

# THE BATTLE OF NEIGHBORHOODS IN GENOVA



Matteo Zangani

# PROBLEM

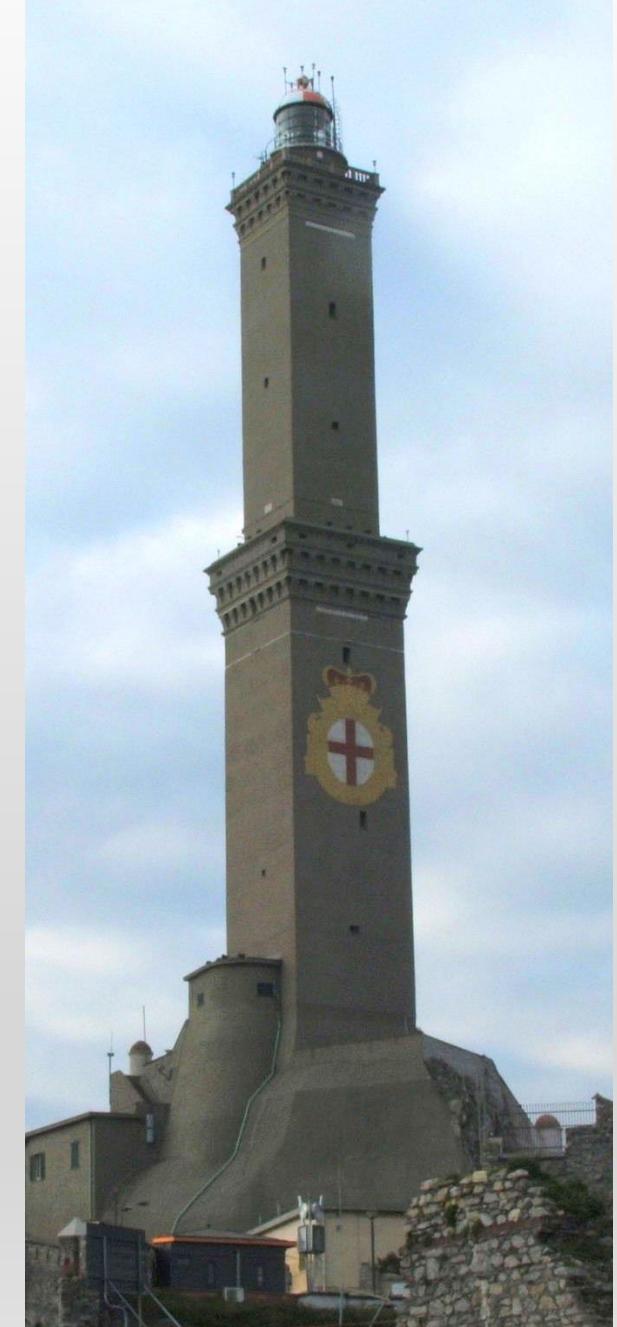
The goal of this project is designed to analyze the opportunity of relocation for our family in one of the neighborhood in the city of Genova since a new job opportunity arise.

Each neighborhood will be explored to analyze their features and try to be able to identify the best places to relocate.

