

Relocation in Genova

Part of the Capstone project



Genova City center

Matteo Zangani

Spring 2021

Relocation in Genova

The Capstone project



Didascalia

Scope of work

The current research aims to help facilitate the choice for a relocation place in Genova as more job opportunity arise in the recent times.

Introduction of Genova region

is the capital of the Italian region of Liguria and the sixth-largest city in Italy. In 2015, 594,733 people lived within the city's administrative limits. As of the 2011 Italian census, the Province of Genova, which in 2015 became the Metropolitan City of Genova had 855,834 resident persons. Over 1.5 million people live in the wider metropolitan area stretching along the Italian Riviera.

On the Gulf of Genoa in the Ligurian Sea, Genoa has historically been one of the most important ports on the Mediterranean: it is currently the busiest Italy and in the Mediterranean Sea and twelfth-busiest in the European Union.

Genoa was the capital of one of the most maritime republics for over seven centuries, from the 11th century to 1797. Particularly from the 12th century to the 15th century, the city played a leading role in the commercial trade in Europe, becoming one of the largest naval powers of the continent and considered among the wealthiest cities in the world. It was also nicknamed la Superba ("the proud one") by Petrarch due to its glories on the seas and impressive landmarks.

The city has hosted massive shipyards and steelworks since the 19th century, and its solid financial sector dates back to the Middle Ages. The Bank of Saint George, founded in 1407, is the oldest known state deposit bank in the world and has played an important role in the city's prosperity since the middle of the 15th century.

Problem

Data can help to determine relocation decision. This project aims to predict where a suitable place can be found by exploring each Genova Neighborhood and housing prices.

Data

1. Data acquisition

The data has been extracted by scanning the web pages of <https://www.mercato-immobiliare.info/liguria/genova/genova.html>, this website contains a table with the list of Genova Neighborhood and the housing price. On wikipedia website <https://en.wikipedia.org/wiki/Genoa> there is the list of Municipalities and the information about the city that we used in this report.

Foursquare API was used to retrieve the important information regarding venues in each Neighborhood to be able to model the main characteristics.

1.Data preparation

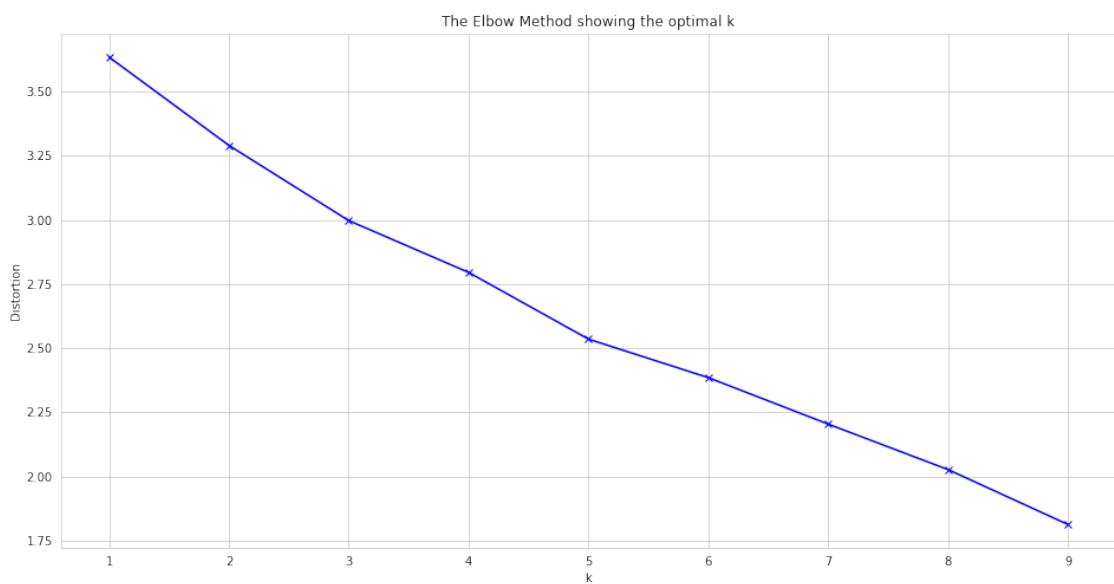
After importing the table few issues had to be addressed.

