# Coursera Report: Battle of the Neighbourhoods

# Business problem:

- Is there some objective data that the general public (or tour operators) can use to make recommendations instead of relying on the often deceitful marketing material issued by the cities themselves or by the cities' tourist boards?
- Can we use data science to tell us whether it is worth visiting an English city without being disappointed or mislead by advertising?

# Data

- list of cities and towns in England
- https://www.thegeographist.com/uk-cities-population-1000
- scrape the data and then clean it, removing localities in Wales and Scotland and taking the top 500 or so remaining places to analyze

# Data (2)

- Data example:

1 1 London London London 8,907,918

2 1 Birmingham West Midlands West Midlands 1,153,717

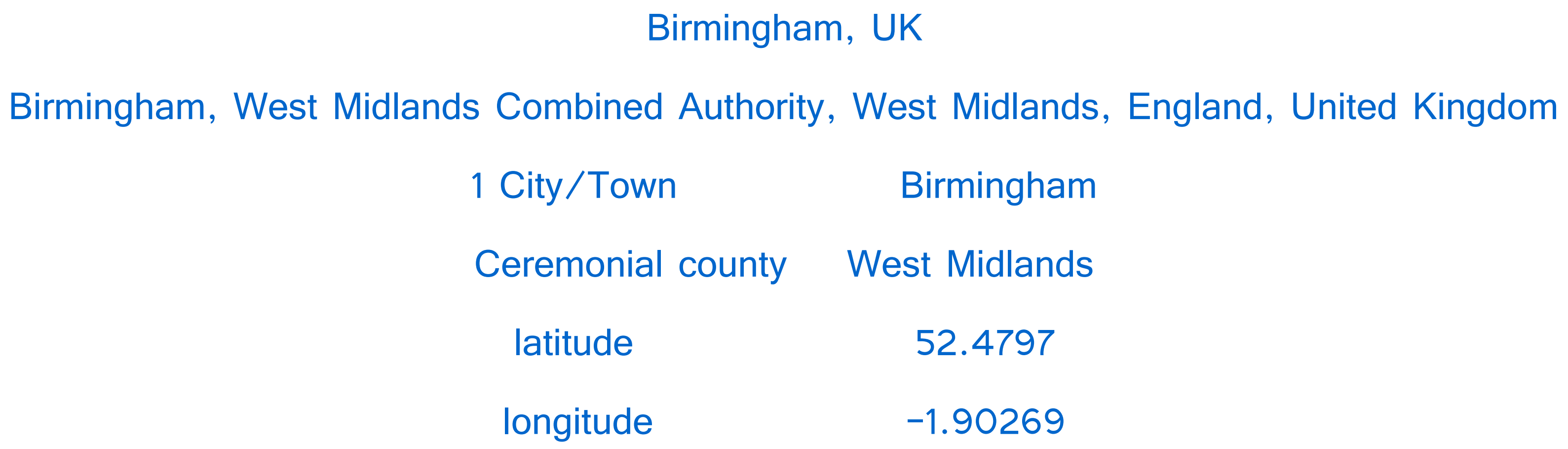
3 1 Glasgow Glasgow Scotland 612,040

4 1 Liverpool Merseyside North West 579,256

5 1 Bristol Bristol South West 571,922

# Data (3)

- Use a geocoder service to find the latitude and longitude of the center of these localities
- Example:



Birmingham, UK

Birmingham, West Midlands Combined Authority, West Midlands, England, United Kingdom

| | |
|---|---|
| 1 City/Town | Birmingham |
| Ceremonial county | West Midlands |
| latitude | 52.4797 |
| longitude | -1.90269 |

# Data (4)

- Use FourSquare venue explorer to find what venues are available within a reasonable distance to the city center
- Example output:

```
0    A    52.449601    -1.819154  Costa Coffee    52.446411    -1.822441    Café
1    A    52.449601    -1.819154  Wilko    52.446645    -1.823626    Store
2    A    52.449601    -1.819154  Sainsbury's    52.445749    -1.820319  Supermarket
3    A    52.449601    -1.819154  Subway    52.446418    -1.823159  Restaurant
4    A    52.449601    -1.819154  The Spread Eagle (Wetherspoon)    52.446488
```

# Methodology

- Looked at how accurate the data was at representing English localities
- Locality names were misspelled ("Corsett"="Consett")
- Others ambiguous (representing localities in Scotland instead of England) "Ferryhill"="Ferryhill, County Durham"
- The data was cleaned to correct these errors