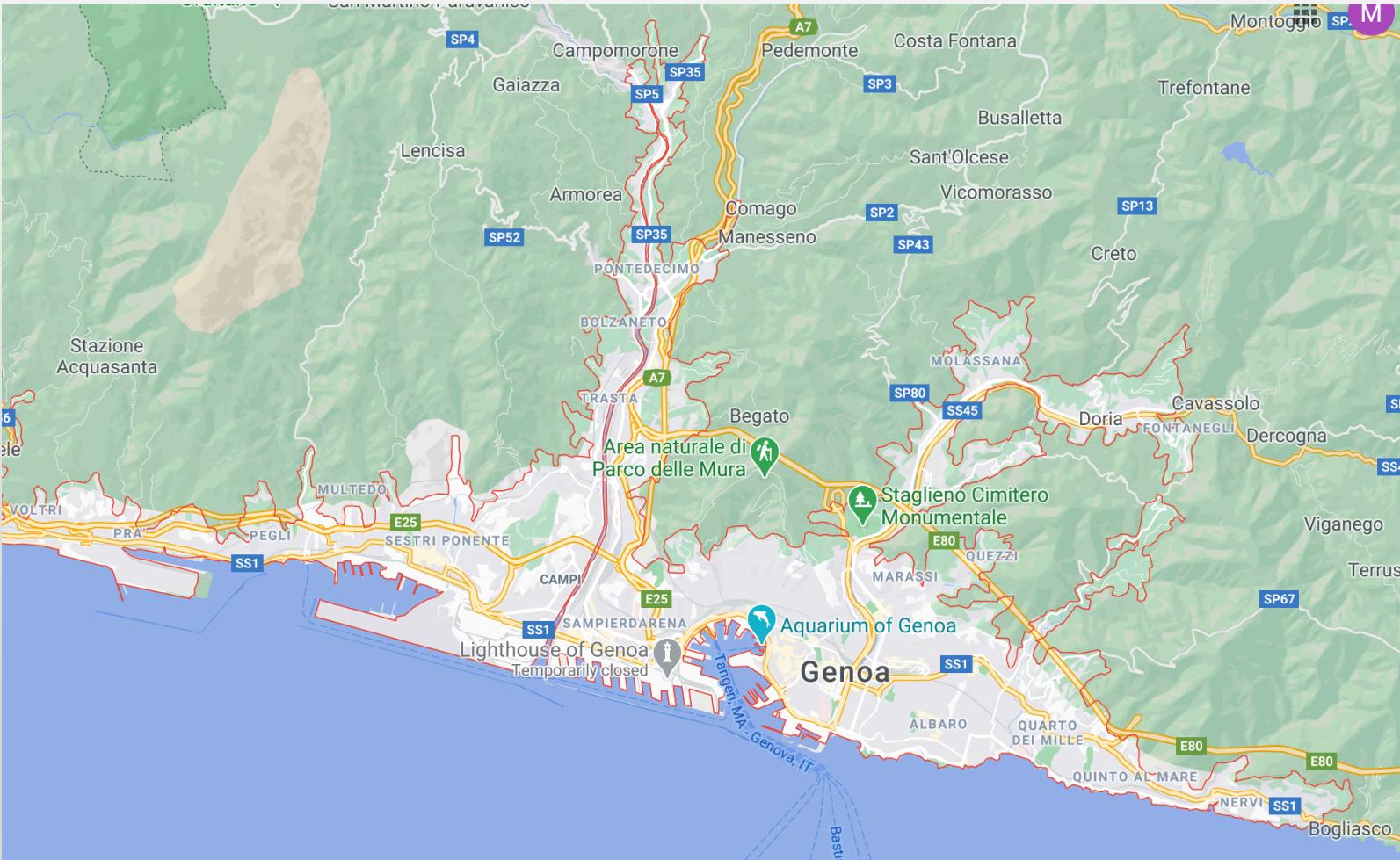
# INTRODUCTION

Genova is a city in the North of Italy with sustainable industrial development.

The city is on the Mediterranean Sea and has a wonderful climate to live.



# MAP OF GENOVA



Data from google map

# DATA

The name and number of Neighborhoods in Genova will be taken from the website [mercato-immobiliare.info](http://mercato-immobiliare.info)

The website also gives the average sell prices for houses, which will help the final decision on which area is more convenient within the chosen clusters.

From wikipedia website we can assign the municipalities values to each Neighborhood

Using Foursquare we will explore venues of each Neighborhood and understand what are the main characteristics of each area.

# METHODOLOGY

The data will be imported and analyzed to identify the most interesting areas of Genova.

