NEW CINEMA LOCATION

BATTLE OF THE NEIGHBOURHOODS CAPSTONE DATA SCIENCE PROJECT

August 2021

Catherine Brown

INTRODUCTION

Scenario:

I have been tasked by a client with finding a suitable location in the Nottinghamshire area for a new cinema to be built or for an existing building to be converted. The Nottinghamshire area already has many cinema venues, so the location has to be carefully selected to ensure that it will be a successful business venture. My client already has a chain of independent cinemas in the North of England and is looking for a new location in the Midlands, with Nottinghamshire being the preferred location. Other individuals or businesses interested in opening a new Cinema in the Nottinghamshire area may also be find this project of use.

The scope of this project within the wider business analysis:

Opening a new cinema requires significant capital investment. Deciding on a location, whilst an important part of the strategic decision, is only one of many factors to be explored. The local demographics of the area is an important factor to consider. Other analysis such as SWOT and 5 Market forces should also be a starting point for decision making. The barriers to entry for building a new cinema are high due to the upfront capital investment and the understanding of the business sector. Streaming providers such as Netflix and Amazon Prime should be considered as a threat to the viability of existing and new entrants to the cinema market. The high supply cost of showing new release Hollywood movies should be considered when deciding what type of cinema to build. The projected revenues and costs should then be computed to determine the actual return on investment to ascertain whether the project is potentially viable. Funding would then need to be sourced. The scope of this project is to determine only which NG (Nottinghamshire) postcode is the most suitable location for a new cinema to be built.

Types of cinemas:

The existing cinemas in Nottinghamshire are Multiplex cinemas such as Odeon and CineWorld, along with smaller chains such as the Savoy cinemas. There is also a popular independent cinema, Broadway, in the centre of Nottinghamshire that specialises in classics and alternative films. The type of cinema to be built would depend on the results of the business analysis. Cinemas such as the Savoy or Broadway are much smaller than the Multiplex cinemas and will perhaps show only a couple of the main release films at one time, offering older films that are cheaper from a supply perspective on the other screens. They offer a more intimate experience than the multiplex cinemas, with Broadway also having a bar and small restaurant. Creating more of an experience is likely to be ever more important when competing for customers with Netflix and Amazon Prime.

DATA USED

1. NG Postal code data:

The Nottingham postal code data will be extracted from the webpage: https://en.wikipedia.org/wiki/NG_postcode_area', using the function dfs=pd.read_html(url) to read the data into a pandas dataframe (df). This data will be used to select from when determining a suitable location for a new cinema. The extracted data will need to have coordinates found, so the postcode district and post town are combined into a full address column:

	Postcode district	Post town	Coverage	Local authority area(s)	Full address
0	NG1	NOTTINGHAM	Nottingham city centre	Nottingham	NG1,NOTTINGHAM
1	NG2	NOTTINGHAM	Nottingham city centre, Colwick Park, Sneinton...	Nottingham, Rushcliffe	NG2,NOTTINGHAM
2	NG3	NOTTINGHAM	Carlton, Sneinton, St Ann's, Mapperley	Gedling, Nottingham	NG3,NOTTINGHAM
3	NG4	NOTTINGHAM	Gedling Village, Netherfield, Carlton, Colwick	Gedling	NG4,NOTTINGHAM
4	NG5	NOTTINGHAM	Sherwood, Arnold, Bestwood, Carrington, Top Va...	Nottingham, Gedling	NG5,NOTTINGHAM
5	NG6	NOTTINGHAM	Bestwood Village, Bulwell, Old Basford	Nottingham, Gedling	NG6,NOTTINGHAM

2. Converting the NG Postal code data into coordinates for mapping:

Using Geopy, the NG Postal code data will be used to find the approximate centre latitude and longitude coordinates of the postal codes. These coordinates will be added to the dataframe.

