

Coursera Capstone Project - The Battle of Neighborhoods

Abdullah Sharaf

2020

Contents

| | |
|---|---|
| The introduction: Business problem..... | 0 |
| Data collection and preparation | 1 |
| Methodology..... | 3 |
| Analysis – KMean clustering | 5 |
| Results and Discussion | 6 |
| Conclusion..... | 6 |

The introduction: Business problem

The project report is prepared as a part of IBM Data Science certification course offered by Coursera. The scope of the project to use the Foursquare API and Machine Learning algorithm to analyze the city of Liverpool, the UK with respect to study and work there in Liverpool and for tourists. Liverpool is an important city in the United Kingdom, which is the 5th largest economy in the world, and Liverpool is very attractive for business, tourists and is the financial city of the United Kingdom. In the final project, I am going to focus on identifying 5 most famous things that Liverpool's do as neighborhood vide and highlight the potential business opportunity. Also, this would help students to complete study and know work opportunities. about tourists to find easily where to shop, eat and sightseeing while enjoying their holidays in the UK. This project has been

chosen mainly to keep students and tourists, provide sufficient evidence about business opportunities.

Data collection and preparation

- a) Getting the information about City of Liverpool
I have used Google, Python pandas and BeautifulSoup library to extract data and grouping into data frames. The city of Liverpool has 21 neighborhood boroughs. I am going to use Foursquare API for extracting the data from the web.
Using Python library – Geopy, neighborhood longitude and latitude have been generated , and has been added to the data frame (created above), see below for neighborhood and longitude/latitude data frames. Refer Table 1 and 2 for Neighborhood and Neighborhood with longitude and latitude respectively.
- b) Foursquare API to get venue locations.
Foursquare API has been used to get large amount of data about city of Liverpool in the json file format. Refer table 3 for most popular locations for each neighborhood within a radius of 1000 m.
- c) I would like to use one of the simple and famous machine learning algorithm – KMeans clustering - (unsupervised machine learning algorithm) segregating the based on similar characteristics.

Table 1: Neighborhood data frame

| Neighborhood | |
|--------------|-----------------------|
| 0 | Liverpool City Centre |
| 1 | Liverpool Waterfront |
| 2 | Baltic Triangle |
| 3 | Fazakerley |
| 4 | Norris Green |
| 5 | Belle Vale |
| 6 | Orrell Park |
| 7 | Ropewalks |
| 8 | Cavern Quarter |
| 9 | Georgian Quarter |
| 10 | Cantrill Farm |
| 11 | St George's Quarter |
| 12 | Warbreck |
| 13 | Knowledge Quarter |
| 14 | Islington |
| 15 | Sefton Park |
| 16 | Vauxhall |
| 15 | Sefton Park |
| 16 | Vauxhall |

Table 2: Neighborhood with longitude and latitude data frame

| | Neighborhood | Latitude | Longitude |
|----|-----------------------|-----------|------------|
| 0 | Liverpool City Centre | 14.678012 | 121.097637 |
| 1 | Liverpool Waterfront | 53.602502 | -2.843294 |
| 2 | Baltic Triangle | 53.395745 | -2.980043 |
| 3 | Fazakerley | 53.468984 | -2.936971 |
| 4 | Norris Green | 53.446428 | -2.925689 |
| 5 | Belle Vale | 53.395074 | -2.864178 |
| 6 | Orrell Park | 53.462029 | -2.962716 |
| 7 | Ropewalks | 53.403172 | -2.981180 |
| 8 | Cavern Quarter | 53.407214 | -2.987353 |
| 9 | Georgian Quarter | 53.398541 | -2.969100 |
| 10 | Cantrill Farm | 43.731338 | -79.804172 |
| 11 | St George's Quarter | 53.647359 | -1.786240 |
| 12 | Warbreck | 53.839476 | -3.041475 |
| 13 | Knowledge Quarter | 53.405647 | -2.969243 |
| 14 | Islington | 51.538429 | -0.099905 |
| 15 | Sefton Park | 53.381829 | -2.936697 |
| 16 | Vauxhall | 53.710648 | -1.333605 |

| | | | |
|----|------------|------------|------------|
| 17 | Fincham | 52.628797 | 0.489968 |
| 18 | Knotty Ash | 53.415467 | -2.896207 |
| 19 | Stanley | -51.695057 | -57.849169 |
| 20 | Swanside | 53.414617 | -2.881365 |

