OPENING A NEW BUSINESS IN GENOVA/ITALY

A DATA CENTRIC APPROACH

BY

A. BURCU ÖZKAPTAN

ANALYZING VENUES AND OTHER DATA IN A CITY IS IMPORTANT FOR DECIDING A NEW BUSINESS OPPORTUNITY

How to Decide What Sort of a Business to open and Where?

- Whether you are 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, hotels, sports halls, shopping venues, etc...

- 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.

- Methodology: Leveraging the Foursquare location data to explore or compare neighborhoods in the city of Genova/Italy and using the Foursquare location data to solve a business problem of opening a new business.

ASKING THE RIGHT QUESTION

- 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 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.
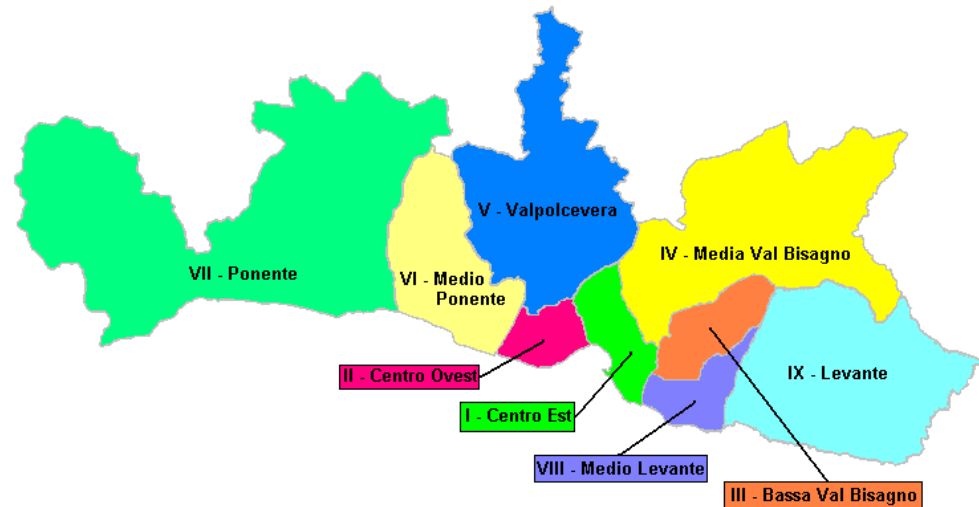
DATA DESCRIPTION

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

- The list of municipalities and neighborhoods together with the population data of each municipality,

- The coordinates of each neighborhood,

- The venues of each neighborhood that were queried from Foursquare API.

- The real-estate prices obtained from the Revenue Services (Agenzia Entrate)

DATA DESCRIPTION

- 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 has some irregularities, data cleansing was required.

DATA DESCRIPTION

- After building a dataframe including the municipalities, population of each municipality along with the neighborhood names in each municipality, in order to utilize the Foursquare location data, the latitude and the longitude coordinates of each neighborhood had to be obtained by using Geocoder Python package: 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, ArcGIS geocoder was chosen, one of the most popular geocoding APIs with a high accuracy in obtaining the right coordinates.

- Finally, the neighborhoods in Genova were explored by leveraging the functionalities of Foursquare API and were clustered by using K-Means algorithm.

- On a final note, DBSCAN algorithm was also used however, due to high number of features with 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.

DATA DESCRIPTION

- The real estate prices (mean quotation price / m² 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

