Similarity Comparisons of Luxembourg Towns

Curtis Thompson

Introduction

Background

Luxembourg is a small European country located between France, Belgium, and Germany. With an area of 2,586km^2 it is the seventh smallest country in Europe, and its population of 628,381 makes it the eighth least populous country in Europe [1]. 91.5% of the population lives in urban areas.

There are twelve legally recognised towns in Luxembourg; Diekirch, Differdange, Dudelange, Echternach, Esch-sur-Alzette, Ettelbruck, Grevenmacher, Luxembourg City, Remich, Rumelange, Vianden, and Wiltz. Luxembourg City is the largest town with a population over 100,000, however the small size of the country means that the remaining towns do not have the qualities needed to be regarded as cities in the international usage of the word.

Problem

The small nature of the towns means that they are difficult to compare to settlements in other European countries, as well as to each other. This project attempts to compare the towns of Luxembourg and use this to build a recommender system that finds the most similar town to any town in Luxembourg.

Interest

This project would be of interest for the people of Luxembourg, and particularly the government of Luxembourg who will want to understand the current state of towns in Luxembourg as well as their similarities and differences. This project may also be of interest to people looking to emigrate to Luxembourg and are looking for the most similar town to their current place of residence.

Data

Data Sources

Wikipedia has been used to scrape basic data on each town such as name, region, population, area, latitude, and longitude [2].

The Foursquare API has been used to obtain data on places and businesses within each town [3]. These places can be found by looking for places within the vicinity of the latitude and longitude data scraped from Wikipedia. Data obtained on each place includes name, latitude, longitude, address, neighbourhood, and type of establishment.

Weather data was considered for use in this project however there are only three weather stations in Luxembourg as suggested by The Weather Underground [4]. This meant that differences in

weather between manty of the towns were negligible and would not be useful in the recommender system.

Wikipedia Data

Data scraped from Wikipedia contains the following features related to each town: canton, area (in kilometres squared), population as of 2016, date of law, English name, Luxembourgish name, and Wikipedia link to individual town page. Via the individual pages the latitude and longitude of the towns was scraped. There are 12 rows and 9 columns in the dataset.

Foursquare Data

Data obtained from the Foursquare API contains the following features for each establishment; unique Foursquare ID, place name, latitude, longitude, postal code, country code, city/town, state, country, formatted address, address, cross street, place category, place top-level category, neighbourhood, unique Foursquare page ID, and neighbourhood. Establishments within 10 kilometres of any town in Luxembourg were obtained.

Methodology

Data Analysis

The nearest town to each establishment was calculated using a nearest neighbour classifier. This nearest neighbour classifier was trained on the latitude and longitude data for each town, so any prediction for an establishment gave the nearest Luxembourg town measured by Euclidean distance.

Establishment data was then grouped by nearest town. For any establishment with a nearest town, the establishment is considered to be in that town for the remainder of the report. The number of establishments found in each town is shown in Figure 1. Luxembourg City has far more establishments than any other town, but this is expected as Luxembourg City has an area at least twice the size of any other town and a population approximately four times greater than the next most populous town.