First one was that some Neighborhoods cannot be associated to Municipalities data set because of different spelling. The problem has been fixed manually. Second one was that some Neighborhoods were not listed at all into Municipalities because very small. Those were rejected for the computation. One Neighborhood was positioned outside the city boundary, that outlier was removed. This left us with a total of 35 Neighborhoods.

Regarding the exploration of the venues in each neighborhood a searching radius of 1000m radius has been used to allow a decent number of venues to be retrieved.

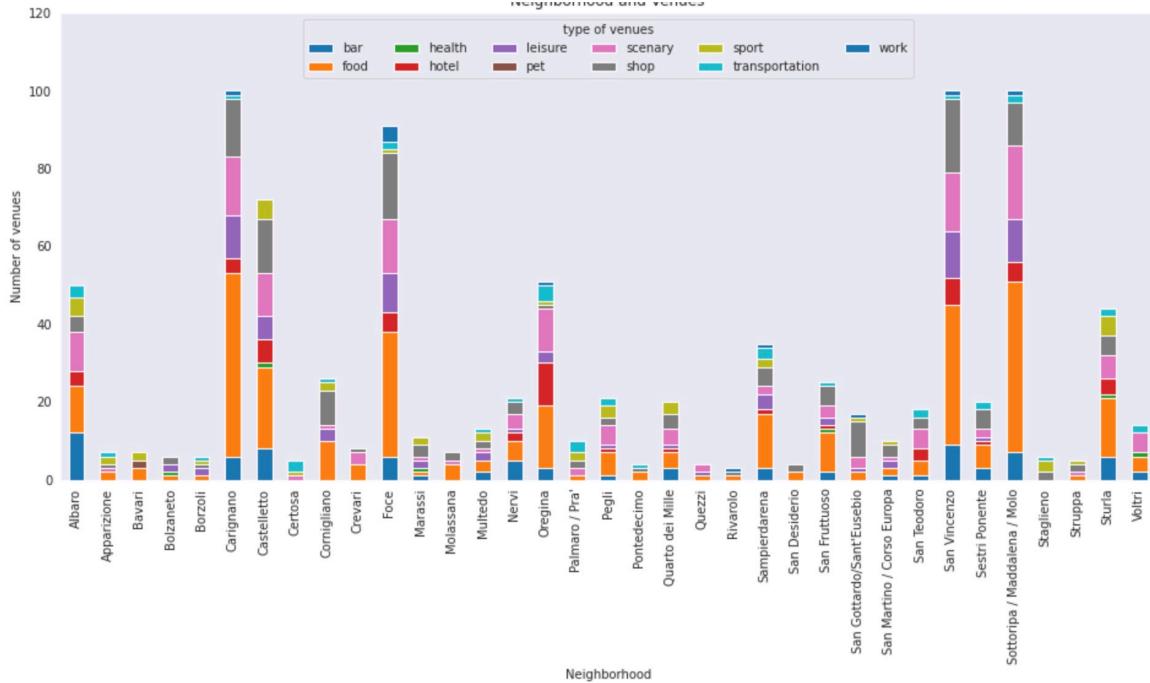
Last but not least, the number of venues returned was 141 of unique categories and they were very diversified by name.

Here below is the elbow test for K-NN modeling if all the venue category were used. It makes very difficult to identify an elbow point for best K value estimation.

