026	E3
MK1 1AD	E99999999	E06000042	E05009406	E18000009	E40000007	E92000001	E12000008	E14000822	E15000008	E24000026	E3
MK1 1AE	E99999999	E06000042	E05009407	E18000009	E40000007	E92000001	E12000008	E14000822	E15000008	E24000026	E3
MK1 1AF	E99999999	E06000042	E05009407	E18000009	E40000007	E92000001	E12000008	E14000822	E15000008	E24000026	E3
MK1 1AG	E99999999	E06000042	E05009406	E18000009	E40000007	E92000001	E12000008	E14000822	E15000008	E24000026	E3
MK1 1AH	E99999999	E06000042	E05009406	E18000009	E40000007	E92000001	E12000008	E14000822	E15000008	E24000026	E3

for our purposes, so some data wrangling and aggregation will be in order. ONS publish these files regularly in CSV format. We will get our data from here:
https://geoportal.statistics.gov.uk/datasets/national-statistics-postcode-lookup-may-2020

Venue Information

Finally, for the **venue and point of interest information**, our good friend **Foursquare** can come to the rescue, where we will leverage their API to perform venue exploration around a given longitude and latitude to see what we find.
(e.g.https://foursquare.com/explore?mode=url&ne=52.093746%2C-0.52434&q=Food&sw=51.965952%2C-0.846891)



Data Preparation

There is going to need to be considerable cleaning, filtering and aggregation of data before we can model the outcomes.

First problem will be the scope of the area we look at. Milton Keynes and its surrounding towns cover approximately 45 postcode areas, including Bedford town, Buckingham, and Olney. Now whilst it might be useful to do the analysis at some point on all this data, we will set the scope to cover the main Milton Keynes conurbation, covering postal codes MK1 through MK19.

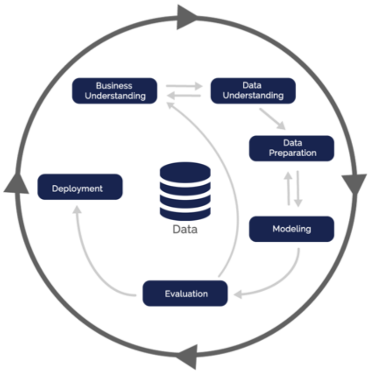
Second problem will be to exclude postcode that are not areas, but are in-fact businesses. The Royal Mail / Post Office (as the originators of the postal code scheme in 1959) decided to allocate significant businesses which received huge amounts of post, their own postcode. We are not interested in these postcodes / businesses. Luckily, then can be easily identified, as the ONS allocated them a null lat / long reference.

Third problem will be to normalise the data so that we can work with the same level of granularity. We are going to analyse our insights across approximately 20 regions in Milton Keynes, but there are actually over 21,000 unique postcodes in Milton Keynes. So we will need to aggregate or group by the major post code, in order to summarise at the major 20 key postcodes. This also poses another problem. The latitude and longitude of each minor postcode will need to be averaged out to give an approximation of the central longitude and latitude of the major postcode. It will be sufficient for our analysis, but more complex methods could be used to determine the actual central longitude and latitude, such as Euclidian distance calculations.

Finally, the ONS data contains massive amounts of reference data, that we will not be using, so for efficiency and speed reasons, we shall be dropping 30-40 columns of data that are just not relevant for our analysis.

Methodology

We will follow the standard CRISP-DM iterative methodology as shown in the following diagram;



Tooling

Its also important to consider the tooling, and we will be using the following;

- Watson Studio, for cloud-based Notebooks
- IBM Cloud Object Storage for any files we load (and save)
- Github for sharing the published results (in my personal repo https://github.com/brian-naylor/Coursera_Capstone)
- Within Notebooks, we're likely to make extensive use of Pandas for dataframe manipulation, Folium for visualisation, and Geocode for location plotting.

Modelling

Since we are trying to understand what parts of the city are best to visit for specific categories, we are going to need to apply some unsupervised clustering techniques to the venues, to determine their common characteristics, and what would make sense to group them together into a set of clusters. Also, the number of clusters is going to be far less than the number of boroughs/regions, otherwise we might as well just say "go to postcode region XX for the best Chinese food restaurants."