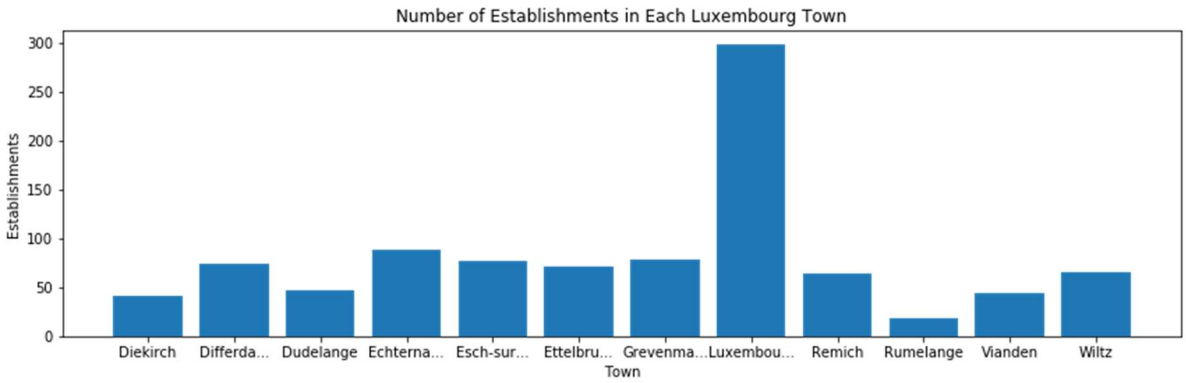

Figure 1: The number of establishments near each town in Luxembourg according to Foursquare data.

The category of each establishment can also be grouped using the category hierarchy provided in the API. Grouping of establishments by both town and top-level category was then performed. For each establishment category, towns could be compared by the number of establishments in that

category. These counts are shown in Figure 2. A majority of the towns have at least one of each type of establishment. The exceptions to this are Event establishments and Residence establishments, however most resident buildings such as houses are not retrievable establishments on Foursquare.

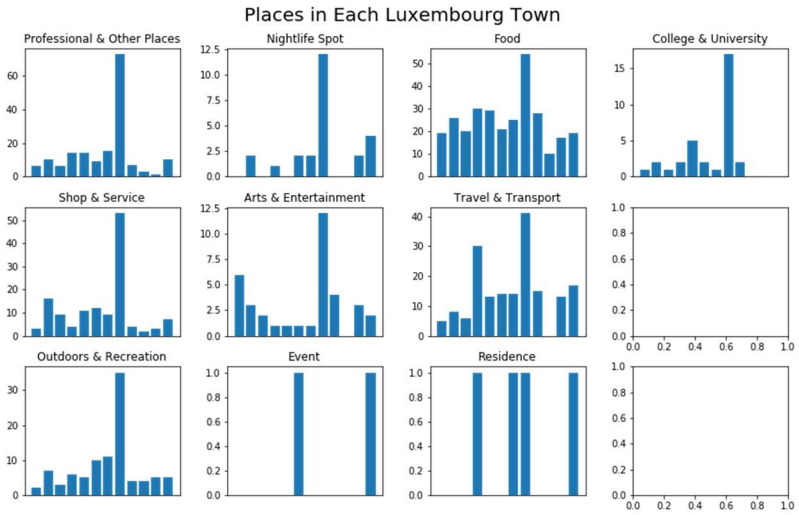

Figure 2: The number of establishments of each Foursquare category in each town in Luxembourg according to Foursquare data.

Towns are also compared on proportion of establishments in each top-level category. This can be seen in Figure 3. Food establishments are the most frequent type of establishment in each town except for Luxembourg City, where Professional & Other Places is the most common category, and Echternach, where Travel & Transport is the most common category.

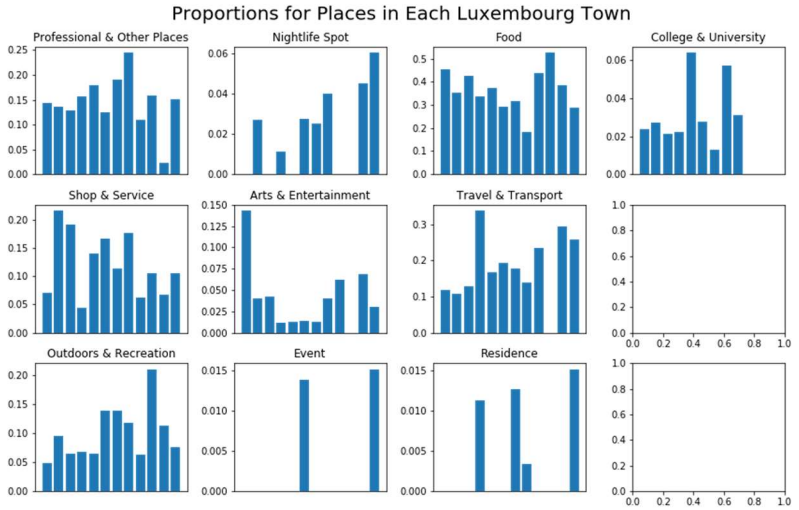

Figure 3: The proportion of category establishments in comparison to the top number of establishments in each town in Luxembourg.