EXPLORING THE NEIGHBORHOODS IN GENOVA

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 Ponent   6
- Bassa Val Bisagno   6
- Media Val Bisagno   5
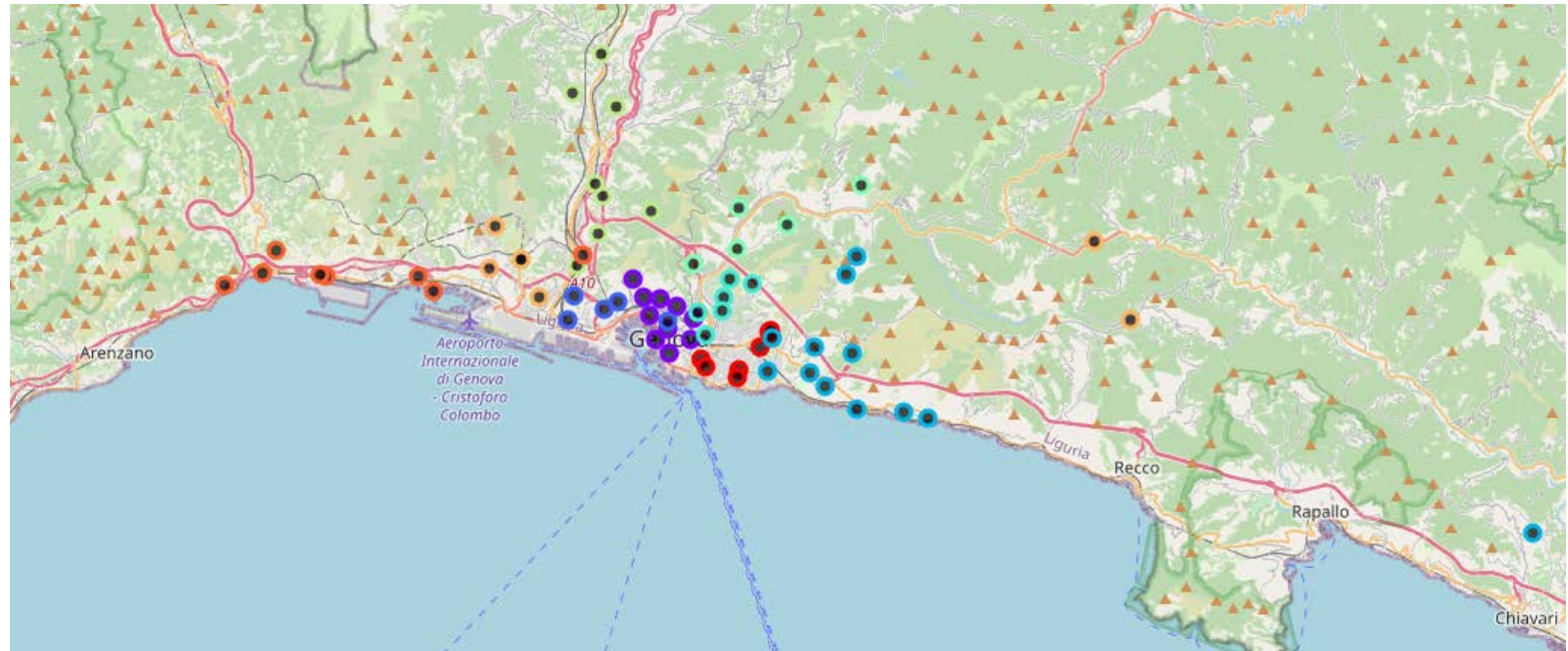
EXPLORING THE VENUES WITH FOURSQUARE

- By using the Foursquare API to retrieve data from the Foursquare database, the venues around points of interests, being the neighborhoods, were retrieved:

  - 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 barplots. 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.
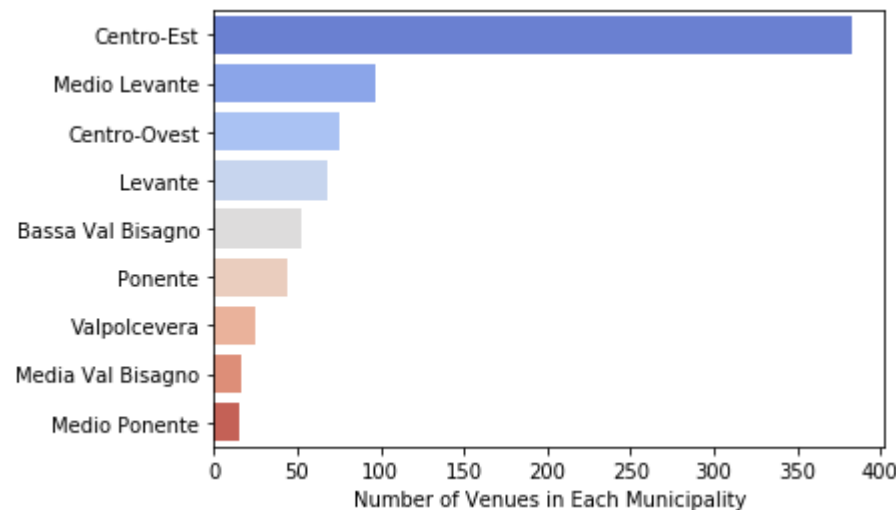
EXPLORING THE VENUES WITH FOURSQUARE

