OPENING A NEW BUSINESS IN GENOA

A DATA CENTRIC APPROACH

Ayse Burcu Özkaptan

May 18, 2021

Table of Contents

Table of Figures

1 INTRODUCTION

1.1 Problem Description and Background Information

1.1.1 Problem and Objective

Whether one is new to the city or not, deciding on a business is no easy task. This decision involves careful analysis of the city, it's surroundings, real estate prices and all the business around the city such the restaurants, bars, bakeries, bookstores, hotels, sports halls, shopping venues, in short, anywhere that has foot traffic.

The main challenge is to pick the right spot in a suitable neighborhood with the highest possible foot traffic, preferably close to the business districts as the tourism and small businesses were hit hard by COVID. Picking the place with the highest foot traffic is particularly important due to the COVID restrictions that enforces the clients to have take-aways only.

The methodology is to leverage the Foursquare location data to explore or compare neighborhoods in the city of Genoa/Italy and use the Foursquare location data followed by K-Means Clustering Algorithm to solve a business problem of opening a new business.

Acquiring and analyzing data is necessary but not sufficient for each business decision. Picturing the problem and physical entities and using imagination by combining the data at hand gives better a understanding of the problem. In our case, we can ask 2 questions:

- "If we would like to open a business in district x, which type of business should I choose?"
- "If we would like to open a y-type of business, where would be the best place?"

For Example, if in a particular district there are too many sports venues and some ice-cream shops, it might be a good idea to open up a restaurant or café that serves healthy snacks, such as a salad bar or sandwich joint.

Using a similar approach, if we would like to open up a salad bar, we can pick a district where salad bars can make profit together with other types of businesses.

1.1.2 The City of Genoa

Genoa 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 Genoa, which in 2015 became the Metropolitan City of Genoa, 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 in Italy and in the Mediterranean Sea and twelfth-busiest in the European Union.

Genoa was the capital of one of the most powerful maritime republics for over seven centuries, from the 11^{th} century to 1797. Particularly from the 12^{th} century to the 15^{th} 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.

The historical center, also known as old town, of Genoa is one of the largest and most-densely populated in Europe. Part of it was also inscribed on the World Heritage List (UNESCO) in 2006 as Genoa: Le Strade Nuove and the system of the Palazzi dei Rolli.

Genoa is also home to the University of Genoa, which has a history going back to the 15th century, when it was known as Genuense Athenaeum. The city's rich cultural history in art, music and cuisine allowed it to become the 2004 European Capital of Culture. It is the birthplace of Guglielmo Embriaco, Christopher Columbus, Andrea Doria, Niccolò Paganini, Giuseppe Mazzini, Renzo Piano and Grimaldo Canella, founder of the House of Grimaldi, among others.

Genoa, which forms the southern corner of the Milan-Turin-Genoa industrial triangle of Northwest Italy, is one of the country's major economic centers. A number of leading Italian companies are based in the city, including Fincantieri, Leonardo, Ansaldo Energia, Ansaldo STS, Edoardo Raffinerie Garrone, Piaggio Aerospace, Mediterranean Shipping Company and Costa Cruises.

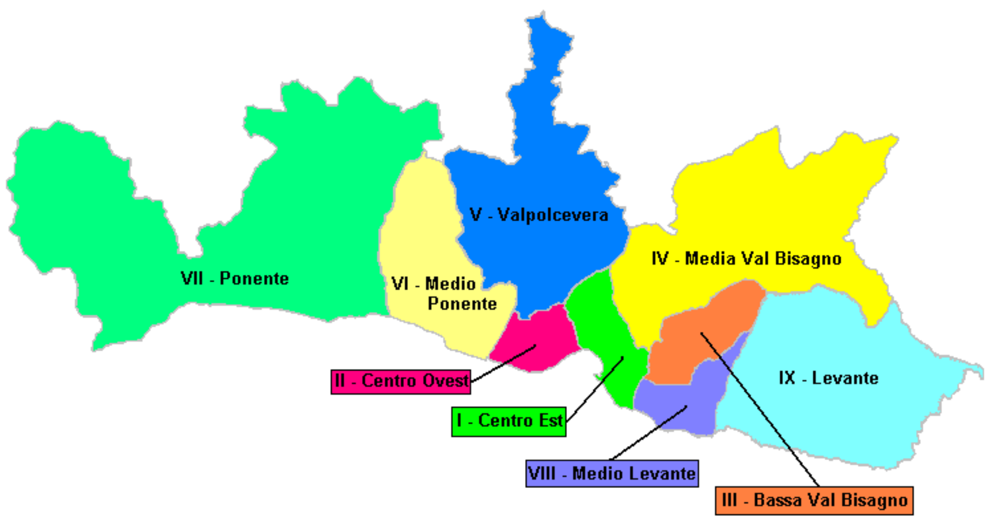


Figure 1: Map of Genoa and its Municipalities

1.2 Data

In order for us to perform the analysis for a given city, we needed

- The list of municipalities and neighborhoods together with the population data of each municipality (via web-scraping from Genoa's Wikipedia page)
- The coordinates of each neighborhood (via ARCGIS geocoder),
- The venues of each neighborhood that were queried from Foursquare API.
- The real-estate prices obtained from the Revenue Services (Agenzia Entrate)

The city of Genova is subdivided into 9 municipalities, as approved by the Municipal Council in 2007. The necessary data to obtain the municipalities and neighborhoods was **scraped** from https://en.wikipedia.org/wiki/Genoa#Municipal_government. Given that the table had some irregularities, data cleansing was required. All the information retrieved via web scraping was placed inside a Pandas dataframe called df.

After building a dataframe including the municipalities, population of each municipality and neighborhoods in each municipality; in order to utilize the Foursquare location data for obtaining the venues, we need to get the latitude and the longitude coordinates of each neighborhood. Leveraging the Google Maps Geocoding API to get the latitude and longitude was not possible as Google started charging for the API. Therefore, I used Geocoder Python package instead: https://geocoder.readthedocs.io/index.html.

In this package Nominatim API could provide a suitable free solution to our problem, but it was observed that it is not always possible to obtain the coordinates for each neighborhood. As a result, I chose to use **ArcGIS geocoder**, one of the most popular geocoding APIs with a high accuracy in obtaining the right coordinates. The latitudes and longitudes of all neighborhoods obtained by means of ArcGIS were placed inside the dataframe df.

The real estate prices (mean quotation price / m^2 for year 2019) were retrieved from page 14 of the Regional Statistics of Liguria pdf file and combined with other information scraped from the Wikipedia as a csv file.

	Municipality	Neighbourhoods	Population	% of Total Population	Mean Quotation €/m2 2019
0	Centro-Est	Prè, Molo, Maddalena, Oregina, Lagaccio, San N...	91402	15.0	1895
1	Centro-Ovest	Sampierdarena, Belvedere, Campasso, San Bartol...	66626	10.9	1163
2	Bassa Val Bisagno	San Fruttuoso, Sant'Agata, Marassi, Quezzi, Fe...	78791	12.9	1369
3	Media Val Bisagno	Staglieno (Parenzo, San Pantaleo), Molassana, ...	58742	9.6	1430
4	Valpolcevera	Rivarolo, Borzoli Est, Certosa, Teglia, Begato...	62492	10.3	1127
5	Medio Ponente	Sestri, Borzoli Ovest, San Giovanni Battista, ...	61810	10.1	2059
6	Ponente	Voltri, Crevari, Pra', Palmaro, Ca' Nuova, Peg...	63027	10.3	2028
7	Medio Levante	Foce, Brignole, San Martino, Chiappeto, Albaro...	61759	10.1	2680
8	Levante	Sturla, Quarto, Quartara, Castagna, Quinto al ...	66155	10.8	3762

Figure 2: Municipalities, Neighborhoods, Population and Real Estate Prices in Genova

The mean quotation prices were then added to our main dataframe df.

- Data Cleansing:

After taking a quick look at the geolocation data, it has been observed that there were some mismatches due to some name similarities of the neighborhoods having multiple occurrences in Genova. The wrong information was corrected by hand within the context of data cleansing.

```
df.loc[df.Neighborhood == 'Belvedere', 'Latitude'] = 44.41278
df.loc[df.Neighborhood == 'Belvedere', 'Longitude'] = 8.93302

df.loc[df.Neighborhood == 'Molo', 'Latitude'] = 44.4067775
df.loc[df.Neighborhood == 'Molo', 'Longitude'] = 8.9280807
```

Figure 3: Some examples of data cleansing by hand to correct the location data.

