CAPSTONE PROJECT: Investment/Touristic Neighborhoods in Carlow - Ireland

Introduction

Ireland may have 40 shades of green, but Carlow can cater for 40 shades of green fingers. The Carlow Garden Trail is one of the island's best, while the wild, natural-style gardens at Altamont House are home to an explosion of snowdrops every February.

Dive into the heart of Ireland's Ancient East and wander the ancient yew tree walk at Huntington Castle, where the ghosts of monks have been known to mosey. Then there are the Delta Sensory Gardens on the outskirts of Carlow town, a therapeutic and multi-sensory oasis for visitors of all abilities.

The county, over the years, has been a huge attraction to tourists. The influx of people into the county also brings in money into local businesses like restaurants, cinemas, pubs, fast foods etc.

You can make outings to noble estates on the outskirts of Carlow, like Duckett's Grove, which has the dreamlike ruins of a Gothic Revival house that burnt down in 1933. Although not geared for tourism, Carlow has real charm, especially on the banks of the River Barrow and the old center at Montgomery and Burin Streets, which have rows of Georgian houses.

However, these varying businesses are found in different neighborhoods. Some neighborhoods have more of a business than others. The distribution of different businesses in the respective neighborhoods could be of great interest to investors who are new to the city. The problem that this project aims to solve is to facilitate the investors' choice of neighborhood and the type of business that could do well in that neighborhood and also to point out which will be the best neighborhoods for tourist to stay in when visiting Carlow with easy access to various activities.

Data Description

To solve the mentioned problem, geolocation data on all the businesses was obtained. Foursquare was the location data provider of choice from which the data for this analysis was retrieved. The data used was obtained from Irish townlands data with 326 baronies, 2,507 civil parishes, 3,474 electoral divisions, 61,108 townlands currently mapped on OpenStreetMap. The various columns in the data set are defined below:

OSM_ID	Integer. The id of the object in the OSM database. If it's positive, it's a way; if it's negative, it's a relation. (Consult the OSM data model for more information). IDs are not shared across objects (e.g. there will never be an ED and a townland with the same OSM_ID).
NAME	String. The name of the object. Should be the "common name". Almost certainly in English, but may be in Irish. (NB In the KML/KMZ file, this is NAME2 due to how ogr2ogr converts things. Suggestions welcome for how to fix this.)
NAME_GA	String. The name of the object in Irish.
NAME_EN	String. The name of the object in English. Many objects don't have this, in which case presume that NAME is the English name
ALT_NAME	String. Alternative name.
ALT_NAME_G	String. Alternative Irish name. (it's short for "alt_name ga")
AREA	Float. Area in square metres
LATITUDE	Float. Latitude of the centre of this area
LONGITUDE	Float. Longitude of the centre of this area.
OSM_USER	String. Username of the OSM user who mapped this
OSM_TIMEST	String. ISO formatted datetime of when the object was added to OSM
epoch_tstmp	Integer. Unix 'epoch' time of when the object was added to OSM.
t_ie_url	String. URL of this object on Townlands.ie (this site)
LOGAINM_RE	String. Value of logainm ref tag. Logainm ID of this object.
co_name	String. Name of the county this object is in, if known. Not in 'county' or 'civil_parish' files.
co_osm_id	Integer. OSM_ID of the county this object is in, if known. Not in 'county' or 'civil_parish' files.
co_names	String. Comma separated list of names of the counties this object is in, if known. Orderd by the county's osm_id. Only in the 'civil_parish' files.

co_osm_ids	String. Comma separated list of osm_ids of the counties this object is in, if known. Orderd by the county's osm_id. Only in the 'civil_parish' files.
co_logainm	String. Logainm ref of the county.
ed_name	String. Name of the ED this object is in, if known. Only in 'townland' files.
ed_osm_id	Integer. OSM_ID of the ED this object is in, if known. Only in 'townland' files.
ed_logainm	String. Logainm ref of the ED.
cp_name	String. Name of the civil parish this object is in, if known. Only in 'townland' files.
cp_osm_id	Integer. OSM_ID of the civil parish this object is in, if known. Only in 'townland' files.
cp_logainm	String. Logainm ref of the civil parish.
bar_name	String. Name of the barony this object is in, if known. Only in 'townland' files.
bar_osm_id	Integer. OSM_ID of the barony this object is in, if known. Only in 'townland' files.
bar_logain	String. Logainm ref of the barony this object is in.
attributio	String. Value of the 'attribution' tag (if any)

Out of all these columns, just 4 of them were needed – That is CO_NAME, which is the County, ED_NAME which are the neighborhoods in that county and the latitude and longitudes.

Using Foursquare a total of 20 neighborhoods were randomly retrieved from the database such as Rathrush, Templepeter, Old Leighlin, Clogrenan etc. within a 2500m radius. Out of the 20 neighborhoods, there was a total of 130 business categories currently active in Carlow as. The most occurring business categories are Italian restaurant, Coffee shop, Burger joint, flower shop etc.