- The 25 most popular venue categories in the city
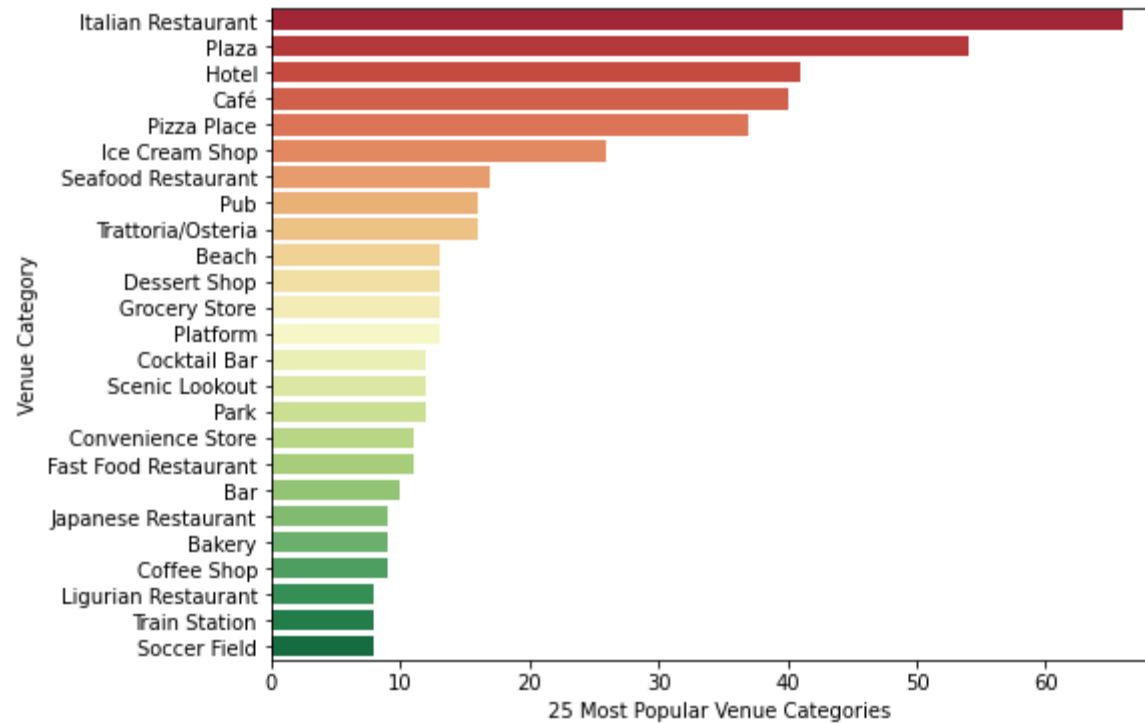
MUNICIPALITY COMPARISONS