	Municipality	Population	% of Total Population	Neighborhood	Mean Quotation €/m2 2019	Latitude	Longitude
0	Centro-Est	91,402	15.0%	Prè	1895	44.414615	8.924991
1	Centro-Est	91,402	15.0%	Molo	1895	44.406777	8.928081
2	Centro-Est	91,402	15.0%	Maddalena	1895	44.410290	8.933340
3	Centro-Est	91,402	15.0%	Oregina	1895	44.420400	8.922152

Figure 4: Our Main Dataframe df Before Exploratory Data Analyses.

2 METHODOLOGY

2.1 Exploring the Neighborhoods in Genoa

Genova has 9 municipalities and 69 neighborhoods. The number of the neighborhoods in each municipality is as follows:

Levante	11
Centro-Est	10
Valpolcevera	9
Ponente	8
Medio Levante	8
Centro-Ovest	6
Medio Ponente	6
Bassa Val Bisagno	6
Media Val Bisagno	5

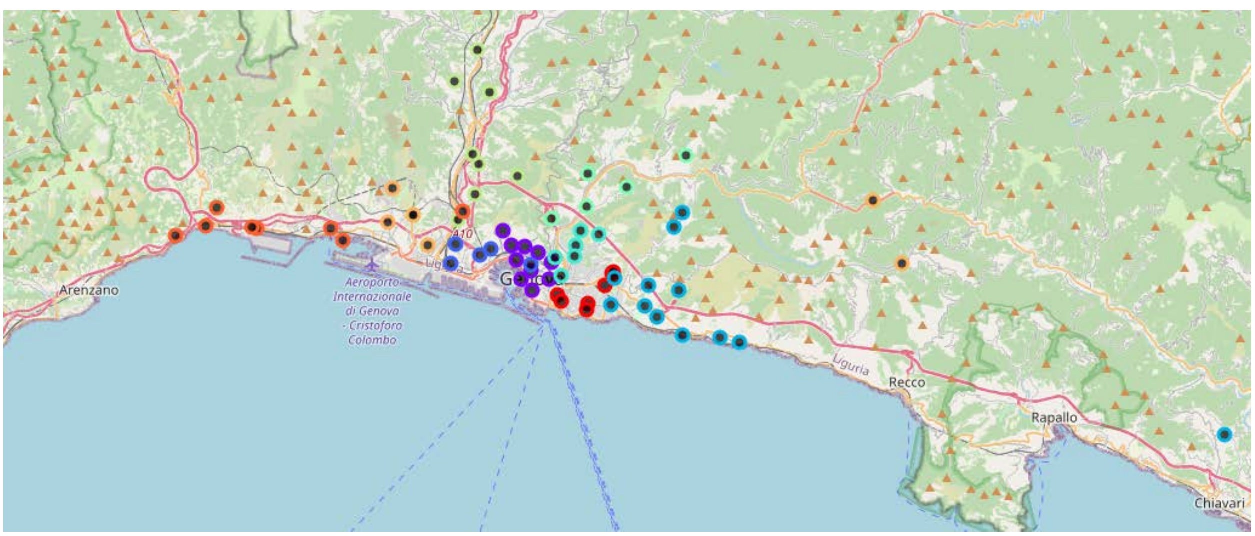


Figure 5: Neighborhoods in Genoa. Each neighborhood is color-coded according to the municipality it is belonging to.

Next step is to retrieve and explore the venues from all the neighborhoods in Genoa. I created a function called `getNearbyVenues(mun_names, ngh_names, latitudes, longitudes, radius=500)` for doing the trick which takes the following input parameters:

- mun_names: Names of the municipalities
- ngh_names: Names of the neighborhoods

- latitudes: Latitudes of the neighborhoods,
- longitudes: Longitudes of the neighborhoods
- radius: Radius of search. A default value of 500 meters was used.

```python
def getNearbyVenues(mun_names, ngh_names, latitudes, longitudes, radius=500):

    venues_list=[]
    for mun, ngh, lat, lng in zip(mun_names, ngh_names, latitudes, longitudes):
        # print(name)

        # create the API request URL
        url = ('https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}').format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            mun,
            ngh,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Municipality',
                  'Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

