

Applied Data Science Capstone

IBM Data Science Capstone

**Battle of the Neighborhoods:
Finding the best neighborhood
to purchase a home in
Barcelona, Spain**

Kristopher Zapata

December 2020

Table of Contents

1. Introduction.....	4
1.1. Business Problem	5
1.2. Target Audience for this Project	5
2. Data	6
2.1. Data Wrangling.....	6
3. Methodology.....	9
3.1. Exploratory Data Analysis.....	9
3.2. Clustering Neighborhoods	11
3.3. Mapping out Cluster Neighborhoods	12
4. Results.....	17
4.1. Cluster Comparison	17
4.2. Cluster 0 - Upper Middle Class:.....	18
4.3. Cluster 1 – Outskirts.....	20
4.4. Cluster 2 – Lower Middle Class	22
4.5. Cluster 3 – High End	25
5. Discussion.....	27
6. Conclusion	28
Bibliography.....	29

Table of Figures

FIGURE 1: TOP 10 MOST EXPENSIVE NEIGHBORHOODS IN BARCELONA.....	11
FIGURE 2: TOP 10 LEAST EXPENSIVE NEIGHBORHOODS IN BARCELONA	11
FIGURE 3: TOP 10 NEIGHBORHOODS WITH THE HIGHEST CAGR	11
FIGURE 4: TOP 10 NEIGHBORHOODS WITH THE LOWEST CAGR.....	11
FIGURE 5: ELBOW METHOD PLOT	12
FIGURE 6: ORIGINAL BARCELONA MAP WITH NEIGHBORHOOD MARKERS	12
FIGURE 7: CLUSTER MARKER MAP	13
FIGURE 8: CHOROPLETH MAP OF AVERAGE HOUSING PRICE.....	14
FIGURE 9: CHOROPLETH MAP OF CAGR.....	15
FIGURE 10: CHOROPLETH MAP OF NEIGHBORHOOD CLUSTER.....	16
FIGURE 11: BUBBLE PLOT OF CLUSTER COMPARISON.....	18
FIGURE 12: SCATTER PLOT OF CLUSTER 0 NEIGHBORHOODS	20
FIGURE 13: SCATTER PLOT OF CLUSTER 1 NEIGHBORHOODS	22
FIGURE 14: SCATTER PLOT OF CLUSTER 2 NEIGHBORHOODS	25
FIGURE 15: SCATTER PLOT OF CLUSTER 3 NEIGHBORHOODS	27

Table of Figures

TABLE 1: HOUSING PRICE INDEX.....	7
TABLE 2: HOUSING PRICE INDEX WITH COORDINATES	7
TABLE 3: VENUES DATA FRAME	7
TABLE 4: NEW JOINT-DATA FRAME OF HOUSING PRICE INDEX AND VENUES.....	8
TABLE 5: DATA FRAME USED FOR K-MEANS CLUSTERING	8
TABLE 6: TOP 10 MOST COMMON VENUES PER NEIGHBORHOOD.....	8
TABLE 7: MERGED DATA FRAME	9
TABLE 8: STATISTICAL ANALYSIS	10
TABLE 9: CLUSTER COMPARISON.....	17
TABLE 10: CLUSTER 0 - UPPER MIDDLE CLASS NEIGHBORHOODS.....	19
TABLE 11: CLUSTER 0 - STATISTICAL ANALYSIS.....	19
TABLE 12: CLUSTER 0 - STATISTICAL ANALYSIS OF CATEGORICAL VARIABLES ...	20
TABLE 13: CLUSTER 1 - OUTSKIRTS.....	21
TABLE 14: CLUSTER 1 - STATISTICAL ANALYSIS.....	21
TABLE 15: CLUSTER 1 - STATISTICAL ANALYSIS OF CATEGORICAL VARIABLES ...	22
TABLE 16: CLUSTER 2 - LOWER MIDDLE CLASS NEIGHBORHOODS.....	23
TABLE 17: CLUSTER 2 - STATISTICAL ANALYSIS.....	24
TABLE 18: CLUSTER 2 - STATISTICAL ANALYSIS OF CATEGORICAL VARIABLES ...	24
TABLE 19: CLUSTER 3 - HIGH END NEIGHBORHOODS.....	26
TABLE 20: CLUSTER 3 - STATISTICAL ANALYSIS.....	26
TABLE 21: CLUSTER 3 - STATISTICAL ANALYSIS OF CATEGORICAL VARIABLES ...	26

1. Introduction

Barcelona is one of the most popular cities in the European Union. It is home to 1.6 million of which around 20% are foreigners, composed of 179 different nationalities, making it a cosmopolitan city filled with diversity. (Record number of the foreign population in Barcelona) Barcelona is also a famous dorm city, hosting more than 188,000 students per year, contributing to the youth and vibrancy of the city. (Explorer Publishing, 2020)

Due to Barcelona's richness in arts, architecture, culture and privileged location, it is also one of the most visited cities in the world, with more than 12 million visitors in 2019. (BarcelonaYellow, 2020)

With the rise of business hubs cities in Europe, following suite with Silicon Valley, Barcelona has not fallen behind, positioning itself as the sixth most important European hub for startups and innovations. (Novoselteva, 2019) Besides startups, big companies such as Hewlett Packard have established their presence in Barcelona, contributing to Barcelona positioning itself as one of the most important business hubs around Europe. There are many reasons behind Barcelona positioning itself as a major business hub, such as connectivity, local and international talent, and more competitive salaries as compared to other major European cities.

Barcelona is also an example of urban planning development, dating back from the times of Ildefons Cerdà to present day. Cerdà is mainly responsible for the design of the current district of l'Eixample, which till this day is considered an urban development model of excellence. In essence, his idea was to combine the advantages of rural living (green space, fresh air and food, community) with the advantages of urban living (commerce, culture, free flow of goods and ideas). (Roberts, 2019) Although initially his ideas were rejected, they were later implemented and have become the guidelines for city authorities for developing a greener, sustainable city focused on improving the quality of life of its residents.

Although Barcelona is one of the leading global examples of urban development, It's also one of the most densely populated cities in Europe with over 16,000 people per square kilometer on average. The highest density district is l'Eixample, where there are 36,000 people per square kilometer, and the highest density neighborhood is La Sagrada Familia, with 50,000 people per square kilometer. (PopulationStat, 2020)

Considering all the factors mentioned above such as tourist arrivals and student arrivals, it's only natural that Barcelona also suffers of overpriced real estate. Most home purchasers/renters struggle to find good value for their money when it comes to finding a home to live in. One of the main causes has been the rise of vacation rentals in the city, which are far more lucrative for the tenants than permanent home rentals. This has led to many protests from the local tenants who have been driven out of their homes due to many landlords opting for temporary vacation rentals. Despite the restrictions and limitations brought upon by the local authorities, vacation rental properties continued to open up and home rental prices continued to rise. (O'Sullivan, 2018)

Due to the COVID-19 pandemic ongoing worldwide, one of its many effects has been felt exceptionally in the real estate market in many countries, such as Spain. Leading to lowering prices of real estate in the big metropolitan areas, like Barcelona. Although rental prices have decreased by -3%, home purchases have actually increased by +1,3%, focused mostly on the outer layers of the Barcelona Metropolitan area. This phenomenon can be explained by several factors such as:

- As working from home becomes an option many people have chosen to move back to their hometowns or cities in order to reduce their costs of living and expend the cost of rent.
- As more online educational programs rise the arrivals of international student has decreased impacting the demand for accommodations
- Due to rise in unemployment and people taking pay cuts in order to keep their jobs, the government has allowed people to ask their tenants for a temporary rent reduction, which they are obliged to give given the current state of affairs.
- The decrease in tourist arrivals, which has led to various vacation rental properties being turned into permanent rental properties, broadening the home rental offer. (Idealista, 2020)
- People or businesses who had savings before the COVID-19 pandemic were waiting for home prices to decrease in order to get a good deal and have therefore capitalized on the opportunity. It's important to note that the increase in home purchases have mostly been in the outer smaller towns relatively close to Barcelona relatively close to the urban core, which means working from home is a growing trend being adopted permanently for many companies.

1.1. Business Problem

Considering the elements presented above, we want to determine the best neighborhood for purchasing a house in Barcelona, considering certain, but not all, relevant factors that can determine its success. Now, determining the best neighborhood for purchasing a home depends on the requirements and needs of the home purchaser, therefore we'll want to create clusters of neighborhoods that share certain similarities, which would help narrow down the search as well as provide a list of neighborhoods that share similar features desired by the home purchaser such as:

- Venue density in each neighborhood, providing convenient services such as supermarkets, pharmacies, bars, restaurants etc...
- Housing price index per neighborhood
- Compound Annual Growth Rate of the Housing Price Index per neighborhood.