By segmenting and clustering the different Neighborhoods we will be able to identify common characteristics between areas which will give us more options to choose the best places

Housing price and location will drive the final decision on which is the best area for searching a new house

# DATA EXAMPLE

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Albaro	Café	Hotel	Pizza Place	Beach	Plaza	Ice Cream Shop	Cocktail Bar	Restaurant	Harbor / Marina
1	Apparizione	Pizza Place	Bakery	Sports Club	Gym / Fitness Center	Convenience Store	Gym Pool	Plaza	Soccer Field	Electronics Store
2	Bavari	Italian Restaurant	Diner	Stables	Café	Food & Drink Shop	Garden	Furniture / Home Store	Fried Chicken Joint	French Restaurant
3	Bolzaneto	Pizza Place	Toy / Game Store	Stadium	Coffee Shop	Health Food Store	Wine Shop	Flea Market	Fried Chicken Joint	French Restaurant
4	Borzoli	Soccer Field	Rock Club	Train Station	Furniture / Home Store	Bakery	Stadium	Garden	Fried Chicken Joint	French Restaurant

After the data has been cleaned and imported, all the available venues have been retrieved and merged. Here is a table summary with displayed the most common venues. A searching radius of 1000m have been used from the center of each Neighborhood  
The price value was not used for modeling as it would interfere too much in the clustering process

# CLUSTERING

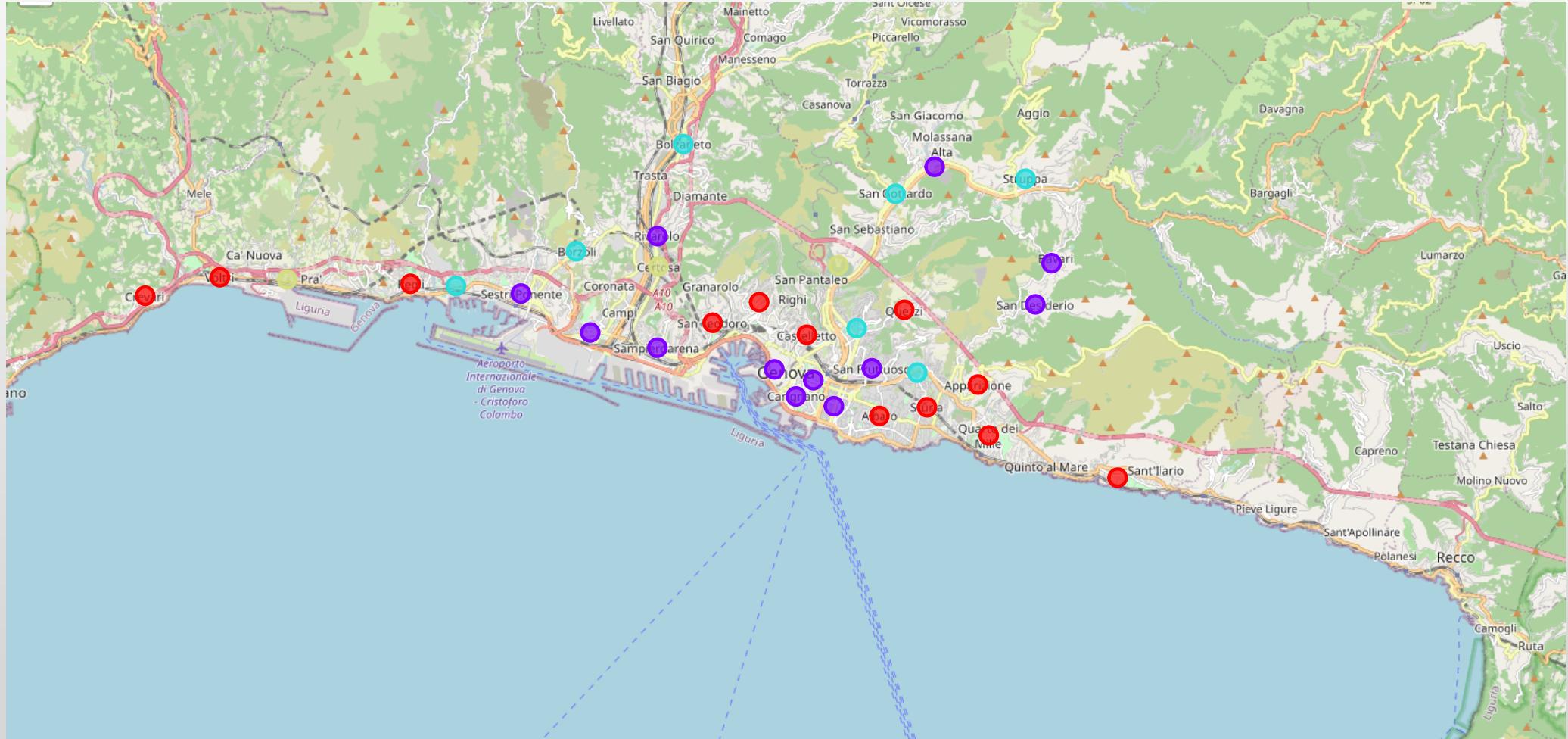
The algorithm used is **k-nearest neighbors algorithm (k-NN)** is a non-parametric classification method. The k-nearest neighbors algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems.