	Postcode district	Post town	Coverage	Local authority area(s)	Full address	Longitude	Latitude
0	NG1	NOTTINGHAM	Nottingham city centre	Nottingham	NG1,NOTTINGHAM	-1.141801	52.955513
1	NG2	NOTTINGHAM	Nottingham city centre, Colwick Park, Sneinton...	Nottingham, Rushcliffe	NG2,NOTTINGHAM	-1.165191	52.941864
2	NG3	NOTTINGHAM	Carlton, Sneinton, St Ann's, Mapperley	Gedling, Nottingham	NG3,NOTTINGHAM	-1.108363	52.959671
3	NG4	NOTTINGHAM	Gedling Village, Netherfield, Carlton, Colwick	Gedling	NG4,NOTTINGHAM	-1.101884	52.966431
4	NG5	NOTTINGHAM	Sherwood, Arnold, Bestwood, Carrington, Top Va...	Nottingham, Gedling	NG5,NOTTINGHAM	-1.158969	52.989743
5	NG6	NOTTINGHAM	Bestwood Village, Bulwell, Old Basford	Nottingham, Gedling	NG6,NOTTINGHAM	-1.212583	53.005388

3. Mapping the NG Postal codes

The postal codes will then be mapped using Folium to provide a visual of the distribution of the postal codes.

4. Obtaining existing Cinema venues around the Nottinghamshire area

Foursquare will be used to obtain existing cinema venues around the Nottinghamshire area. This information will be stored in a dataframe, the data cleaned and the resulting location venues will be mapped onto the same map as the NG Postal codes. This will provide an initial visual of postal areas that do not already have a nearby cinema venue. It is better to identify a location that does not already have a cinema so that there is less competition for customers.

5. Clustering the NG Postal codes using popular venues

Foursquare will be used again to find the nearby 100 venues for each postal area. The data is then grouped by postal code area and sorted into the top 10 venues within each postal code area. A clustering method called KMeans will be applied to this data in order to cluster the postal code areas based on top local venues. To find the optimal k for the algorithm, the Elbow method is used. The information produced from this clustering exercise will provide an idea as to which postal code areas are similar types of Neighbourhoods based on nearby venues.

6. Mapping the clusters on the same map and determining the best location

With each area assigned to a particular cluster and imposed on the map, this will then provide an additional visual of which clusters tend to have cinemas and help identify if there are any postal codes that fall in the same cluster but that do not have a nearby cinema. Such a location could be a potentially good location for a new cinema.

7. Limitations of the scope of this project

It is important to note again that proximity to existing cinemas and similarity of neighbourhoods based on popular venues are only a very small part of the overall picture when deciding where to build a new cinema in Nottinghamshire and that the results of this project should be used in conjunction with a vast array of other supporting material as discussed in the wider business analysis section.

METHODOLOGY

The following methodology was followed in this project:

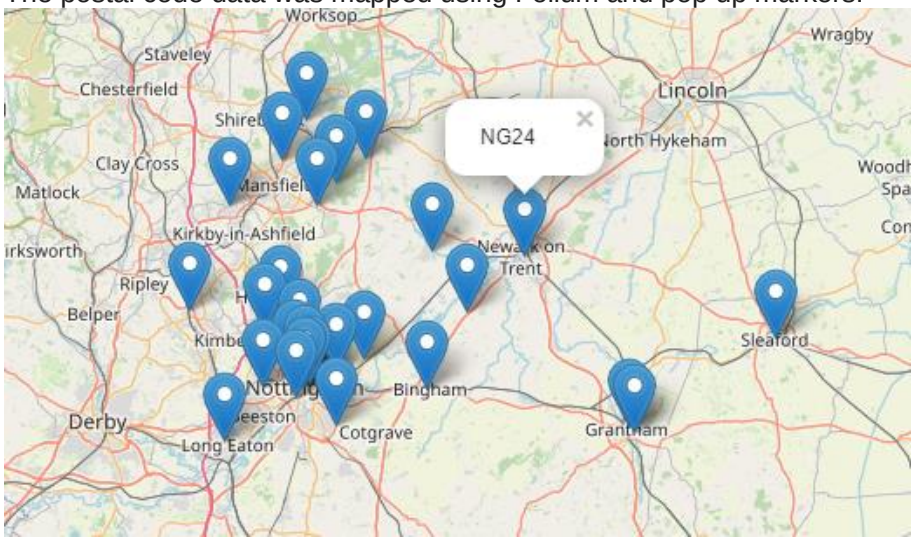
1. The Nottingham postal code data was extracted from the webpage: https://en.wikipedia.org/wiki/NG_postcode_area', using the function dfs=pd.read_html(url) to read the data into a pandas dataframe. In order to use this data to find the coordinates of the postcode area, it was necessary to combine the potcode district and post town into a full address. This full address column was then added to the dataframe:

	Postcode district	Post town	Coverage	Local authority area(s)	Full address
0	NG1	NOTTINGHAM	Nottingham city centre	Nottingham	NG1,NOTTINGHAM
1	NG2	NOTTINGHAM	Nottingham city centre, Colwick Park, Sneinton...	Nottingham, Rushcliffe	NG2,NOTTINGHAM
2	NG3	NOTTINGHAM	Carlton, Sneinton, St Ann's, Mapperley	Gedling, Nottingham	NG3,NOTTINGHAM
3	NG4	NOTTINGHAM	Gedling Village, Netherfield, Carlton, Colwick	Gedling	NG4,NOTTINGHAM
4	NG5	NOTTINGHAM	Sherwood, Arnold, Bestwood, Carrington, Top Va...	Nottingham, Gedling	NG5,NOTTINGHAM
5	NG6	NOTTINGHAM	Bestwood Village, Bulwell, Old Basford	Nottingham, Gedling	NG6,NOTTINGHAM

2. Geopy was used to find the latitude and longitude coordinates of the postcode districts. Two of the postcode districts did not return any information, but Geopy was used to find the latitude and longitude values for these areas (on re-running the code, Geopy did not work for this), and the NaN data was replaced with the correct coordinates:

	Postcode district	Post town	Coverage	Local authority area(s)	Full address	Longitude	Latitude
0	NG1	NOTTINGHAM	Nottingham city centre	Nottingham	NG1,NOTTINGHAM	-1.141801	52.955513
1	NG2	NOTTINGHAM	Nottingham city centre, Colwick Park, Sneinton...	Nottingham, Rushcliffe	NG2,NOTTINGHAM	-1.165191	52.941864
2	NG3	NOTTINGHAM	Carlton, Sneinton, St Ann's, Mapperley	Gedling, Nottingham	NG3,NOTTINGHAM	-1.108363	52.959671
3	NG4	NOTTINGHAM	Gedling Village, Netherfield, Carlton, Colwick	Gedling	NG4,NOTTINGHAM	-1.101884	52.966431
4	NG5	NOTTINGHAM	Sherwood, Arnold, Bestwood, Carrington, Top Va...	Nottingham, Gedling	NG5,NOTTINGHAM	-1.158969	52.989743
5	NG6	NOTTINGHAM	Bestwood Village, Bulwell, Old Basford	Nottingham, Gedling	NG6,NOTTINGHAM	-1.212583	53.005388

3. The postal code data was mapped using Folium and pop up markers:



4. Foursquare was used to find existing cinema venues within a 50,000m radius of the central coordinates of the Nottinghamshire location on Geopy:

```
# Define an instance of the Geocoder
address = 'Nottinghamshire, UK'

geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print(latitude, longitude)
```