```python
genova_venues = getNearbyVenues(df['Municipality'], df['Neighborhood'], df['Latitude'], df['Longitude'])
```

Figure 6: getNearbyVenues Function and the Function Call for Retrieving the Venues from Foursquare API.

Our main dataframe df had all the latitude and longitude data to make Foursquare queries for each of the neighborhoods. Our dataframe df has 69 observations, i.e 69 neighborhoods, so we make 69 calls to Foursquare in the for loop and parse the json results including with all the venues in that neighborhood into `venues_list` list containing the name, latitude, longitude and category of each venue. After populating this list with all the venues in all neighborhoods, we place them inside a Pandas dataframe together with the municipality and neighborhood names, neighborhood latitude and longitudes where the venue is located. As a result, the returned dataframe would have municipality name, neighborhood name, neighborhood latitude and longitude, venue name, venue latitude and longitude and venue category columns. In short, the call of `getNearbyVenues` function returns us all the venues in Genoa.

	Municipality	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Centro-Est	Prè	44.414615	8.924991	Trattoria dell'Acciughetta	44.414603	8.924658	Trattoria/Osteria
1	Centro-Est	Prè	44.414615	8.924991	Sommergibile Nazario Sauro	44.414077	8.924242	Boat or Ferry
2	Centro-Est	Prè	44.414615	8.924991	I 2 Truogoli	44.415413	8.924814	Diner
3	Centro-Est	Prè	44.414615	8.924991	Galata Museo del Mare	44.414149	8.923971	Science Museum
4	Centro-Est	Prè	44.414615	8.924991	Ristorante Beluga	44.414449	8.922484	Seafood Restaurant

Figure 7: First Few Rows of the Dataframe genova_venues Returned by getNearbyVenues Funcion.

By using the Foursquare API to retrieve data from the Foursquare database, the venues around points of interests, being the neighborhoods, were retrieved and according to this information there are;

- 775 venues
- 143 unique venue categories

Let's elaborate on the number of venues to get some understanding about the city, and possibly about the foot traffic in the next 2 bar plots.

The 25 most popular venue categories in the city can be summarized with the following bar plot:

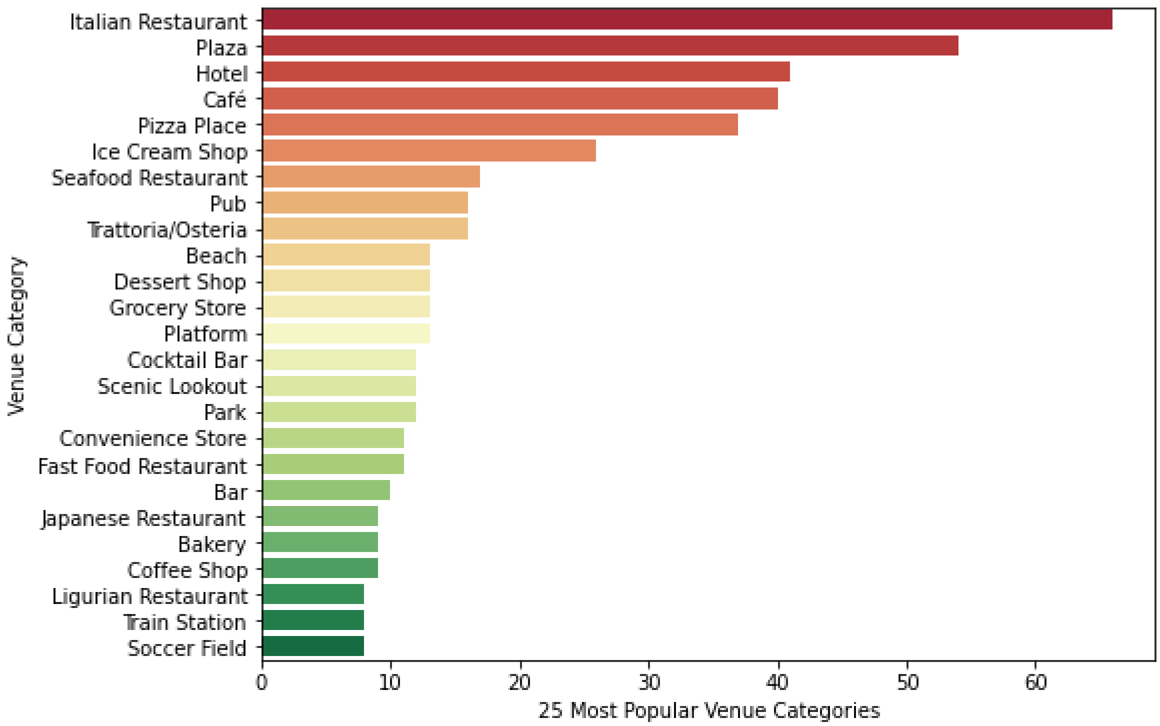


Figure 8: 25 Most Popular Venue Categories in Genoa

Here are the numbers of venues in each municipality. According to this plot, the majority of the venues are situated from the coastal center to the east coast of the city.

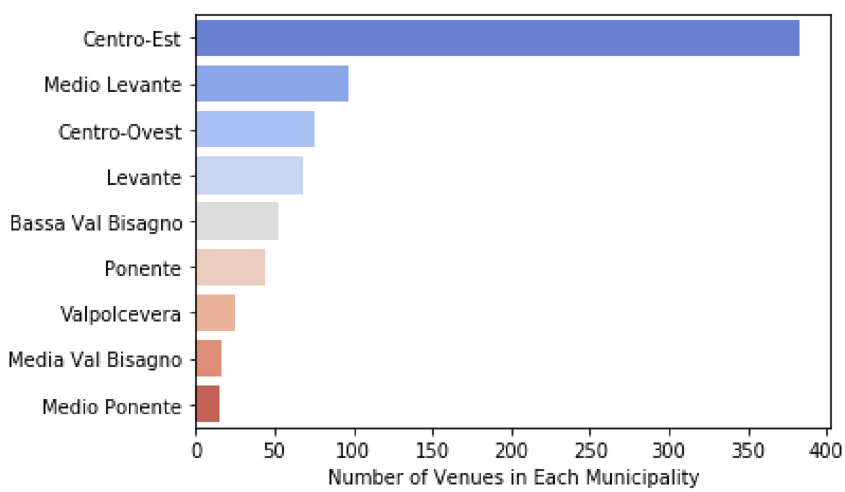
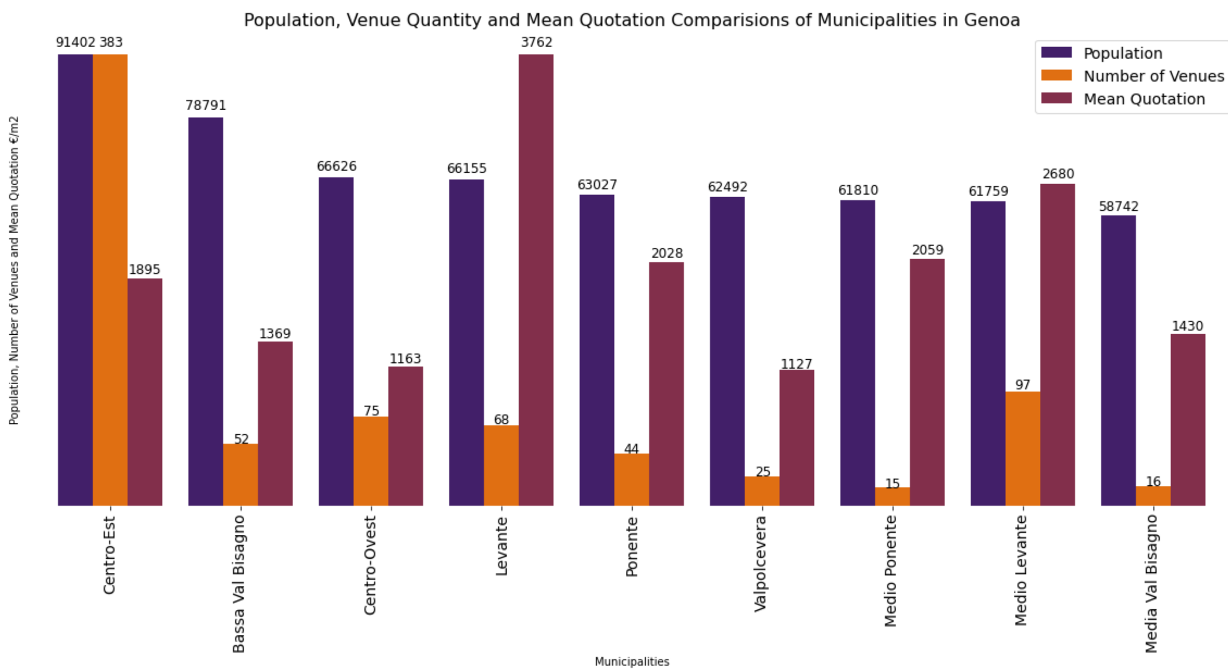


Figure 9: Number of Venues in Each Municipality

Now that we have the number of venues in each municipality, we can have a brief comparative summary of the municipalities:

	Municipality	Neighbourhoods	Population	% of Total Population	Mean Quotation €/m2 2019	Number of Venues
0	Centro-Est	Prè, Molo, Maddalena, Oregina, Lagaccio, San N...	91402	15.0	1895	383
1	Centro-Ovest	Sampierdarena, Belvedere, Campasso, San Bartol...	66626	10.9	1163	75
2	Bassa Val Bisagno	San Fruttuoso, Sant'Agata, Marassi, Quezzi, Fe...	78791	12.9	1369	52
3	Media Val Bisagno	Staglieno (Parenzo, San Pantaleo), Molassana, ...	58742	9.6	1430	16
4	Valpolcevera	Rivarolo, Borzoli Est, Certosa, Teglia, Begato...	62492	10.3	1127	25
5	Medio Ponente	Sestri, Borzoli Ovest, San Giovanni Battista, ...	61810	10.1	2059	15
6	Ponente	Voltri, Crevari, Pra', Palmaro, Ca' Nuova, Peg...	63027	10.3	2028	44
7	Medio Levante	Foce, Brignole, San Martino, Chiappeto, Albaro...	61759	10.1	2680	97
8	Levante	Sturla, Quarto, Quartara, Castagna, Quinto al ...	66155	10.8	3762	68

Figure 10: Populations, Mean Quotation Prices, Number of Venues for Each Municipality in Genoa.



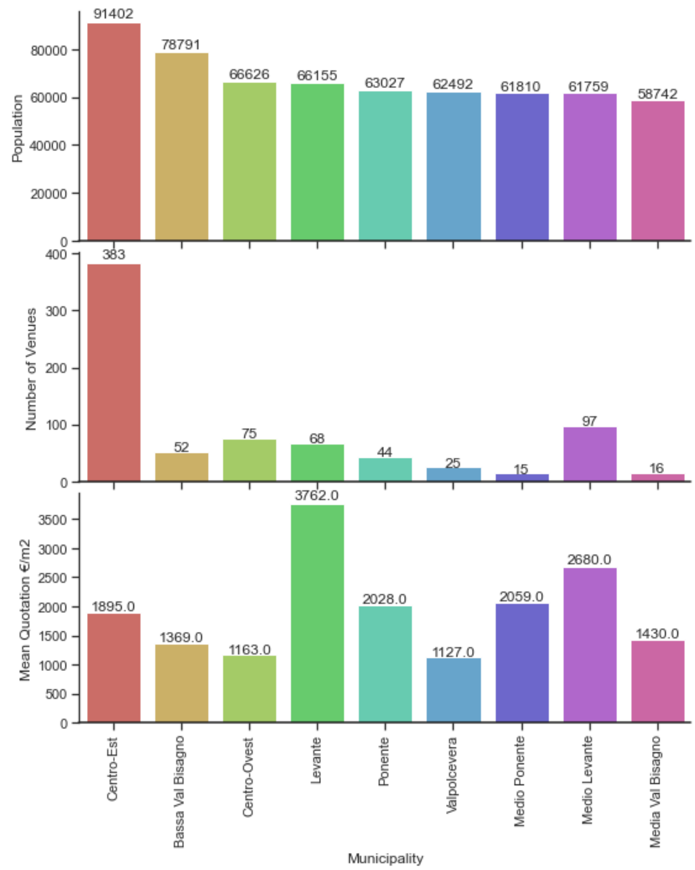


Figure 11: Population, Venue Quantity and Mean Quotation Comparisons of Municipalities in Genoa.

Going one step further, we will now take a look at the distribution of the venue categories with respect to municipalities and neighborhoods, both in tabular form and in Sankey diagram.

Municipality	Neighborhood	Venue
Bassa Val Bisagno	Fereggiano	13
	Forte Quezzi	1
	Marassi	4
	Quezzi	4
	San Fruttuoso	13
	Sant'Agata	17
Centro-Est	Carignano	33
	Castelletto	5
	Lagaccio	6
	Maddalena	93
	Manin	8
	Molo	98
	Oregina	21
	Prè	44
	San Nicola	5
	San Vincenzo	70
Centro-Ovest	Angeli	8
	Belvedere	41
	Campasso	7
	Sampierdarena	13
	San Teodoro	6
Levante	Apparizione	3
	Bavari	3
	Borgoratti	4
	Nervi	12
	Quartara	6
	Quarto	8
	Quinto al Mare	7
