Manchester housing market clustering

By Andrei Staradubets

November 21, 2019

1. Introduction

1.1 Background

Nowadays that is alwais a problem to find a good place in new city for living. This project helps us as seekers for property in a new city to find interesting wards to consider within our budget.

1.2 Problem

In this project we will try to find all optimal UK wards within possible price for a client who wants to buy a property. Specifically, this report will be targeted to people, who wants to find all different kinds of offers on Manchester housing market with price below 175000.

1.3 Interest

People, who want to find ward of their dreams, Real estate agencies.

2. Data acquisition and cleaning

2.1 Data sources

Main sources were provided by data.gov.uk as two datasets

"**National Statistics Postcode Lookup UK**“ (614529 rows × 16 columns),

“Land Registry 2019 Open Dataset” (1753031 rows × 4 columns).

Also,

Foursquare API provides 74 different types of venues

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into one table. All data fortunately was clean, so cleaning was not necessary.

2.3 Feature selection

After joining all data in one table we have 84 wards with their name, coordinates and Mean property price and 74 features as different types of venues.

	Ward Name	Price	Latitude	Longitude
3	Atherleigh	133690.242537	53.520424	-2.504809
5	Audenshaw	168528.720000	53.473169	-2.128313
7	Barton	151272.531481	53.478252	-2.357463
8	Besses	157128.371795	53.547343	-2.281292
11	Bucklow-St Martins	156611.996528	53.418071	-2.427692

	Ward	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	Bakery	Bar	Breakfast Spot	Bus Line	Bus Stop	...	Tailor Shop	Tanning Salon	Theater	Track	Trail	Train Station	Tram Station
0	Atherleigh	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.0	0.0	0.0	0.0
1	Audenshaw	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.0	0.0	0.0	0.0
2	Barton	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.0	0.0	0.0	0.0
3	Besses	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.0	0.0	0.0	0.0
4	Bucklow-St Martins	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.25	0.0	0.0	0.0	0.0	0.0	0.0

3. Exploratory Data Analysis

In this project we will direct our efforts on detecting differences between different possible wards in Manchester, UK

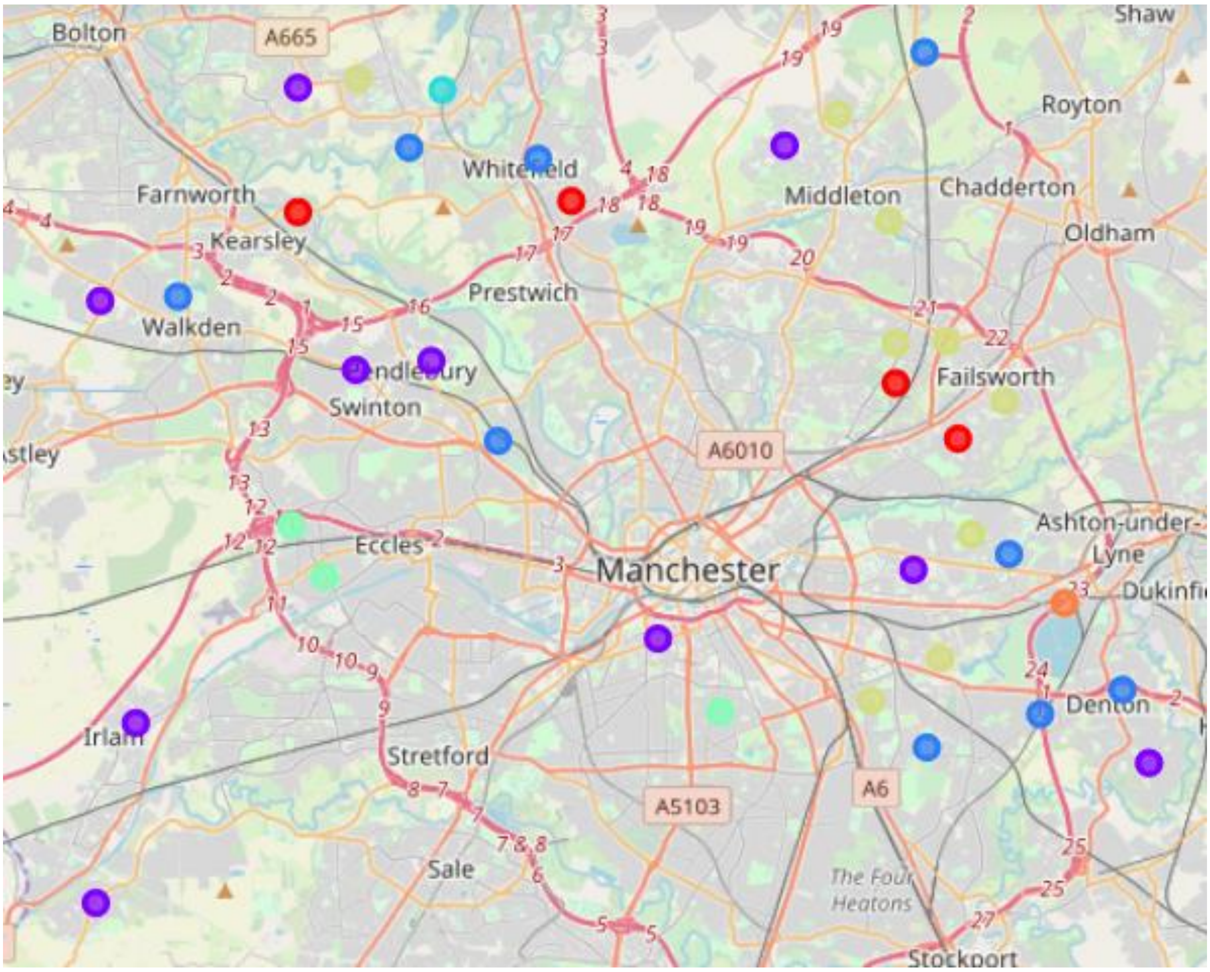
In first step we have collected the required data: location and type (category) of every venue within 1km from each ward center.