After inspecting the dataset several k segmentation values were tested.  
An optimal segmentation that yields a sensitive result is k=4

Data was segmented by using all the venues which reflect the character of each Neighborhood

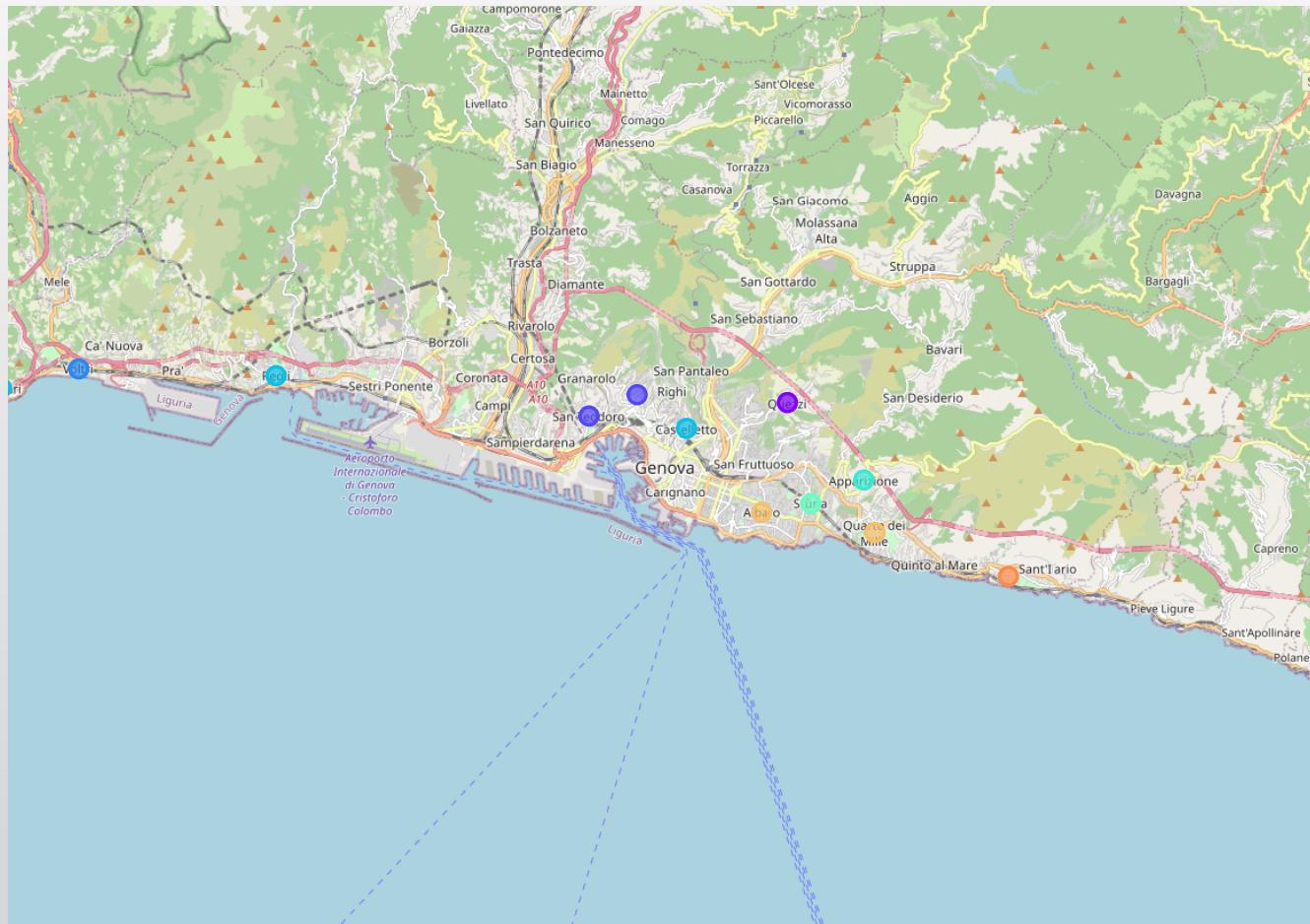
**A full description of the data and methodology is included in the final report**