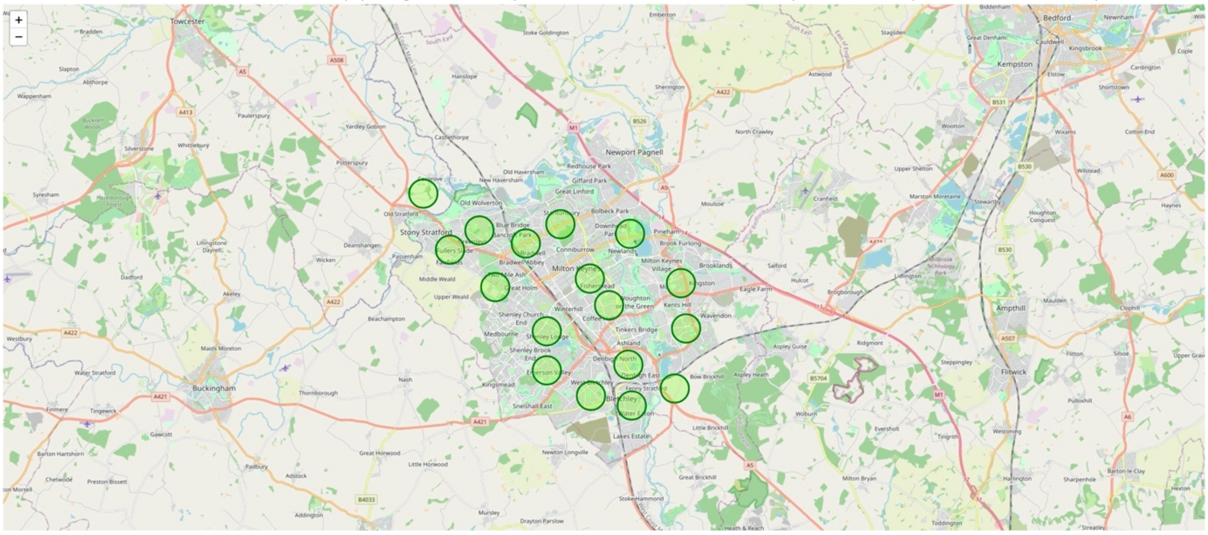
So we will need to identify all the relevant venues within the regions of interest, determine their relative importance or density within each region, and then apply clustering techniques to group them together. We will obviously need to plot them to visualise the results, and then evaluate to see whether they make sense.

Exploratory Data Analysis

One of the first datasets we had to create and cleanse was our base postcode regions, joined together with the location details of longitude and latitude, looking as follows;

	Postcode	Town	Borough	Local authority area(s)	lat	long
0	MK1	MILTON KEYNES	Denbigh, Mount Farm	Milton Keynes	52.007169	-0.727622
1	MK2	MILTON KEYNES	Brickfields, Central Bletchley, Fenny Stratfor...	Milton Keynes	51.991884	-0.725566
2	MK3	MILTON KEYNES	Church Green, Far Bletchley, Old Bletchley, Ne...	Milton Keynes, Aylesbury Vale	51.995494	-0.750159
3	MK4	MILTON KEYNES	Emerson Valley, Furzton, Kingsmead, Oxley Park...	Milton Keynes, Aylesbury Vale	52.004916	-0.776778
4	MK5	MILTON KEYNES	Crownhill, Elfield Park, Grange Farm, Oakhill,...	Milton Keynes	52.019522	-0.776946

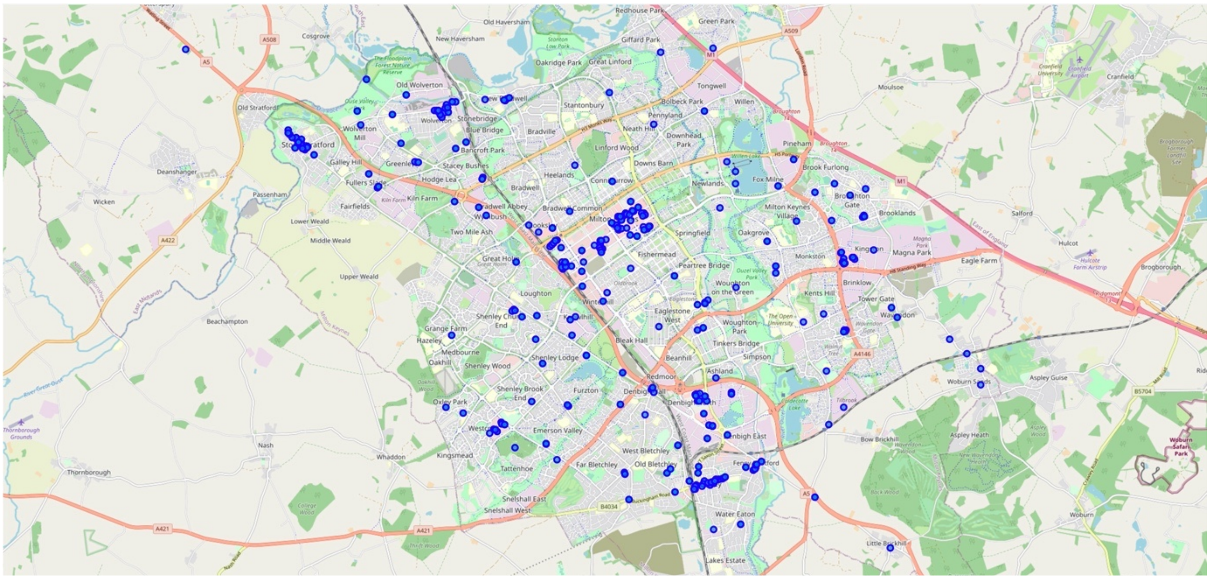
We could then use our mapping techniques to visualise this upon a map of Milton Keynes:



For each of the postcode regions, which we will refer to as 'boroughs' we used Foursquare location API, to find all the food venues within the borough. We actually found the surprising amount in excess of 830 venues, as follows;

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Denbigh, Mount Farm	52.007169	-0.727622	IKEA Food Hall	52.005148	-0.730610	Scandinavian Restaurant
1	Denbigh, Mount Farm	52.007169	-0.727622	McDonald's	51.991932	-0.698971	Fast Food Restaurant
2	Denbigh, Mount Farm	52.007169	-0.727622	Hungry Horse - Furzton Lake	52.017923	-0.767102	Restaurant
3	Denbigh, Mount Farm	52.007169	-0.727622	Domino's Pizza	51.997529	-0.733617	Pizza Place
4	Denbigh, Mount Farm	52.007169	-0.727622	Papa John's Pizza	51.993874	-0.734646	Pizza Place

Its worth seeing what this looks like upon a map, so we plotted the location of each venue, just in case there were obvious gaps in the data, or self defining regional clusters. This is what we found…



When we look across the different types of food venues, we find that there are over 60 different categories from Fast Food Trucks, Pizza outlets, Fish and Chips to fine dining in Italian Restaurants.

In order to perform cluster analysis, first we will perform one-hot-encoding, which counts the occurrence of each venue category by the borough, and then we will aggregate up across all the different venues, to give us the occurrence or density of categories, in a table such as;

	Borough	Afghan Restaurant	African Restaurant	American Restaurant	Asian Restaurant	BBQ Joint	Bakery	Bar	Bistro	Brasserie	...	Scandinavian Restaurant	Snack Place	Souvlaki Shop	Steakhouse	Sushi Restaurant	Tea Room	Thai Restaurant	Turkish Restaurant	Vegetarian / Vegan Restaurant	Wings Joint
0	Ashland, Beanhill, Bleak Hall, Coffee Hall, Ea...	0.000000	0.00	0.000000	0.020408	0.000000	0.000000	0.020408	0.00	0.000000	...	0.020408	0.020408	0.000000	0.000000	0.020408	0.000000	0.000000	0.020408	0.000000	0.000000
1	Aspley Guise, Battlesden, Bow Brickhill, Drayt...	0.000000	0.00	0.000000	0.020408	0.000000	0.000000	0.000000	0.00	0.000000	...	0.020408	0.000000	0.020408	0.000000	0.000000	0.000000	0.000000	0.020408	0.020408	0.000000
2	Bancroft/Bancroft Park, Blue Bridge, Bradwell ...	0.000000	0.00	0.000000	0.020408	0.020408	0.000000	0.020408	0.00	0.000000	...	0.000000	0.020408	0.000000	0.000000	0.020408	0.000000	0.000000	0.040816	0.000000	0.020408
3	Beachampton, Calverton, Castlethorpe Cosgrove,...	0.000000	0.02	0.020000	0.000000	0.000000	0.040000	0.000000	0.02	0.020000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.020000	0.020000	0.040000	0.000000	0.000000
4	Blakelands, Conniburrow, Downs Barn, Giffard P...	0.000000	0.00	0.000000	0.020408	0.020408	0.000000	0.020408	0.00	0.000000	...	0.000000	0.020408	0.000000	0.000000	0.040816	0.000000	0.000000	0.020408	0.000000	0.020408
5	Bolbeck Park, Downhead Park, Fox Milne, Newlan...	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	...	0.000000	0.020000	0.000000	0.000000	0.020000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Brickfields, Central Bletchley, Fenny Stratfor...	0.000000	0.00	0.000000	0.040816	0.000000	0.020408	0.020408	0.00	0.000000	...	0.020408	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040816	0.020408	0.000000
7	Brinklow, Broughton, Kingston, Middleton, Monk...	0.000000	0.00	0.040816	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	Browns Wood, Caldecotte, Kents Hill, Tilbrook,...	0.020833	0.00	0.041667	0.000000	0.000000	0.020833	0.000000	0.00	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	Central Milton Keynes, Campbell Park	0.000000	0.00	0.000000	0.020408	0.000000	0.000000	0.020408	0.00	0.000000	...	0.000000	0.020408	0.000000	0.000000	0.040816	0.000000	0.000000	0.020408	0.000000	0.000000
10	Church Green, Far Bletchley, Old Bletchley, Ne...	0.000000	0.00	0.000000	0.040000	0.000000	0.020000	0.020000	0.00	0.000000	...	0.020000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.000000	0.000000
11	Crownhill, Elfield Park, Grange Farm, Oakhill,...	0.000000	0.00	0.000000	0.040816	0.000000	0.020408	0.020408	0.00	0.000000	...	0.000000	0.020408	0.000000	0.040816	0.020408	0.000000	0.000000	0.020408	0.000000	0.000000
12	Crownhill, Grange Farm, Great Holm, Hazeley, L...	0.000000	0.00	0.000000	0.000000	0.021277	0.000000	0.000000	0.00	0.000000	...	0.000000	0.021277	0.000000	0.021277	0.000000	0.000000	0.021277	0.021277	0.000000	0.000000
13	Denbigh, Mount Farm	0.000000	0.00	0.000000	0.040816	0.000000	0.020408	0.020408	0.00	0.000000	...	0.020408	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040816	0.000000	0.000000
14	Emerson Valley, Furzton, Kingsmead, Oxley Park...	0.000000	0.00	0.000000	0.020833	0.000000	0.041667	0.020833	0.00	0.000000	...	0.000000	0.000000	0.000000	0.020833	0.000000	0.000000	0.020833	0.020833	0.000000	0.000000
15	Fairfields, Fullers Slade, Galley Hill, Kiln F...	0.000000	0.00	0.020408	0.000000	0.020408	0.040816	0.000000	0.00	0.020408	...	0.000000	0.020408	0.000000	0.000000	0.000000	0.020408	0.020408	0.020408	0.000000	0.000000
16	Greenleys, Hodge Lea, Old Wolverton, Stacey Bu...	0.000000	0.00	0.000000	0.000000	0.020833	0.041667	0.000000	0.00	0.020833	...	0.000000	0.020833	0.000000	0.000000	0.000000	0.020833	0.000000	0.020833	0.000000	0.020833

