The Battle of Neighborhoods

Summary

Transforming data into usable information is a clever way to make informed decisions. Here we show how to manipulate data to make them available in a beautiful way that is easy to understand. More specifically, we show how to manipulate data which is segmented by geographical location and project it into maps.

We have developed a pipeline to dissect cities of any size in small areas which are relevant from the administrative standpoint. Then, specific data from each location is gathered and processed. When a city is dissected and data organized by location, clustering techniques help us in finding similarities among city areas and in making better decisions.

Keywords: Machine Learning, Competitive intelligence, Data Science, Foursquare, Clustering Techniques, Smart Cities, Girona.

Background

Cities, as large human settlements, flourished in parallel with the development of agriculture and have been characterized by the specialization and social division of labor. Great cities have existed in all civilizations of the human history and they have been planned and managed by bureaucracy systems which needed data science to do the job. One great example that links cities and data science is the development of the concept of the number zero. Even if in some of them the number zero was not invented, most of the ancient civilizations developed the concept of zero and used it for calculations in many areas like accounting, astronomy, geography etc.

In cities people has taken decisions that have reshaped the Earth surface and the future of humanity lies in the hands of city boys. Cities are great places to enhance human interaction both for the good and for the bad. In both cases data science is paramount in improving human life. In the last couple of centuries there has been a clear trend to concentrate people in cities that have become megalopolys. We believe, however, that networks of small to middle sized cities will have a much better capacity to face the challenges of a crowded world.

This is why at Global Digital Management we concentrate on delivering analytics services to businesses and governmental agencies operating in these small or middle sized cities that are dynamic and innovative.

In this post we show a short example of what clustering techniques can deliver by applying a socieconomic analysis to Girona, a small catalan city which is the capital of a world famous tourist destination, La Costa Brava.

Methodology

There are three key aspects when dealing with competitive intelligence:

Consistent and trustable raw data

Here we are using raw data from official and well stablished sources. This will ensure data quality from the beginnig.

* Data from the Cartographic Institute of Catalonia ICC
* Data from Girona’s local council that can be obtained either from Girona Open Data GOD or from l’Observatori.
* Data from Foursquare

Accurate data processing

When dealing with geographical data we find that geographic areas are far from having regular shapes. For instance, we are using census sections in Girona. Each of these administrative units have a unique surface and a unique shape that can be represented by an irregular polygon. In order to have a perfect fit between the Foursquare data and the geographical data we have performed complex calculations in the background.

Carefull choice of machine learning techniques

Clustering techniques are unsupervised machine learning techniques which compare a bunch of parameters of several samples to finally cluster samples together in different groups. One of the limitations of most clustering algorithms is that they we have to suplly them a parameter, the final number of clusters *k*. There is no proper way to optimize the number of clusters so the result is always dependent on the initial choice of the analyst. We have developed a special technique to optimize the parameter *k*clusters in order to have the best results.

Results and Discussion

Neighborhoods by venues

The total amount of venues recorded in the Foursquare database from Girona is less than 300, a number which is far below the actual number of venues. Girona is a very dynamic town and the commercial center of a rather big geographical area. It accumulates many governmental offices and it hosts a university which has had a great impact on the city over the last 30 years. However underrepresented, the Foursquare venues are able to capture the special profile of the different neighborhoods. There is a great cluster containing four neighborhoods, Centre, Eixample, Santa Eugènia and Oest. They are the four neighborhoods that constitute what we can call the center of the city. Although the socioeconomic profile of these four neighborhoods is not necessarily similar in the streets there is a complex mixture of restaurants, supermarkets, groceries, hairdressers etc. All four neighborhoods are places with all you need to live and, at the same time, in most of them there are locations were people from other neighborhoods may eventually go shopping. Next cluster is formed by the neighborhoods Est and Sud. They are very different neighborhoods in term of socioeconomic landscape. While Est is a low-income neighborhood, Sud has areas with people with very high income. However, the venues profile is similar and it is basically characterized by almost no retailers nor personal services, and the presence of some municipal equipment. Finally, we have three clusters with only one neighborhood. Montjuïc is a residential area with some gardens. Nord is a mixture of working-class immigration and working-class native population which is located a bit far away from the city center. It is by no means a commercial place, but there you can find some specialized businesses. The las one, Mas Xirgu, is a very special one, we see it coming alone in every clustering because it is basically an industrial area with special business like car dealers, garages, pet shops and clinics, and industrial providers.

Neighborhoods by education level

When looking at Girona by its venues profile the main feature was the presence of a great center comprising four neighborhoods. When we look at the education level, the city is split into three clear areas and the center is also split. Now, Montjuïc joins the Centre, Eixample and Sud neighborhoods. These are the places where people with the highest education is living. Next, with middle education levels we find the west crown of the city, Santa Eugènia, Oest and Nord. Finally, neighborhoods Est and Mas Xirgu cluster together even if their profile is not exactly the same. Both have a lot of people without primary studies or just with primary studies. This is probably what makes them clustering together. But Mas Xirgu has half of people under 16 as compared with Est and Mas Xirgu. Instead Mas Xirgu has almost 16% of people with university studies while in Est neighborhood it is a bare 3%.

Neighborhoods by number of people in households

Clustering by the number of people in households is often a good way to reveal socioeconomic features of neighborhoods and a good way to reveal familiar structures if any. Clustering Girona according to number of people in households yields a rather atomized map, with three clusters with two neighborhoods in it and three neighborhoods that come alone. First of all, the neighborhood Centre is a very special one, one of those which has experienced a process of gentrification due to the pressure of tourism and university students. Here 65% of houses have one or two residents, while only 7% of the houses have more than 4 residents. Montjuïc is another singularity, it has the lowest proportion of houses with a single resident and 75% of houses have between 2 and 4 residents, it is the prototypical middle-class family neighborhood. Mas Xirgu comes again as a special case, it has the highest number of houses with only one resident and together with houses with two residents make up 75% of households. Then there is a 22% of households with four or five residents, showing the presence of two very different areas in the same neighborhood. Neighborhoods Est and Santa Eugènia probably cluster together because they are the two neighborhoods with the highest proportion of houses with 5 residents or more with more than 16% of households. Neighborhoods Sud and Oest are family neighborhoods with more than 90% of households with 1 to four residents, the class with two residents per house being the one with more occurrences. In neighborhoods Nord and Eixample the leading categories are the hoses with one or two residents and then, at a certain distance, families with 3 or 4 members. In summary, the profile of Girona is that of a modern city were families are getting smaller and most people live alone or just in a couple. However, slight differences are relevant and show that the map of the city is quite diverse.

6 Conclusion

Foursquare might not be the most comprehensive database to make a venues profile of the city of Girona. In the future this should be complemented with data from other databases like google maps.

Girona is a small town of 102 thousand people but it is a quite diverse city we can say that neighborhoods matter, they matter in terms of venues and in terms of socioeconomic profile. Clustering them by different criteria gives different results showing that having the right information might be crucial for city management.

Future directions.

Complementing Foursquare data with data from other databases

Adding data from other socioeconomic variables.

Define a socioeconomic index might help a lot on delivering a very focused information to stakeholders.

Tracking venues to follow food traffic in time series would be of great interest for stakeholders but this will be done after confinement measures are relieved.