# Methodology (2)

- FourSquare output was analysed
- For very small localities the FourSquare data was too sparse
- The data was filtered to take the top 300 localities instead of the top 1000

# Methodology (3)

- K-means clustering algorithm was used to attempt to cluster the data in a meaningful way
- Another problem with FourSquare: inconsistency of nomenclature

  Example: “art gallery” and “art museum”, “coffe shop” and “cafe” being different categories
- Algorithm corrected to ensure consitency

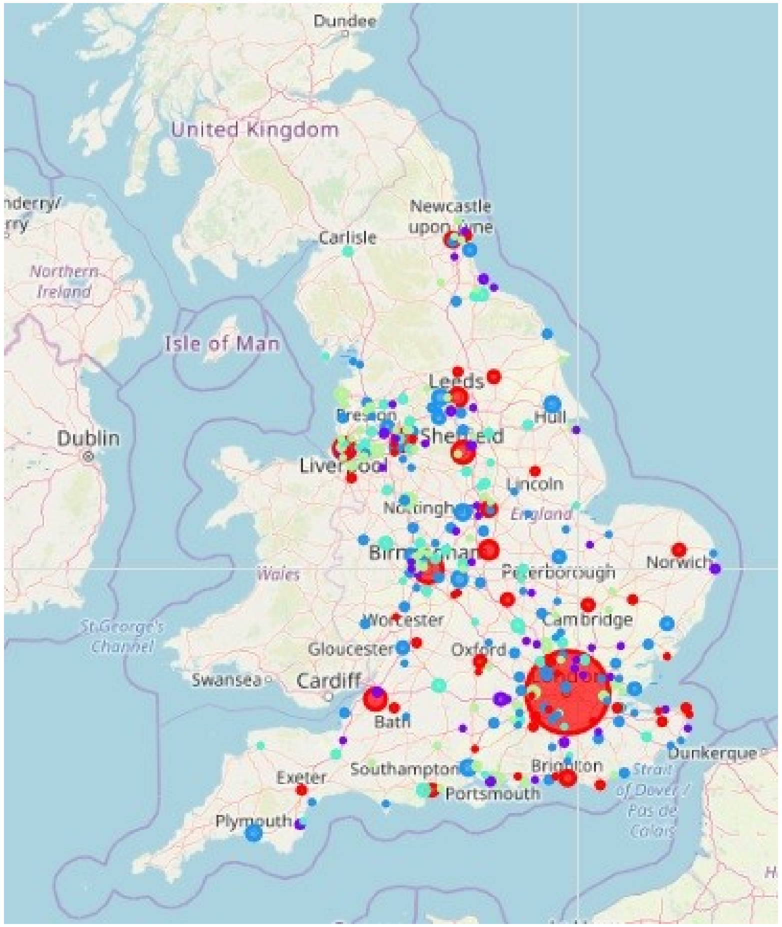
# Methodology (4)

- Elbow Method was used to get an idea of the optimal number of clusters.
- k=3 as the optimal number
- Seemed very low so experiments were carried out using both k=3 and k=5

# Results

- Venue data was gathered from FourSquare.
- Abundance of information for particular locations...
- BUT
- It was evident that the venue information was incomplete
- Large number of venues were not contained within the FourSquare data.