We then transform this into the most important categories in each Borough, giving us a table as follows;

	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	Ashland, Beanhill, Bleak Hall, Coffee Hall, Ea...	Coffee Shop	Restaurant	Fast Food Restaurant	Sandwich Place	Breakfast Spot	Fish & Chips Shop	Food Truck	Café	Indian Restaurant	Burger Joint
1	Aspley Guise, Battlesden, Bow Brickhill, Drayt...	Fast Food Restaurant	Café	Indian Restaurant	Pizza Place	Coffee Shop	Chinese Restaurant	Sandwich Place	Fish & Chips Shop	Restaurant	Italian Restaurant
2	Bancroft/Bancroft Park, Blue Bridge, Bradwell ...	Coffee Shop	Fast Food Restaurant	Sandwich Place	Fish & Chips Shop	Indian Restaurant	Burger Joint	Pizza Place	Restaurant	Breakfast Spot	Turkish Restaurant
3	Beachampton, Calverton, Castlethorpe Cosgrove,...	Indian Restaurant	Pizza Place	Sandwich Place	Coffee Shop	Chinese Restaurant	Café	Fast Food Restaurant	Middle Eastern Restaurant	Pub	Fish & Chips Shop
4	Blakelands, Conniburrow, Downs Barn, Giffard P...	Coffee Shop	Fast Food Restaurant	Sandwich Place	Restaurant	Burger Joint	Breakfast Spot	Sushi Restaurant	Asian Restaurant	Latin American Restaurant	Fish & Chips Shop
5	Bolbeck Park, Downhead Park, Fox Milne, Newlan...	Coffee Shop	Sandwich Place	Burger Joint	Fish & Chips Shop	Food Truck	Ice Cream Shop	Pizza Place	Restaurant	Fast Food Restaurant	English Restaurant
6	Brickfields, Central Bletchley, Fenny Stratfor...	Fast Food Restaurant	Café	Pizza Place	Chinese Restaurant	Coffee Shop	Indian Restaurant	Restaurant	Sandwich Place	Italian Restaurant	Asian Restaurant
7	Brinklow, Broughton, Kingston, Middleton, Monk...	Coffee Shop	Chinese Restaurant	Café	Restaurant	Sandwich Place	Gastropub	Fish & Chips Shop	Food Truck	Indian Restaurant	Pizza Place
8	Browns Wood, Caldecotte, Kents Hill, Tilbrook,...	Fast Food Restaurant	Coffee Shop	Café	Restaurant	Sandwich Place	Chinese Restaurant	Fish & Chips Shop	American Restaurant	Gastropub	Indian Restaurant
9	Central Milton Keynes, Campbell Park	Coffee Shop	Fast Food Restaurant	Restaurant	Sandwich Place	Fish & Chips Shop	Food Truck	Indian Restaurant	Italian Restaurant	Breakfast Spot	Burger Joint
10	Church Green, Far Bletchley, Old Bletchley, Ne...	Café	Pizza Place	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Turkish Restaurant	Coffee Shop	Asian Restaurant	Indian Restaurant	Sandwich Place
11	Crownhill, Elfield Park, Grange Farm, Oakhill,...	Coffee Shop	Sandwich Place	Fast Food Restaurant	Chinese Restaurant	Indian Restaurant	Steakhouse	Asian Restaurant	Italian Restaurant	Restaurant	Food Truck
12	Crownhill, Grange Farm, Great Holm, Hazeley, L...	Coffee Shop	Fish & Chips Shop	Café	Sandwich Place	Indian Restaurant	Fast Food Restaurant	Chinese Restaurant	Food Truck	Steakhouse	Turkish Restaurant
13	Denbigh, Mount Farm	Café	Fast Food Restaurant	Pizza Place	Chinese Restaurant	Indian Restaurant	Italian Restaurant	Asian Restaurant	Coffee Shop	Turkish Restaurant	Sandwich Place
14	Emerson Valley, Furzton, Kingsmead, Oxley Park...	Chinese Restaurant	Café	Fast Food Restaurant	Pizza Place	Food Truck	Fish & Chips Shop	Restaurant	Indian Restaurant	Italian Restaurant	Portuguese Restaurant
15	Fairfields, Fullers Slade, Galley Hill, Kiln F...	Indian Restaurant	Fast Food Restaurant	Fish & Chips Shop	Coffee Shop	Café	Pizza Place	Chinese Restaurant	Sandwich Place	Pub	Bakery
16	Greenleys, Hodge Lea, Old Wolverton, Stacey Bu...	Indian Restaurant	Fast Food Restaurant	Fish & Chips Shop	Coffee Shop	Chinese Restaurant	Middle Eastern Restaurant	Bakery	Sandwich Place	Pub	Pizza Place