1.2. Target Audience for this Project

There are several target audiences for this project, depending on the objectives established by the home purchaser. The profiles are defined as follows:

- People interested in buying a home to live in the city.
- People or businesses interested in purchasing properties in Barcelona as an investment, whether its to rent as a vacation rentals or permanent rental, House Flipping or as a long-term investment to sell in the future.

2. Data

In order to find a solution to the problem presented above, the following data will be used:

- GeoJson files of the city of Barcelona separated by district, extracted from a github repository from Martin Gonzalez, which were originally extracted from the Ayuntamiento of Barcelona website. These files have all the coordinates of Barcelona and will be used to create the various choropleth maps of Housing Sales Price Index of Barcelona. (González, Github, 2020)
- Foursquare API to get the most common venues of given neighborhoods of Barcelona and as well as retrieve the geographical coordinates for each.
- The Housing Price Index provided by the Ayuntamiento de Barcelona, the municipal office of Barcelona, which in turn obtained the data from a study provided by Idealista, the most popular real estate website in Spain. From this source, I was able to retrieve the average price per square meter in each neighborhood and I also calculated the Compound Annual Growth Rate of the average price per square meter for each neighborhood over the last 8 years. (Ajuntament de Barcelona. Departament d'Estadística i Difusió de Dades, 2020)

Data about different venues in different neighborhoods of that specific neighborhood is needed as well. In order to retrieve that information, Foursquare API is used, which is a location data provider with information about all manner of venues and events within an area of interest.

After providing the list of neighborhoods, a connection is established to the Foursquare API to gather information about venues in each and every neighborhood within a radius of 500 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue is as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue Name
5. Venue Latitude
6. Venue Longitude
7. Venue Category

2.1. *Data Wrangling*

In this project, the starting data set contains the average price per square meter of the last 8 years for all the neighborhoods of Barcelona. Since the initial data set is downloaded from the Barcelona municipal office website as an excel file, the compound annual growth rate is calculated in the excel file. The compound annual growth rate can be calculated in python, but for the case of easing up the data

wrangling process, it was done in the excel file. The historical prices were also dropped since they had already served their purpose in calculating the CAGR. If the project was about building a future price predictor using a regression model, the historical prices would have been kept, but since this project was about clustering, they were discarded.

As seen in the figure below, there are missing values within the Average Price per square meter (sqm)sqm and the Compound Annual Growth Rate, therefore, these missing values are replaced with the mean value of each column, considering it would be a mistake to drop the rows with missing values since each row represents a neighborhood.

TABLE 1: HOUSING PRICE INDEX

	Neighborhood	Avg Price per m2	CAGR
0	Baró de Viver	NaN	NaN
1	Can Baró	3339.0	5.894268
2	Can Peguera	NaN	NaN
3	Canyelles, Nou Barris, Barcelona	NaN	NaN
4	Ciutat Meridiana	1690.0	2.456182

In order to add the coordinates for each neighborhood into the data frame, which will be needed for retrieving the most common venues and mapping, the geocoder function is applied to the neighborhood column, extracting the location, point, latitude, longitude and altitude. However, since only the location, latitude and longitude columns are needed, the other two columns are dropped.

TABLE 2: HOUSING PRICE INDEX WITH COORDINATES

	Neighborhood	Avg Price per m2	CAGR	Location	Latitude	Longitude
0	Baró de Viver	3695.106061	4.185477	(Baró de Viver, Besòs, la Trinitat Vella, Sant...	41.449947	2.199539
1	Can Baró	3339.000000	5.894268	(Can Baró, Horta-Guinardó, Barcelona, Barcelon...	41.416092	2.162402
2	Can Peguera	3695.106061	4.185477	(Can Peguera, Nou Barris, Barcelona, Barcelonè...	41.434857	2.165785
3	Canyelles, Nou Barris, Barcelona	3695.106061	4.185477	(Canyelles, Nou Barris, Barcelona, Barcelonès,...	41.442835	2.165228
4	Ciutat Meridiana	1690.000000	2.456182	(Ciutat Meridiana, Nou Barris, Barcelona, Barc...	41.460918	2.174567

An additional data frame is generated from the output of the function created to extract the most common venues per neighborhood, from which the following information is retrieved:

- Venue Name
- Venue Category
- Venue Latitude
- Venue Longitude

TABLE 3: VENUES DATA FRAME

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Baró de Viver	41.449947	2.199539	The 1982 Birres & Burgers	41.451200	2.204559	Burger Joint
1	Baró de Viver	41.449947	2.199539	Bar Rincón Gallego	41.451982	2.202147	Café
2	Baró de Viver	41.449947	2.199539	Mercadona	41.452754	2.202675	Supermarket
3	Baró de Viver	41.449947	2.199539	Parc del Nus de la Trinitat	41.449112	2.195828	Park
4	Baró de Viver	41.449947	2.199539	METRO Baró de Viver	41.449728	2.199813	Metro Station

Both data frames are joined into one single data frame, eliminating redundancies and facilitating further data manipulation.

TABLE 4: NEW JOINT-DATA FRAME OF HOUSING PRICE INDEX AND VENUES

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Avg Price per m2	CAGR
0	Baró de Viver	41.449947	2.199539	The 1982 Birres & Burgers	41.451200	2.204559	Burger Joint	3695.106061	4.185477
1	Baró de Viver	41.449947	2.199539	Bar Rincón Gallego	41.451982	2.202147	Café	3695.106061	4.185477
2	Baró de Viver	41.449947	2.199539	Mercadona	41.452754	2.202675	Supermarket	3695.106061	4.185477
3	Baró de Viver	41.449947	2.199539	Parc del Nus de la Trinitat	41.449112	2.195828	Park	3695.106061	4.185477
4	Baró de Viver	41.449947	2.199539	METRO Baró de Viver	41.449728	2.199813	Metro Station	3695.106061	4.185477

The column which is relevant for the project is Venue Category, since one of the objectives is to determine the most common venues in each neighborhood, therefore, One Hot Encoding will be applied to this specific column to create a new data frame which will be the main data frame used for the K-means clustering algorithm.

One hot encoding is a process by which categorical variables are converted into a form that can be processed by machine learning algorithms, providing a better execution.

Once the data frame has been generated, the neighborhoods column is added back as it is needed for grouping rows by Neighborhood and calculating the mean of the frequency of occurrence of each venue category. This data frame is the main data frame used in the K-means clustering algorithm. Considering that only numerical columns can be used for executing the algorithm, the data frame is ready to be used once the columns for the Average Price per sqm and the Compound Annual Growth Rate are added, and the neighborhood column is dropped,

TABLE 5: DATA FRAME USED FOR K-MEANS CLUSTERING

	Yoga Studio	Accessories Store	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	BBQ Joint	Ba Stc
0	0.0	0.0	0.0	0.0	0.0	0.0	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	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

In order to generate the final data frame that would harbor all the major components to be analyzed, a data frame containing the top 10 most common venues per neighborhood was generated.