Table 3 – Formatted data from json file using Foursquare API

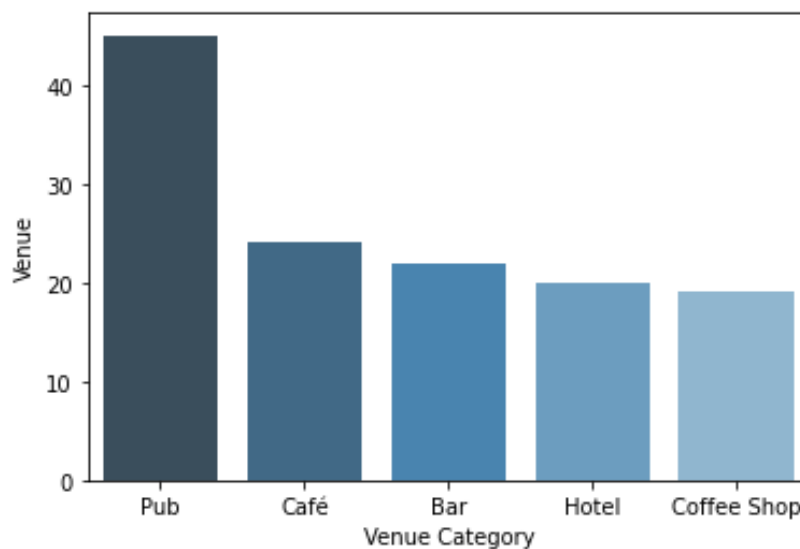
| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|-----|-----------------------|-----------------------|------------------------|--|----------------|-----------------|----------------------|
| 0 | Liverpool City Centre | 14.678012 | 121.097637 | Filinvest 1 Clubhouse | 14.681498 | 121.095315 | Club House |
| 1 | Liverpool City Centre | 14.678012 | 121.097637 | FILHAI Cooperative Store, Filinvest I | 14.681688 | 121.094703 | Grocery Store |
| 2 | Liverpool City Centre | 14.678012 | 121.097637 | Burger Machine | 14.680151 | 121.091390 | Burger Joint |
| 3 | Liverpool City Centre | 14.678012 | 121.097637 | Filinvest 1 Covered Basketball Court | 14.677475 | 121.089207 | Basketball Court |
| 4 | Liverpool City Centre | 14.678012 | 121.097637 | Filinvest 1 Block 1 Covered Court | 14.678945 | 121.089111 | Basketball Court |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 869 | Swanside | 53.414617 | -2.881365 | Burger King | 53.419822 | -2.883981 | Fast Food Restaurant |
| 870 | Swanside | 53.414617 | -2.881365 | LloydsPharmacy | 53.414107 | -2.874789 | Pharmacy |
| 871 | Swanside | 53.414617 | -2.881365 | Aldi | 53.419324 | -2.886415 | Supermarket |
| 872 | Swanside | 53.414617 | -2.881365 | Kwik-Fit | 53.418180 | -2.891934 | Auto Garage |
| 873 | Swanside | 53.414617 | -2.881365 | Bowring Park Road/Chelwood Avenue Bus Stop | 53.406477 | -2.885822 | Bus Stop |

874 rows x 7 columns

Methodology

d) Visualization and Data Exploration

Further to collecting the data and created data frames, a quick analysis has been done to understand the data on top 5 venues in each neighborhood. See below chart for top 5 venues. It can be seen from below Bar chart that Pub is the popular venues in most of the neighborhood areas.



There are 168 unique venue categories in total and top 5 are Pub, Café, Bar, Hotel and Coffee shop. A data frame has been created for venue, grouped by neighborhood. It can be seen that there are plenty of neighborhoods more that 100 venues (put max 100 as limit), shows that there is significant contribution to UK economy.

| | Neighborhood | Venue |
|----|-----------------------|-------|
| 0 | Baltic Triangle | 100 |
| 1 | Belle Vale | 15 |
| 2 | Cantrill Farm | 23 |
| 3 | Cavern Quarter | 100 |
| 4 | Fazakerley | 6 |
| 5 | Fincham | 2 |
| 6 | Georgian Quarter | 96 |
| 7 | Islington | 100 |
| 8 | Knotty Ash | 13 |
| 9 | Knowledge Quarter | 100 |
| 10 | Liverpool City Centre | 8 |
| 11 | Liverpool Waterfront | 8 |
| 12 | Norris Green | 5 |
| 13 | Orrell Park | 11 |
| 14 | Ropewalks | 100 |
| 15 | Sefton Park | 33 |
| 16 | St George's Quarter | 92 |
| 17 | Stanley | 4 |