Extracted Features

The features described in this section were extracted for use in the similar town recommender.

Simple town statistics were extracted; area, population, latitude, and longitude. Each of these features were scaled using the standard scaler.

One-hot encoding was used on the canton of each town to extract binary features. These binary features provide additional locational features besides the common latitude and longitude measurements.

Establishment counts for each top-level category of establishment were extracted. These features were then scaled using the standard scaler. Overall, there were ten features extracted in this group of features.

Establishment proportions for each top-level category in relation to the total number of establishments in the town were also extracted, however these were not scaled. This amounted to ten further features for each town.

Overall, 32 features were extracted.

Similar Town Recommender

To find similar towns for every town, a distance matrix was calculated using Euclidean distance. This distance matrix was calculated using the 32 features previously extracted.

To find the most similar town to any other town, the entry was the lowest distance for that town is the most similar town. The least distance town must be excluded from this however, as the least distance town for any town will be itself (with a Euclidean distance of 0).

For any extra town, where the most similar Luxembourg town is to be found, it can be compared to each Luxembourg town using Euclidean distance. The least distance town in this case will be the most similar Luxembourg town.

Results

The distance matrix for each Luxembourg town, as defined in the Similar Town Recommender section of the report, can be found in Figure 4.

	Diekirch	Differdange	Dudelange	Echternach	Esch-sur-Alzette	Ettelbruck	Grevenmacher	Luxembourg City	Remich	Rumelange	Vianden	Wiltz
Diekirch	0.000000	9.979070	4.028559	3.265391	4.702032	4.084962	4.649879	6.556426	4.937570	4.750778	3.659995	4.123144
Differdange	9.979070	0.000000	8.892508	10.282682	9.037458	10.364365	9.825300	10.993741	11.610629	9.549099	9.452105	10.342833
Dudelange	4.028559	8.892508	0.000000	4.612957	2.610017	4.066737	3.757865	5.129291	4.157874	2.991348	3.634314	4.116790
Echternach	3.265391	10.282682	4.612957	0.000000	4.565158	4.239311	3.425727	5.959162	4.247383	4.689266	4.508925	4.406556
Esch-sur-Alzette	4.702032	9.037458	2.610017	4.565158	0.000000	3.185155	3.553049	4.725635	3.793213	1.846377	4.591893	3.504174
Ettelbruck	4.084962	10.364365	4.066737	4.239311	3.185155	0.000000	3.489011	5.296944	3.561725	3.066864	4.144549	2.381826
Grevenmacher	4.649879	9.825300	3.757865	3.425727	3.553049	3.489011	0.000000	5.497930	3.284683	3.391146	3.740393	4.142573
Luxembourg City	6.556426	10.993741	5.129291	5.959162	4.725635	5.296944	5.497930	0.000000	5.991381	5.774971	6.628008	5.411300
Remich	4.937570	11.610629	4.157874	4.247383	3.793213	3.561725	3.284683	5.991381	0.000000	3.583664	4.967501	4.289091
Rumelange	4.750778	9.549099	2.991348	4.689266	1.846377	3.066864	3.391146	5.774971	3.583664	0.000000	4.201178	3.647245
Vianden	3.659995	9.452105	3.634314	4.508925	4.591893	4.144549	3.740393	6.628008	4.967501	4.201178	0.000000	3.879994
Wiltz	4.123144	10.342833	4.116790	4.406556	3.504174	2.381826	4.142573	5.411300	4.289091	3.647245	3.879994	0.000000

Figure 4: The distance matrix using Euclidean distance for each town in Luxembourg.