TABLE 6: TOP 10 MOST COMMON VENUES PER NEIGHBORHOOD

	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	10th Most Common Venue
0	Baró de Viver	Plaza	Track Stadium	Park	Supermarket	Café	Metro Station	Burger Joint	Planetarium	Palace	Nightclub
1	Can Baró	Spanish Restaurant	Scenic Lookout	Bar	Chinese Restaurant	Soccer Field	Grocery Store	Café	Pool	Soccer Stadium	Breakfast Spot
2	Can Peguera	Park	Escape Room	Café	Basketball Stadium	Gym / Fitness Center	Tapas Restaurant	Hostel	Grocery Store	Soccer Field	Brewery
3	Canyelles	Soccer Field	Market	Mediterranean Restaurant	Grocery Store	Tapas Restaurant	Food	Skate Park	Hot Spring	Food & Drink Shop	Brewery
4	Ciutat Meridiana	Metro Station	Park	Mediterranean Restaurant	Supermarket	Plaza	Grocery Store	Outdoors & Recreation	Paper / Office Supplies Store	Palace	Paella Restaurant

The data frame with the most common venues per neighborhood is now merged with the initial data frame containing the housin price index and coordinates for each neighborhood. An additional column is also added containing the Cluster Labels created using the Kmeans Algorithm resulting on the the final data frame as detailed below:

TABLE 7: MERGED DATA FRAME

	Neighborhood	Avg Price per m2	CAGR	Location	Latitude	Longitude	Cluster Labels	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
0	Baró de Viver	3695.106061	4.185477	(Baró de Viver, Besòs, la Trinitat Vella, Sant...	41.449947	2.199539	2	Plaza	Track Stadium	Park	Supermarket	Café	Metro Station	Burger Joint	Plane
1	Can Baró	3339.000000	5.894268	(Can Baró, Horta-Guinardó, Barcelona, Barcelon...	41.416092	2.162402	2	Spanish Restaurant	Scenic Lookout	Bar	Chinese Restaurant	Soccer Field	Grocery Store	Café	
2	Can Peguera	3695.106061	4.185477	(Can Peguera, Nou Barris, Barcelona, Barcelonè...	41.434857	2.165785	2	Park	Escape Room	Café	Basketball Stadium	Gym / Fitness Center	Tapas Restaurant	Hostel	C

3. Methodology

3.1. Exploratory Data Analysis

In order to gain a broader understanding of the data set, it's important to execute preliminary exploratory data analysis. As seen in the figure below, the average price per square meter for homes is 3.695€ and this metric has a compound annual growth rate of 4,2% over the last 8 years. It's important to note that these figures already include the 2020 prices impacted by COVID-19, therefore, conclusions can be made that, despite the current pandemic crisis, housing prices have been growing at a steady pace over the last 8 years. The housing price range extends between 1.690€ and 6.460€ per square meter and some neighborhoods register a CAGR as high as 8,6%, which is highly encouraging for investors.

TABLE 8: STATISTICAL ANALYSIS

	Avg Price per m2	CAGR
count	73.000000	73.000000
mean	3695.106061	4.185477
std	1010.993902	1.934642
min	1690.000000	-3.473222
25%	2994.000000	3.447772
50%	3695.106061	4.185477
75%	4312.000000	5.015778
max	6460.000000	8.631825

In the table below, we find the best and worst rated neighborhoods in regards to the average price per square meter and the compound annual growth rate (CAGR). In the case of the Top 10 least expensive neighborhoods, it's a common mistake to invest in these neighborhoods due to their low price without factoring in the CAGR. From an investment standpoint, it's best to look into the neighborhoods with the highest CAGR that just look at the least expensive neighborhoods. To substantiate this affirmation El Carmel, which is among the Top 10 list of least expensive neighborhoods has a higher CAGR than the city's average of 4,2%. As much as 50% of these neighborhoods are also in the Top 10 neighborhoods with the lowest CAGR.

FIGURE 1: TOP 10 MOST EXPENSIVE NEIGHBORHOODS IN BARCELONA

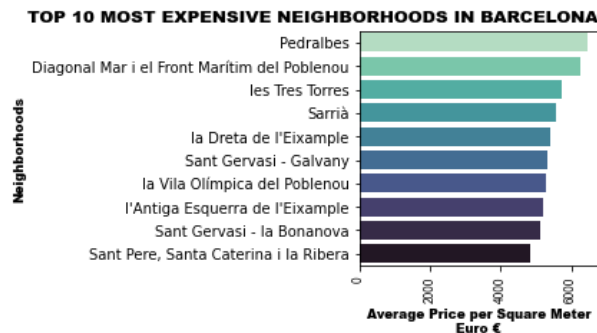


FIGURE 2: TOP 10 LEAST EXPENSIVE NEIGHBORHOODS IN BARCELONA

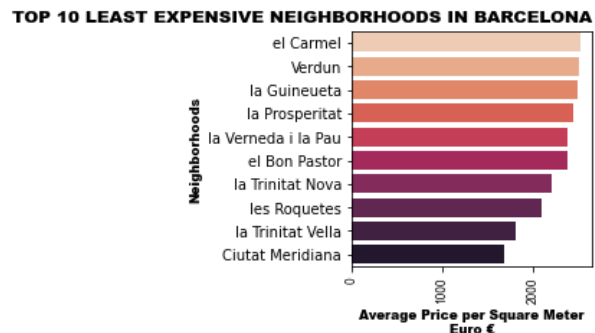


FIGURE 3: TOP 10 NEIGHBORHOODS WITH THE HIGHEST CAGR

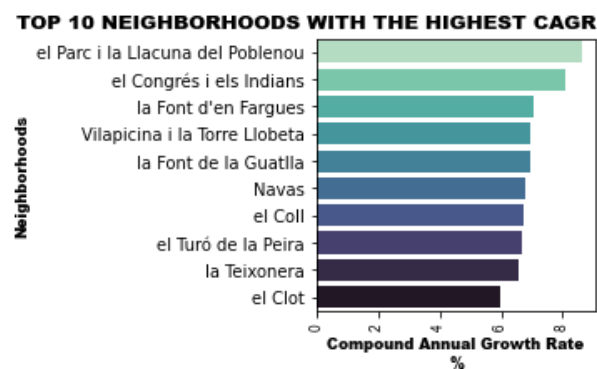
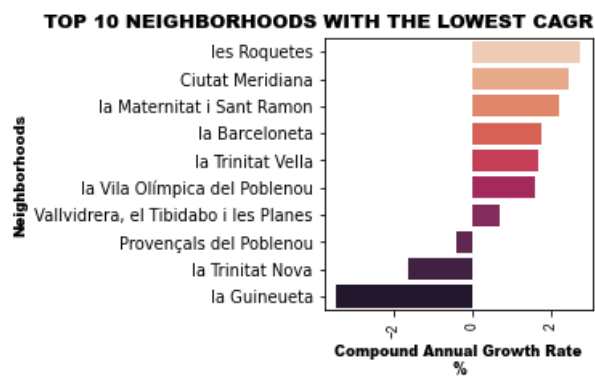


FIGURE 4: TOP 10 NEIGHBORHOODS WITH THE LOWEST CAGR

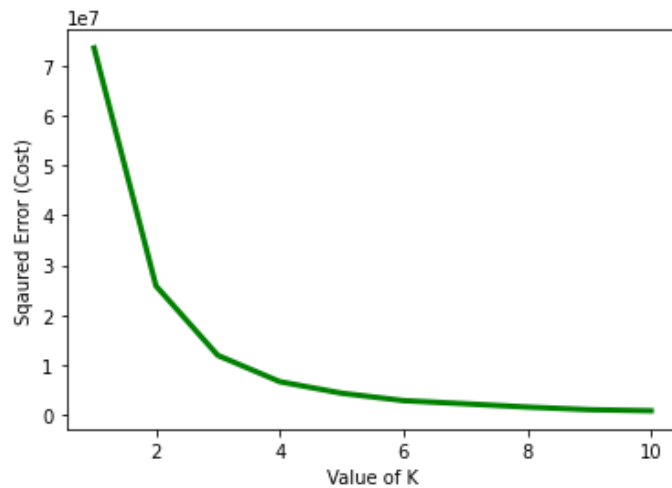


3.2. Clustering Neighborhoods

In order to create the clusters for the neighborhoods, we'll be applying the K-means clustering algorithm. K-means clustering is a type of unsupervised machine learning algorithm, which is used when you have data without defined categories or groups. The objective for using this algorithm is to find groups or clusters in the data, with the number of groups represented by the variable K. The algorithm assigns each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. (Trevino, 2016)

Before applying the K-means clustering algorithm, we need to define the number of clusters or the value of K, for which we'll be using the elbow method, which consist in plotting the mean distance to the centroid as a function of K and identifying the "elbow point," where the rate of decrease sharply shifts. (Trevino, 2016) In the figure below, we find that the "elbow point" is at K=4, therefore, the number of clusters that are retrieved from the K-means algorithm will be 4.