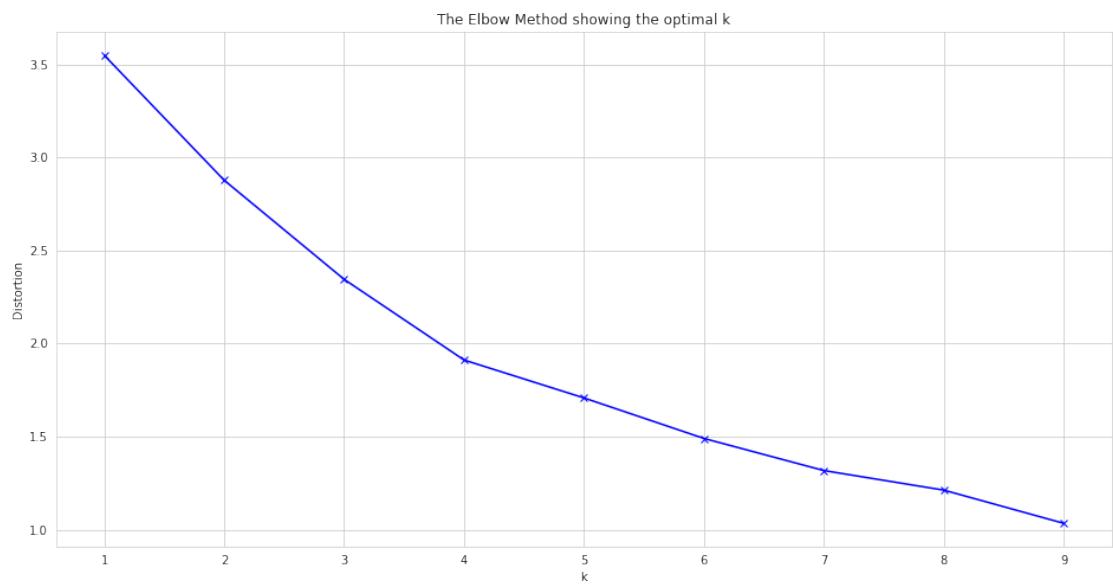


Elbow method for best K estimation on raw dataset

So the venue categories had to be regrouped based on their characteristics to allow a better modeling. Total main 11 groups were identified



Histogram with venues regrouped and neighborhoods



Elbow method for best K estimation after regrouping

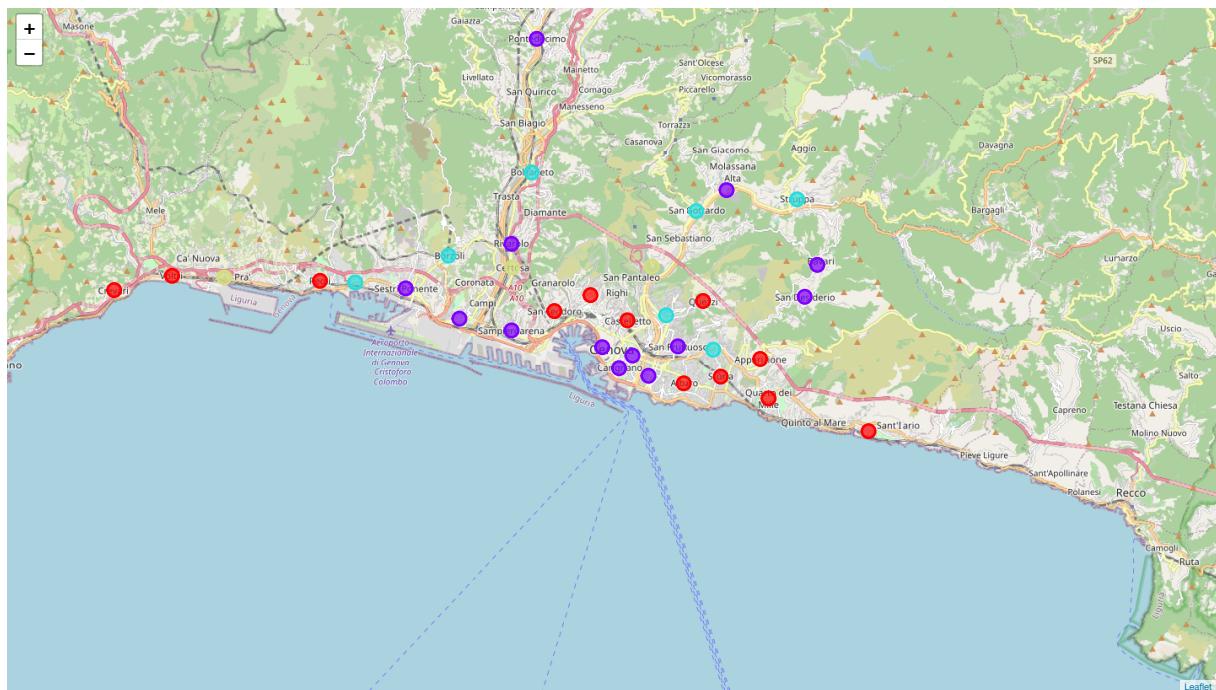
After regrouping the distortion plot looks like picture below, so it's easier to identify a small elbow at $K = 4$

Data Analysis

In the data analysis part we used cluster segmentation method to identify different clusters and being able to bring forward the best information for the final decision.

The algorithm used is ***k*-nearest neighbors algorithm (*k*-NN)** is a non-parametric classification method. The *k*-nearest neighbors algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems.