This is the really important source data telling us which categories occur most in each Borough, and is the basis that we use for the machine learning clustering.

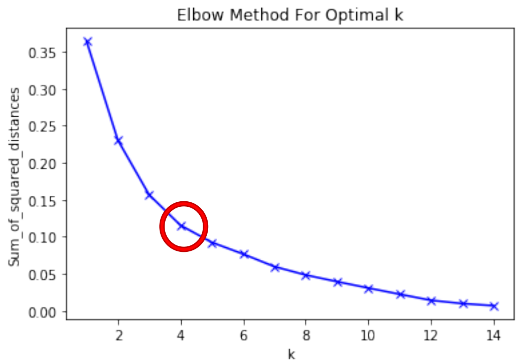
Clustering

We are going to use a machine learning algorithm called ***k-means*** (see below) to determine the optimal number of clusters and grouping for all the venues.

> *k-means clustering* is a method of vector quantization, originally from signal processing, that aims to partition n observations into *k clusters* in which each observation belongs to the *cluster* with the nearest *mean* (*cluster* centres or *cluster* centroid), serving as a prototype of the *cluster*.

One of the challenges with ***k-means***, is selecting the ***k*** value for number of clusters. It is effectively an iterative process, where the results from different ***k*** values are examined to determine the optimal value for ***k***.

We can use the 'elbow method' to plot the different values of ***k***, against their error rate, and the point on the plot where there is greatest change in angle, is likely to be the optimal value of ***k***. This is what we found;



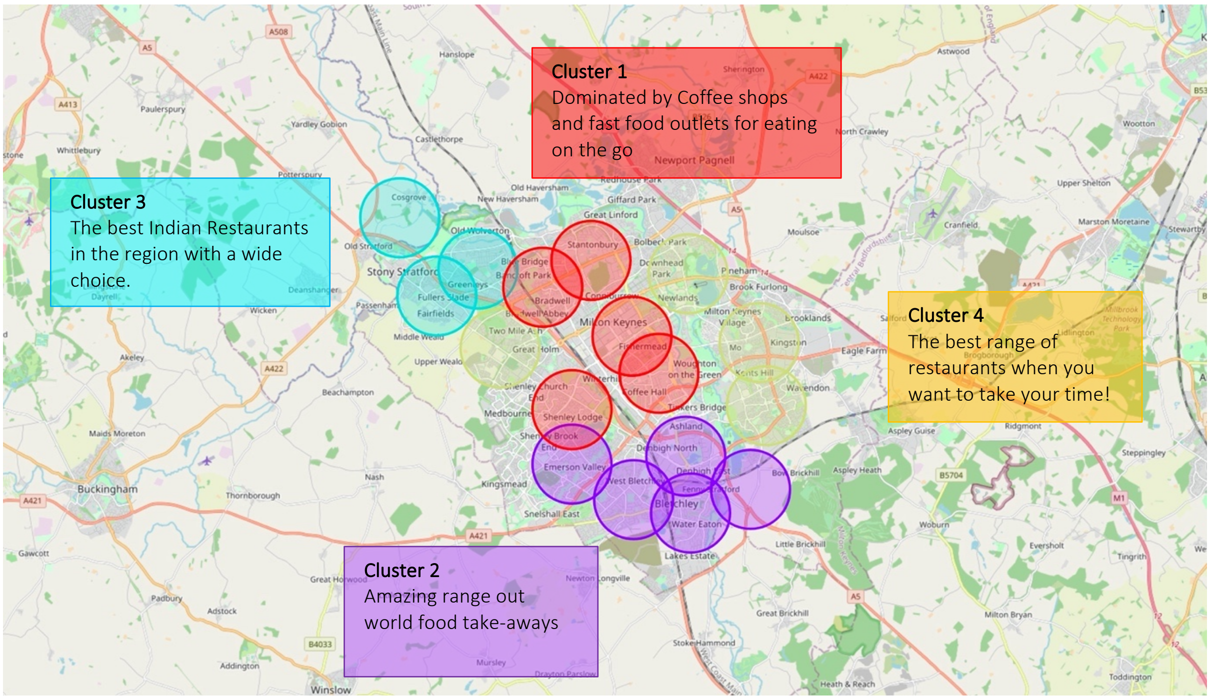
So it appears, that the optimal value of ***k = 4***.