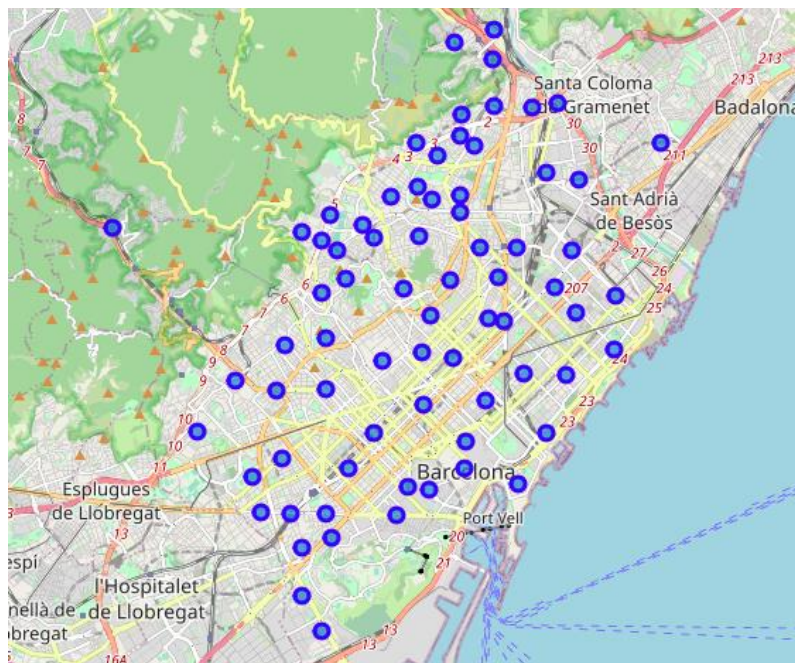
FIGURE 5: ELBOW METHOD PLOT



3.3. Mapping out Cluster Neighborhoods

In order to create the maps for displaying the multiple measurement indicators, the Folium library is imported. Folium is a Python library used for visualizing geospatial data. It is easy to use and powerful. Folium works as a Python wrapper for Leaflet.js which is a leading open-source JavaScript library for plotting interactive maps. Having extracted the coordinates for Barcelona and each of its neighborhoods, the initial folium map is created, displaying markers for each neighborhood.

FIGURE 6: ORIGINAL BARCELONA MAP WITH NEIGHBORHOOD MARKERS

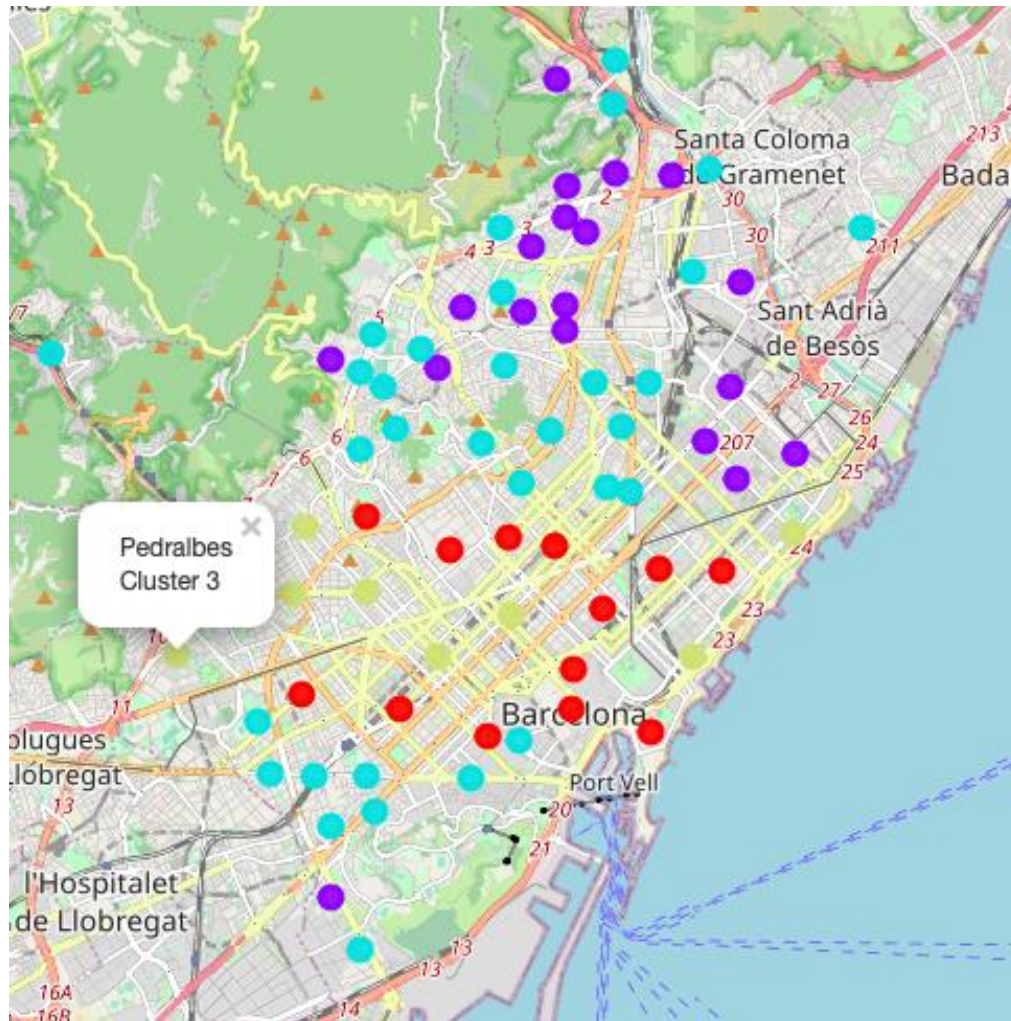


In the map below, we've mapped the city with markers in each neighborhood and the color indicating the cluster it was assigned to by the K-means clustering algorithm. The colors represent the following:

- Red: Cluster 0 – Upper Middle-Class Neighborhoods
- Purple: Cluster 1 – Outskirts
- Cyan: Cluster 2 – Lower Middle-Class Neighborhoods
- Green: Cluster 3 – High End Neighborhoods

As seen below, when you click in any markers, a pop-up label is displayed featuring the neighborhood name and the cluster it's assigned to.

FIGURE 7: CLUSTER MARKER MAP



SOURCE:

https://nbviewer.jupyter.org/github/kriszapata/Coursera_Capstone/blob/master/map_clusters.html

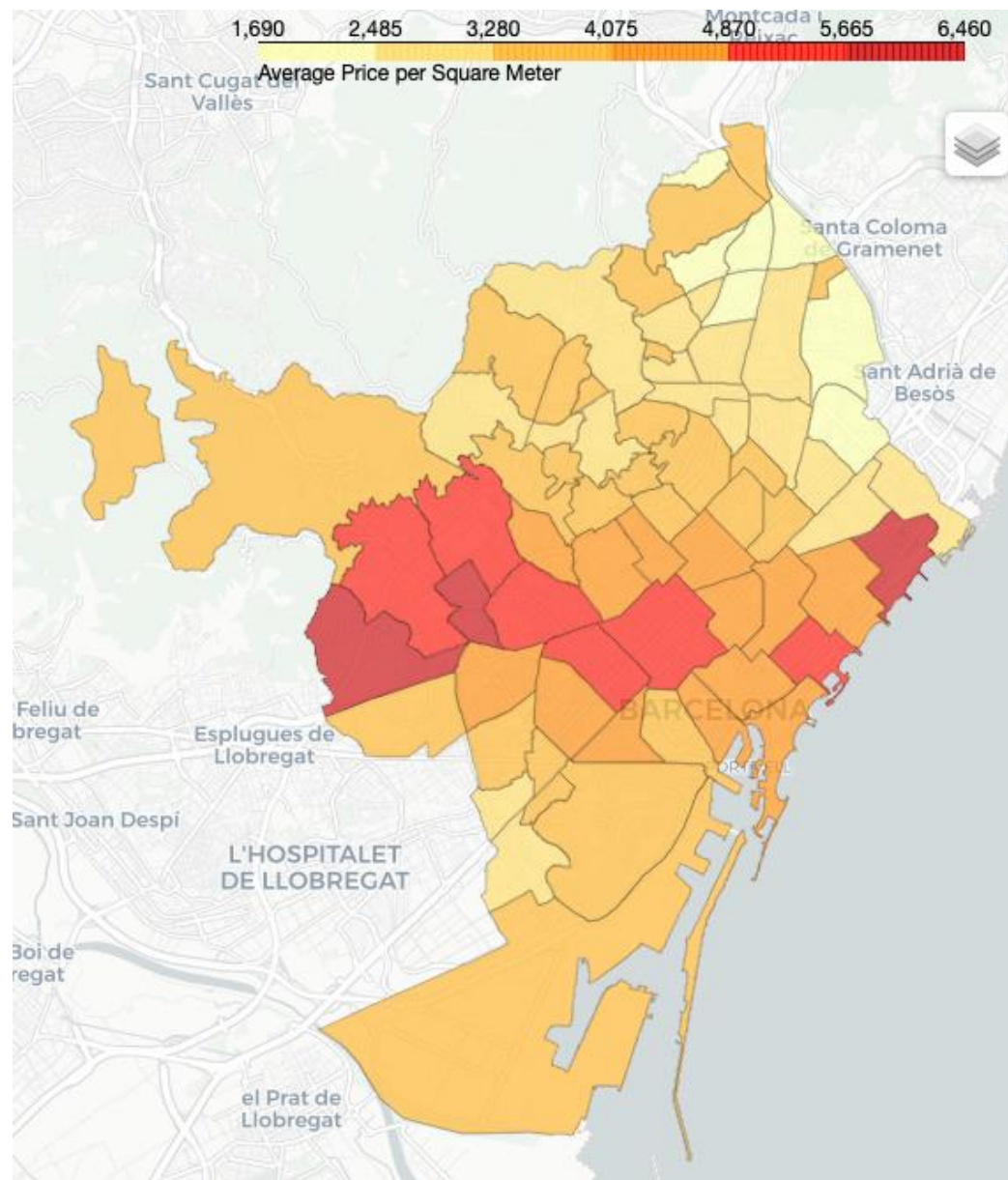
Having successfully created a map displaying the neighborhoods with their respective clusters, in order to display different features such as the average price per square meter and the compound annual growth rate, choropleth maps are created with each feature as the main indicator measured.