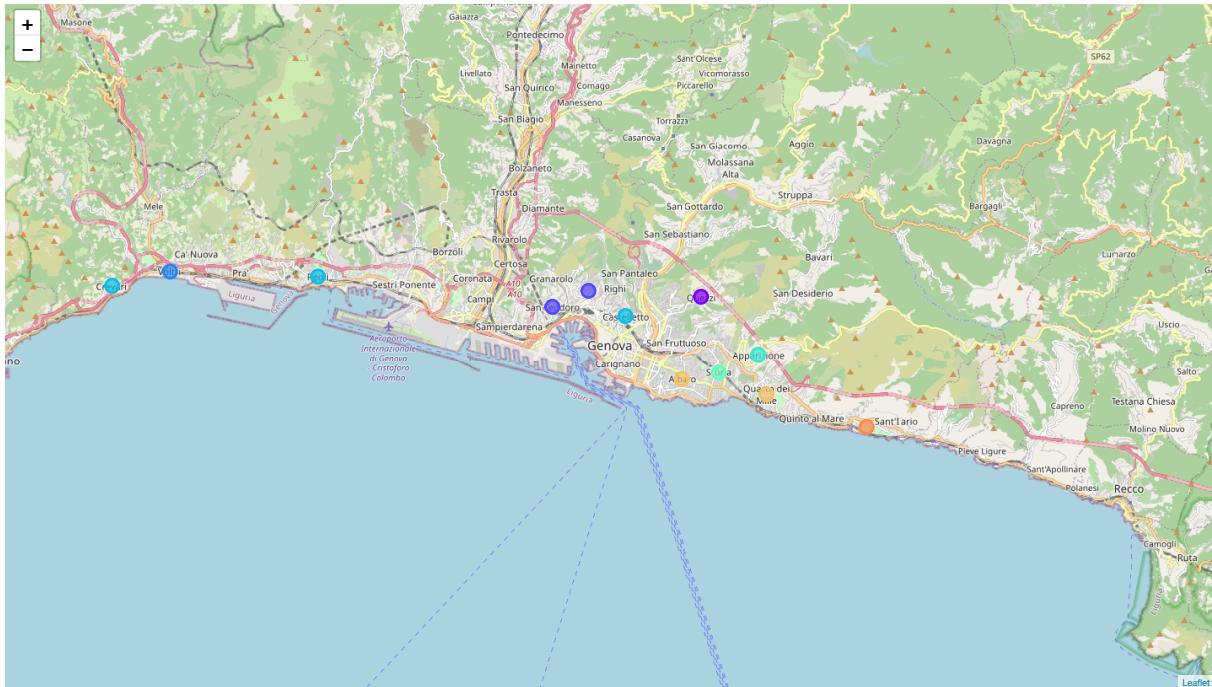
Below we represent the map obtain with the module folium of all the 4 Clusters



Display of all Cluster found, color code represent different cluster

1. Cluster #1

This cluster based on venues category shows the general character of residential area, although the housing price is very wide.