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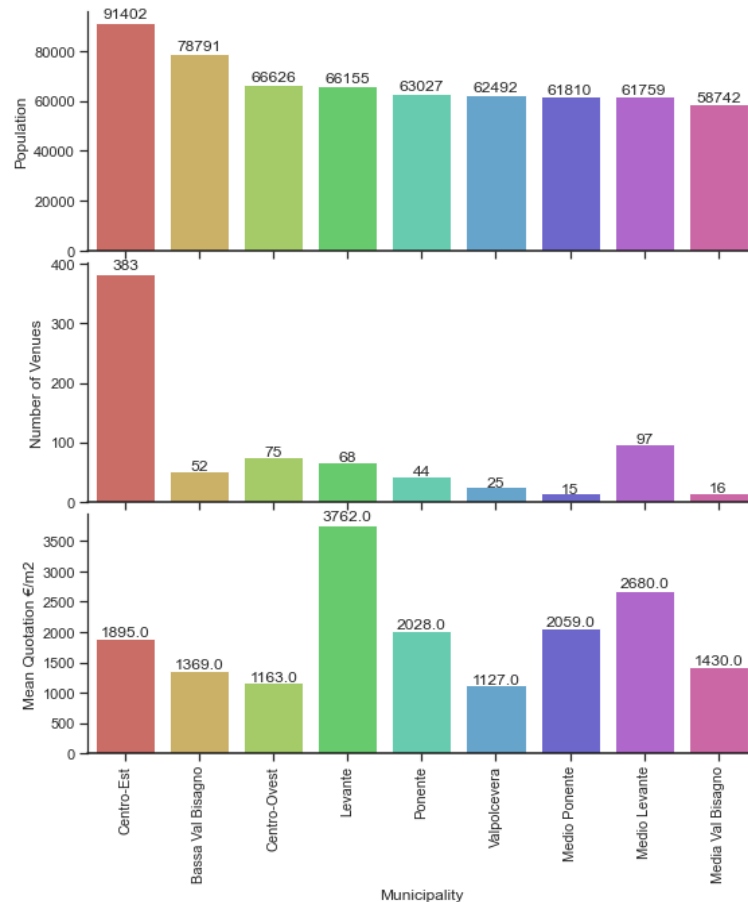

	Municipality	Neighbourhoods	Population	% of Total Population	Mean Quotation €/m2 2019	Number of Venues
0	Centro-Est	Prè, Molo, Maddalena, Oregina, Lagaccio, San N...	91402	15.0	1895	383
1	Centro-Ovest	Sampierdarena, Belvedere, Campasso, San Bartol...	66626	10.9	1163	75
2	Bassa Val Bisagno	San Fruttuoso, 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

DISTRIBUTION OF VENUES IN EACH MUNICIPALITY

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

Municipality	Neighborhood	Venue
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

Municipality	Neighborhood	Venue
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

DISTRIBUTION OF VENUES IN EACH MUNICIPALITY

- Even though this 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.

Distribution of Venues in Each Municipality
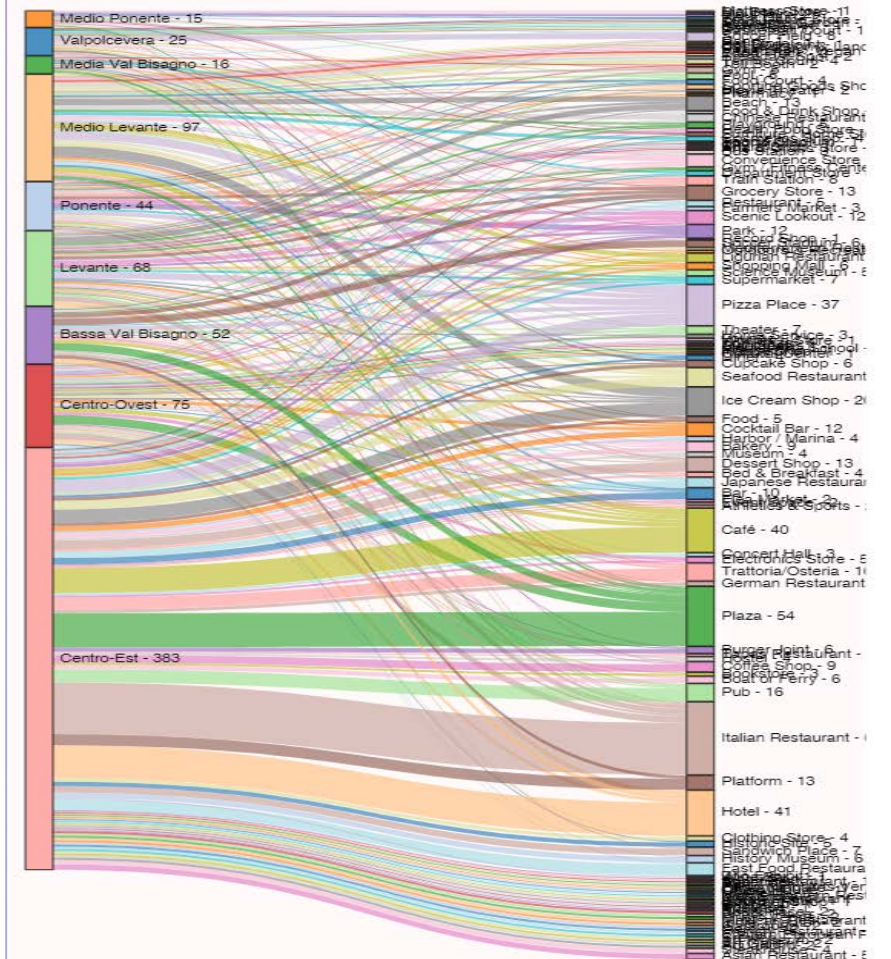
DISTRIBUTION OF VENUES WITH RESPECT TO NEIGHBORHOODS AND MUNICIPALITIES