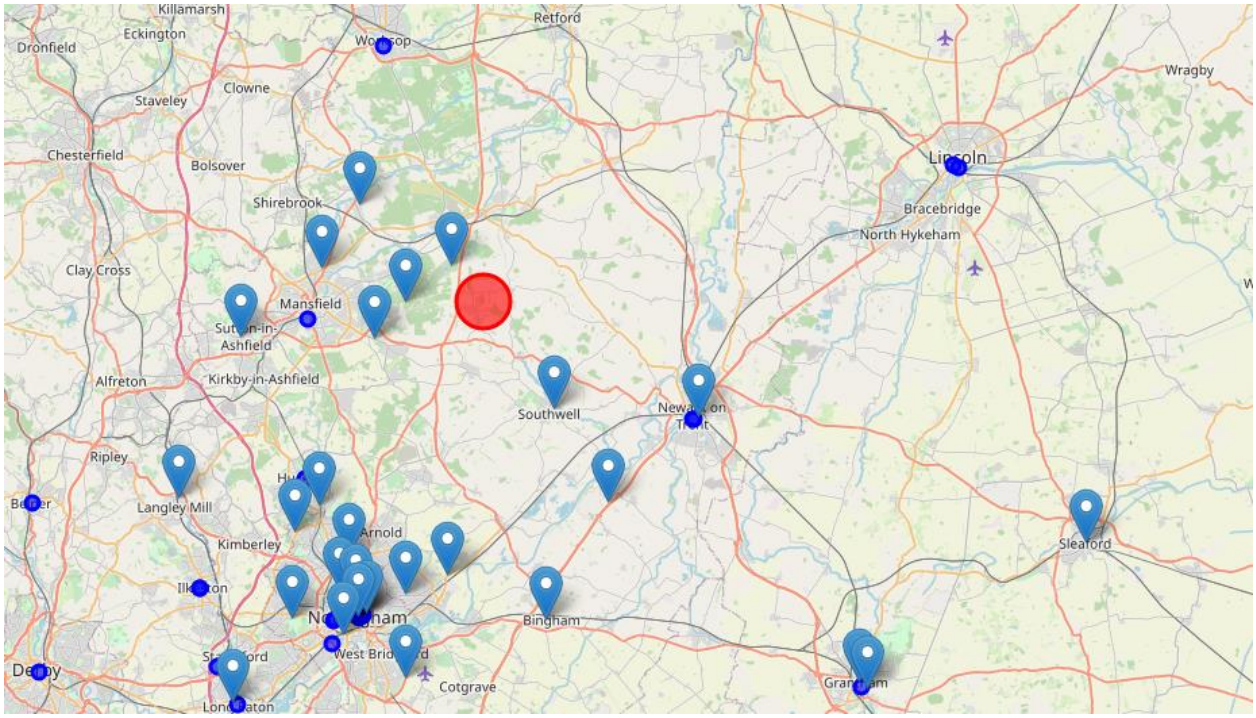
```
53.1459288 -1.0214971168122484
```

```
# Search for specific venue category
search_query = 'Cinema'
radius = 50000
```

5. The results were read in a json file and stored in a dataframe calles df_fil. The data required cleaning – certain venues found using the search query cinema were not actually cinema venues when the category was checked, but some were coffee shops or bus stops. These were removed from the dataframe to leave only real cinema venues:

	name	categories	address	lat	lng
0	Savoy Cinema	Multiplex	Bridge St.	53.301454	-1.123811
1	Showcase Cinema	Multiplex	Redfield Way	52.936903	-1.175230
2	Broadway Cinema	Movie Theater	14-18 Broad St	52.954308	-1.143878
3	Showcase Cinema de Lux	Multiplex	inTu Shopping Centre	52.919420	-1.472932
4	The Arc Cinema - Hucknall	Movie Theater	8 High Street	53.038040	-1.202630
6	Odeon	Multiplex	Park Lane	53.134927	-1.200858

6. The existing cinemas were then mapped onto the existing map shown in Step 3, with the large red circle marking the centre of the Nottinghamshire area:



7. An exploration of the different Nottinghamshire postal codes was then made using Foursquare. The following function was created to extract nearby venues:

```
# Explore the Nottinghamshire postcode Neighbourhoods

# Create a function to extract nearby venues
def getNearbyVenues(names, latitudes, longitudes, radius=4000, LIMIT=100):
```

This function was applied to search the original postal code dataframe from stages 1&2 (df) and a new dataframe was formed (Notts_venues). The Notts_venues dataframe was then grouped by Postal code and there were found to be 150 unique categories of venue.