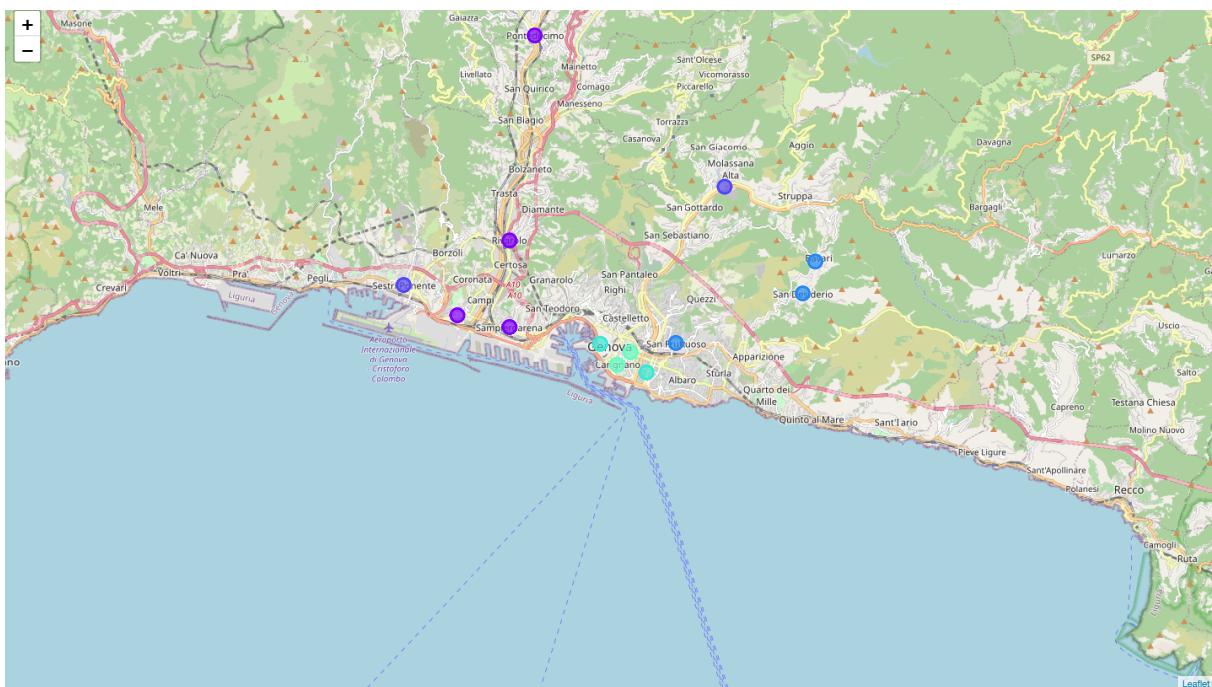


Cluster #1, color code represent housing price

The data within the cluster seems away from the working areas and from the main motorways and intersections

The Neighbor included are mostly located on the seaside, which makes it very interesting for relocation.

2. Cluster #2

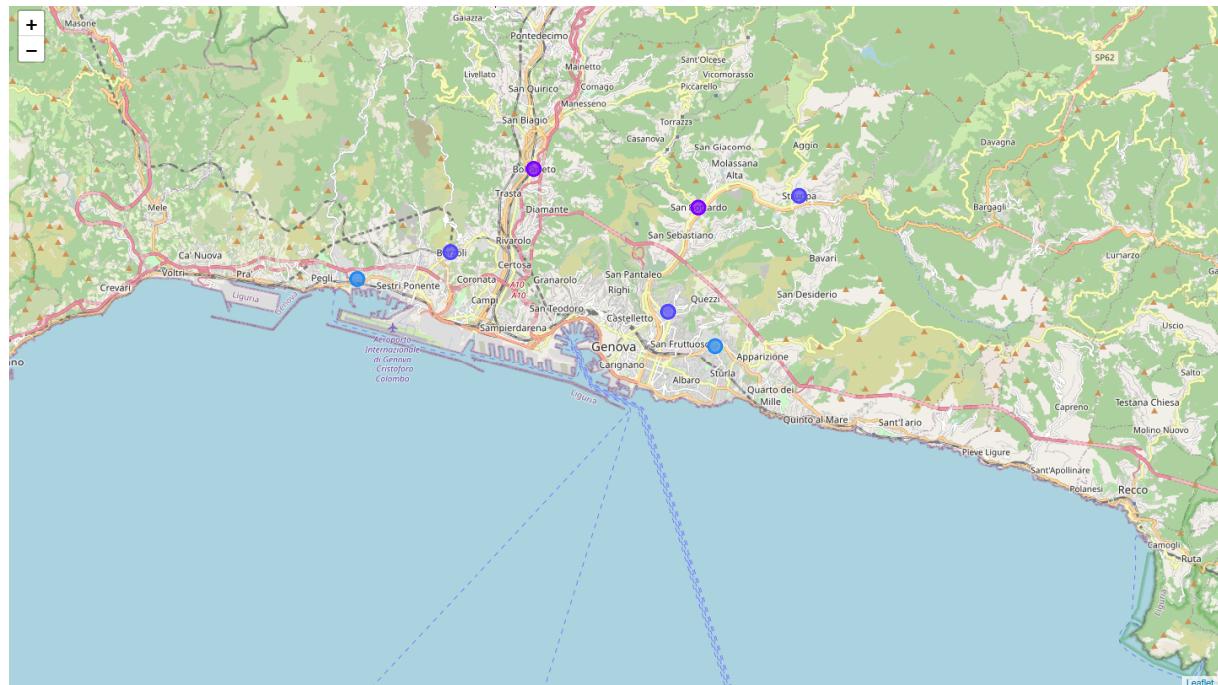


This cluster #2 based on venues category shows the general character of office/ industrial area, the housing price is narrow and clusters in two distinct blocks:

- 1 city center offices
- 2 Industrial areas

The data within the cluster are generally not too far from sea side except few neighborhoods, however they are mostly located within the harbor area which makes impossible to go to the beach.