This distance matrix can then be used to find the most similar town for every town. These most similar towns are shown in Table 1.

Town	Most Similar Town
Diekirch	Echternach
Differdange	Dudelange
Dudelange	Esch-sur-Alzette
Echternach	Diekirch
Esch-sur-Alzette	Rumelange
Ettelbruck	Wiltz
Grevenmacher	Remich
Luxembourg City	Esch-sur-Alzette
Remich	Grevenmacher
Rumelange	Esch-sur-Alzette
Vianden	Dudelange
Wiltz	Ettelbruck

Table 1: The most similar town to every town in Luxembourg.

After finding the most similar town for every town, the towns can be plotted on a graph with edges where one node is the most similar town for the other node. This graph is found in Figure 5.

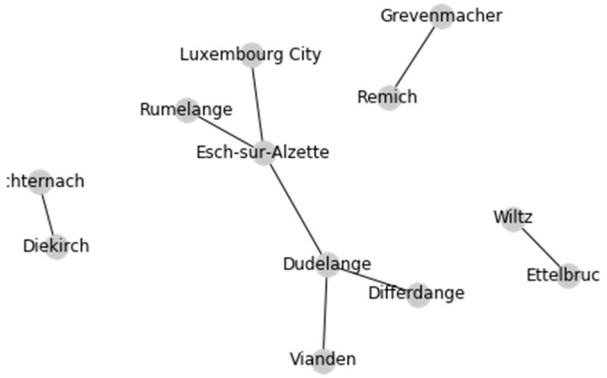

Figure 5: The town similarity graph produced from most similar town predictions for each town in Luxembourg.

Discussion

The most similar town predictions can be split into four distinct groups based upon the most similar town predictions produced, as shown in Figure 5. The first group contains Echternach and Diekirch. The second group contains Grevenmacher and Remich. The third group contains Wiltz and Ettelbruck. The fourth group contains the remaining towns; Luxembourg City, Rumelange, Esch-sur-Alzette, Dudelange, Differdange, and Vianden.

The first group contains Echternach and Diekirch. Although not in the same canton, these two towns can both be found in the east of Luxembourg. Both can be characterised as lacking Event and Residence establishments. While this is true for other towns in Luxembourg, these two towns also

have a low proportion of their establishments falling in the Shop & Service and Outdoors & Recreation categories.

The second group contains Grevenmacher and Remich. These two towns can also be found in the east of Luxembourg but more southern than the first group. These two towns are remarkably similar in terms of Food, Travel & Transport, and Shop & Service proportions.

The third group contains Wiltz and Ettelbruck. These two towns can both be found in the north of Luxembourg. The reason for this grouping is like the second group; the two towns are remarkably similar. The most similar establishment categories for these towns are Professional & Other Places, Food, and Outdoors & Recreation.

The fourth group contains the remaining towns; Luxembourg City, Rumelange, Esch-sur-Alzette, Dudelange, Differdange, and Vianden. It can be presumed that these towns make up the remaining group as there is no great similarity to the other three groups.

Conclusion

There are twelve towns in Luxembourg that can be compared to each other based upon establishment data gathered from Foursquare and statistical data gathered from Wikipedia. By grouping establishment data into top-level categories, 32 features can be extracted to be used in a distance matrix with Euclidean distance. Four groupings can be made using the distance matrix. The first group is Echternach and Diekirch. The second group is Grevenmacher and Remich. The third group is Wiltz and Ettelbruck. The fourth group contains the remaining towns. Each of these groups can be identified by locational data and establishment category data.

References

[1] https://www.cia.gov/library/publications/the-world-factbook/geos/lu.html (Accessed at 2020/04/29)

[2] https://en.wikipedia.org/wiki/List_of_towns_in_Luxembourg (Accessed at 2020/04/28)

[3] https://developer.foursquare.com/docs/places-api/ (Accessed at 2020/04/28)

[4] https://www.wunderground.com/ (Accessed at 2020/04/29)