8. One hot coding was used to add the unique categories of venue by instance and postal code to a new dataframe (Notts_onehot). This dataframe was then grouped by Postal code and the mean frequency of each nearby venue was found (Notts_grouped).
9. The Notts_grouped dataframe was sorted to show the top 10 venues by postal code (Notts_venues_sorted):

	Postal code	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	NG1	Pub	Bar	Indian Restaurant	Coffee Shop	Café	Italian Restaurant	Bookstore	Plaza	Brewery	Gym / Fitness Center
1	NG10	Pub	Grocery Store	Coffee Shop	Supermarket	Gym / Fitness Center	Hotel	Restaurant	Discount Store	Bar	Gastropub
2	NG11	Pub	Bar	Indian Restaurant	Coffee Shop	Café	Bookstore	Cocktail Bar	Italian Restaurant	Burger Joint	Movie Theater
3	NG12	Pub	Bar	Indian Restaurant	Coffee Shop	Café	Bookstore	Cocktail Bar	Italian Restaurant	Burger Joint	Movie Theater
4	NG13	Pub	Italian Restaurant	Chinese Restaurant	Gym	Grocery Store	Café	Coffee Shop	Train Station	Garden Center	Supermarket
5	NG14	Grocery Store	Pub	Supermarket	Train Station	Coffee Shop	American Restaurant	English Restaurant	Clothing Store	Pet Store	Park

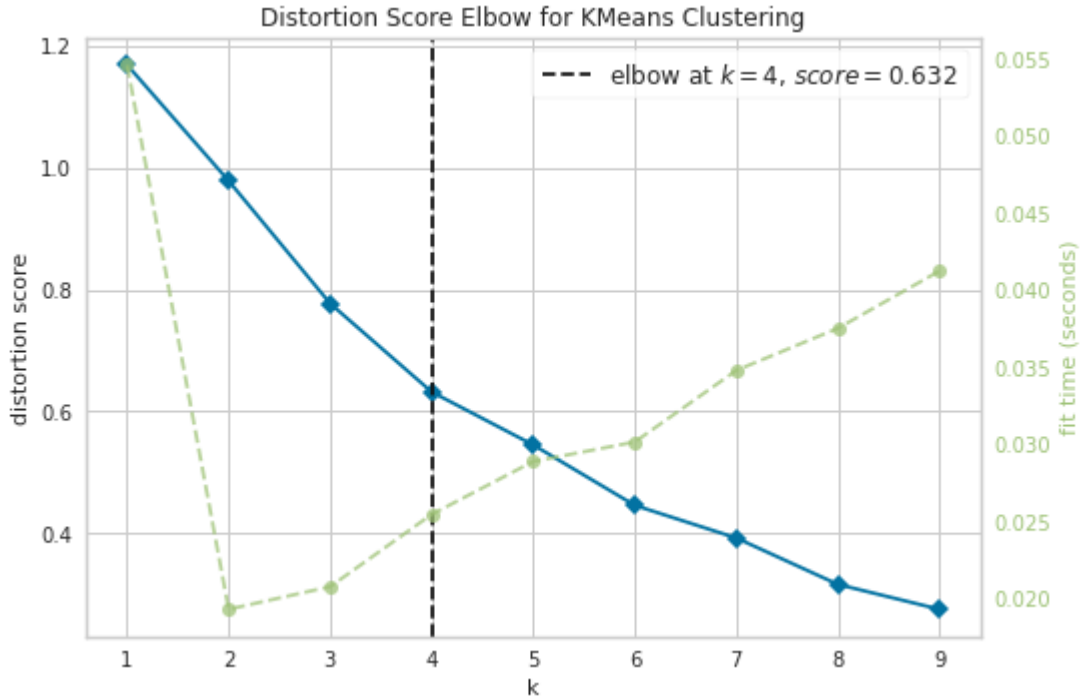
10. A Notts_clustering dataframe was created based on the Notts_grouped data above. The unsupervised machine learning algorithm K-Means was applied to the data to group the postal codes into clusters according to their nearby venues:

```
from sklearn.cluster import KMeans
! pip install yellowbrick
from yellowbrick.cluster import KElbowVisualizer

Notts_clustering = Notts_grouped.drop('Postal code',1)
```

K-Means forms the clusters by minimizing the sum of the distance of points from their respective cluster centroids.