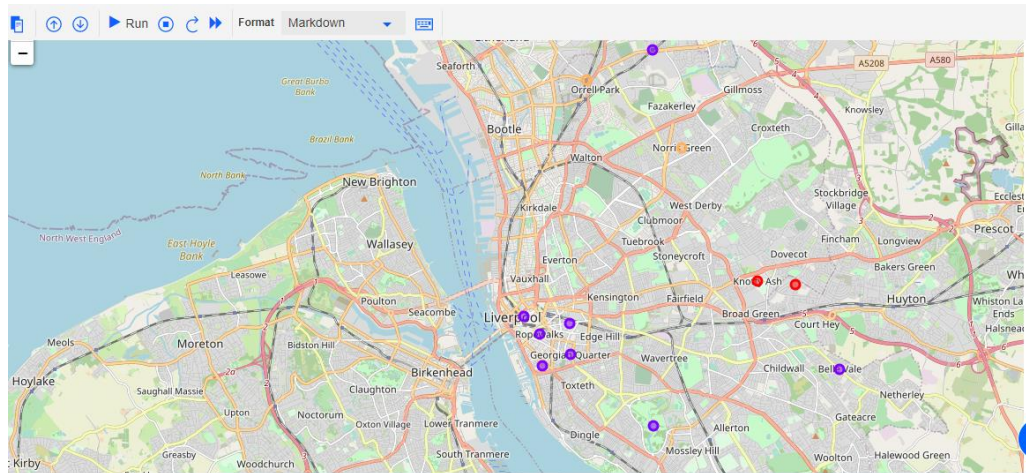
| | | |
|----|----------|----|
| 18 | Swanside | 7 |
| 19 | Vauxhall | 39 |
| 20 | Warbreck | 12 |

Data has been normalized used one hot encode normalization technique to reduce the impact of bias in analysis.

| Neighborhood | Afghan Restaurant | American Restaurant | Art Gallery | Art Museum | Arts & Crafts Store | Asian Restaurant | Athletics & Sports | Austrian Restaurant | Auto Garage | ... | Tram Station | Turkish Restaurant | Vegetarian / Vegan Restaurant | Video Game Store | Vietnamese Restaurant | Warehouse Store |
|--------------|-----------------------|---------------------|-------------|------------|---------------------|------------------|--------------------|---------------------|-------------|-----|--------------|--------------------|-------------------------------|------------------|-----------------------|-----------------|
| 0 | Baltic Triangle | 0.00 | 0.010000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.010000 | 0.020000 | 0.000000 | 0.00 | 0.000000 |
| 1 | Belle Vale | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.066667 |
| 2 | Cantrill Farm | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 3 | Cavern Quarter | 0.00 | 0.010000 | 0.01 | 0.010000 | 0.00 | 0.000000 | 0.01 | 0.000000 | ... | 0.000000 | 0.010000 | 0.010000 | 0.000000 | 0.00 | 0.000000 |
| 4 | Fazakerley | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 5 | Fincham | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 6 | Georgian Quarter | 0.00 | 0.010417 | 0.00 | 0.010417 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.010417 | 0.020833 | 0.000000 | 0.00 | 0.000000 |
| 7 | Islington | 0.01 | 0.000000 | 0.00 | 0.010000 | 0.02 | 0.000000 | 0.00 | 0.010000 | ... | 0.000000 | 0.000000 | 0.010000 | 0.000000 | 0.01 | 0.000000 |
| 8 | Knotty Ash | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.076923 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 9 | Knowledge Quarter | 0.00 | 0.010000 | 0.01 | 0.010000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.010000 | 0.020000 | 0.000000 | 0.00 | 0.000000 |
| 10 | Liverpool City Centre | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 11 | Liverpool Waterfront | 0.00 | 0.125000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 12 | Norris Green | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.200000 |
| 13 | Orrell Park | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 14 | Ropewalks | 0.00 | 0.020000 | 0.01 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.010000 | 0.020000 | 0.000000 | 0.00 | 0.000000 |
| 15 | Sefton Park | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.030303 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 16 | St George's Quarter | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.01087 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.010870 | 0.00 | 0.000000 |
| 17 | Stanley | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 18 | Swanside | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.142857 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 19 | Vauxhall | 0.00 | 0.051282 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.000000 | 0.000000 | 0.000000 | 0.025641 | 0.00 | 0.000000 |
| 20 | Warbreck | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | ... | 0.166667 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |

Analysis – KMean clustering

Machine learning algorithm – KMean clustering is used to cluster all neighborhood areas based on venue category and is divided into 5 clusters. By using Folium library – Liverpool map has been created with clustered neighborhood and following tables show details of the each cluster (1 to 5)



Cluster 1

| | Neighborhood | 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 |
|----|--------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 18 | Knotty Ash | Supermarket | Fast Food Restaurant | Chinese Restaurant | Grocery Store | Bus Stop | Sandwich Place | Auto Garage | Sports Club | Pub | Park |
| 20 | Swanside | Fast Food Restaurant | Supermarket | Pharmacy | Auto Garage | Grocery Store | Bus Stop | Discount Store | Fish & Chips Shop | Farmers Market | English Restaurant |

