Clustering and Segmentation of wards in Cambridge, UK

INTRODUCTION

Description and discussion of the background

Cambridge in England is a very special city. On one hand, it is very small. Located northeast of London, the capital, it has a population of just about 125,000 [1]. On the other hand, it is a quite international place. With the University of Cambridge and many world famous companies, e.g., Microsoft, Apple, Amazon, ARM, located in the city, Cambridge is attracting talents from all over the world in recent years. Therefore, some people even call it Silicon Fen by analogy with Silicon Valley in California. The arrival of a great many of knowledge workers has changed the demography and different functionalities of Cambridge.

As a resident of Cambridge, UK, I have decided to study the city using data analysis techniques, trying to look more closely into the different areas of the city I live, out of curiosity. In this study, I am trying to first analyse the different wards, find the similarities between them. Then, I hope this study can also shed some light on the different functionalities of each ward.

Potential audiences

The potential audiences of this study can be the people who are trying to relocate to or settle in Cambridge.

DATA DESCRIPTION

Data will be acquired from the following sources to conduct the study.

1. The divisions of Cambridge wards are obtained from [2]
2. Foursquare API is employed to get the common venues at different wards of Cambridge to clustering them based on similarities
3. The geocoder from geopy is used to convert addresses into latitude and longitude values

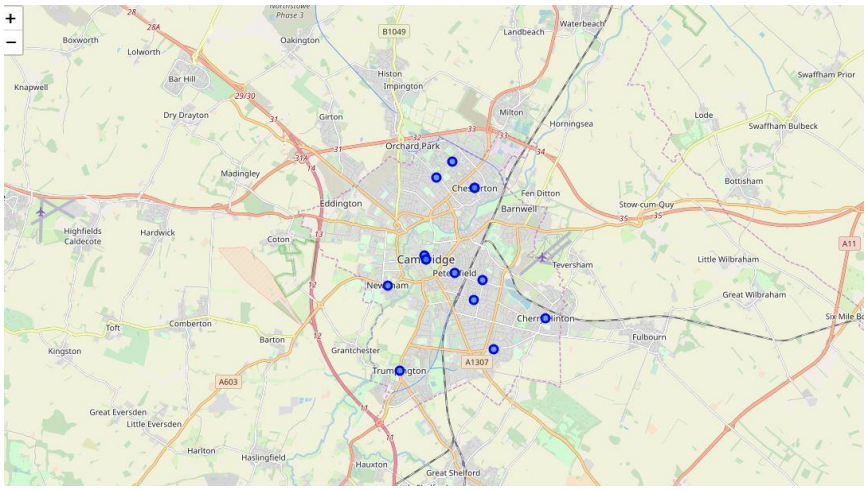
METHODOLOGY

Exploratory data analysis & Inferential statistical testing

The data were loaded from https://mapit.mysociety.org/area/2263/children to get the ward names of Cambridge. According to the names of the wards, their location is obtained from geocoder

	Ward	Latitude	Longitude
0	Abbey	52.204406	0.122944
1	Arbury	52.225036	0.128007
2	Castle	52.203482	0.123582
3	Cherry Hinton	52.187843	0.175241
4	Coleridge	52.192661	0.144213
5	East Chesterton	52.222288	0.144592
6	King's Hedges	52.229237	0.135074
7	Market	52.203482	0.123582
8	Newnham	52.196542	0.107044
9	Petersfield	52.199810	0.135933
10	Queen Edith's	52.179638	0.152737
11	Romsey	52.197962	0.148062
12	Trumpington	52.173992	0.112116
13	West Chesterton	52.222288	0.144592

A map of Cambridge is generated using Folium package with wards superimposed on it



Foursquare API is utilized to explore the wards and segment them. The limit was set as 100 and the radius was set as 500 m considering the dimension of the studied area. In total, 398 venues were extracted. Below are the heading of the different venues extracted and the count of venues of different wards.

	Ward	Ward Latitude	Ward Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbey	52.204406	0.122944	Arts Picturehouse	52.202927	0.123748	Indie Movie Theater
1	Abbey	52.204406	0.122944	Savino's	52.204327	0.123427	Café
2	Abbey	52.204406	0.122944	Hilton Cambridge City Centre	52.203379	0.121676	Hotel
3	Abbey	52.204406	0.122944	Pint Shop	52.204269	0.119238	Bar
4	Abbey	52.204406	0.122944	John Lewis & Partners	52.203671	0.122583	Department Store

Ward	Ward Latitude	Ward Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Abbey	100	100	100	100	100	100
Arbury	4	4	4	4	4	4
Castle	100	100	100	100	100	100
Cherry Hinton	6	6	6	6	6	6
Coleridge	8	8	8	8	8	8
East Chesterton	4	4	4	4	4	4
King's Hedges	7	7	7	7	7	7
Market	100	100	100	100	100	100
Newnham	4	4	4	4	4	4
Petersfield	35	35	35	35	35	35
Queen Edith's	4	4	4	4	4	4
Romsey	16	16	16	16	16	16
Trumpington	6	6	6	6	6	6
West Chesterton	4	4	4	4	4	4

Here shows the top 10 most popular venues of the wards of Cambridge.