	San Bartolomeo	8
	San Desiderio	3
	Sant'Ilario	5
	Sturla	9
Media Val Bisagno	Molassana	4
	Montesignano	6
	Sant'Eusebio	4
	Staglieno	2
Medio Levante	Albaro	10
	Chiappeto	5
	Foce	27
	Lido	21
	Puggia	5
	San Giuliano	22
	San Martino	7
Medio Ponente	Borzoli Ovest	3
	Calcinara,	3
	Campi	1
	Cornigliano	2
	San Giovanni Battista	5
	Sestri	1
Ponente	Ca' Nuova	1
	Castelluccio	2
	Crevari	3
	Multedo	4
	Palmaro	4
	Pegli	12
	Pra'	5
	Voltri	13
Valpolcevera	Bolzaneto	4
	Borzoli Est	3
	Certosa	4
	Pontedecimo	5
	Rivarolo	4
	Teglia	5

Figure 12: Distribution of the Venue Categories with respect to Municipalities and Neighborhoods (Tabular Form).

Even though the following Sankey Diagram is a bit complicated, it tells us how diverse the venues are and how the diversity varies from one municipality to another.

Despite the “trending” endpoint of Foursquare was not producing any results as it is no longer in use, we still can have a rough idea about the foot traffic judging by the number and types of the venues in a municipality.

Centro-Est, Medio Levante, Levante and Centro-Ovest are the highest-ranking contenders.

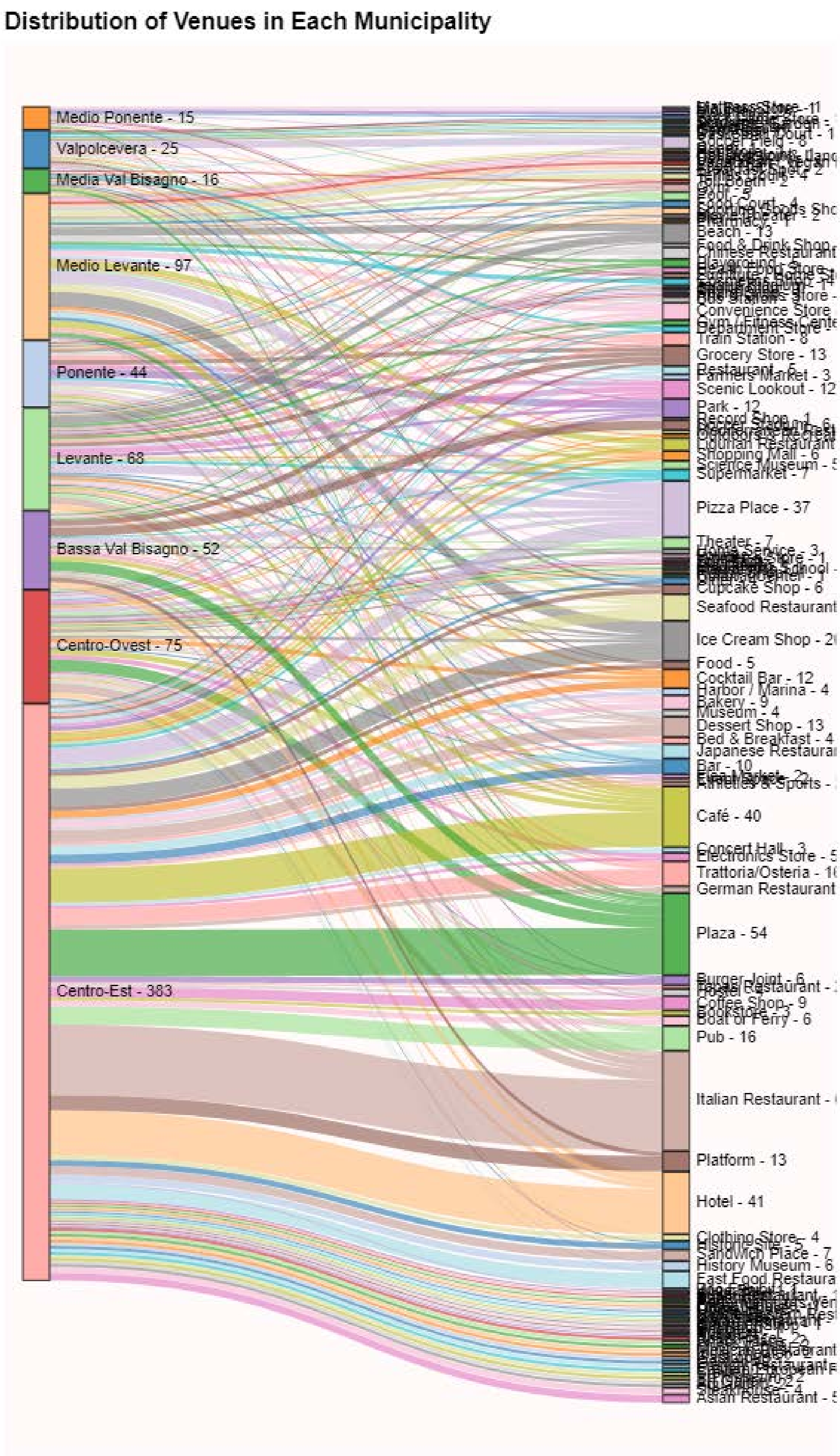


Figure 13: Distribution of Venue Categories in each Municipality

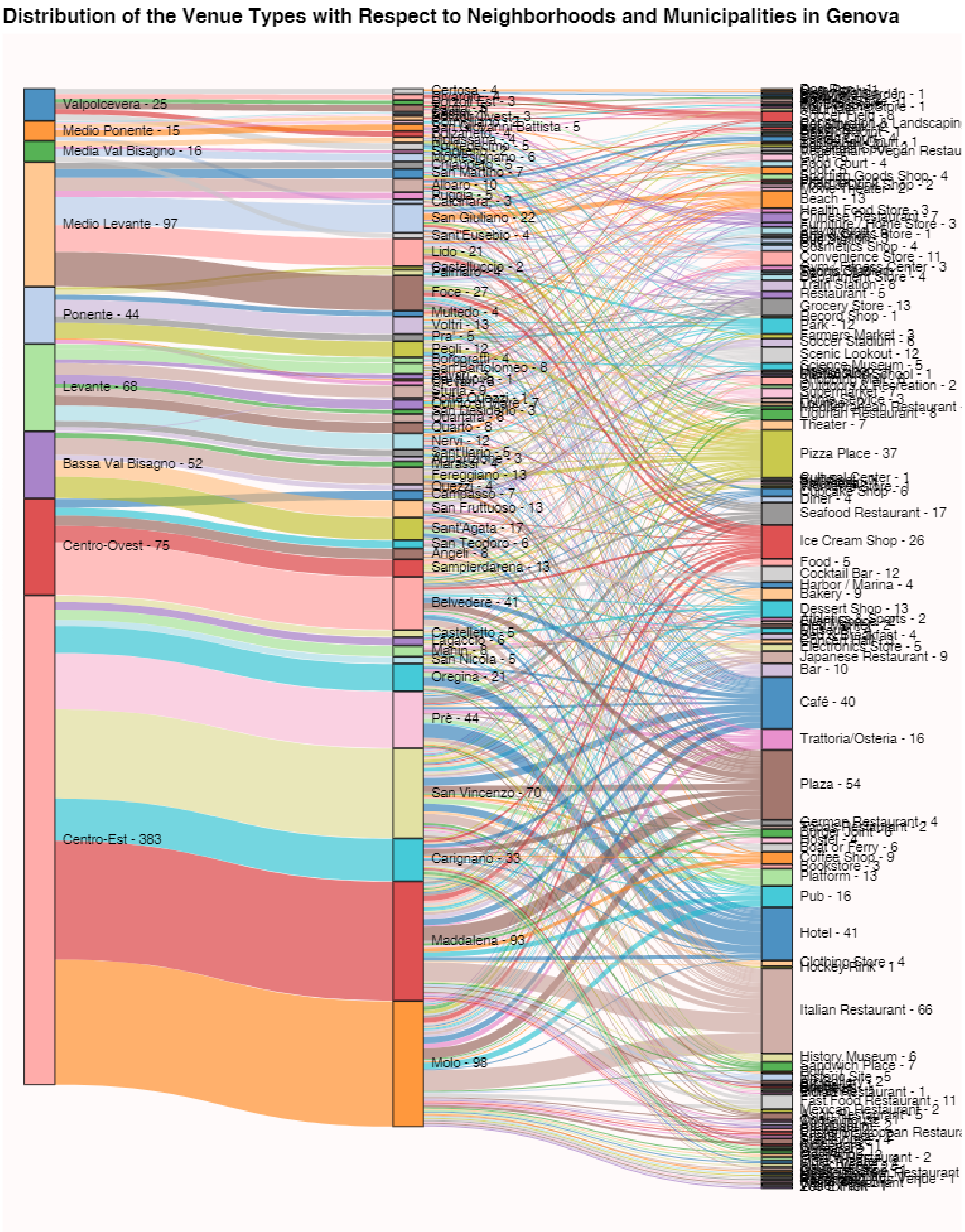


Figure 14: Distribution of the Venue Categories with respect to Municipalities and Neighborhoods (Sankey Diagram).

2.2 Analysis of Each Neighborhood

2.2.1 Feature Engineering

In order for us to arrive at a solution for our business problem, we need to divide the city into clusters by using K-Means Algorithm. In order for us to do that, we need to decide on the features of our dataset. A meaningful choice would be to characterize each neighborhood according to the density of each venue category by using one-hot encoding.

We work on the venue category of `genova_venues` dataframe and have the 143 venue categories converted into one-hot encoding for all 775 venues. The resulting dataframe genova_onehot, then has a shape of 775 x 144 (143 categories + 1 column for the neighborhood name).

	Neighborhood	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Bakery	Bar	Basketball Court	Beach	Beach Bar	Bed & Breakfast	Big Box Store	Bistro	Board Shop	Boat or Ferry	Bookstore	Boutique	Breakfast Spot	Brewery	Bridge
0	Prè	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Prè	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
2	Prè	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Prè	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Prè	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 15: Part of the Dataframe After One-hot Encoding.

Next, I grouped the rows by neighborhood and by taking the mean of the frequency of occurrence of each category and assigned the result to `genova_grouped`. What this actually does is that, for each neighborhood, we divide the number of venues in each category to the total number of venues in that neighborhood. For example, there is 1 bar in Albaro out of 10 venues. Hence, the frequency of it in the table below is 1/10 = 0.1. This dataframe has 63 rows as opposed to 69, meaning that 6 out of 69 neighborhoods do not have any venues at all.

	Neighborhood	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Bakery	Bar	Basketball Court	Beach	Beach Bar	Bed & Breakfast	Big Box Store	Bistro	Board Shop	Boat or Ferry	Bookstore	Boutique	Breakfast Spot
0	Albaro	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.100000	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.000000	0.000000	0.000000
1	Angeli	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.125000	0.000000	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.125000	0.000000	0.000000	0.000000
2	Apparizione	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.000000	0.000000	0.000000
3	Bavari	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.000000	0.000000	0.000000

Figure 16: Part of the Dataframe genova_grouped. Feature Engineering. The neighborhoods are grouped and the meain of the frequency of occurence of each category is taken.

2.2.2 Getting to Know the Neighborhoods – Top 10 Venue Categories

In order to have an idea about the neighborhoods, the top 10 venue categories of the neighborhoods can be examined. I first wrote a function called `return_most_common_venues(row, num_top_venues)` which takes the following parameters as inputs:

- The row from the dataframe `genova_grouped`
- The number of top venue categories we wish to see in ranking.

