

Capstone Project - The Battle of Neighborhoods

Final presentation

Introduction

"How to select the best home location in Milan?"

My customer asked me to find a suitable place to buy a home in Milan. He wants to select the best place to live, according to the nearby venues. In particular, he wants to select the place that offers leisure time possibilities, such as Cinemas, theaters, shops and restaurants and transportation services, such as Metro, Tram and Bus stations. He has a clear idea on which are his priorities and he asked me to analyze the neighborhood for each of the five home location candidates.

Data section

The data available for fulfilling the task are:

- Milan city centre coordinates and the radius from the centre to determine the searching zone
- the list of the venue categories that are important for the neighborhood analysis
- the weight that every venue category has in the decision

Methodology

This section describes the steps done to find the best home location. Basically the venues around Milan city center are identified. Then a k-means clustering algorithm is used to locate the centroid of the obtained clusters. At the end the best cluster is chosen according to the venues number and weights of each venue belonging to each cluster and the final best position is identified.

Results

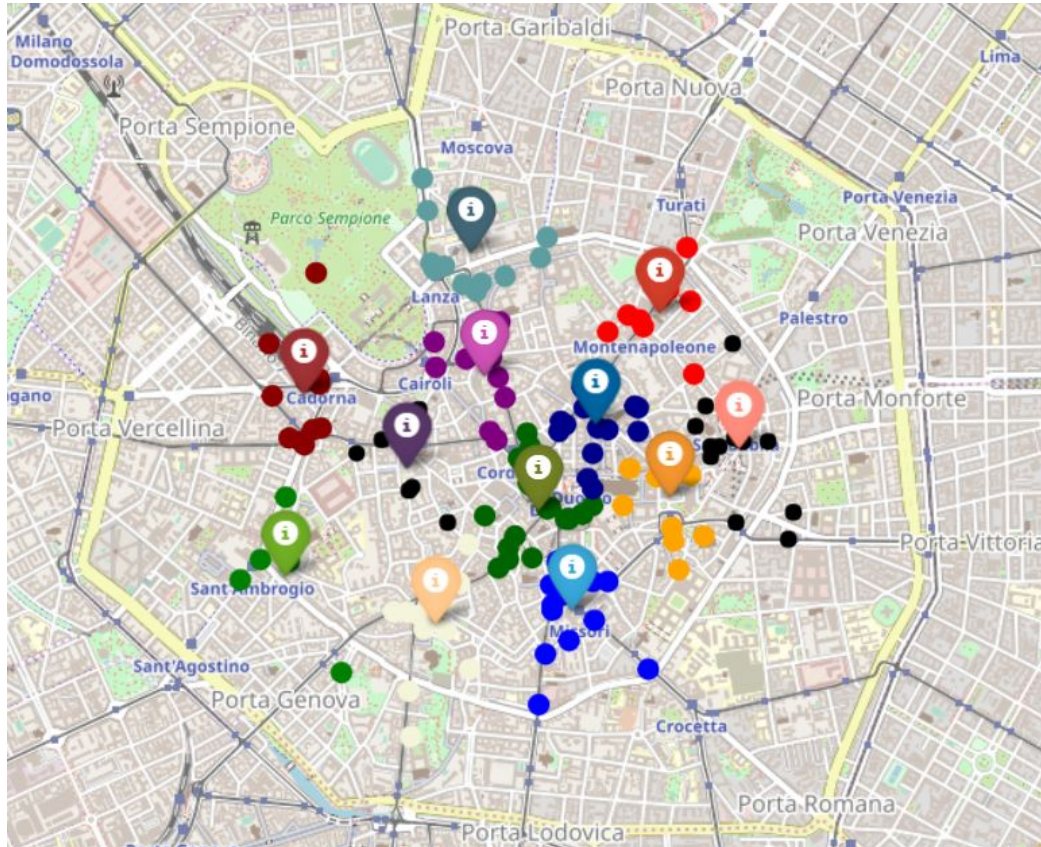
Starting from Milan city center, the following venues were taken into account:

'Food', 'Shop & Service', 'Bus Stop', 'Metro Station', 'Cinema', 'Theater', 'Tram'