	Ward	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	Abbey	Pub	Coffee Shop	Sandwich Place	Café	Italian Restaurant	Clothing Store	Sushi Restaurant	Burger Joint	Science Museum	Restaurant
1	Arbury	Bed & Breakfast	Chinese Restaurant	Bus Station	Wings Joint	Gastropub	Department Store	Eastern European Restaurant	Electronics Store	English Restaurant	Fast Food Restaurant
2	Castle	Pub	Coffee Shop	Italian Restaurant	Sandwich Place	Café	Burger Joint	Sushi Restaurant	Science Museum	Pizza Place	Bar
3	Cherry Hinton	Pharmacy	Plaza	Indian Restaurant	Pub	Gastropub	Restaurant	Electronics Store	Coffee Shop	Cosmetics Shop	Deli / Bodega
4	Coleridge	Coffee Shop	Playground	Gym	Bar	Bookstore	Park	Platform	English Restaurant	Cosmetics Shop	Deli / Bodega

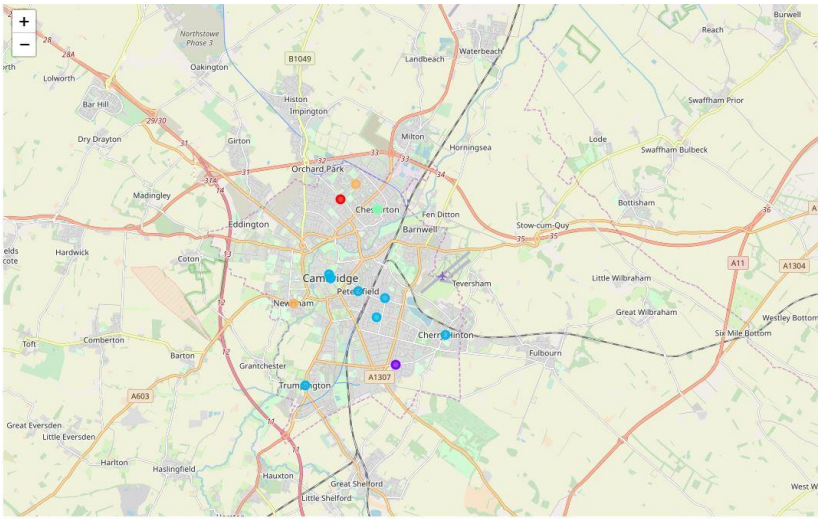
Machine learning techniques and why?

The k-means clustering was chosen to cluster the wards for the reason of simplicity and efficiency. Below shows the results of the clustering.

	Ward	Latitude	Longitude	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	Abbey	52.204406	0.122944	2	Pub	Coffee Shop	Sandwich Place	Café	Italian Restaurant	Clothing Store	Sushi Restaurant	Burger Joint	Science Museum	Restaurant
1	Arbury	52.225036	0.128007	0	Bed & Breakfast	Chinese Restaurant	Bus Station	Wings Joint	Gastropub	Department Store	Eastern European Restaurant	Electronics Store	English Restaurant	Fast Food Restaurant
2	Castle	52.203482	0.123582	2	Pub	Coffee Shop	Italian Restaurant	Sandwich Place	Café	Burger Joint	Sushi Restaurant	Science Museum	Pizza Place	Bar
3	Cherry Hinton	52.187843	0.175241	2	Pharmacy	Plaza	Indian Restaurant	Pub	Gastropub	Restaurant	Electronics Store	Coffee Shop	Cosmetics Shop	Deli / Bodega
4	Coleridge	52.192661	0.144213	2	Coffee Shop	Playground	Gym	Bar	Bookstore	Park	Platform	English Restaurant	Cosmetics Shop	Deli / Bodega

RESULTS

The clustering results were displayed on the maps as shown below with different clusters marked with different colours. We can see clearly the similarities/dis-similarities of different wards.



DISCUSSION

From this study, it can be seen that the wards of Cambridge can be clustering with different functionalities. However, this study is by no means complete. Instead, it is intended to serve as a preliminary study of the Cambridge area and many things can be involved in the future work, some of which are summarized as the following:

1. Cambridge is a small-scale city, and the radius was set to 500m when retrieved the venues from different wards. This value might too big for a city with the size as Cambridge. This might cause overlapping for different wards and the value should be better-tuned.
2. The best number of clusters should be decided from elbow method while in this study, it was set to 5 for the consideration of total 13 wards.
3. Other clustering methods can also be explored to study the wards.
4. Choropleth maps might be adopted together with supplementary data such as population density, average income, housing price index, etc to give richer information of each ward.

CONCLUSION

This study briefly clustered and segmented the wards in Cambridge. It is expected to give people who are coming to or trying to settling in Cambridge a taste of different wards. In the future, more data as mentioned above can be added to it to enrich its value.

REFERENCES

[1] https://en.wikipedia.org/wiki/Cambridge

[2] https://mapit.mysociety.org/area/2263/children