```python
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

Figure 17: return_most_common_venues Function

Using the `return_most_common_venues` function, we create a new dataframe called `neighborhoods_venues_sorted` and display the top 10 venues for each neighborhood. This representation gives us a good idea about the general preferences of the people living in each neighborhood as these top 10 businesses were able to thrive together. This also acts as a precursor to clustering as to which categories go along with other categories.

2.3 Clustering

As the final step of the methodology, K-means clustering was used. I ran K-means algorithm for cluster sizes ranging from 2 to 50 and plotted the sum of square errors as performance metric vs cluster size to pick the right cluster size for the analysis.

```python
from sklearn.metrics import silhouette_samples, silhouette_score

# set the maximum number of clusters
max_kclusters = 50

# performance metrics
SSE = [] #sum of squared error
no_of_clusters = []

# Drop the column with neighborhood names as it cannot be used in clustering.
genova_grouped_clustering = genova_grouped.drop('Neighborhood', 1)

for kclusters in range(2, max_kclusters+1):
    # run k-means clustering
    kmeans  = KMeans(init = "k-means++", n_clusters=kclusters, random_state=0).fit(genova_grouped_clustering)

    SSE.append(kmeans.inertia_)
    no_of_clusters.append(kclusters)

plt.plot(no_of_clusters, SSE)
plt.xlabel('Number of Clusters')
plt.ylabel('Sum of Squared Error')
```

Figure 18: Running the K-Means Algorithm for Different Cluster Sizes.

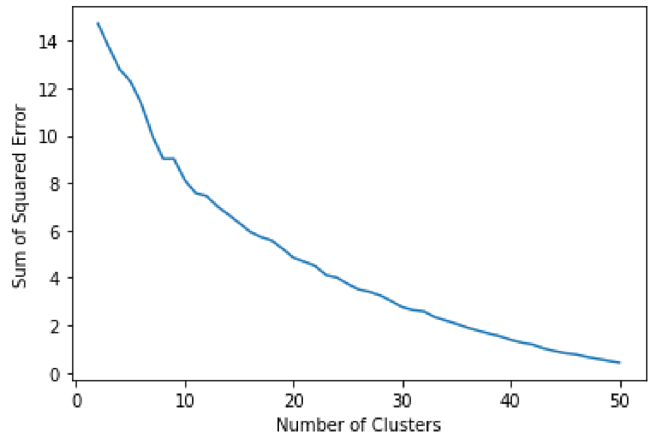


Figure 19: Choosing the Right Cluster Size: Sum of Squred Error vs Cluster Sizes.

Looking at the sum of squared error of K-Means clustering, it is observed that there is not a clear elbow point. Given that there are 63 neighborhoods used in the analyses, choosing a high cluster size would obviously reduce the error, since many of the clusters would have only 1 member. However, this makes the idea of clustering a bit useless in a case where the number of clusters is close to the number of neighborhoods. For this reason, we should use Silhouette Score to measure the goodness of our clustering.

The Silhouette Coefficient is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample. The Silhouette Coefficient for a sample is (b - a) / max(a, b). To clarify, b is the distance between a sample and the nearest cluster that the sample is not a part of. Note that Silhouette Coefficient is only defined if number of labels is 2 <= n_labels <= n_samples - 1.

The best value is 1 and the worst value is -1. Values near 0 indicate overlapping clusters. Negative values generally indicate that a sample has been assigned to the wrong cluster, as a different cluster is more similar. The Silhouette Coefficient is generally higher for convex clusters than other concepts of clusters, such as density-based clusters like those obtained through DBSCAN.

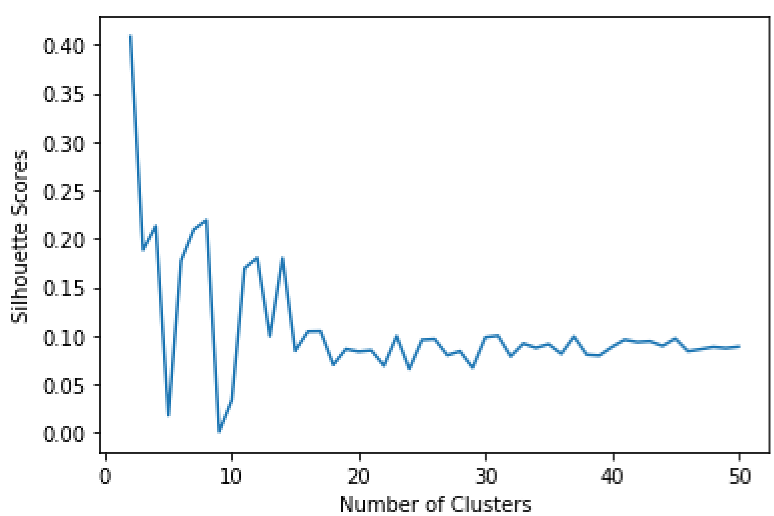


Figure 20: Silhoutte Scores vs Number of Clusters

Selecting the number of clusters as 8 would be a good choice in our case. After running the algorithm once more, this time for a cluster size of 8, we get 63 labels for the clusters. Next, we merge the labels with the dataframe `neighborhoods_venues_sorted` and then remove the rows where the label is NaN signifying that the neighborhood does not have any venues and was not considered in clustering. In other words, the neighborhoods that do not have any venues do not belong to any cluster, resulting in NaN as cluster values. Those neighborhoods should be removed as they are not useful in our analyses.