11. The Elbow method of finding the most appropriate value for k was used:



12. A dataframe Notts_merged was created to add the cluster labels to the Notts_venues_sorted dataframe:

	Postcode district	Post town	Coverage	Local authority area(s)	Full address	Longitude	Latitude	Cluster	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	NG1	NOTTINGHAM	Nottingham city centre	Nottingham	NG1,NOTTINGHAM	-1.141801	52.955513	1	Pub	Bar	Indian Restaurant	Coffee Shop	Café	Italian Restaurant	Bookstore	Plaza	Brewery	Gym / Fitness Center
1	NG2	NOTTINGHAM	Nottingham city centre, Colwick Park, Sneinton...	Nottingham, Rushcliffe	NG2,NOTTINGHAM	-1.165191	52.941864	1	Bar	Pub	Indian Restaurant	Coffee Shop	Café	Bookstore	Italian Restaurant	French Restaurant	Hotel	Cocktail Bar
2	NG3	NOTTINGHAM	Carlton, Sneinton, St Ann's, Mapperley	Gedling, Nottingham	NG3,NOTTINGHAM	-1.151815	52.965092	1	Pub	Bar	Indian Restaurant	Café	Coffee Shop	Italian Restaurant	Bookstore	Movie Theater	Seafood Restaurant	Park
3	NG4	NOTTINGHAM	Gedling Village, Netherfield, Carlton, Colwick	Gedling	NG4,NOTTINGHAM	-1.101884	52.966431	1	Pub	Bar	Indian Restaurant	Coffee Shop	Park	Café	Bookstore	Italian Restaurant	Burger Joint	Cocktail Bar
4	NG5	NOTTINGHAM	Sherwood, Arnold, Bestwood, Carrington, Top Va...	Nottingham, Gedling	NG5,NOTTINGHAM	-1.158969	52.989743	1	Pub	Park	Bar	Cocktail Bar	Café	Indian Restaurant	Grocery Store	Gym / Fitness Center	Concert Hall	Seafood Restaurant
5	NG6	NOTTINGHAM	Bestwood Village, Bulwell, Old Basford	Nottingham, Gedling	NG6,NOTTINGHAM	-1.214203	53.005101	2	Pub	Supermarket	Grocery Store	Hotel	Fast Food Restaurant	Warehouse Store	Discount Store	Tram Station	Restaurant	Furniture / Home Store

13. The following code was added in order to change the Cluster labels to integers from floats so that they could be plotted on the map:

```
Notts_merged['Cluster']= Notts_merged['Cluster'].astype(int)
```

14. The postal codes were clustered into 4 clusters and the cluster results plotted onto the map using Folium to get a good overall visual image of the best location(s) to build a new cinema.

RESULTS

The Postal codes were clustered into 4 clusters, this being the most appropriate value of k for the kmeans clustering, with the top 5 most common venues defining their name and shown as follows:

Red cluster – Sports shop area

```
# Red cluster - Sports shop area
Notts_merged.loc[Notts_merged['Cluster'] == 0, Notts_merged.columns[[0] + list(range(5, Notts_merged.shape[1]))]]
```

1]:

	Postcode district	Longitude	Latitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
20	NG20	-1.148304	53.204385	0	Sporting Goods Shop	Pub	Bar	Insurance Office	Clothing Store

Purple cluster – Pubs and Grocery stores areas

```
# Purple cluster - Pubs and Grocery Stores area
Notts_merged.loc[Notts_merged['Cluster'] == 1, Notts_merged.columns[[0] + list(range(5, Notts_merged.shape[1]))]]
```

5]:

	Postcode district	Longitude	Latitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	NG6	-1.214203	53.005101	1	Pub	Supermarket	Grocery Store	Fast Food Restaurant	Hotel
9	NG10	-1.277511	52.900998	1	Pub	Grocery Store	Supermarket	Coffee Shop	Gym / Fitness Center
13	NG14	-1.058502	52.978257	1	Grocery Store	Pub	Supermarket	Coffee Shop	Train Station
14	NG15	-1.188919	53.021228	1	Pub	Grocery Store	Supermarket	Hotel	Fast Food Restaurant
15	NG16	-1.331449	53.025449	1	Grocery Store	Pub	Fast Food Restaurant	Supermarket	Park
17	NG17	-1.269579	53.123772	1	Grocery Store	Pub	Clothing Store	Supermarket	Gastropub
18	NG18	-1.101884	53.144768	1	Grocery Store	Burger Joint	Gastropub	Resort	Coffee Shop
21	NG21	-1.132632	53.123200	1	Pub	Grocery Store	Park	Bar	Gastropub
24	NG24	-0.803577	53.075369	1	Pub	Grocery Store	Coffee Shop	Supermarket	Clothing Store
26	NG31	-0.642605	52.912789	1	Grocery Store	Clothing Store	Pub	Pizza Place	Discount Store
27	NG32	-0.632345	52.908662	1	Hotel	Grocery Store	Pub	Clothing Store	Coffee Shop
29	NG34	-0.409453	52.998419	1	Supermarket	Fast Food Restaurant	Gastropub	Hardware Store	Hotel