Two most “crowded” wards are Denton North East and Hulme with 12 venues per ward. The least ones are Woodhouse Park, West Middleton and Radcliffe East with 1 venue per ward.

4. Clustering

Second step in our analysis is clustering wards according to their venues - we have used SciKitLearn K-means method to divide wards on 7 different categories.

Possible wards on the map:



Clusters description:

Cluster 1

```
clusters_analysis.loc[clusters_analysis['Cluster'] == 0, clusters_analysis.columns[[1] + list(range(2, clusters_analysis.
```

	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
3	Shop & Service	Pub	Women's Store	Food & Drink Shop	Discount Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop
16	Pub	Sporting Goods Shop	Women's Store	Food & Drink Shop	Discount Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop
22	Pub	Italian Restaurant	Food Truck	Discount Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop
27	Soccer Stadium	Pub	Women's Store	Food & Drink Shop	Discount Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop
40	Pub	Deli / Bodega	Discount Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	Women's Store

As we can see, Cluster 1 is the most "**pubbed**" (sorry) cluster with a lot of venues for good friday evening

Cluster 2

```
clusters_analysis.loc[clusters_analysis['Cluster'] == 1, clusters_analysis.columns[[1] + list(range(2, clusters_analysis.shape[1
```

	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	Cluster
4	Grocery Store	IT Services	Tailor Shop	Gym / Fitness Center	Gym	Halal Restaurant	Gastropub	Garden Center	Furniture / Home Store	Convenience Store	1
8	Supermarket	Discount Store	Sandwich Place	Grocery Store	Food & Drink Shop	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	1
10	Grocery Store	Convenience Store	Indian Restaurant	Auto Garage	Furniture / Home Store	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	1
19	Grocery Store	Convenience Store	Tanning Salon	Pharmacy	Performing Arts Venue	Park	Café	Food Truck	Deli / Bodega	Garden Center	1
20	Grocery Store	Convenience Store	Stationery Store	Indian Restaurant	Fast Food Restaurant	Dive Bar	Farm	Fish & Chips Shop	Flower Shop	Food & Drink Shop	1
23	Asian Restaurant	Grocery Store	Indian Restaurant	Auto Garage	Furniture / Home Store	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	1
24	Pub	Athletics & Sports	Grocery Store	Food & Drink Shop	Discount Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	1
28	Indian Restaurant	Chinese Restaurant	Grocery Store	Fish & Chips Shop	Food Truck	Dive Bar	Farm	Fast Food Restaurant	Flower Shop	Food & Drink Shop	1
34	Park	Grocery Store	Bar	Flower Shop	Women's Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Food & Drink Shop	1
38	Grocery Store	Women's Store	Food Truck	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	Furniture / Home Store	1

The idea of the cluster 2 is small stores with everything. Let's call it "**stored**"

Cluster 3

```
ters_analysis.loc[clusters_analysis['Cluster'] == 2, clusters_analysis.columns[[1] + list(range(2, clusters_analysis.shape[1]))]]
```

	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	Cluster
5	Women's Store	Canal Lock	Garden Center	Athletics & Sports	Home Service	Halal Restaurant	Gym / Fitness Center	Gym	Grocery Store	Gastropub	2
9	Supermarket	Pub	Mobile Phone Shop	Post Office	Pharmacy	Clothing Store	Italian Restaurant	Outlet Store	Candy Store	Shopping Plaza	2
11	Pub	Hotel	Intersection	Fast Food Restaurant	Supermarket	Gym / Fitness Center	Gym	Grocery Store	Gastropub	Garden Center	2
12	Supermarket	Tram Station	Pharmacy	Italian Restaurant	Soccer Stadium	Flower Shop	Discount Store	Dive Bar	Farm	Fast Food Restaurant	2
21	Supermarket	Food & Drink Shop	Café	Bus Stop	Food Truck	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	2
31	Indian Restaurant	Gastropub	Fast Food Restaurant	Furniture / Home Store	Gym / Fitness Center	Gym	Grocery Store	Halal Restaurant	Garden Center	Deli / Bodega	2
32	Asian Restaurant	Farm	Auto Workshop	Furniture / Home Store	Dive Bar	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	Food Truck	2
33	Tram Station	Supermarket	Coffee Shop	Food Truck	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	2
36	Gym / Fitness Center	Bar	Sandwich Place	Park	Supermarket	Italian Restaurant	Pizza Place	Tram Station	Gym	Garden Center	2
37	Coffee Shop	Trail	Pharmacy	Shopping Mall	Fast Food Restaurant	Supermarket	Grocery Store	Gym / Fitness Center	Gastropub	Garden Center	2

