

Re-defining Italian regions using... the venues types!

The initial, rejected project

This was the previous goal:

"This idea came to my mind randomly, and I like it: I wonder if it's possible to predict if a business is going to be successful or not, just by checking 'a few' data regarding the surroundings. Also, as Italy is quite a wide reality, it may be interesting to check as well the impact of the position on the prediction.

Apart from the 'success' goal, it would be interesting to check if we could predict the business type using just the surrounding environment as input."

Why I couldn't do it:

To have the average rating of a venue, on Foursquare, you need to do a premium call. 500 premium calls is the free limit, and it would not be enough for the project.

Introduction: business problem

I would like to group all the cities in Italy, using the information gathered from the venues and the geographical location. The idea is to verify how big the cultural gap between the different regions is (still).

So it will be treated as a clustering problem. The coordinates of the different cities will be part of the data used by the algorithm.

This theory would be confirmed if, plotting all the data on a map, we were to see a similarity between the new regions and the true ones.

Data

The data we will need for this project are:

- Every town in Italy
- The coordinates of every city
- An Italian map, to plot on

Sources

- | | | |
|-------------------------------------|----|--|
| • For every town in Italy | -> | This dataset |
| • For the coordinates of every city | -> | This dataset |
| • For the Italian map | -> | We use Folium, a library for plotting over maps! |

Methodology

Results

Discussion

Conclusion