Clustering around our venue data, we get the following data table, which defines our clusters, and the most populous venue categories in each.

	Postcode	Town	Borough	Local authority area(s)	lat	long	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	MK1	MILTON KEYNES	Denbigh, Mount Farm	Milton Keynes	52.007169	-0.727622	1	Café	Fast Food Restaurant	Pizza Place	Chinese Restaurant	Indian Restaurant	Italian Restaurant	Asian Restaurant	Coffee Shop	Turkish Restaurant	Sandwich Place
1	MK2	MILTON KEYNES	Brickfields, Central Bletchley, Fenny Stratfor...	Milton Keynes	51.991884	-0.725566	1	Fast Food Restaurant	Café	Pizza Place	Chinese Restaurant	Coffee Shop	Indian Restaurant	Restaurant	Sandwich Place	Italian Restaurant	Asian Restaurant
2	MK3	MILTON KEYNES	Church Green, Far Bletchley, Old Bletchley, Ne...	Milton Keynes, Aylesbury Vale	51.995494	-0.750159	1	Café	Pizza Place	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Turkish Restaurant	Coffee Shop	Asian Restaurant	Indian Restaurant	Sandwich Place
3	MK4	MILTON KEYNES	Emerson Valley, Furzton, Kingsmead, Oxley Park...	Milton Keynes, Aylesbury Vale	52.004916	-0.776778	1	Chinese Restaurant	Café	Fast Food Restaurant	Pizza Place	Food Truck	Fish & Chips Shop	Restaurant	Indian Restaurant	Italian Restaurant	Portuguese Restaurant
4	MK5	MILTON KEYNES	Crownhill, Elfield Park, Grange Farm, Oakhill,...	Milton Keynes	52.019522	-0.776946	0	Coffee Shop	Sandwich Place	Fast Food Restaurant	Chinese Restaurant	Indian Restaurant	Steakhouse	Asian Restaurant	Italian Restaurant	Restaurant	Food Truck

Results

We generated the following clusters and overlaid them on the map of Milton Keynes



Cluster 1 - Coffee Shops and Fast Food if you have time to stop

This is where you want to go to just hang out, be cool, meet your friends and socialise over lunch or a coffee. Plenty of coffee shops and places to eat. This place is dominated by coffee shops, fast food and sandwich places.

	Town	Postcode	long	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	MILTON KEYNES	MK5	-0.776946	0	Coffee Shop	Sandwich Place	Fast Food Restaurant	Chinese Restaurant	Indian Restaurant	Steakhouse	Asian Restaurant	Italian Restaurant	Restaurant	Food Truck
5	MILTON KEYNES	MK6	-0.739299	0	Coffee Shop	Restaurant	Fast Food Restaurant	Sandwich Place	Breakfast Spot	Fish & Chips Shop	Food Truck	Café	Indian Restaurant	Burger Joint
8	MILTON KEYNES	MK9	-0.750867	0	Coffee Shop	Fast Food Restaurant	Restaurant	Sandwich Place	Fish & Chips Shop	Food Truck	Indian Restaurant	Italian Restaurant	Breakfast Spot	Burger Joint
12	MILTON KEYNES	MK13	-0.789564	0	Coffee Shop	Fast Food Restaurant	Sandwich Place	Fish & Chips Shop	Indian Restaurant	Burger Joint	Pizza Place	Restaurant	Breakfast Spot	Turkish Restaurant
13	MILTON KEYNES	MK14	-0.768787	0	Coffee Shop	Fast Food Restaurant	Sandwich Place	Restaurant	Burger Joint	Breakfast Spot	Sushi Restaurant	Asian Restaurant	Latin American Restaurant	Fish & Chips Shop

- MK5, Crownhill, Elfield Park, Grange Farm, Oakhill, Knowlhill, Loughton, Medbourne, Shenley Brook End, Shenley Church End, Shenley Lodge, Shenley Wood
- MK6, Ashland, Beanhill, Bleak Hall, Coffee Hall, Eaglestone, Fishermead, Leadenhall, Netherfield, Oldbrook, Peartree Bridge, Redmoor, Springfield, Tinkers Bridge, Woughton on the Green, Woughton Park, Simpson, Winterhill
- MK9, Central Milton Keynes, Campbell Park
- MK13 Bancroft/Bancroft Park, Blue Bridge, Bradwell and Bradwell Abbey, Bradwell Common, New Bradwell, Heelands, Stonebridge
- MK14, Blakelands, Conniburrow, Downs Barn, Giffard Park, Great Linford, Linford Wood, Oakridge Park, Stantonbury, Neath Hill, Redhouse Park

Cluster 2 - World Foods at your fingertips

This is where you want to go to find a range of Fast Food and restaurants covering a diverse range of cuisines.