Finally, the resulting clusters were visualized together with real estate prices (mean quotation €/m2) overlayed on each municipality. In order to visualize the municipalities, a bit of an unorthodox approach was exercised since there was no data available for plotting the municipality boundaries of Genoa. From the website of Genoa's city hall, the geojson file of the historical areas (http://dati.comune.genova.it/dataset/zone-storiche-di-genova) was downloaded. After editing the geojson file to match today's municipalities, we now have a good approximation of today's municipality boundaries. The boundaries of the "historical areas" trace the limits of the Municipalities and Sestieri of Genoa existing before the construction of the Grande Genova, resulting from the annexes of 1874 and 1926. However, they have been redesigned as an aggregation of census sections, therefore there may be slight differences with respect to the real borders of the previous municipalities.

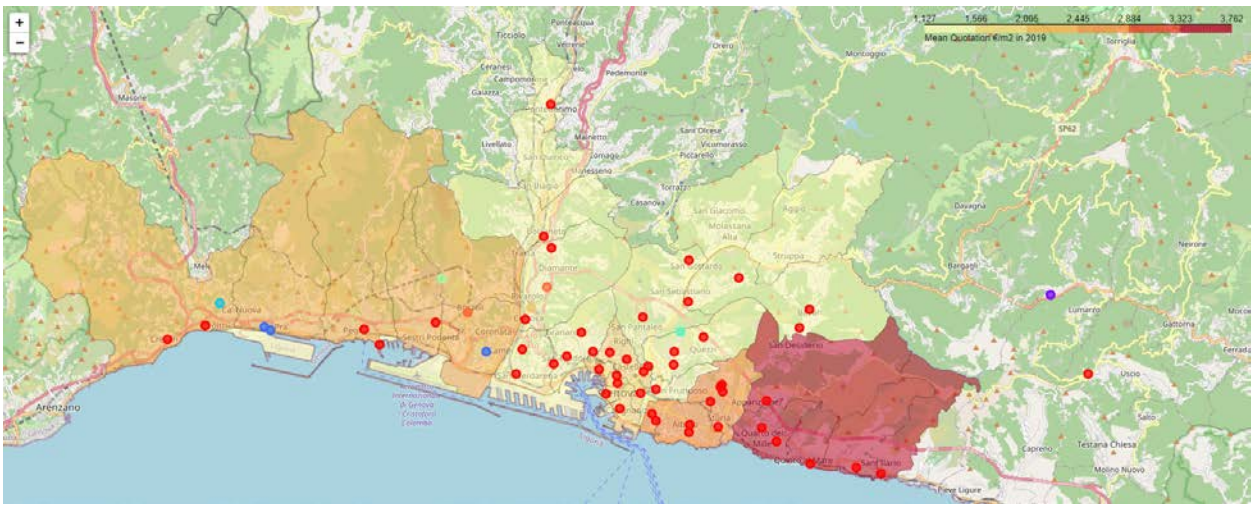


Figure 21: Resulting Clusters with Real Estate Prices Overlayed on Municipalities.

As another clustering algorithm, DBSCAN was also tried to see if the resulting clusters would make more sense. After tweaking the epsilon and number of minimum samples, it can be concluded that DBSCAN does not give promising results due to high number of features but low number of samples. The algorithm mainly picked 1 cluster with many neighborhoods and too much noise, rendering K-Means as a more suitable choice for our application.

```python
epsilon = 0.6
minimumSamples = 4
db = DBSCAN(eps=epsilon, min_samples=minimumSamples).fit(genova_grouped_clustering)
labels = db.labels_
labels
```

```
array([ 0,  0,  0,  0,  0,  0,  0, -1, -1, -1,  0,  0, -1,  0,  0, -1,  0,
        0, -1,  0,  0,  0, -1,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0, -1,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0, -1, -1,  0,  0,  0], dtype=int64)
```

3 RESULTS

Based on the defining categories, we can have an idea about the anatomy of the cluster and see which type of venues go along with each other.

1. Cluster 1: The first cluster is the most crowded and diverse cluster obtained in the results spanning all 9 municipalities. Centro-Est is the place where the majority of the population is living and also the most touristic and densely populated area by the sea, hosting eateries, hotels and other touristic attractions. Since there are already too many Italian restaurants, pizza places and trattorias/osterias, opening a similar venue might not be the best choice due to competition. Opening a bar by the seaside offering breakfasts in the morning, and aperitivo in the evening to respond a wide range of needs for the whole day, street food or ice-cream joint or a high-end Asian restaurant can be promising businesses. Due to the lively ambient generated by tourism and business world, opening a bed-and-breakfast can also be a good idea. Given that the area is densely packed, presumably with a considerable amount of foot traffic, even a clothing or shoe store is a good idea.

	Municipality	Population	% of Total Population	Neighborhood	Mean Quotation €/m2 2019	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	10th Most Common Venue
0	Centro-Est	91,402	15.0%	Prè	1895	Hotel	Seafood Restaurant	Boat or Ferry	Italian Restaurant	Platform	Trattoria/Osteria	Café	History Museum	Lounge	Concert Hall
1	Centro-Est	91,402	15.0%	Molo	1895	Italian Restaurant	Plaza	Café	Pub	Trattoria/Osteria	Ice Cream Shop	Fast Food Restaurant	Pizza Place	Ligurian Restaurant	Steakhouse
2	Centro-Est	91,402	15.0%	Maddalena	1895	Italian Restaurant	Plaza	Café	Pub	Ice Cream Shop	Fast Food Restaurant	Bar	Hotel	Coffee Shop	Food
3	Centro-Est	91,402	15.0%	Oregina	1895	Hotel	Platform	Café	Trattoria/Osteria	Plaza	Italian Restaurant	Asian Restaurant	Train Station	Sandwich Place	Outdoors & Recreation
4	Centro-Est	91,402	15.0%	Lagaccio	1895	Concert Hall	Park	Furniture / Home Store	Fast Food Restaurant	Hockey Rink	Pizza Place	Pier	Performing Arts Venue	Pet Store	Pharmacy
5	Centro-Est	91,402	15.0%	San Nicola	1895	Pizza Place	History Museum	Italian Restaurant	Mexican Restaurant	Piadineria	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy
6	Centro-Est	91,402	15.0%	Castelletto	1895	Dessert Shop	Convenience Store	Cocktail Bar	Pizza Place	Pool	Plaza	Port	Playground	Platform	Other Nightlife
7	Centro-Est	91,402	15.0%	Manin	1895	Pub	Art Gallery	Event Space	Athletics & Sports	Hotel	Plaza	Supermarket	Flea Market	Aquarium	Pharmacy
8	Centro-Est	91,402	15.0%	San Vincenzo	1895	Italian Restaurant	Hotel	Café	Plaza	Platform	Fast Food Restaurant	Dessert Shop	Pizza Place	Japanese Restaurant	Sandwich Place
9	Centro-Est	91,402	15.0%	Carignano	1895	Italian Restaurant	Plaza	Café	Ice Cream Shop	Historic Site	Bookstore	Coffee Shop	German Restaurant	Soup Place	Gastropub

Centro-Ovest is a bit different than the close neighbor Centro-Est, populated more with shops and less with eateries. A hardware or music store can be good. If the business owner prefers an eatery, a bakery serving breakfasts and coffee in the mornings and aperitivo in the evenings to the tourists is usually a good idea in this particular neighborhood housing the boat and ferry terminals and bus station.

10	Centro-Ovest	66,626	10.9%	Sampierdarena	1163	Electronics Store	Theater	Women's Store	Pool	Italian Restaurant	Plaza	Pizza Place	Convenience Store	Clothing Store	Bookstore
11	Centro-Ovest	66,626	10.9%	Belvedere	1163	Plaza	Italian Restaurant	Café	Hotel	Cocktail Bar	Ice Cream Shop	Hostel	Seafood Restaurant	Burger Joint	Park
12	Centro-Ovest	66,626	10.9%	Campasso	1163	Hobby Shop	Intersection	Electronics Store	Martial Arts School	Scenic Lookout	Beach	Pizza Place	Pier	Park	Performing Arts Venue
13	Levante	66,155	10.8%	San Bartolomeo	3762	Bed & Breakfast	Scenic Lookout	Board Shop	Pizza Place	Beach	Gym	Italian Restaurant	Arts & Crafts Store	Plaza	Pool
14	Centro-Ovest	66,626	10.9%	San Teodoro	1163	Food & Drink Shop	Plaza	Cultural Center	Boat or Ferry	Home Service	Health Food Store	Piadineria	Park	Performing Arts Venue	Pet Store
15	Centro-Ovest	66,626	10.9%	Angeli	1163	Bus Station	Plaza	Department Store	Bakery	Surf Spot	Home Service	Toll Plaza	Boat or Ferry	Pharmacy	Other Nightlife

Both Bassa Val Bisagno and Media Val Bisagno are far from the bustling city center, housing more sports venues, outdoors and performance arts. Opening a sports goods shop or a pub would go nice with the top venues, especially given that the real estate prices are relatively low.