In order to create choropleth maps, geojson files of the city of Barcelona separated by district are needed, which are extracted from a GitHub repository from Martin Gonzalez. (González, GitHub, 2020) This particular file is loaded into the notebook and its later converted into a geopandas file utilizing the geopandas function, which turns

the data found in the geojson file into a data frame, making it easier to plot the figure onto the existing maps.

In the choropleth map below, we see the average price per sqm represented in each neighborhood, where the darker the color, the more expensive the neighborhood is. As observed in the map, most of the more expensive neighborhoods are located in the southwestern part of the city, with two exceptions along the beach-side.

FIGURE 8: CHOROPLETH MAP OF AVERAGE HOUSING PRICE



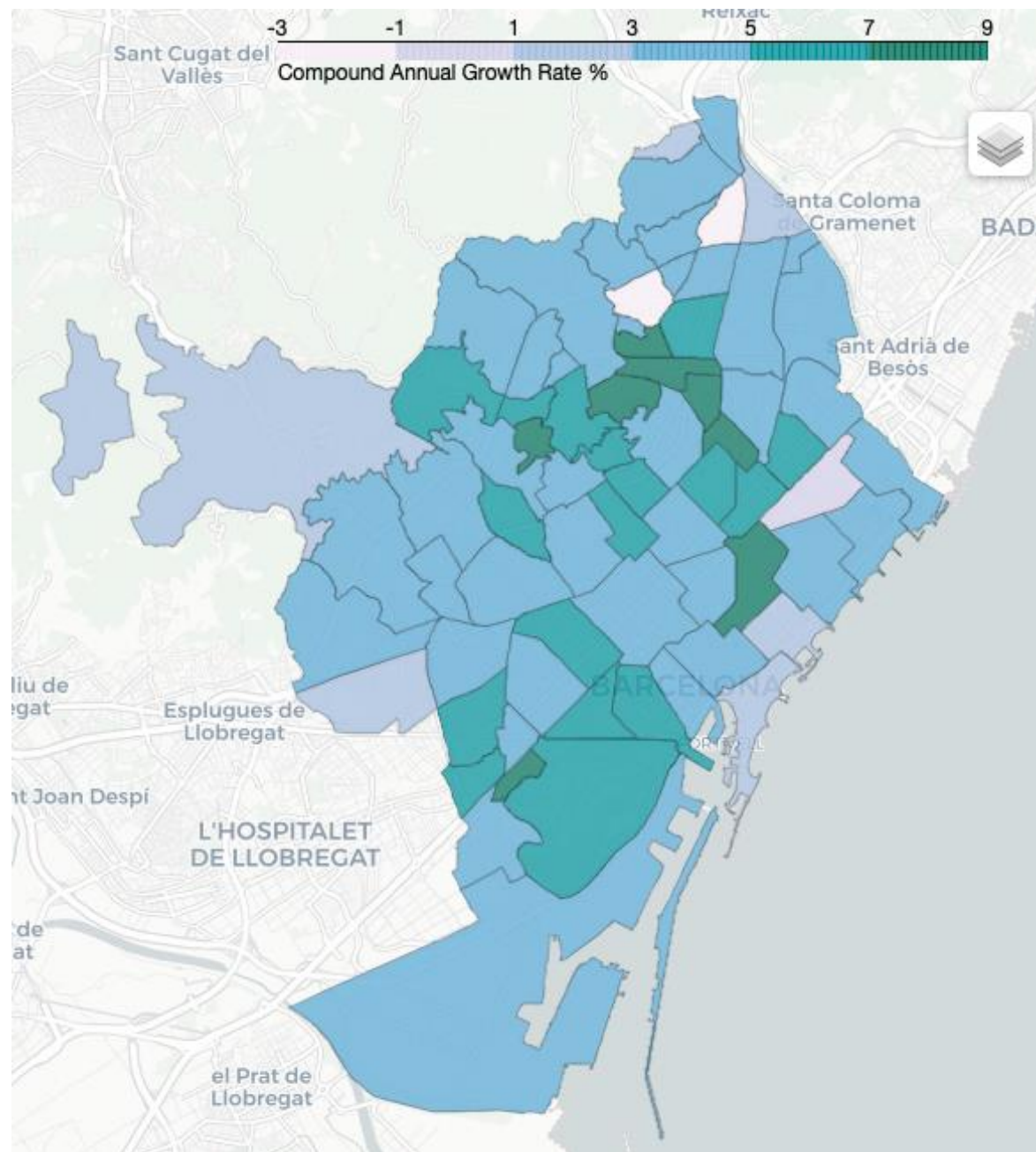
SOURCE:

https://nbviewer.jupyter.org/github/kriszapata/Coursera_Capstone/blob/master/AVG_PRICE_BCN.html

In the map below, we observe the compound annual growth rate per neighborhood, where most of the neighborhoods with the highest CAGR are close together. It's worth noting that these neighborhoods are close to Plaça de les Glòries, which is currently under construction but would turn into the new city center of Barcelona, as idealized by the famous Catalan urban planner Ildefons Cerdà, connecting all the major streets

of Barcelona as well as offering a green area destined to improve the quality of life of the city residents. (Transforming Barcelona: The future of Les Glòries, 2021)

FIGURE 9: CHOROPLETH MAP OF CAGR



SOURCE:

https://nbviewer.jupyter.org/github/kriszapata/Coursera_Capstone/blob/master/CAGR_BCN.html