# MAP OF NEIGHBORHOODS AFTER CLUSTERING



Map of Genova with Clusters plotted

# DATA ANALYSIS

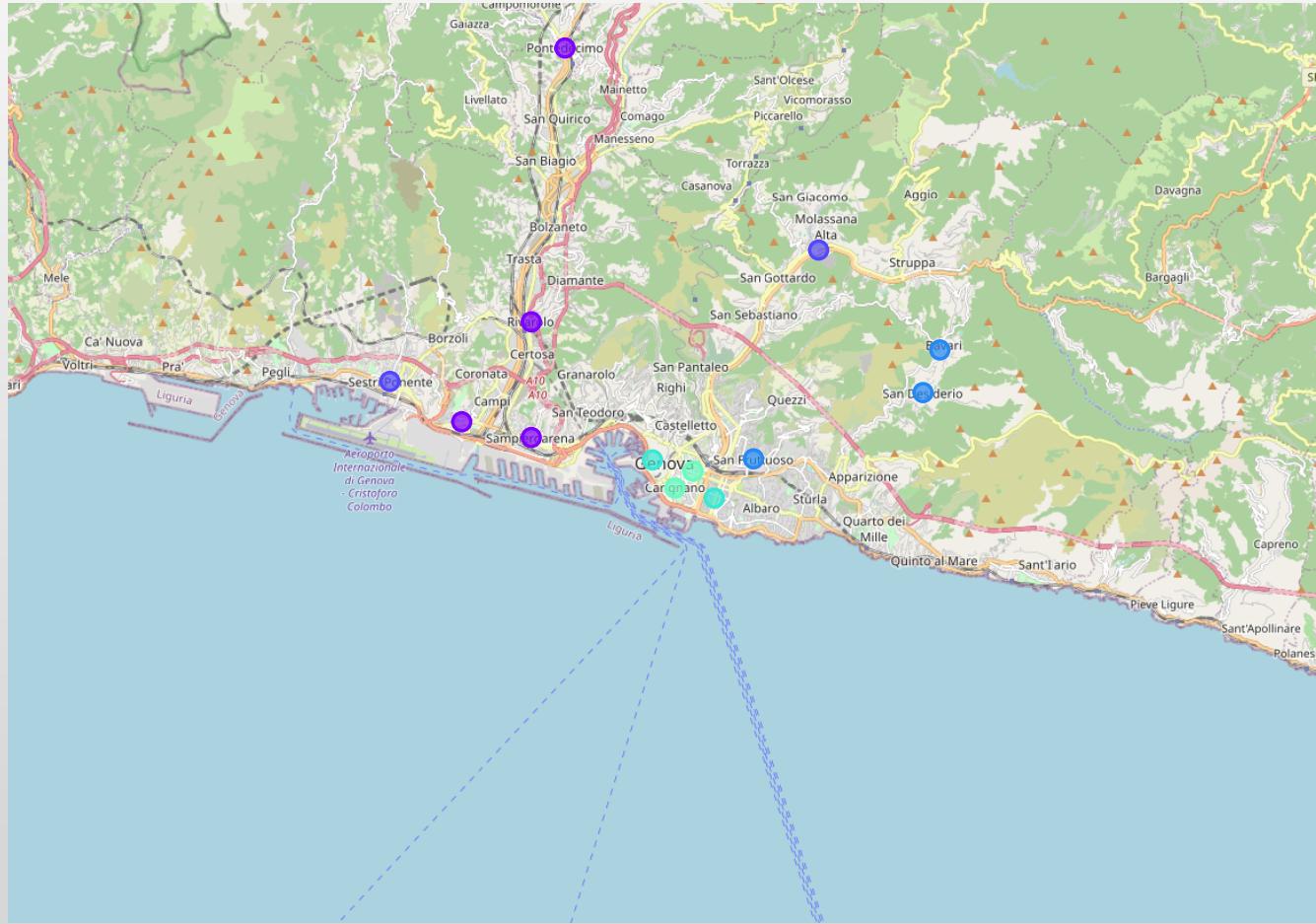


Cluster 1

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.