16	Bassa Val Bisagno	78,791	12.9%	San Fruttuoso	1369	Soccer Stadium	Grocery Store	Supermarket	Pub	Italian Restaurant	Plaza	Flea Market	Sporting Goods Shop	Event Space	Hotel
17	Bassa Val Bisagno	78,791	12.9%	Sant'Agata	1369	Platform	Hotel	Plaza	Italian Restaurant	Sandwich Place	Convenience Store	Theater	Café	Bed & Breakfast	Bakery
18	Bassa Val Bisagno	78,791	12.9%	Marassi	1369	Grocery Store	Café	Convenience Store	Pizza Place	Aquarium	Pharmacy	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store
19	Bassa Val Bisagno	78,791	12.9%	Quezzi	1369	Plaza	Grocery Store	Home Service	Aquarium	Pharmacy	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store
20	Bassa Val Bisagno	78,791	12.9%	Fereggiano	1369	Pizza Place	Soccer Stadium	Plaza	Record Shop	Grocery Store	Playground	Café	Coffee Shop	Pet Store	Park
22	Media Val Bisagno	58,742	9.6%	Staglieno	1430	Toll Booth	Café	Aquarium	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy	Piadineria
23	Media Val Bisagno	58,742	9.6%	Molassana	1430	Department Store	Grocery Store	Furniture / Home Store	Pizza Place	Aquarium	Pharmacy	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store
24	Media Val Bisagno	58,742	9.6%	Sant'Eusebio	1430	Italian Restaurant	Plaza	Basketball Court	Diner	Pool	Port	Playground	Platform	Pub	Opera House
25	Media Val Bisagno	58,742	9.6%	Montesignano	1430	Department Store	Farmers Market	Plaza	Food	Outdoors & Recreation	Chinese Restaurant	Aquarium	Piadineria	Performing Arts Venue	Pet Store

For Valpolcevera, mainly populated by music and performance arts and sports activities, the people would most likely to being seated longer than in a fast-food restaurant or the like. Opening a pub, or a steak house can go very well along these activities.

For Ponente and Medio Ponente, situated by the sea side where outdoor and sports activities are enjoyed. Opening a small business such as an ice-cream shop or truck, a sandwich/focaccia joint, a bar by the sea or a bed and breakfast would be a good idea especially due to a high number of local and foreign tourists.

28	Valpolcevera	62,492	10.3%	Certosa	1127	Dog Run	Sculpture Garden	Gym Pool	Metro Station	Plaza	Playground	Platform	Pizza Place	Pool	Opera House
29	Valpolcevera	62,492	10.3%	Teglia	1127	Stadium	Sporting Goods Shop	Restaurant	Soccer Field	Shopping Mall	Pet Store	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue
30	Valpolcevera	62,492	10.3%	Bolzaneto	1127	Theater	Train Station	Toy / Game Store	Coffee Shop	Aquarium	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store
31	Valpolcevera	62,492	10.3%	Pontedecimo	1127	Dessert Shop	Cupcake Shop	Mediterranean Restaurant	Pizza Place	Park	Performing Arts Venue	Pet Store	Pharmacy	Piadineria	Pier
34	Medio Ponente	61,810	10.1%	San Giovanni Battista	2059	Seafood Restaurant	Pool	Gym	Furniture / Home Store	Health Food Store	Aquarium	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store
37	Medio Ponente	61,810	10.1%	Calcinara,	2059	Italian Restaurant	Plaza	Mountain	Aquarium	Piadineria	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy
38	Ponente	63,027	10.3%	Voltri	2028	Seafood Restaurant	Pizza Place	Beach	Supermarket	Park	Pharmacy	Café	Scenic Lookout	Bus Station	Food Court
39	Ponente	63,027	10.3%	Crevari	2028	Italian Restaurant	Bakery	Scenic Lookout	Aquarium	Pier	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy
43	Ponente	63,027	10.3%	Pegli	2028	Pizza Place	Dessert Shop	Scenic Lookout	Grocery Store	Supermarket	Harbor / Marina	Toll Booth	Café	Train Station	Plaza
44	Ponente	63,027	10.3%	Multedo	2028	Beach	Pier	Italian Restaurant	Soccer Field	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy	Piadineria

Medio Levante, being between the bustling touristic city center and a calmer yet still touristic Levante is a mixture of Centro-Est and Levante, both in terms of venues and real estate prices. A safe bet would be opening a bed and breakfast or a bakery/bar offering traditional Italian food and aperitivo or a seafood restaurant by the seaside.

46	Medio Levante	61,759	10.1%	Foce	2680	Ice Cream Shop	Seafood Restaurant	Plaza	Pizza Place	Ligurian Restaurant	Burger Joint	German Restaurant	Chinese Restaurant	Breakfast Spot	Dessert Shop
47	Medio Levante	61,759	10.1%	San Martino	2680	Pizza Place	Food Court	Pet Store	Chinese Restaurant	Convenience Store	Science Museum	Park	Performing Arts Venue	Pharmacy	Piadineria
48	Medio Levante	61,759	10.1%	Chiappeto	2680	Food Court	Chinese Restaurant	Science Museum	Convenience Store	Construction & Landscaping	Aquarium	Performing Arts Venue	Pet Store	Pharmacy	Piadineria
49	Medio Levante	61,759	10.1%	Albaro	2680	Tennis Court	Gym	Restaurant	Food Truck	Bar	Café	Ice Cream Shop	Ligurian Restaurant	Pool	Plaza
50	Medio Levante	61,759	10.1%	San Giuliano	2680	Beach	Tennis Court	Ice Cream Shop	Road	Lounge	Hot Dog Joint	Gym	Pizza Place	Food Truck	Plaza
51	Medio Levante	61,759	10.1%	Lido	2680	Pizza Place	Ice Cream Shop	Seafood Restaurant	Japanese Restaurant	Ligurian Restaurant	Vegetarian / Vegan Restaurant	Café	Breakfast Spot	Convenience Store	Cosmetics Shop
52	Medio Levante	61,759	10.1%	Puggia	2680	Sporting Goods Shop	Café	Cosmetics Shop	Convenience Store	Shopping Mall	Piadineria	Park	Performing Arts Venue	Pet Store	Pharmacy

Levante, situated by the sea side where outdoor and sports activities are enjoyed. Due to the high real estate prices in this municipality, the business should be able to meet a high demand or a small place needs to be preferred to reduce the initial startup expense (acquisition or rental of the immovable). In contrast, opening a small business such as an ice-cream shop or truck, a sandwich/focaccia joint, a bar by the sea or a bed and breakfast would be a good idea especially due to a high number of local and foreign tourists. If a higher budget can be allocated, a disco or a nightclub by the sea might be a different but profitable decision.

53	Levante	66,155	10.8%	Sturla	3762	Italian Restaurant	Plaza	Grocery Store	Pizza Place	Train Station	Gym	Gym / Fitness Center	Sporting Goods Shop	Cocktail Bar	Aquarium
54	Levante	66,155	10.8%	Quarto	3762	Pub	Sports Club	Hotel	Grocery Store	Supermarket	Pizza Place	Gym / Fitness Center	Harbor / Marina	Performing Arts Venue	Other Nightlife
55	Levante	66,155	10.8%	Quartara	3762	Grocery Store	Trattoria/Osteria	Supermarket	Tennis Stadium	Scenic Lookout	Hotel	Platform	Playground	Pizza Place	Pier
56	Levante	66,155	10.8%	Quinto al Mare	3762	Pizza Place	Italian Restaurant	Park	City	Beach	Ice Cream Shop	Plaza	Playground	Platform	Pool
57	Levante	66,155	10.8%	Nervi	3762	Beach	Café	Cocktail Bar	Park	Ice Cream Shop	Bakery	Bar	Trattoria/Osteria	Train Station	Hotel
58	Levante	66,155	10.8%	Apparizione	3762	Italian Restaurant	Farmers Market	Scenic Lookout	Aquarium	Piadineria	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy
59	Levante	66,155	10.8%	Borgoratti	3762	Food Court	Shopping Mall	Science Museum	Health Food Store	Aquarium	Pharmacy	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store
60	Levante	66,155	10.8%	San Desiderio	3762	Italian Restaurant	Movie Theater	Playground	Aquarium	Piadineria	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy
61	Levante	66,155	10.8%	Bavari	3762	Diner	Café	Soccer Field	Piadineria	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy	Aquarium
62	Levante	66,155	10.8%	Sant'Ilario	3762	Pub	Playground	Pizza Place	Beach	Hotel	Pool	Plaza	Platform	Port	Office

2. **Cluster 2 - Outdoors:** Having only 1 member, the second cluster is where outdoor activities are enjoyed. Preferring a fast-food restaurant, food truck or ice cream shop would be a good choice especially for seasonal attractions and also for lower start-up expenses. If our potential business owner can dedicate a higher budget, a sporting goods store would also be fitting.