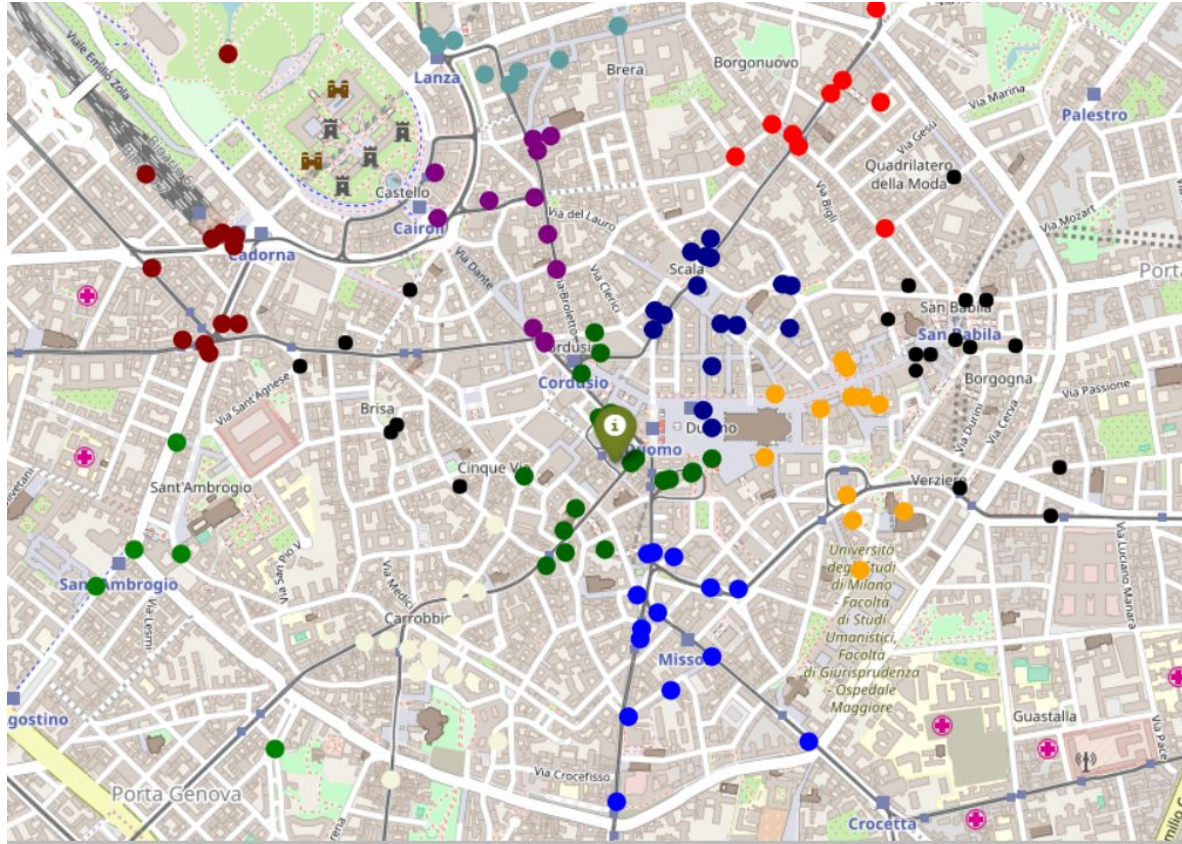
And the result is:

```
32 "Food" venues are found within 1000m of location "Milan"  
44 "Shop & Service" venues are found within 1000m of location "Milan"  
5 "Bus Stop" venues are found within 1000m of location "Milan"  
10 "Metro Station" venues are found within 1000m of location "Milan"  
17 "Cinema" venues are found within 1000m of location "Milan"  
1 "Theater" venues are found within 1000m of location "Milan"  
37 "Tram" venues are found within 1000m of location "Milan"
```

The 146 venues are then grouped by k-means algorithm into 12 clusters.



And the best cluster is obtained (the one with the marker in the picture).



Discussion

We were able to find the best place according to the algorithm chosen. There are however some critical points. First of all, 12 cluster groups were chosen. This number was obtained empirically, since the use of train and data set to select the right k value is meaningless in this case. If this number is changed also the position of the best place will also change. We did some experiments, but at the end the position is more or less in the same zone.

In addition to this, how weights are chosen plays a role in the final position.

It is however to be said that the methodology used is able to find the best position in the city because it select the center of a zone full of venues. Even changing the number of clusters and the weight of each venue, the result will be quite close to the position obtained as the best.

Conclusion

Using the information obtained from Foursquare and using machine learning algorithm, we were able to identify the best place to buy a Home in Milan.