# DATA ANALYSIS



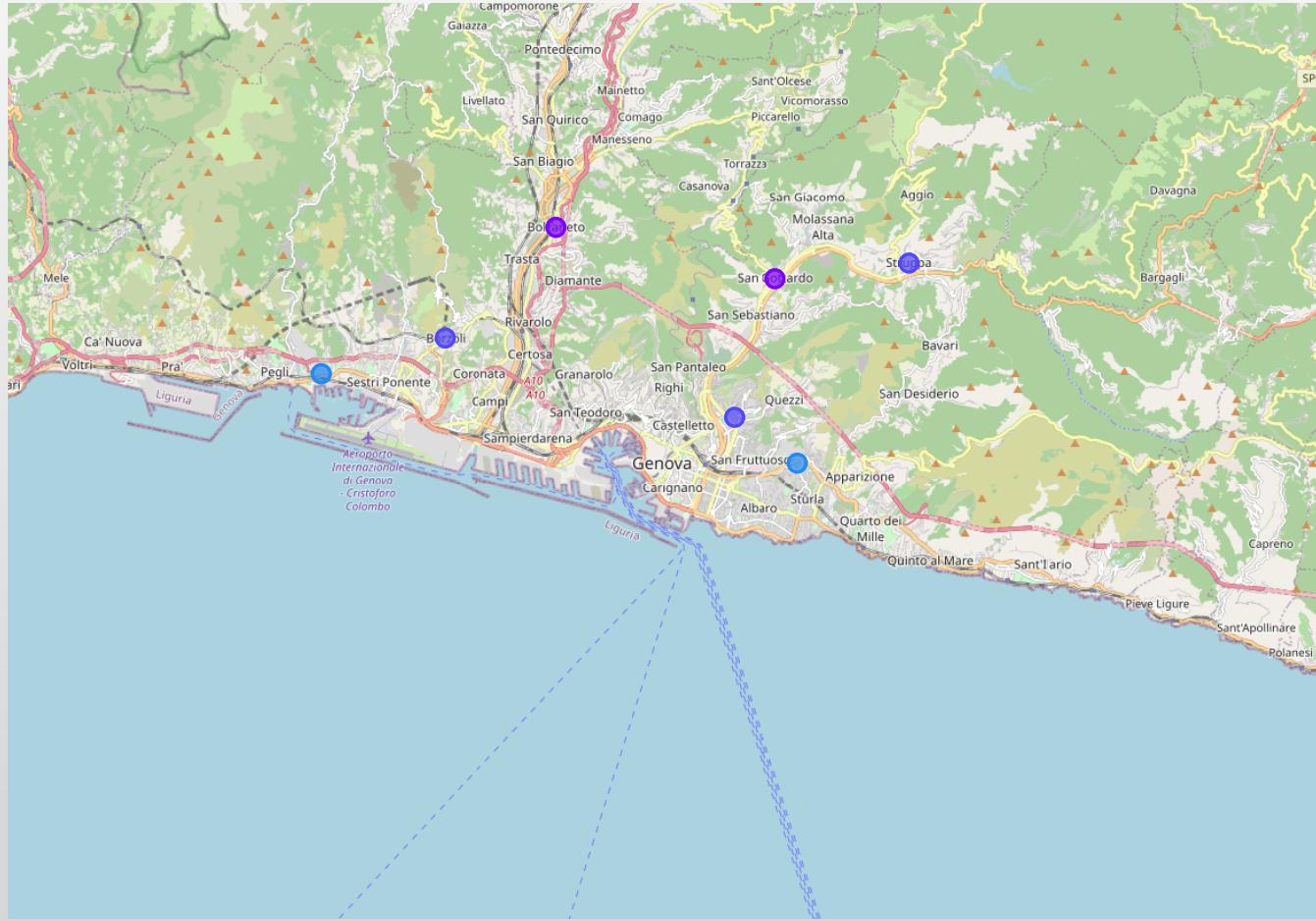
Cluster 2

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.