- This version of the Sankey diagram is a bit more complicated but explains the whole distribution of the venues in the neighborhoods.

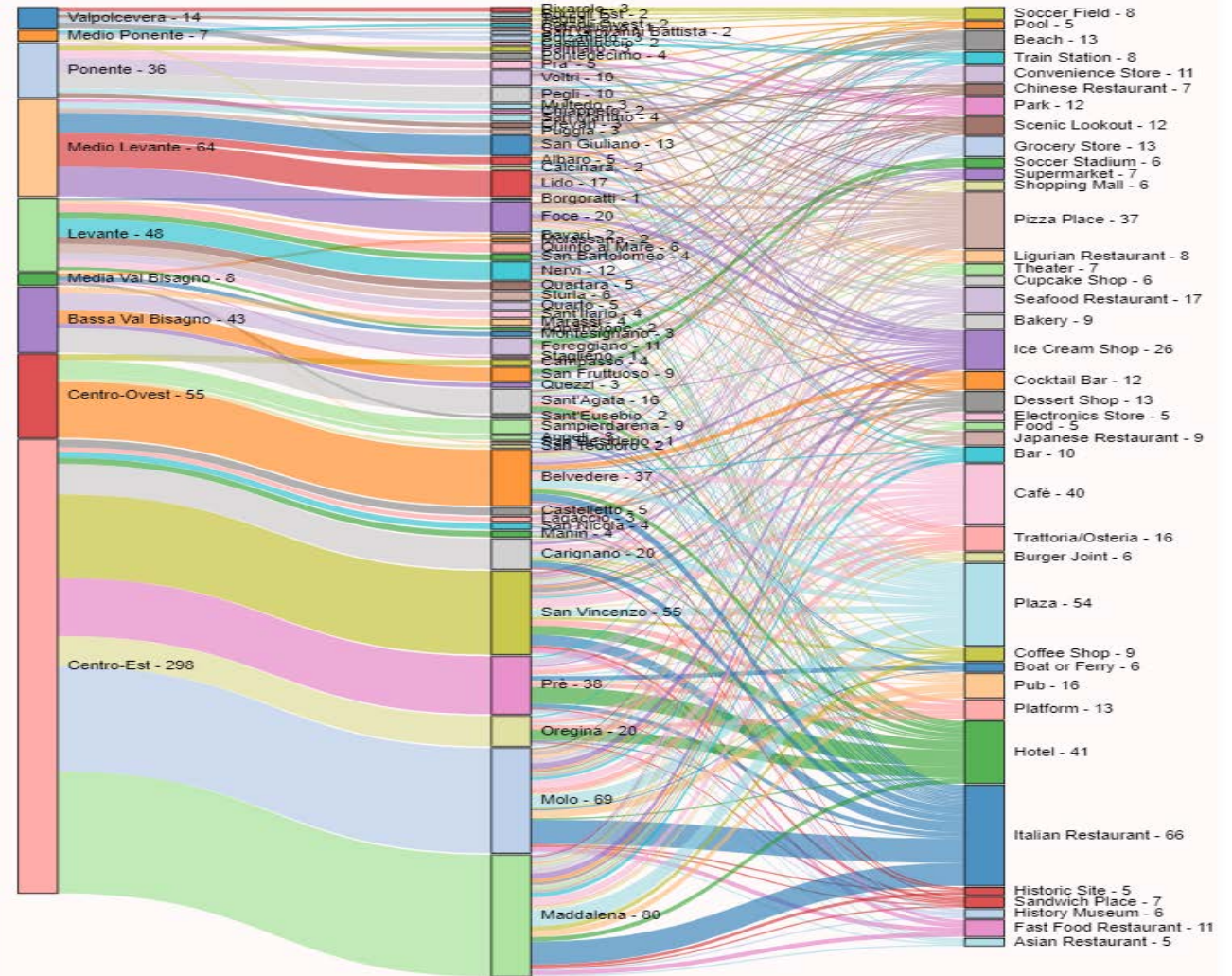Distribution of