3. Cluster #3

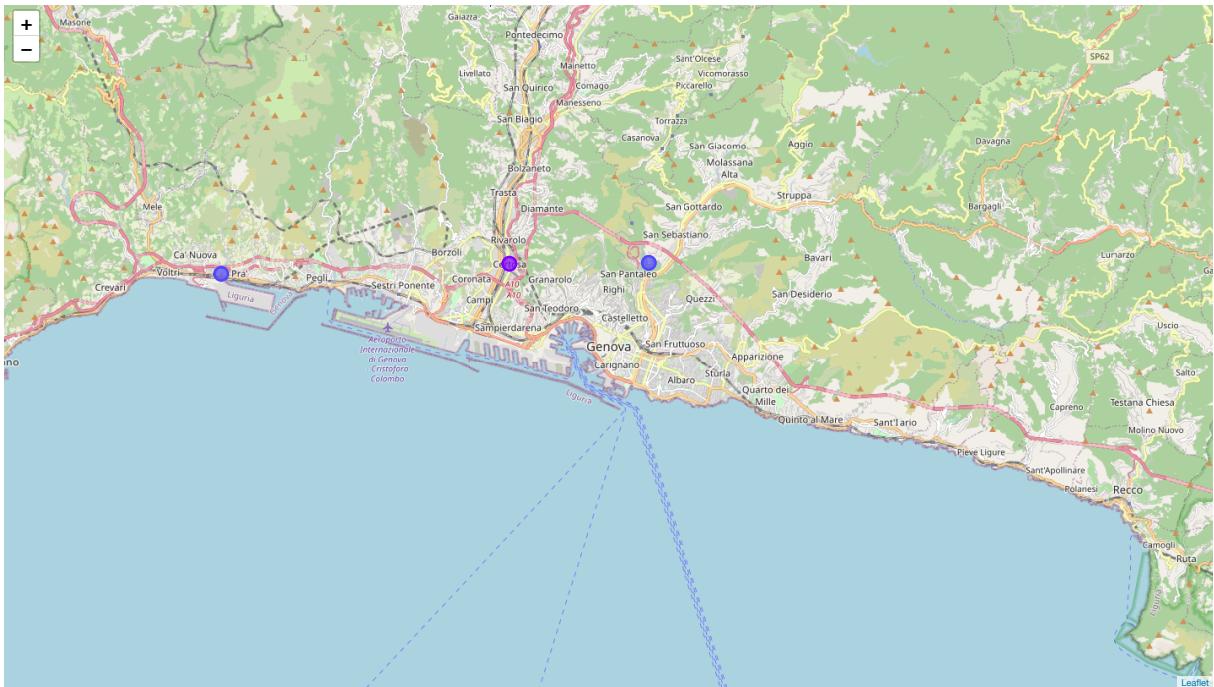


Cluster #3, color code represent housing price

This cluster based on venues category shows the general character of mixed area, the housing price is very uniform and low.

The Neighborhoods included are mostly located away the seaside, and the venues ranges don't make it very appealing.

4. Cluster #4



Cluster #4, color code represent housing price

This cluster based on venues category shows that is located around main intersection and transportation area.