Cluster 2

| | Neighborhood | 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 | Liverpool Waterfront | Café | American Restaurant | Harbor / Marina | Soccer Field | Pub | Train Station | Restaurant | Yoga Studio | English Restaurant | Electronics Store |
| 2 | Baltic Triangle | Bar | Coffee Shop | Pub | Café | Hotel | Restaurant | Italian Restaurant | Burger Joint | Vegetarian / Vegan Restaurant | Indian Restaurant |
| 3 | Fazakerley | Supermarket | Coffee Shop | Grocery Store | Train Station | Golf Course | Café | Diner | Fast Food Restaurant | Farmers Market | English Restaurant |
| 5 | Belle Vale | Clothing Store | Supermarket | Fast Food Restaurant | Sandwich Place | Warehouse Store | Coffee Shop | Bus Stop | Discount Store | Optical Shop | Bakery |
| 7 | Ropewalks | Coffee Shop | Café | Bar | Hotel | Pub | Burger Joint | Beer Bar | Italian Restaurant | Department Store | Caribbean Restaurant |
| 8 | Cavern Quarter | Coffee Shop | Hotel | Café | Bar | History Museum | Burger Joint | Pub | Italian Restaurant | Beer Bar | Sandwich Place |
| 9 | Georgian Quarter | Bar | Pub | Coffee Shop | Café | Grocery Store | Restaurant | Hotel | Vegetarian / Vegan Restaurant | Italian Restaurant | Sandwich Place |
| 10 | Cantrill Farm | Pizza Place | Sandwich Place | Pet Store | Bar | Shopping Mall | Fast Food Restaurant | Breakfast Spot | Coffee Shop | Mediterranean Restaurant | Beer Store |
| 11 | St George's Quarter | Coffee Shop | Supermarket | Bar | Café | Pub | Indian Restaurant | Clothing Store | Furniture / Home Store | Tennis Court | Gym |
| 12 | Warbreck | Tram Station | Gastropub | Hotel | Pet Store | Casino | Sandwich Place | Beach | Thrift / Vintage Store | Train Station | Pizza Place |
| 13 | Knowledge Quarter | Coffee Shop | Bar | Pub | Café | Hotel | Restaurant | Theater | Burger Joint | Cocktail Bar | Comic Shop |
| 14 | Islington | Pub | French Restaurant | Mediterranean Restaurant | Gastropub | Japanese Restaurant | Bakery | Coffee Shop | Boutique | Fish Market | Cocktail Bar |
| 15 | Sefton Park | Pub | Grocery Store | Café | Pharmacy | Park | Hotel | Outdoor Sculpture | Sandwich Place | Restaurant | Playground |
| 16 | Vauxhall | Coffee Shop | Sporting Goods Shop | American Restaurant | Fast Food Restaurant | Italian Restaurant | Pizza Place | Sandwich Place | Pub | Hotel | Restaurant |

Cluster 3

| | Neighborhood | 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 |
|----|--------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------------|------------------------|
| 17 | Fincham | Hotel | Italian Restaurant | Diner | Fish & Chips Shop | Fast Food Restaurant | Farmers Market | English Restaurant | Electronics Store | Eastern European Restaurant | Donut Shop |

Cluster 4

| | Neighborhood | 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 |
|----|--------------|-----------------------|-----------------------|-----------------------|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------------|
| 19 | Stanley | Diner | Pub | Grocery Store | Tourist Information Center | Yoga Studio | Fast Food Restaurant | Farmers Market | English Restaurant | Electronics Store | Eastern European Restaurant |

Cluster 5

| | Neighborhood | 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 | Liverpool City Centre | Grocery Store | Basketball Court | Club House | Burger Joint | Golf Driving Range | Seafood Restaurant | Yoga Studio | Discount Store | Farmers Market | English Restaurant |
| 4 | Norris Green | Fish & Chips Shop | Warehouse Store | Pharmacy | Park | Discount Store | Diner | Fast Food Restaurant | Farmers Market | English Restaurant | Electronics Store |
| 6 | Orrell Park | Pizza Place | Supermarket | Train Station | Pub | Park | Sandwich Place | Discount Store | Pharmacy | Gym / Fitness Center | Grocery Store |

Results and Discussion

With above K-mean clustering, we can observe that cluster 2 are group of neighborhoods commonly known for pub, café and coffee shop and hotel and it is the biggest cluster. Cluster 1 is mainly for Supermarket and restaurant, Grocery store, Fish & chips shop and pharmacies. cluster 3 is having unique venue – watch-shop.

By using the data collected from Foursquare, I have completed a small project to understand most popular venues in the City of Liverpool neighborhood areas. This project report could be helpful to students and tourists, to know what we have in Liverpool city , and could be used as guide to tourist to find location for pubs, to eat out and explore city.

Conclusion

There is a potential opportunity to further explore the data to know more insight into the data from a business perspective like start a restaurant, hotel businesses, and travel companies, etc.