Turquoise clusters – Pubs, bars and restaurant areas

```
# Turquoise cluster - Pubs, bars and restaurant area
Notts_merged.loc[Notts_merged['Cluster'] == 2, Notts_merged.columns[[0] + list(range(5, Notts_merged.shape[1]))]]
```

]:

	Postcode district	Longitude	Latitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	NG1	-1.141801	52.955513	2	Pub	Bar	Coffee Shop	Indian Restaurant	Café
1	NG2	-1.165191	52.941864	2	Bar	Pub	Indian Restaurant	Coffee Shop	Café
2	NG3	-1.151815	52.965092	2	Pub	Bar	Indian Restaurant	Café	Coffee Shop
3	NG4	-1.101884	52.966431	2	Pub	Bar	Indian Restaurant	Coffee Shop	Café
4	NG5	-1.158969	52.989743	2	Pub	Park	Bar	Grocery Store	Indian Restaurant
6	NG7	-1.168569	52.969543	2	Pub	Bar	Indian Restaurant	Coffee Shop	Café
7	NG8	-1.216109	52.952063	2	Pub	Coffee Shop	Gym / Fitness Center	Sandwich Place	Grocery Store
8	NG9	-1.149646	52.953419	2	Pub	Bar	Coffee Shop	Indian Restaurant	Café
10	NG11	-1.149646	52.953419	2	Pub	Bar	Coffee Shop	Indian Restaurant	Café
11	NG12	-1.149646	52.953419	2	Pub	Bar	Coffee Shop	Indian Restaurant	Café
12	NG13	-0.957896	52.950935	2	Pub	Italian Restaurant	Gym	Grocery Store	Train Station
19	NG19	-1.186113	53.165889	2	Pub	Bar	Discount Store	Clothing Store	Pharmacy
23	NG23	-0.887753	53.018881	2	Lake	Café	Pub	Gastropub	Spa
25	NG25	-0.950388	53.080154	2	Pub	Train Station	Ice Cream Shop	Tea Room	Racecourse
28	NG33	-1.101884	52.915371	2	Pub	Grocery Store	Restaurant	Park	Golf Course