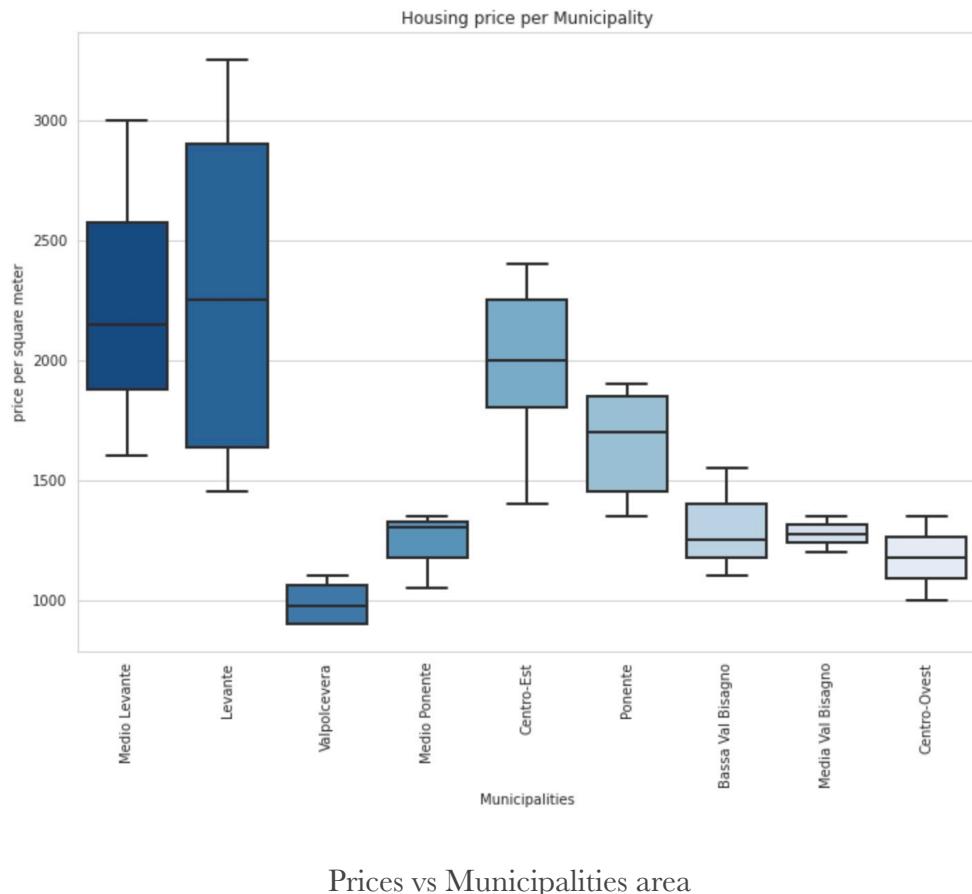
This characteristic makes it not suitable for relocation.

Data Interpretation

After sewing the main data modeling only cluster #1 was suitable for relocation purposes.

However the price ranges span widely across the area.

Price ranges and areas were analyzed to get the best compromise.