the Venue Types with Respect to Neighborhoods and Municipalities in Genova
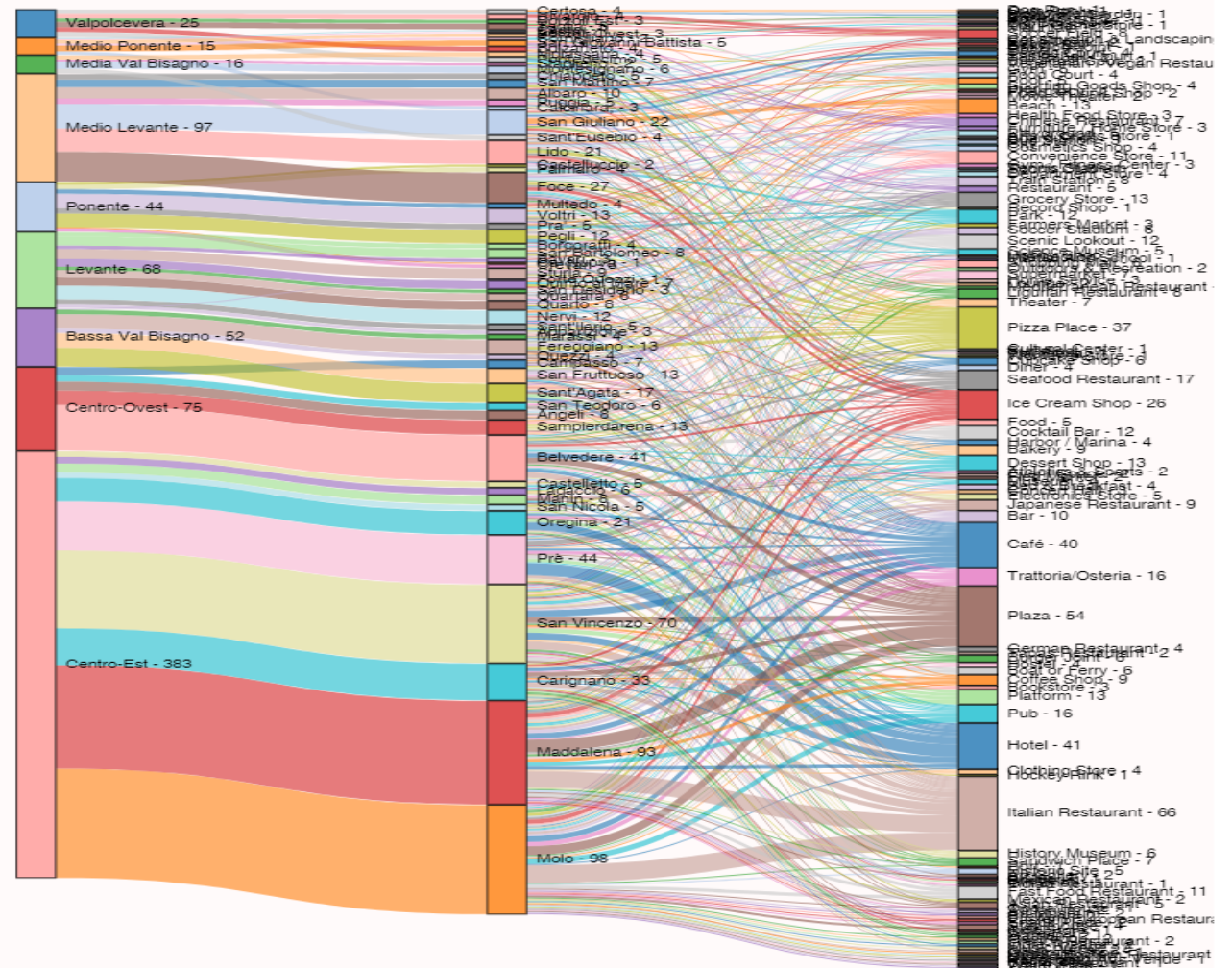
DISTRIBUTION OF TOP 40 VENUES WITH RESPECT TO NEIGHBORHOODS AND MUNICIPALITIES