Methodology

Data pre-processing

The data used was downloaded from Irish townlands data and the columns needed were selected as described above. The County selected was Carlow because it is a very small county with a population slightly over 24000 people with very beautiful touristic sites especially for Christians

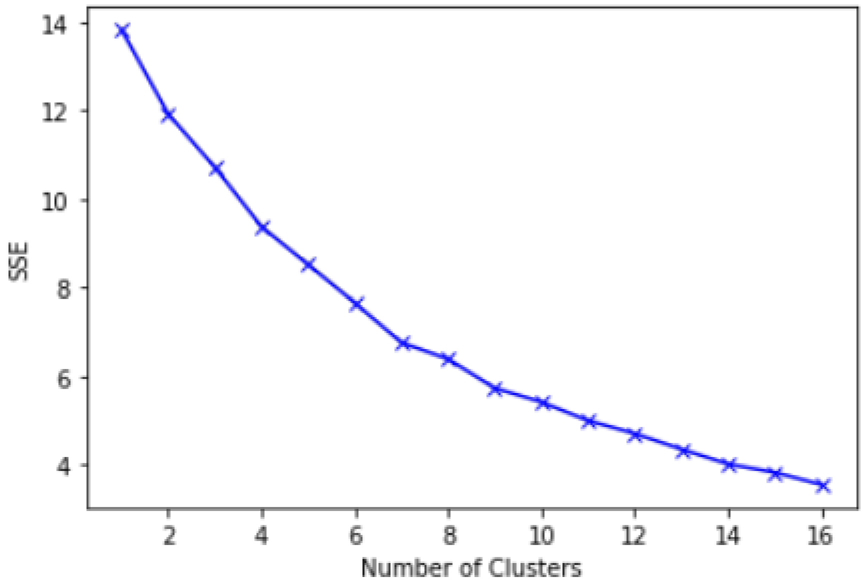
The data which was initially obtained from Foursquare was received in a JSON format. Therefore, the information which was important to solve the business problem was extracted from the JSON file and saved into a dataframe which was easy for use. The information on neighborhoods extracted were Neighborhood name, Neighborhood Latitude, Neighborhood Longitude, Business name, Business Latitude, Business Longitude, Business Category. Dummy variables were created using binary coding out of the business category. The data was then grouped by neighborhoods and their means per neighborhood taken. The latter was then used for the clustering process.

Figure 1: Business Categories in Carlow

K-MEANS CLUSTERING

K-means clustering was used to group the similar neighborhoods into clusters using the variables created in the pre-processing step. To decide how many centers for the clustering to use, several

K-means models were created with 1 to 15 centers and their respective within sum of squares errors (SSE) were plotted against their number of centers (clusters). This can be seen in the figure below. The elbow point was not very distinct. However, this plot seems to show a great drop in the SSE from the 1 cluster to 8 clusters. After 8 clusters, the SSE seemingly didn't drop that much compared to the previous. Therefore, a K-means clustering model specified with 8 centers was implemented.



RESULTS AND DISCUSSION

All the neighborhoods returned by Foursquare where grouped into the 8 clusters as seen in table below. Cluster 2 and 5 had the greatest number of neighborhoods and cluster 3 had the least. Note that these neighborhoods are not the absolute number of neighborhoods in Carlow.

	Cluster Labels	Neighborhood
0	0	5
1	1	3
2	2	16
3	3	2
4	4	3
5	5	13
6	6	2
7	7	6

Table 1: Neighborhood Distribution Across Clusters

The clusters mentioned are highly influenced by different venues. Some venues are more influential than others. This influence was based on the cluster mean values per feature. Table 2 below shows the most influential features per cluster. The venues shown here have been ordered in descending order of importance. Cluster 1 had a total of 8 venues; cluster 2 had a total of 5 venues; cluster 3 had a total of 39 venues; cluster 4 had a total of 4 venues, cluster 6 had a total of 5 venues, cluster 6 had 25 venues, cluster 7 had 3 venues and cluster 8 had 18 venues.

Cluster	**Venues**
Cluster 1	Garden Center', 'Business Service', 'Canal Lock', 'Gas Station', 'Indian Restaurant', 'Racetrack', 'Sporting Goods Shop', 'Supermarket'
Cluster 2	'Asian Restaurant', 'Chinese Restaurant', 'Convenience Store', 'Racetrack', 'River'
Cluster 3	'Convenience Store', 'Racetrack', 'Hill', 'Food Court', 'Fish & Chips Shop', 'Canal Lock', 'Indian Restaurant', 'Coffee Shop', 'River'
Cluster 4	'Asian Restaurant', 'Shopping Mall', 'Sporting Goods Shop'
Cluster 5	Asian Restaurant', 'Auto Garage', 'Garden', 'Golf Driving Range', 'Hotel', 'Wine Shop'
Cluster 6	'Gym / Fitness Center', 'Electronics Store', 'Gym', 'Deli / Bodega', 'Gas Station', 'Farm', 'Rugby Pitch', 'Gastropub', 'Movie Theater'
Cluster 7	'Asian Restaurant', 'Construction & Landscaping', 'River'
Cluster 8	'Plaza', 'Auto Garage', 'Asian Restaurant', 'Rugby Pitch', 'Racetrack', 'Park', 'Indian Restaurant', 'Gym / Fitness Center', 'Golf Driving Range'

Table 2: Summary of influential venues per cluster

CONCLUSION

The top venue or business categories which could interest investors and tourist can be found in table 2. These business categories seem to be the most economically active in respective neighborhoods. Based on the stated top economic activities, investors could profit if they invest in eating& drinking as we clearly see that there is at least a restaurant all neighborhoods regardless of clusters. Moreover, I can’t help but notice the lack of a pharmacy in the clusters, hence investors who are interested in investing in health (pharmacies) can look into that and also an eye opener for tourist to take along the medications they may need in case if any ill health. Also, in case investors are highly interested in fun & entertainment, neighborhoods in cluster 6 is preferable.