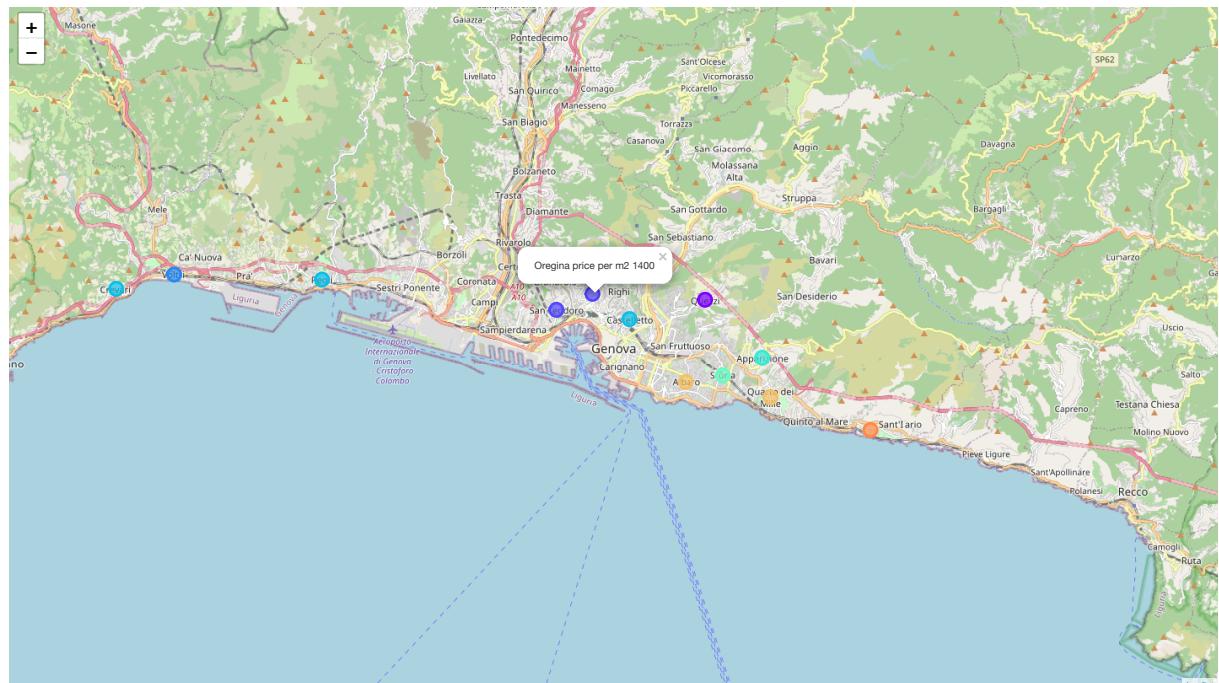
Conclusions

As we can see the Eastern bounds are far more expensive (Levante, Medio Levante, Centro-Est).

In this case we are left with two interesting areas.

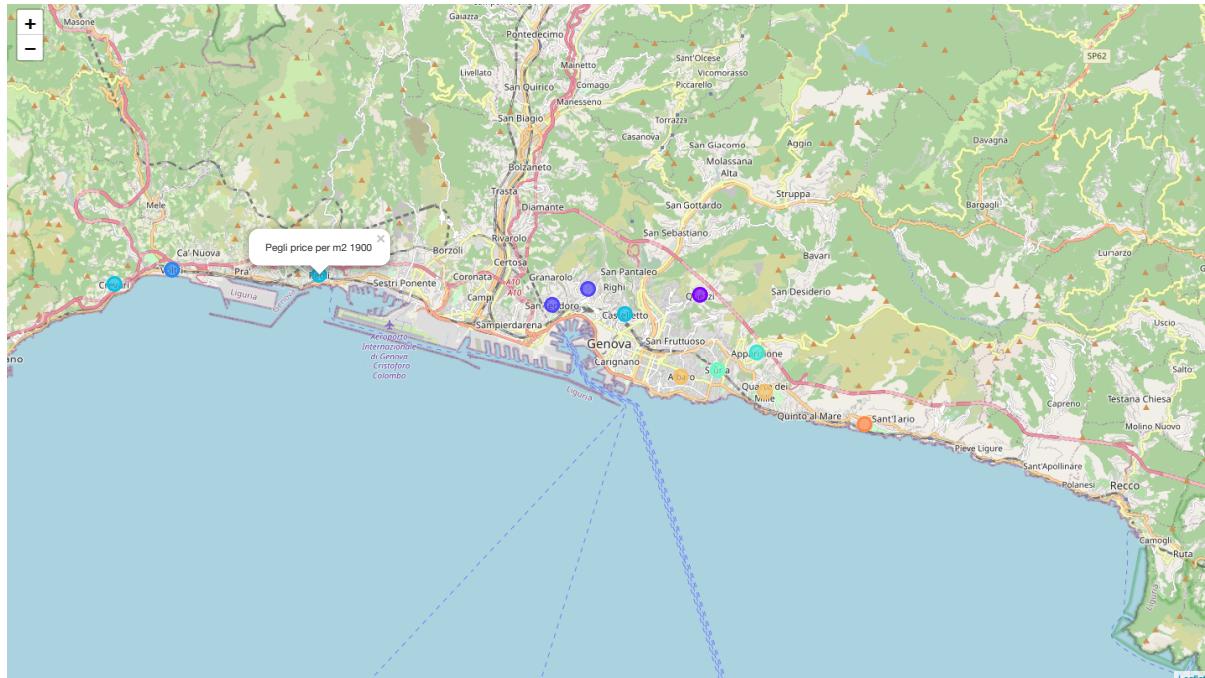
First one is Oregina, which is located in the Center side.

According to the web search has a wonderful view, however roads are narrow and is not located near any beach



Oregina

On the western Side we still have one interesting area which is located not too far from the city center and still within reasonable price range. This location is Pegli and is marked in the map below.



Pegli

This Particular location has nice view, beach access and all the transportation needed. Price range in not too low but still affordable, this would be our final choice.



Image of Pegli Neighborhood

Considerations and Improvements

First of all by using KNN clustering might not be the best choice, the result is slightly biased by a small dataset. Moreover the KNN method yields different results at every run and the seeds are randomly positioned.

More data can be retrieved to create a better model, like social media posts and 'likes', schools/universities and traffic information data (that can help on understanding some noisy areas vs quiet places)