That cluster is more for daily routine life-style. Let's call it "**routine**"

Cluster 4

```
ters_analysis.loc[clusters_analysis['Cluster'] == 3, clusters_analysis.columns[[1] + list(range(2, clusters_analysis.shape[1]))]]
```

	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	Cluster
29	Auto Garage	Women's Store	Furniture / Home Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	Food Truck	3

For car owners, for whom car is more then anything else. **Driver**

Cluster 5

```
clusters_analysis.loc[clusters_analysis['Cluster'] == 4, clusters_analysis.columns[[1] + list(range(2, clusters_analysis.shape[1
```

	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	Cluster
0	Park	Recreation Center	Gastropub	Food & Drink Shop	Discount Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	4
2	Chinese Restaurant	Park	Pub	Gym	Pizza Place	Sandwich Place	Fish & Chips Shop	Discount Store	Dive Bar	Farm	4
26	Halal Restaurant	Shop & Service	Pizza Place	Park	Flower Shop	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Food & Drink Shop	4
39	Track	Park	Miscellaneous Shop	Café	Women's Store	Food & Drink Shop	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	4

Recreation cluster

Cluster 6

```
clusters_analysis.loc[clusters_analysis['Cluster'] == 5, clusters_analysis.columns[[1] + list(range(2, clusters_analysis.shape[1
```

	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	Cluster
6	Hotel	Pizza Place	Malay Restaurant	Pub	Food & Drink Shop	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	5
7	Construction & Landscaping	Train Station	Bus Line	Breakfast Spot	Convenience Store	Gym	Grocery Store	Gastropub	Gym / Fitness Center	Halal Restaurant	5
13	Business Service	Chinese Restaurant	Playground	Convenience Store	Pub	Grocery Store	Gastropub	Garden Center	Furniture / Home Store	Food Truck	5
14	Chinese Restaurant	Indian Restaurant	Home Service	Construction & Landscaping	Gym / Fitness Center	Food & Drink Shop	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	5
15	Pub	Gym / Fitness Center	Construction & Landscaping	Bakery	Theater	Fast Food Restaurant	Chinese Restaurant	Food & Drink Shop	Dive Bar	Farm	5
17	Hotel	Gym / Fitness Center	Gym	Bakery	Sandwich Place	Market	Food & Drink Shop	Dive Bar	Farm	Fast Food Restaurant	5
18	Convenience Store	Locksmith	Home Service	Auto Garage	Furniture / Home Store	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	5
25	Café	Construction & Landscaping	Bar	Women's Store	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	Food Truck	5
30	Dive Bar	Photography Studio	Women's Store	Deli / Bodega	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	Food Truck	5
35	Bus Stop	Gym	Bar	Playground	Turkish Restaurant	Gym / Fitness Center	Halal Restaurant	Grocery Store	Gastropub	Garden Center	5

If you want to work near your home and be always in rush of city life, that is your choice. **Business**

Cluster 7

```
clusters_analysis.loc[clusters_analysis['Cluster'] == 6, clusters_analysis.columns[[1] + list(range(2, clusters_analysis.shape[1
```

	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	Cluster
1	Reservoir	Discount Store	Convenience Store	Dive Bar	Farm	Fast Food Restaurant	Fish & Chips Shop	Flower Shop	Food & Drink Shop	Food Truck	6

Cozy cluster for cozy life. **Suburb**

5. Conclusions

Purpose of this project was to divide Manchester wards by types of venues in order to aid customers in narrowing down the search for optimal ward for them. By calculating mean price from open UK Government datasets we define start set of wards for clustering. Then from Foursquare data we have identified content of each ward. Clustering of those locations was then performed in order to create major zones of interest and put it on the map to be used as starting points for final exploration by clients.

Result of all this is 7 zones containing different possible wards for our customer. This, of course, does not imply that those zones are actually optimal locations! Purpose of this analysis was to only provide info on areas, but not prove impossibility of buying property in others wards - it is entirely possible that there will be a very good offer in any of those areas, reasons which would make them suitable. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only good price but also other factors taken into account and all other relevant conditions met.

Final decision on optimal ward will be made by customer based on specific characteristics of ward, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.