- To get a simplified view, we can filter out the top 40 venue categories and get a similar Sankey representation

Venue Category Distribution at Genova for top 40 Categories

ANALYZING NEIGHBORHOODS – 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 have the 143 venue categories converted into one-hot encoding for all 775 venues.

- Next, the rows are grouped by neighborhood and by taking the mean of the frequency of occurrence of each category. 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.

	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

CLUSTERING

- As the final step of the methodology, different clustering algorithms were used.

- K-means algorithm was the preferred algorithm suitable for this application.

- The algorithm was run for cluster sizes ranging from 2 to 50

PERFORMANCE METRICS – SSE?

- 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.
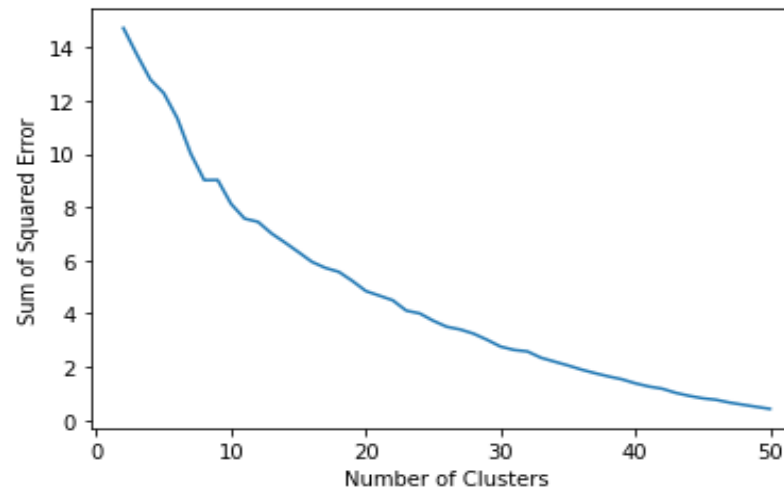
PERFORMANCE METRICS

- 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.
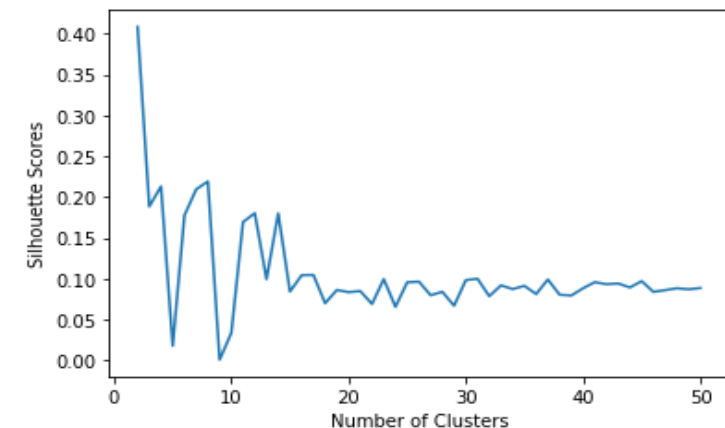
CHOOSING THE RIGHT CLUSTER SIZE

- 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.