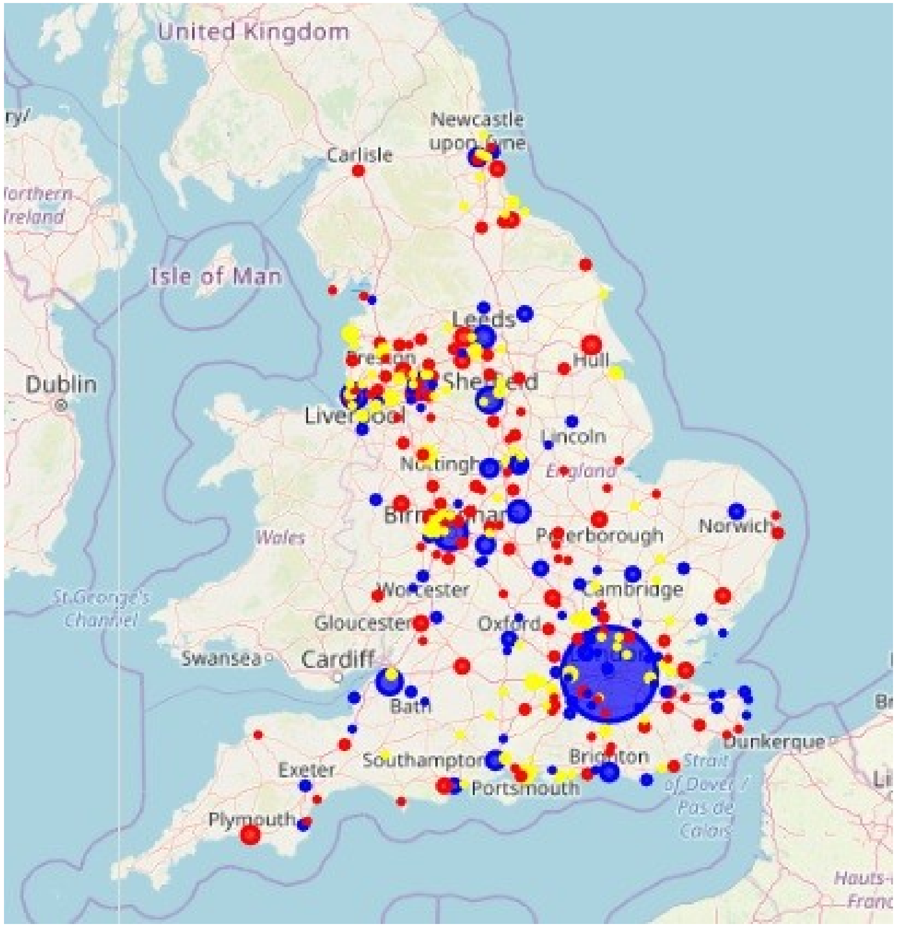
# Results (2): K=5


Dundee
United Kingdom
Newcastle upon Tyne
Carlisle
Northern Ireland
Isle of Man
Leeds
Preston
Hull
Sheffield
Dublin
Liverpool
Lincoln
Nottingham
England
Birmingham
Norwich
Wales
Peterborough
Worcester
Cambridge
St George's Channel
Gloucester
Oxford
Swansea
Cardiff
London
Bath
Dunkerque
Exeter
Southampton
Brighton
Strait of Dover / Pas de Calais
Portsmouth
Plymouth

# Results (3): K=5

1) big cities with a vibrant restaurant and coffee culture as well as some smaller towns with the same vibe.
2) smaller "downmarket" towns with a mix of various shops, pubs and cafes and for which FourSquare has incomplete data
3) slightly more upmarket towns with an interesting mix of various types of shops, restaurants and cafes.
4) towns where the main attraction is an abundance of mainstream shops, complemented by fast food places, pubs and cafes
5) towns where inexplicably there is very little FourSquare data. This inclues large towns such as Portsmouth, Luton and Blackpool which certainly don't lack in venues!

# Results (4): K=3


United Kingdom
Newcastle upon Tyne
Carlisle
Northern Ireland
Isle of Man
Dublin
Leeds
Preston
Hull
Sheffield
Liverpool
Lincoln
Nottingham
England
Birmingham
Norwich
Peterborough
Wales
Worcester
Cambridge
St George's Channel
Gloucester
Oxford
Swansea
Cardiff
London
Bath
Dunkerque
Exeter
Southampton
Brighton
Strait of Dover / Pas de Calais
Portsmouth
Plymouth

# Results (5): K=3

1) Towns where the main attraction is an abundance of mainstream shops, complemented by fast food places, pubs and cafes

2) Big cities with a vibrant restaurant and coffee culture, perhaps with Art and Concert venues as well as some smaller towns with the same vibe.

3) Towns where inexplicably there is very little FourSquare data. This inclues large towns such as Portsmouth, Luton and Blackpool which certainly don't lack in venues!

# Discussion

- Clustering in 5 or in 3 groups didn’t result in any noticeable difference.
- In the case of the 5 clusters, three of the clusters didn’t seem to be distinguishable
- the division in 3 clusters made sense, as predicted by the Elbow Method.
- the FourSquare data was very unreliable, with a very large number of important venues missing even in big cities where there appeared to be a good amount of data.

# Conclusion

- Cities and towns in England are really "identikit", bland city centres as far as venues such as shops, cafes etc. is concerned.
- BUT:
- FourSquare data was hopelessly incomplete, with very little in terms of landmarks and very incomplete data as regards shops and other amenitie

# Conclusion (2)

- If you want to discover something new perhaps you should try one of the many cities for which FourSquare has no data
- Best for "coffee culture" are big cities such as London, Birmingham, Manchester, Leeds.
- Best for a more traditional feel are smaller ones such as Bradford and Plymouth
- FourSquare is woefully incomplete and cannot be relied upon fully for this task.