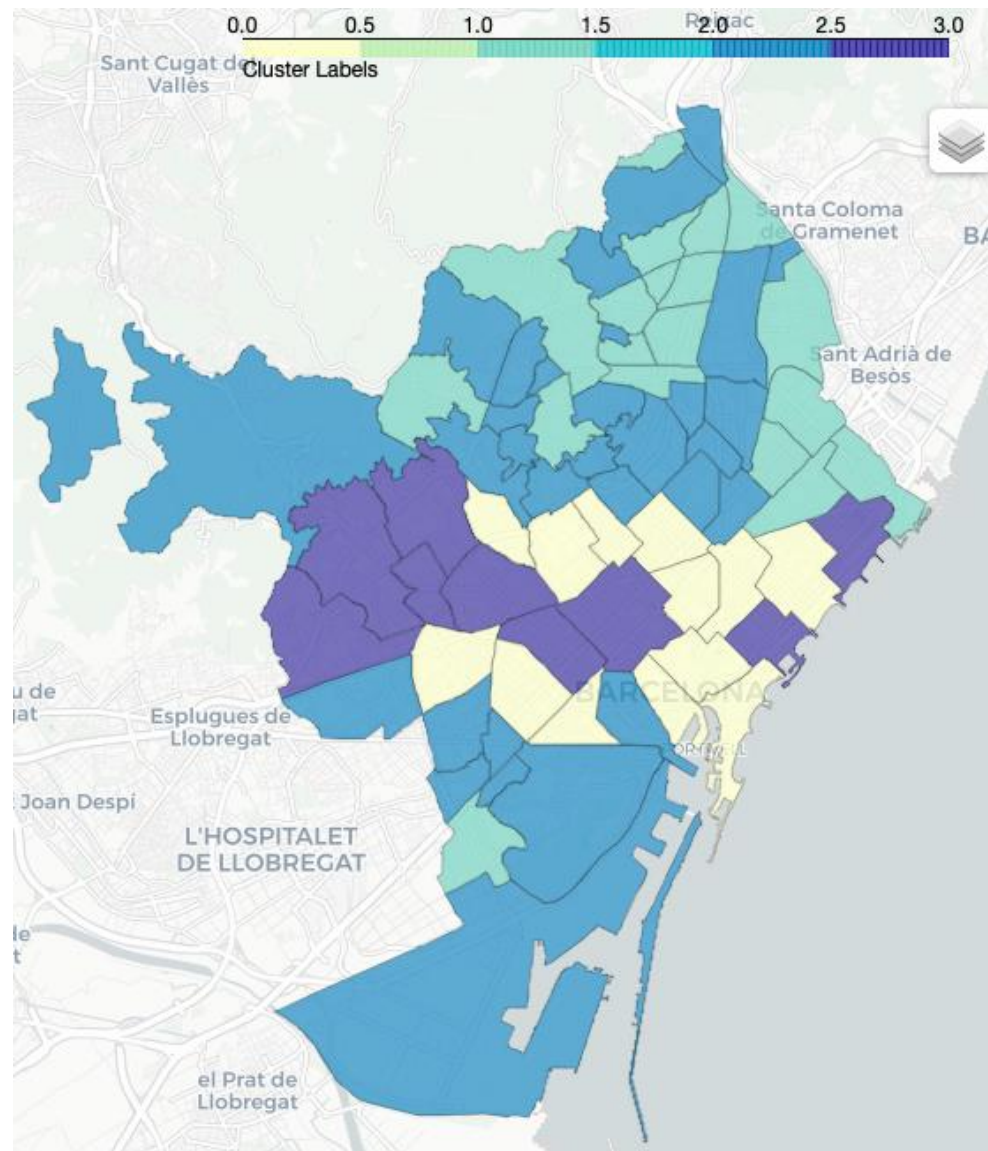
In the final map, the choropleth map is used to display the cluster labels, although the cluster labels represent categorical value, in this case since it's technically a numerical value it can be mapped out. The greatest benefit of this final map is the ability to access all of the most important information regarding each neighborhood such as:

- Neighborhood Name

- Cluster Label
- Average Price per SQM
- Compound Annual Growth Rate (CAGR)

It's important to note that the K-means algorithm used, included features such as average price per sqm and CAGR, as a result, the neighborhoods within each cluster will share similar values.

FIGURE 10: CHOROPLETH MAP OF NEIGHBORHOOD CLUSTER



SOURCE:

https://nbviewer.jupyter.org/github/kriszapata/Coursera_Capstone/blob/master/CLUSTERMAP_BCN.html

4. Results

Having successfully executed the K-means clustering algorithm and having previously explored which neighborhoods are within each cluster, it's time to deep dive into each cluster to get a better understanding of the cluster profile. In order to create the cluster profile, there are 4 main components taken into account:

1. Average price per sqm
2. CAGR
3. Most common venues within each cluster
4. Personal experience of having lived in the city for 3+ years.

4.1. Cluster Comparison

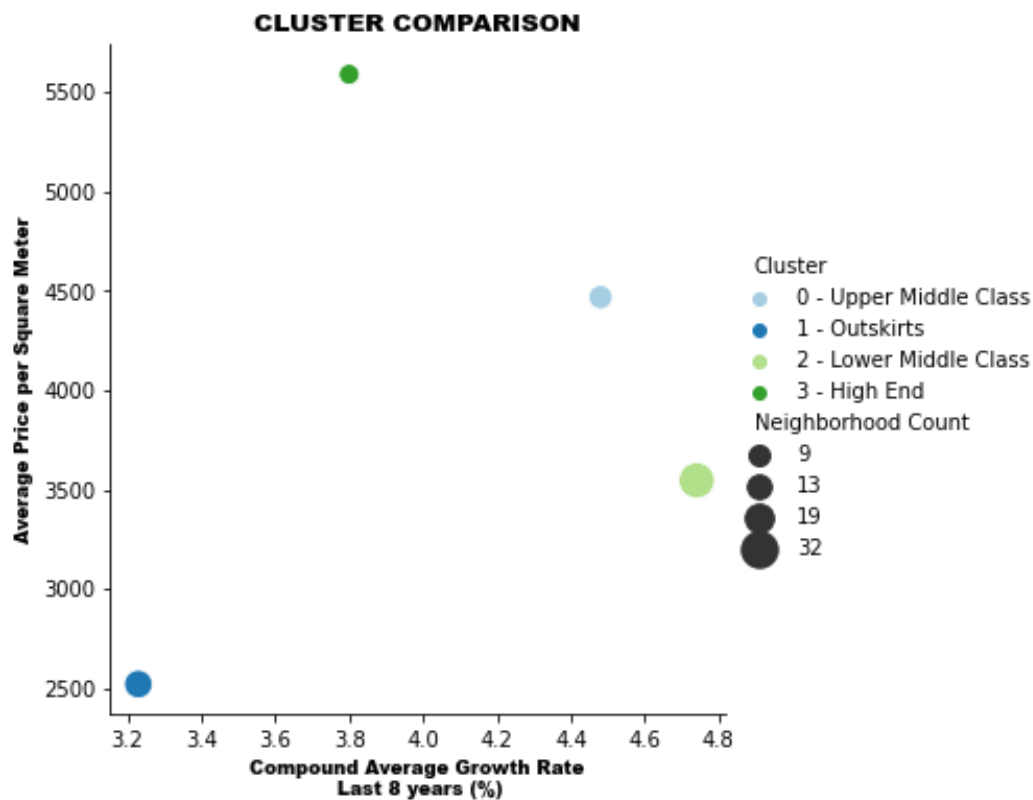
Before examining each cluster and the neighborhoods included within it is necessary to conduct a cluster comparison in order to identify which cluster has the best performance in terms of average price per sqm and CAGR. It is necessary as well to identify the size of each cluster, based on the number of neighborhoods that integrate them. The size of the cluster is representative of the variety, meaning there's more options for the home buyers to consider. On the other hand, it can present a dilemma to home buyers considering there will be bigger differences between the neighborhoods within the cluster.

TABLE 9: CLUSTER COMPARISON

	Cluster	Average Housing Price Index	CAGR	Neighborhood Count
0	0 - Upper Middle Class	4468	4.48	13
1	1 - Outskirts	2521	3.23	19
2	2 - Lower Middle Class	3546	4.74	32
3	3 - High End	5589	3.80	9

As seen in the figures 35 and 36, Cluster 2 is the best option for purchasing a home in terms of investment, considering it's the second least expensive cluster and has the best performance in the CAGR, translating to a low investment with high returns. Although purchasing a house is an investment and home buyers will value this factor highly, it would be a mistake to overlook other factors such as the most common venues within each neighborhood. After all, a deciding factor in the decision process can come down to the venues nearby. For example, a family with kids would not want to live in a neighborhood full of bars and discos, due to the noise and the crowds they attract.

FIGURE 11: BUBBLE PLOT OF CLUSTER COMPARISON



4.2. Cluster 0 - Upper Middle Class:

With the exception of the inclusion of el Putxet i el Farró, most of the neighborhoods included in this cluster are within the city center, located close to tourist attractions such as the Rambla, the Sagrada Familia, and Camp Nou. I would say that the Putxet I el Farró was classified within this cluster due to the average housing price, considering that it's closer to these neighborhoods than to the neighborhoods in cluster 3. Excluding el Putxet i el Farró, the average profile of the people living in this area are: young people, in their 20s and 30s, most of them sharing a flat because rent is too high, working, studying or both, wanting to be near where most of the action happens in the city.