- 6 neighborhoods to not have any venues, therefore not belonging to a cluster.

- Next, we merge the labels with the dataframe 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.

RESULTING CLUSTERS

- Finally, the resulting clusters were visualized together with real estate prices (mean quotation €/m2) overlayed on each municipality.

- 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.
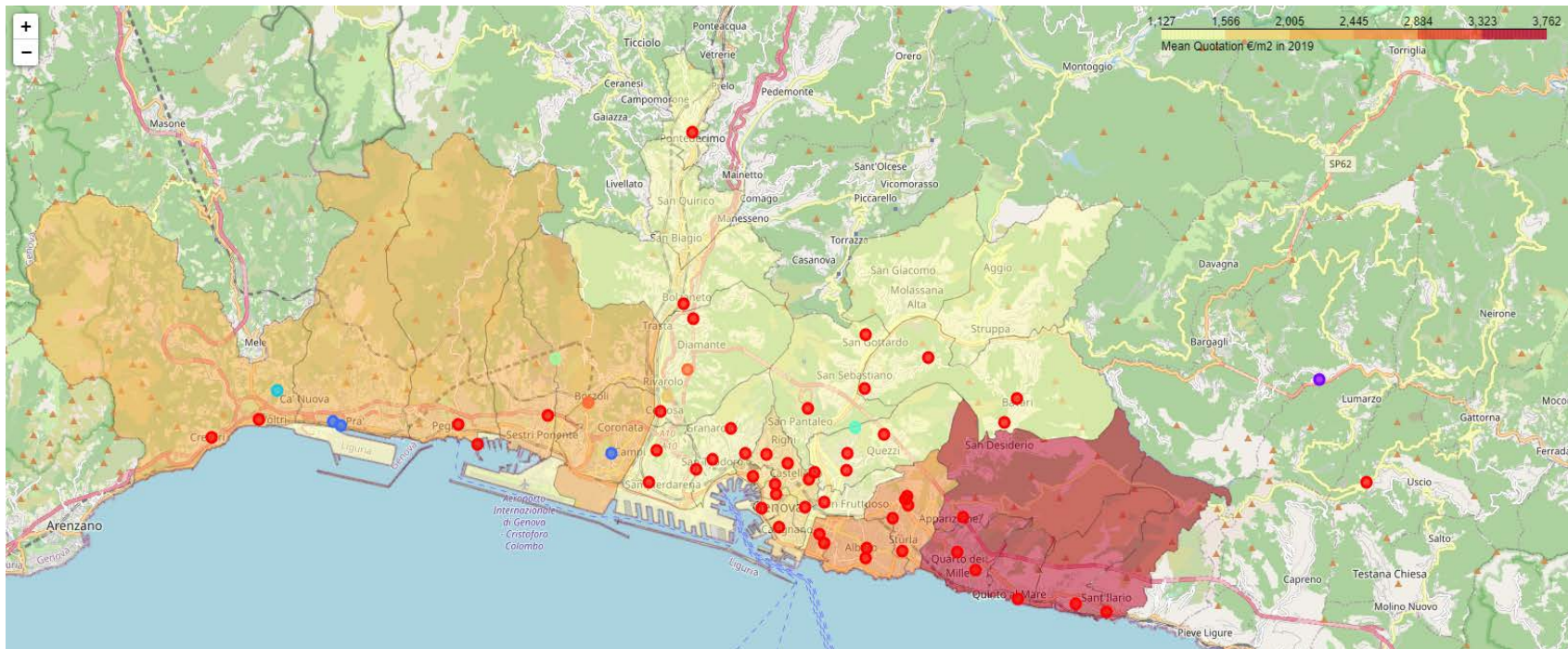
RESULTS – CLUSTER 1

- 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

RESULTS – CLUSTER 1

- **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

RESULTS – CLUSTER 1

- 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

RESULTS – CLUSTER 1

- 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

25

RESULTS – CLUSTER 1

- **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

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

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

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

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

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

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

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

DISCUSSIONS

- 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.

DISCUSSIONS

- 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.

CONCLUSIONS

- 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.

CONCLUSIONS

- 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.

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