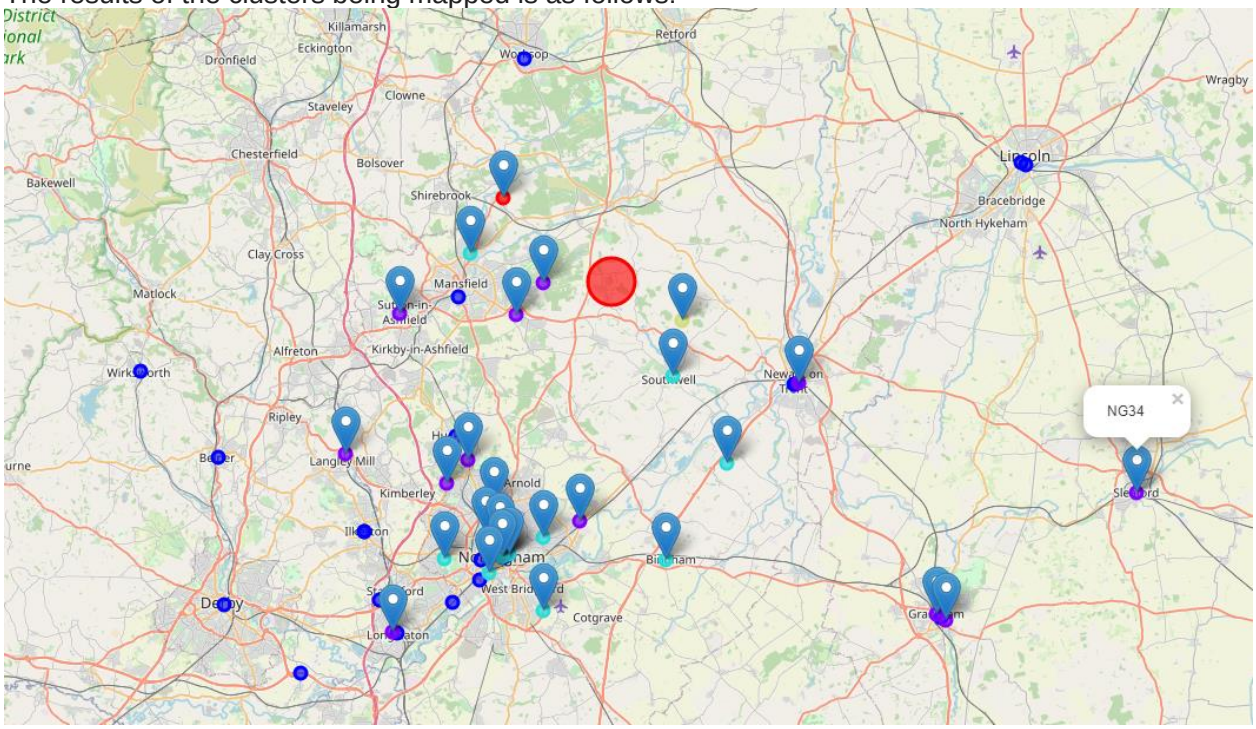
Green cluster – Brewery and Ice cream shop area

```
# Green cluster - Brewery and Ice cream shop area
Notts_merged.loc[Notts_merged['Cluster'] == 3, Notts_merged.columns[[0] + list(range(5, Notts_merged.shape[1]))]]
```

']:

	Postcode district	Longitude	Latitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
22	NG22	-0.939795	53.118346	3	Brewery	Ice Cream Shop	Gastropub	Farmers Market	Zoo

The results of the clusters being mapped is as follows:



DISCUSSION

The map above shows: the Pop up markers containing the Postal codes for the NG postcodes with the corresponding cluster area (out of a possible 4 clusters) shown by the coloured circle at the point of the marker. The map also shows the existing cinema locations marked in dark blue and the centre of the Nottinghamshire area is marked with the large red circle.

This analysis has done the following:

- Taken the Nottingham postal code data and found cinema locations within a 50,000m radius of the Nottinghamshire county central latitude and longitude coordinates.
- Both the postal codes and the nearby existing cinema venues have been plotted on a map of the region.
- The postcode areas have been explored to examine nearby venues and the top 10 venues in a 4,000m radius from the postcode coordinates have been found.
- The postcode areas have been compared and clustered using the kmeans algorithm method.
- The resulting clusters have been plotted as an overlay onto the cinema and postcode map.
- The resulting visual provides a good indication of what type of clusters already have a cinema and therefore which areas may be a good location for a new cinema to be built.

The results above are limited in that they have used only nearby cinemas and clustering of similar Neighbourhoods to determine a suitable location. As can be seen from the results of the clustering, the Neighbourhood clusters in Nottinghamshire are fairly similar - they all contain a lot of pubs, although I have tried to name them according to their most popular venues. There is good coverage with regards to existing cinema venues, with the largest concentration of venues being in the centre of Nottingham as one may expect. Grantham (NG31 and NG32), Long Eaton (NG10) and Newark (NG24) are all found in the purple clusters which has pubs and grocery shops as being the most popular venues. Sleaford (NG34) falls into the purple cluster but does not have a nearby cinema.

CONCLUSION

I recommend that the area of Sleaford is explored further as a potential location for a new cinema to be built or for an existing building to be converted since it has similar other venues in the area to other towns that already have cinemas. It also does not have a cinema in a nearby location.

Since the map shows that Sleaford is not a major city, it is likely that the size of the cinema should be relatively small and further revenue may be gained from providing a food and beverage option, since the most popular venues in the area are a supermarket, a fast-food restaurant and a gastropub.