TABLE 10: CLUSTER 0 - UPPER MIDDLE CLASS NEIGHBORHOODS

	Neighborhood	Avg Price per m2	CAGR	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	el Barri Gòtic	4.717	3,09	Tapas Restaurant	Ice Cream Shop	Plaza	Spanish Restaurant	Italian Restaurant
1	el Camp d'en Grassot i Gràcia Nova	4.399	4,91	Tapas Restaurant	Bakery	Restaurant	Hotel	Spanish Restaurant
2	el Fort Pienc	4.091	4,34	Coffee Shop	Mediterranean Restaurant	Sushi Restaurant	Spanish Restaurant	Restaurant
3	el Parc i la Llacuna del Poblenou	4.567	8,63	Hotel	Bar	Mediterranean Restaurant	Clothing Store	Restaurant
4	el Poblenou	4.557	4,07	Spanish Restaurant	Restaurant	Mediterranean Restaurant	Bakery	Italian Restaurant
5	el Putxet i el Farró	4.532	4,65	Bakery	Italian Restaurant	Café	Wine Bar	Burger Joint
6	la Barceloneta	4.763	1,77	Tapas Restaurant	Mediterranean Restaurant	Paella Restaurant	Bar	Spanish Restaurant
7	la Nova Esquerra de l'Eixample	4.312	4,55	Mediterranean Restaurant	Café	Indian Restaurant	Supermarket	Bakery
8	la Sagrada Família	4.088	4,38	Restaurant	Hotel	Bakery	Plaza	Sandwich Place
9	la Vila de Gràcia	4.343	4,16	Plaza	Mediterranean Restaurant	Bakery	Pizza Place	Bar
10	les Corts	4.654	3,75	Spanish Restaurant	Restaurant	Bakery	Hotel	Café
11	Sant Antoni	4.219	5,37	Café	Cocktail Bar	Mediterranean Restaurant	Bar	Coffee Shop
12	Sant Pere, Santa Caterina i la Ribera	4.837	4,59	Hotel	Tapas Restaurant	Japanese Restaurant	Breakfast Spot	Bistro

As we can see in the Statistical Analysis below, there are 13 neighborhoods included within this cluster. The average price per sqm ranges from 4.088€ to 4.837€, making it the second most expensive cluster but it also has the second highest CAGR from all the clusters, making it highly attractive considering these neighborhoods share privileged locations.

TABLE 11: CLUSTER 0 - STATISTICAL ANALYSIS

	Avg Price per m2	CAGR
count	13	13
mean	4.468	4,48
std	247	1,54
min	4.088	1,77
25%	4.312	4,07
50%	4.532	4,38
75%	4.654	4,65
max	4.837	8,63

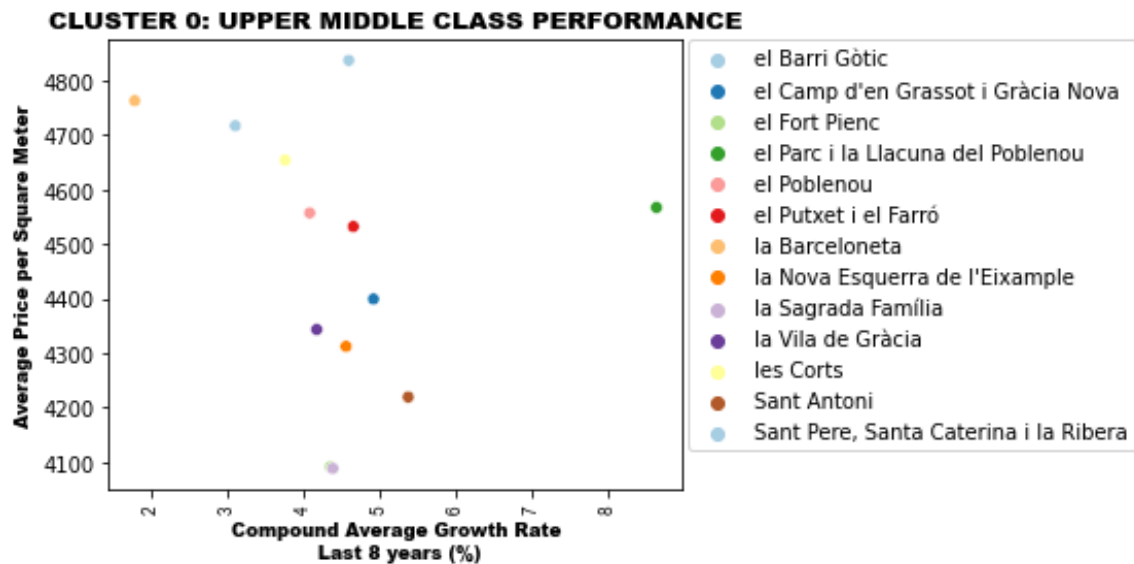
As we can observe in the data frame below, the most common venues seem to be tapas restaurants, Mediterranean or Spanish restaurants, which seems logical since these neighborhoods are popular among tourist due to the proximity to major tourist attractions.

TABLE 12: CLUSTER 0 - STATISTICAL ANALYSIS OF CATEGORICAL VARIABLES

	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	10th Most Common Venue
count	13	13	13	13	13	13	13	13	13	13	13
unique	13	8	10	8	10	10	11	12	11	12	11
top	el Fort Pienc	Tapas Restaurant	Mediterranean Restaurant	Bakery	Spanish Restaurant	Spanish Restaurant	Supermarket	Spanish Restaurant	Restaurant	Supermarket	Tapas Restaurant
freq	1	3	3	3	3	4	2	2	2	2	2

As seen in the scatter plot below, the neighborhoods within this cluster are performing considerably well in terms of CAGR, especially el Parc i la Llacuna del Poblenou which has the highest CAGR of all the neighborhoods in Barcelona. Most of the neighborhoods within this cluster have a higher CAGR than the city average of 4,2%, making it an attractive cluster to look into.

FIGURE 12: SCATTER PLOT OF CLUSTER 0 NEIGHBORHOODS



4.3. Cluster 1 – Outskirts

The most important characteristics of this cluster is the fact that these are the least expensive neighborhoods in Barcelona, mostly located in the outskirts of the city. In this case, it's harder to define the profile of the average person living in these neighborhoods, but, considering the most common venues and the average price per sqm, within this cluster mostly reside low income families, who have jobs in the city, therefore need good connection to the rest of the city.

TABLE 13: CLUSTER 1 - OUTSKIRTS

	Neighborhood	Avg Price per m2	CAGR	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Ciutat Meridiana	1.690	2,46	Metro Station	Park	Grocery Store	Supermarket	Plaza
1	eI Besòs i el Maresme	2.547	4,09	Diner	Hotel	Gym / Fitness Center	Spanish Restaurant	Café
2	eI Bon Pastor	2.380	4,02	Tapas Restaurant	Paper / Office Supplies Store	Café	Athletics & Sports	Organic Grocery
3	eI Carmel	2.532	5,59	Gym	Coffee Shop	Grocery Store	Plaza	Supermarket
4	eI Turó de la Peira	2.534	6,68	Café	Burger Joint	Supermarket	Tapas Restaurant	Pizza Place
5	Horta	2.953	3,29	Tapas Restaurant	Spanish Restaurant	Pizza Place	Mediterranean Restaurant	Restaurant
6	la Guineueta	2.498	- 3,47	Grocery Store	Seafood Restaurant	Plaza	Tapas Restaurant	Breakfast Spot
7	la Marina de Port	2.861	4,15	Spanish Restaurant	Grocery Store	Café	Breakfast Spot	Asian Restaurant
8	la Prosperitat	2.446	2,91	Grocery Store	Spanish Restaurant	Tapas Restaurant	Park	Brewery
9	la Trinitat Nova	2.200	- 1,66	Metro Station	Pharmacy	Park	Breakfast Spot	Music Venue
10	la Trinitat Vella	1.809	1,66	Park	Soccer Field	Tapas Restaurant	Breakfast Spot	Spanish Restaurant
11	la Verneda i la Pau	2.389	3,15	Grocery Store	Food	Coffee Shop	Pedestrian Plaza	Pharmacy
12	les Roquetes	2.097	2,76	Music Venue	Supermarket	Gym	Grocery Store	Falafel Restaurant
13	Porta	2.754	5,39	Tapas Restaurant	Spanish Restaurant	Grocery Store	Café	Clothing Store
14	Provençals del Poblenou	2.994	- 0,41	Spanish Restaurant	Café	Asian Restaurant	Steakhouse	Falafel Restaurant
15	Sant Genís dels Agudells	2.889	5,02	Sports Bar	Grocery Store	Diner	Tennis Court	Tapas Restaurant
16	Sant Martí de Provençals	2.867	4,75	Supermarket	Spanish Restaurant	Grocery Store	Pizza Place	Soccer Field
17	Verdun	2.503	4,06	Grocery Store	Spanish Restaurant	Park	Bar	South American Restaurant
18	Vilapicina i la Torre Llobeta	2.953	6,97	Spanish Restaurant	Tapas Restaurant	Plaza	Grocery Store	Bar

As mentioned before, this cluster is the less expensive cluster to buy a home in, with an average price per sqm of 2.521€. However, it has the lowest CAGR compared to the rest, therefore, investors should be weary of this fact if looking for a sustained return on investment over a prolonged time.