	Town	Postcode	long	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	MILTON KEYNES	MK1	-0.727622	1	Café	Fast Food Restaurant	Pizza Place	Chinese Restaurant	Indian Restaurant	Italian Restaurant	Asian Restaurant	Coffee Shop	Turkish Restaurant	Sandwich Place
1	MILTON KEYNES	MK2	-0.725566	1	Fast Food Restaurant	Café	Pizza Place	Chinese Restaurant	Coffee Shop	Indian Restaurant	Restaurant	Sandwich Place	Italian Restaurant	Asian Restaurant
2	MILTON KEYNES	MK3	-0.750159	1	Café	Pizza Place	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Turkish Restaurant	Coffee Shop	Asian Restaurant	Indian Restaurant	Sandwich Place
3	MILTON KEYNES	MK4	-0.776778	1	Chinese Restaurant	Café	Fast Food Restaurant	Pizza Place	Food Truck	Fish & Chips Shop	Restaurant	Indian Restaurant	Italian Restaurant	Portuguese Restaurant
15	MILTON KEYNES	MK17	-0.699289	1	Fast Food Restaurant	Café	Indian Restaurant	Pizza Place	Coffee Shop	Chinese Restaurant	Sandwich Place	Fish & Chips Shop	Restaurant	Italian Restaurant

- MK1, Denbigh, Mount Farm
- MK2, Brickfields, Central Bletchley, Fenny Stratford, Water Eaton
- MK3, Church Green, Far Bletchley, Old Bletchley, Newton Leys, West Bletchley
- MK4, Emerson Valley, Furzton, Kingsmead, Oxley Park, Shenley Brook End, Snelshall West, Tattenhoe, Tattenhoe Park, Westcroft, Whaddon, Woodhill
- MK17, Aspley Guise, Battlesden, Bow Brickhill, Drayton Parslow, Eversholt, Great Brickhill, Great Horwood, Hulcote, Little Brickhill, Little Horwood, Milton Bryan, Mursley, Nash, Newton Longville, Potsgrove, Salford, Stoke Hammond, Swanbourne, Thornton, * * Tingrith, Wavendon, Whaddon, Woburn, Woburn Sands

Cluster 3 - Curry Central

There are some awesome Indian Restaurants in and around Milton Keynes, and you will find them in these areas;

	Town	Postcode	long	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
10	MILTON KEYNES	MK11	-0.835747	2	Indian Restaurant	Fast Food Restaurant	Fish & Chips Shop	Coffee Shop	Café	Pizza Place	Chinese Restaurant	Sandwich Place	Pub	Bakery
11	MILTON KEYNES	MK12	-0.817932	2	Indian Restaurant	Fast Food Restaurant	Fish & Chips Shop	Coffee Shop	Chinese Restaurant	Middle Eastern Restaurant	Bakery	Sandwich Place	Pub	Pizza Place
16	MILTON KEYNES	MK19	-0.851795	2	Indian Restaurant	Pizza Place	Sandwich Place	Coffee Shop	Chinese Restaurant	Café	Fast Food Restaurant	Middle Eastern Restaurant	Pub	Fish & Chips Shop

- MK11, Fairfields, Fullers Slade, Galley Hill, Kiln Farm, Stony Stratford
- MK12, Greenleys, Hodge Lea, Old Wolverton, Stacey Bushes, Wolverton, Wolverton Mill
- MK19, Beachampton, Calverton, Castlethorpe Cosgrove, Deanshanger, Hanslope, Haversham, Little Linford, Long Street, Old Stratford, Passenham, Upper Weald, Wicken

Cluster 4 - Relax, Chill, & Dine

There are some awesome Indian Restaurants in and around Milton Keynes, and you will find them in these areas. On the surface, this looks similar to Cluster 1, but when you look a little deeper you find that its actually dominated by a range of restaurants, and less frequency of fast food. In other words, you go here to find the best restaurants.