# DATA ANALYSIS

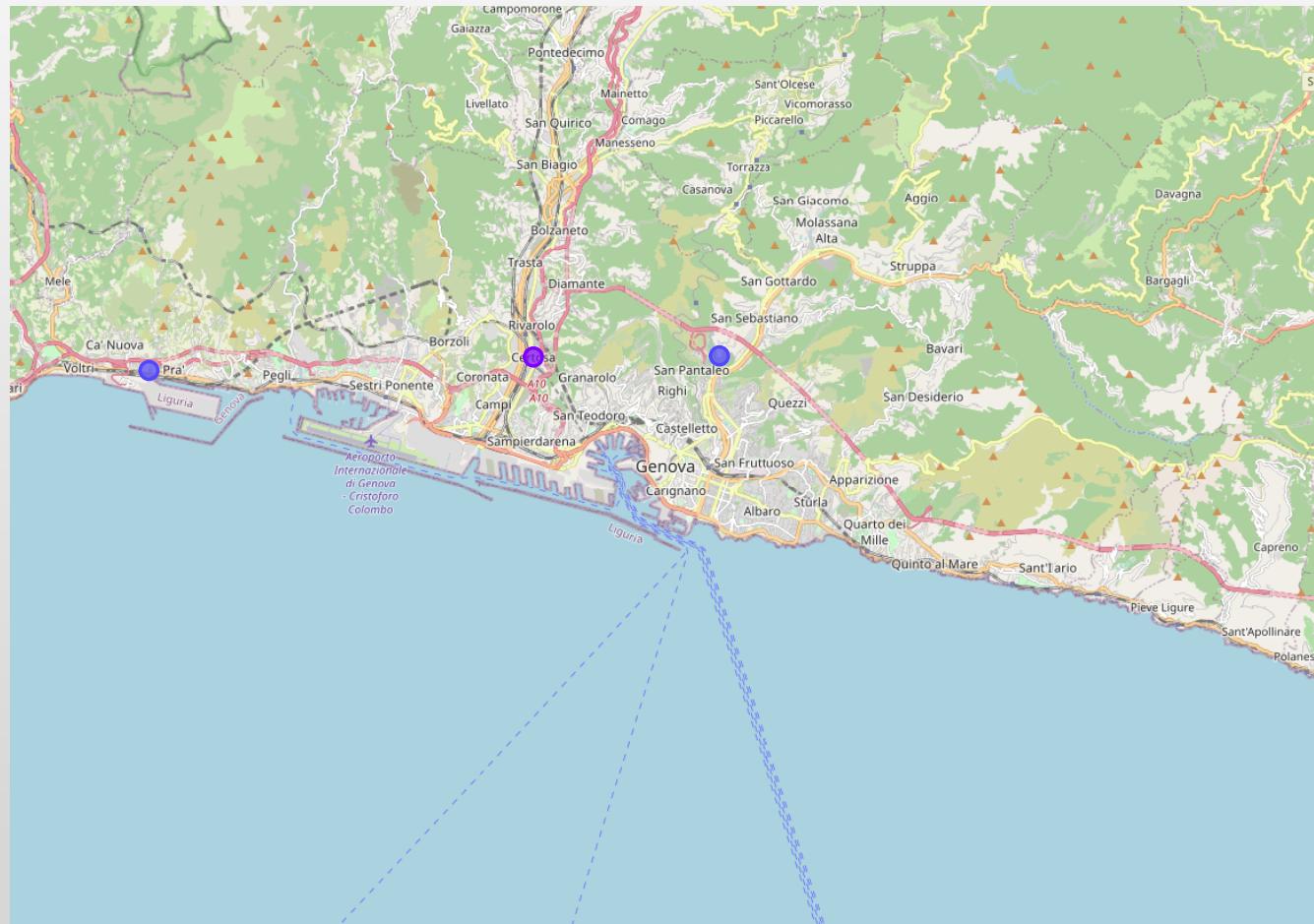


Cluster 3

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.

# DATA ANALYSIS



Cluster 4

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

This characteristic makes it not suitable for relocation.

# CONCLUSIONS

By analyzing the data we can see that cluster 1 is the most interesting and most of neighborhood sits next to the seaside.

However the housing price can be very high and the crowd city center is avoided.

So two interesting areas are identified for relocation.

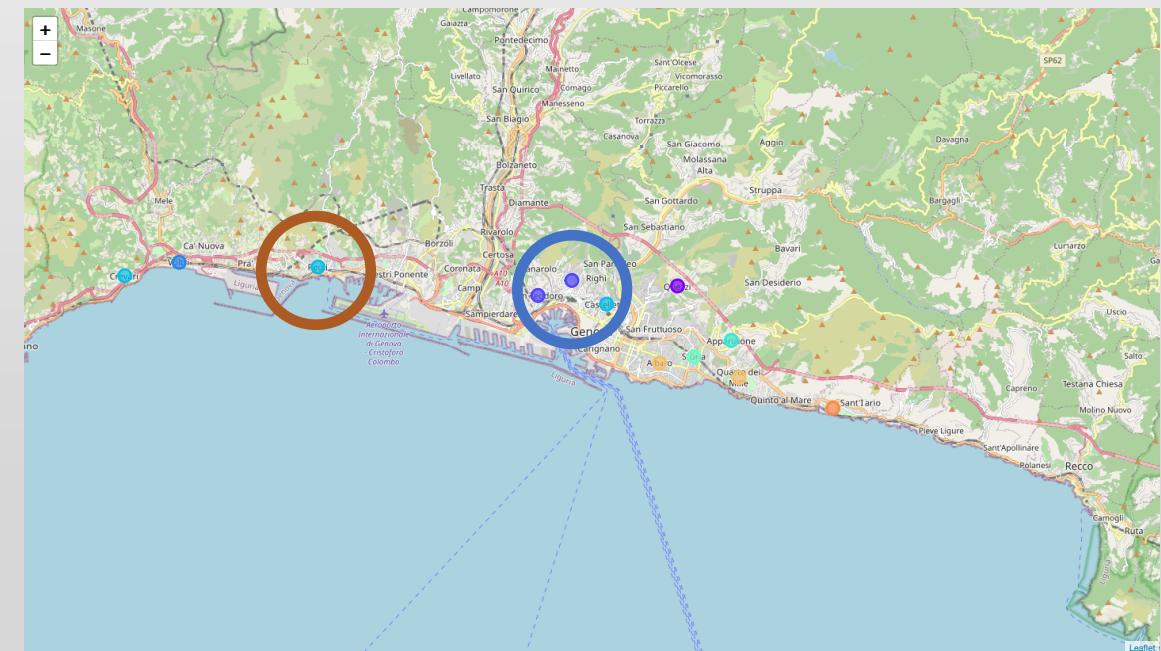
First area of interest is delimited by the blue circle (Oregina Neighborhood).

This area is very interesting: it is close to the city center but a bit away from the main Beaches. Housing price is reasonable.

The second area of interest is marked in the orange circle (Pegli Neighborhood).

It sits outside the main harbor area and is on the seaside. Housing price is far more affordable (1900 Euro/m<sup>2</sup>).

This is our main choice



Map pf Cluster 1 Neighborhoods

# IMPROVEMENTS AND CONSIDERATIONS

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 random positioned.

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