	Municipality	Population	% of Total Population	Neighborhood	Mean Quotation €/m2 2019	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	10th Most Common Venue
36	Medio Ponente	61,810	10.1%	Campi	2059	Mattress Store	Aquarium	Piadineria	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy	Pier

3. **Cluster 3 – Outdoors and Nightlife:** Slightly similar to the second cluster, this cluster is suitable for outdoor, art and nightlife activities. Preferring a fast-food restaurant, food truck, sandwich joint or ice cream shop would be a good choice especially for people who would enjoy grabbing a bite before concerts, performances or sports activities and also for lower start-up expenses. If our potential business owner can dedicate a higher budget, a sporting goods store would also be fitting.

	Municipality	Population	% of Total Population	Neighborhood	Mean Quotation €/m2 2019	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	10th Most Common Venue
35	Medio Ponente	61,810	10.1%	Cornigliano	2059	Park	Big Box Store	Aquarium	Office	Other Nightlife	Outdoors & Recreation	Performing Arts Venue	Pet Store	Pharmacy	Piadineria
40	Ponente	63,027	10.3%	Pra'	2028	Park	Italian Restaurant	Pool	Grocery Store	Aquarium	Pharmacy	Other Nightlife	Outdoors & Recreation	Performing Arts Venue	Pet Store
41	Ponente	63,027	10.3%	Palmaro	2028	Park	Grocery Store	Bus Station	Aquarium	Opera House	Outdoors & Recreation	Performing Arts Venue	Pet Store	Pharmacy	Piadineria

4. **Cluster 4 – Concerts, Arts and Outdoors:** Having only 1 member, the fourth cluster is where concerts, performances and outdoor activities are enjoyed. Preferring a fast-food restaurant, food truck or ice cream shop would be a good choice especially for seasonal attractions and also for lower start-up expenses. If our potential business owner can dedicate a higher budget, a sporting goods store would also be fitting.

	Municipality	Population	% of Total Population	Neighborhood	Mean Quotation €/m2 2019	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	10th Most Common Venue
42	Ponente	63,027	10.3%	Ca' Nuova	2028	Concert Hall	Aquarium	Piadineria	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy	Pier

5. **Cluster 5 – Indoor and Outdoor Sports:** With lower real estate prices in the neighborhood, this cluster would be an ideal place for opening a sporting goods store or another sports center on a different focus, such as archery or skating rink. As the most popular venues are centered around sports activities, the people would like to enjoy meals before or after working out. A salad bar, sandwich joint or a fast-food restaurant could go very well with the other venues in this cluster. Opening another pharmacy close to the sports centers also would a good idea for people injured during workouts.

	Municipality	Population	% of Total Population	Neighborhood	Mean Quotation €/m2 2019	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	10th Most Common Venue
21	Bassa Val Bisagno	78,791	12.9%	Forte Quezzi	1369	Gym / Fitness Center	Aquarium	Office	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy	Piadineria

6. **Cluster 6 – Outdoor and Nightlife:** Having only 1 member, the sixth cluster is where outdoor activities and nightlife are enjoyed. Preferring a fast-food restaurant, food truck or ice cream shop would be a good choice especially for seasonal attractions and also for lower start-up expenses. If our potential business owner can dedicate a higher budget, a camping or sporting goods store would also be fitting, given that the mediocre real estate prices.

	Municipality	Population	% of Total Population	Neighborhood	Mean Quotation €/m2 2019	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	10th Most Common Venue
32	Medio Ponente	61,810	10.1%	Sestri	2059	Mountain	Aquarium	Office	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy	Piadineria

7. **Cluster 7 – Travel, Performance Arts and Nightlife:** Judging by the common venues, this neighborhood in this cluster seems to be frequented by commuters or people enjoying performance arts or nightlife. The business owner can think about opening a bar offering breakfasts in the morning, and aperitivo in the evening to respond a wide range of needs for the whole day.

	Municipality	Population	% of Total Population	Neighborhood	Mean Quotation €/m2 2019	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	10th Most Common Venue
45	Ponente	63,027	10.3%	Castelluccio	2028	Train Station	Aquarium	Piadineria	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy	Pier

8. **Cluster 8 – Performance Arts, Nightlife and Soccer:** Being mainly populated by music and performance arts and sports activities, the people would most likely to being seated longer than in a fast-food restaurant or the like. Opening a pub, or a steak house can go very well along these activities. For Rivarolo and Bozoli Est, even a bowling arena could be a refreshing idea for entrepreneurs.

	Municipality	Population	% of Total Population	Neighborhood	Mean Quotation €/m2 2019	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	10th Most Common Venue
26	Valpolcevera	62,492	10.3%	Rivarolo	1127	Soccer Field	Food & Drink Shop	Pub	Other Nightlife	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy	Aquarium
27	Valpolcevera	62,492	10.3%	Borzoli Est	1127	Rock Club	Convenience Store	Soccer Field	Aquarium	Piadineria	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy
33	Medio Ponente	61,810	10.1%	Borzoli Ovest	2059	Rock Club	Convenience Store	Soccer Field	Aquarium	Piadineria	Outdoors & Recreation	Park	Performing Arts Venue	Pet Store	Pharmacy

4 DISCUSSION

The analyses proved satisfactory results for deciding on the business type. However, there is a lot of room for improvement if I could be able to use another location-based service other than Foursquare, since Foursquare is not used for check-ins anymore. As a consequence, I missed out trending endpoint which would provide invaluable information on real foot-traffic that could be used as one of the most important features in clustering. The details of the venues such as the price range and ratings were subject to a paid premium service which was another downside of Foursquare API.

Having access to the city of our choice is also a major determinant in forming a good dataset. Having the demographics of the city, income and age distribution with respect to neighborhoods would have been very useful. Usually, the bigger the city is, the better (and more) the data is kept. I chose Genoa since I was curious about learning more about a place close to me (and there was enough data to carry out the analyses for La Spezia, which is where I live).

For the sake of a more in-depth analysis, as another clustering algorithm, DBSCAN was also tried to see if the resulting clusters would make more sense. After tweaking the epsilon and number of minimum samples, it can be concluded that DBSCAN does not give promising results due to high number of features but low number of samples. The algorithm mainly picked 1 cluster with many neighborhoods and too much noise, rendering K-Means as a more suitable choice for our application.

5 CONCLUSION

Web-scraping and extensive internet search followed by some hand engineering for correcting the false data were used for preparing the data. The scarcity of data for small cities is a problem for landing on precise decisions.

This project heavily relied on Foursquare API which is no longer the preferred location data provider, thus preventing the acquisition of critical data that would significantly leverage the decisional process. Having the price ranges, ratings and also historical data of the venues would help forming better clusters, thus arriving at better decisions.

Different clustering algorithms were used to achieve the optimal clustering scheme. DBSCAN did not give promising results due to high number of features but low number of samples. The algorithm mainly picked 1 cluster with many neighborhoods and too much noise, rendering K-Means as a more suitable choice for our application.

Again, due to the high number of features and relatively low number of observations, no elbow point was observed on the SSE vs number of clusters graph. Therefore, for accuracy metrics, silhouette score was used for deciding on the number of clusters.

The output of the analyses together with the Sankey diagrams and living as a local in Italy provided good understanding of the surroundings and necessary grounds for answering the business problem. Nevertheless, with more features at hand, more spot-on recommendations could have been made.

6 REFERENCES

[1] Mean quotation prices / m^2 for Genoa (2019)

https://www.agenziaentrate.gov.it/portale/documents/20143/2555996/SR2020_Liguria.pdf/1b86cae7-be56-365e-2734-1a81ca3b8882

[2] Genoa Municipal Government: https://en.wikipedia.org/wiki/Genoa#Municipal_government

[3] Geojson file of the historical areas: http://dati.comune.genova.it/dataset/zone-storiche-di-genova