	Town	Postcode	long	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
6	MILTON KEYNES	MK7	-0.692578	3	Fast Food Restaurant	Coffee Shop	Café	Restaurant	Sandwich Place	Chinese Restaurant	Fish & Chips Shop	American Restaurant	Gastropub	Indian Restaurant
7	MILTON KEYNES	MK8	-0.808131	3	Coffee Shop	Fish & Chips Shop	Café	Sandwich Place	Indian Restaurant	Fast Food Restaurant	Chinese Restaurant	Food Truck	Steakhouse	Turkish Restaurant
9	MILTON KEYNES	MK10	-0.695734	3	Coffee Shop	Chinese Restaurant	Café	Restaurant	Sandwich Place	Gastropub	Fish & Chips Shop	Food Truck	Indian Restaurant	Pizza Place
14	MILTON KEYNES	MK15	-0.726404	3	Coffee Shop	Sandwich Place	Burger Joint	Fish & Chips Shop	Food Truck	Ice Cream Shop	Pizza Place	Restaurant	Fast Food Restaurant	English Restaurant

- MK7, Browns Wood, Caldecotte, Kents Hill, Tilbrook, Walton, Walton Hall, Wavendon, Wavendon Gate, Old Farm Park
- MK8, Crownhill, Grange Farm, Great Holm, Hazeley, Loughton, Loughton Lodge, Two Mile Ash, Whitehouse, Wymbush
- MK10, Brinklow, Broughton, Kingston, Middleton, Monkston, Oakgrove

- MK15, Bolbeck Park, Downhead Park, Fox Milne, Newlands, Pennyland, Pineham, Northfield, Tongwell, Willen Park, Willen, Willen Lake, Great Woolstone, Little Woolstone

Discussion

So does this analysis answer the question? **Where to go in smaller cities when there are no-online guides?** ...and **whether data science could teach me a thing a or two about the place I have lived in for the last 25 years?**

I think the answer is a clear **YES**

Having lived in Milton Keynes for over 25 years, this exercise has taught me somethings I didn't know, and confirmed other things I already assumed. I also didn't know that we have so many coffee shops, that we have a few areas of the city that totally dominated by them, and there are certain areas of the city you go to for certain cuisines. Has the application of Data Science analysis to my home been insightful - YES, but also its been really interesting, and frankly quite easy to do so (once you know the methodology and techniques).

Also, I did iterate around a number of different sized data sets, increasing the area for inspection (by including the areas of Buckingham, Bedford, Olney etc), and also increasing the range of venues retrieved from Foursquare. My conclusions in doing so, it didn't really bring new insight, and in some circumstances reduced the number of optimal recommended clusters to just 2. Adding in more data, or broadening out the range of the search just didn't bring in new statistically relevant information generating new insights. Broadly I considered this as just adding more noise to the signals already in the data. Therefore I converged around a dataset that comprised of the core postcodes within Milton Keynes, and just the food/restaurant venues. This presented the most useful results.

Conclusion

In terms of answering the business problem, we draw the following answers;

- The best places to go to for Fast food, Sandwiches and Coffee Shops can be found in MK9, MK5, MK6 - essentially the Centre of Milton Keynes.
- You will find the most Indian Restaurants in MK11, MK12, and MK19 - especially in Stoney Stratford.
- If you want a choice of world food cuisines, head for MK1, MK2, MK3
- When you want to take your time, relax, chillout and have lunch or dinner in range of good restaurants, head for MK7, MK8, MK10 - especially Woburn Sands.

I think there are multiple conclusions that can be drawn from this exercise, that cover both the intention of the analysis, but also the journey to get to the outcomes.

1. Can you easily use Data Science to determine the best areas of a city to go to when there are no guides or websites available - **YES!**

2. Is this easily re-producible for other cities? **YES** and its actually quite easy to do so. You only need the regions/boroughs, and a way to get the latitude and longitudes to follow the same process.
3. Can Data Science teach me a thing or two about my home town, and strangle the answer has to be **YES**. I never knew we had so many places to go to eat, that some areas are so dominated by coffee bars. But equally, it confirmed the areas where you should go to find the best Indian Restaurants.
4. Can you operate these techniques on smaller cities that don't have the massive depth and range of data available. Again, **YES** as my analysis for Milton Keynes shows.

In terms of the journey of exploring this business problems, what i discovered is how easy it is to both obtain data on pretty much any topics you like, the power of dataframes to be able to manipulate and clean the data so that you can derive insights from it. There are also some really powerful libraries that make visualisation straight forward, and algorithms in libraries that would take weeks (or more) if I were to code them in C++ or whatever language is de jour!

Acknowledgments

Thanks goes to the following sources and inspiration for data, ideas and techniques to make this possible.

- Obviously inspired by the Coursera Data Science professional course itself; https://www.coursera.org/professional-certificates/ibm-data-science

- The source postcode for Milton Keynes from Wikipedia: https://en.wikipedia.org/wiki/MK_postcode_area

- UK Govt, Office for National Statistics: https://www.ons.gov.uk/

- The working Python Notebook, in my github account: https://github.com/brian-naylor/Coursera_Capstone

- Numerous articles on python, data manipulation, code snippets, and examples.

- Github, acting as a fantastic resource to speed up solving a problem within your code.

- Blogs from other students who have completed the Data Science course, such as; https://medium.com/@radialee/capstone-project-the-battle-of-neighborhoods-in-tokyo-restaurants-45a503e65ff

- https://www.linkedin.com/pulse/applied-data-science-capstone-project-restaurant-wagner-mba/?articleId=6670274875946622976

**** END OF REPORT ****