TABLE 14: CLUSTER 1 - STATISTICAL ANALYSIS

	Avg Price per m2	CAGR
count	19	19
mean	2.521	3,23
std	376	2,69
min	1.690	- 3,47
25%	2.385	2,61
50%	2.532	4,02
75%	2.864	4,88
max	2.994	6,97

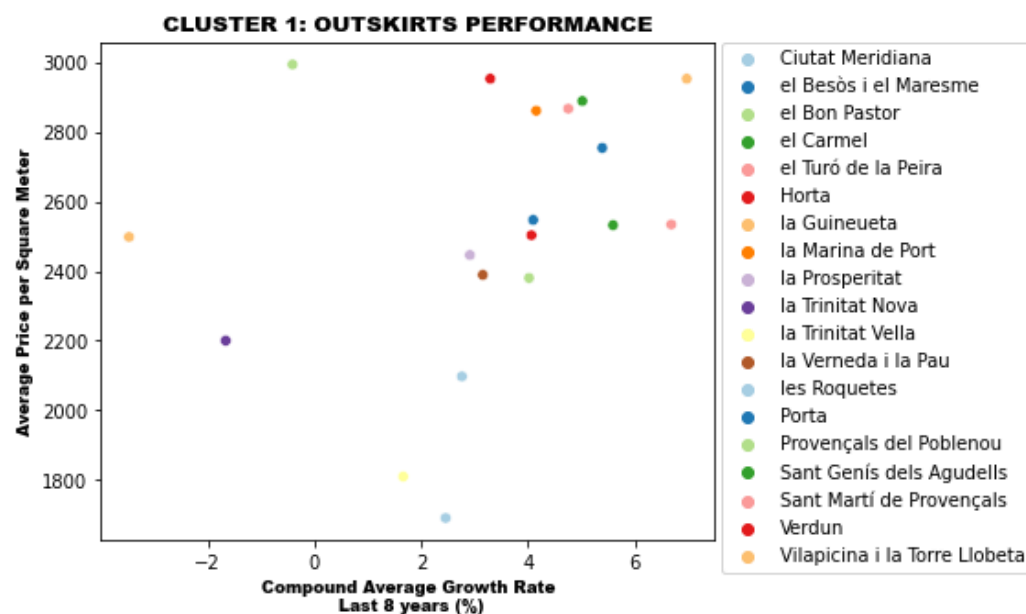
Based on it's most common venues, the cluster profile is a combination of local, low income class families due to the high amount of grocery stores, supermarkets and parks.

TABLE 15: CLUSTER 1 - STATISTICAL ANALYSIS OF CATEGORICAL VARIABLES

	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	10th Most Common Venue
count	19	19	19	19	19	19	19	19	19	19	19
unique	19	9	10	14	15	16	18	17	17	18	17
top	Provençals del Poblenou	Grocery Store	Café	Park	Mediterranean Restaurant	Supermarket	Spanish Restaurant	Supermarket	Other Great Outdoors	Supermarket	Pizza Place
freq	1	5	4	3	3	3	2	2	2	2	2

There are various neighborhoods that have a very high CAGR, like Vilapicina I la Torre Llobeta and el Turó de la Peira. These neighborhoods could be a good investment opportunity due to its performance (low price and high CAGR).

FIGURE 13: SCATTER PLOT OF CLUSTER 1 NEIGHBORHOODS



4.4. Cluster 2 – Lower Middle Class

Based on the average housing price, these neighborhoods target the lower middle class, which is the vast majority of the city population.

TABLE 16: CLUSTER 2 - LOWER MIDDLE CLASS NEIGHBORHOODS

	Neighborhood	Avg Price per m2	CAGR	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Baró de Viver	3.695	4,19	Metro Station	Plaza	Burger Joint	Track Stadium	Supermarket
1	Can Baró	3.339	5,89	Spanish Restaurant	Chinese Restaurant	Grocery Store	Tapas Restaurant	Scenic Lookout
2	Can Peguera	3.695	4,19	Park	German Restaurant	Food & Drink Shop	Café	Sports Club
3	Canyelles	3.695	4,19	Soccer Field	Construction & Landscaping	Food & Drink Shop	Grocery Store	Market
4	eI Baix Guinardó	3.611	4,76	Tapas Restaurant	Restaurant	Japanese Restaurant	Bakery	Bar
5	eI Camp de l'Arpa del Clot	3.555	5,27	Bakery	Spanish Restaurant	Café	Hotel	Pizza Place
6	eI Clot	3.767	5,98	Café	Spanish Restaurant	Restaurant	Hotel	Bakery
7	eI Coll	3.420	6,70	Scenic Lookout	Hotel	Park	Mountain	Supermarket
8	eI Congrés i els Indians	3.271	8,10	Tapas Restaurant	Restaurant	Grocery Store	Spanish Restaurant	Plaza
9	eI Guinardó	3.304	4,22	Café	Tapas Restaurant	Plaza	Breakfast Spot	Restaurant
10	eI Poble Sec	3.716	5,86	Tapas Restaurant	Pizza Place	Café	Mediterranean Restaurant	Cocktail Bar
11	eI Raval	3.587	4,62	Cocktail Bar	Spanish Restaurant	Bar	Tapas Restaurant	Pizza Place
12	Hostafrancs	3.467	4,13	Tapas Restaurant	Spanish Restaurant	Hotel	Mediterranean Restaurant	Pizza Place
13	la Bordeta	3.128	5,08	Restaurant	Theater	Grocery Store	Pizza Place	Burger Joint
14	la Clota	3.695	4,19	Coffee Shop	Supermarket	Museum	Grocery Store	Gym
15	la Font d'en Fargues	3.705	7,06	Italian Restaurant	Supermarket	Restaurant	Cocktail Bar	Vegetarian / Vegan Restaurant
16	la Font de la Guatlla	3.856	6,93	Spanish Restaurant	Plaza	Nightclub	Grocery Store	Tapas Restaurant
17	la Marina del Prat Vermell	3.701	4,19	Spanish Restaurant	Furniture / Home Store	Bakery	Supermarket	Restaurant
18	la Maternitat i Sant Ramon	3.918	2,21	Spanish Restaurant	Pizza Place	Italian Restaurant	Mediterranean Restaurant	Supermarket
19	la Sagrera	3.250	4,28	Spanish Restaurant	Grocery Store	Plaza	Bar	Burger Joint
20	la Salut	3.547	2,97	Clothing Store	Shopping Mall	Miscellaneous Shop	Fast Food Restaurant	Italian Restaurant
21	la Teixonera	3.138	6,55	Spanish Restaurant	Mediterranean Restaurant	Metro Station	Café	Spa
22	la Vall d'Hebron	3.293	4,19	Spanish Restaurant	Soccer Stadium	Athletics & Sports	Diner	Stadium
23	Montbau	3.695	4,19	Park	Plaza	Gym	Soccer Stadium	Soccer Field
24	Navas	3.503	6,79	Supermarket	Bakery	Grocery Store	Tapas Restaurant	Spanish Restaurant
25	Sant Andreu	3.161	4,45	Clothing Store	Tapas Restaurant	Fast Food Restaurant	Spanish Restaurant	Burger Joint
26	Sants	3.705	5,00	Tapas Restaurant	Hostel	Mediterranean Restaurant	Bar	Plaza
27	Sants-Badal	3.239	3,33	Pizza Place	Tapas Restaurant	Hostel	Bakery	Supermarket
28	Torre Baró	3.695	4,19	Train Station	Construction & Landscaping	Metro Station	Yoga Studio	Other Great Outdoors
29	Vallbona	3.695	4,19	Park	Light Rail Station	Train Station	Plaza	Construction & Landscaping
30	Vallcarca i els Penitents	3.845	3,08	Café	Supermarket	Plaza	Grocery Store	Park
31	Vallvidrera, el Tibidabo i les Planes	3.579	0,68	Train Station	Restaurant	BBQ Joint	Yoga Studio	Other Great Outdoors

This is the biggest cluster of all, composed of 32 neighborhoods with an average price per sqm of 3.546€. It's also worth noticing that this neighborhood cluster has the highest CAGR of all the clusters, with a CAGR of 4,7%.

TABLE 17: CLUSTER 2 - STATISTICAL ANALYSIS

	Avg Price per m2	CAGR
count	32	32
mean	3,546	4,74
std	225	1,52
min	3,128	0,68
25%	3,330	4,19
50%	3,599	4,25
75%	3,697	5,87
max	3,918	8,10

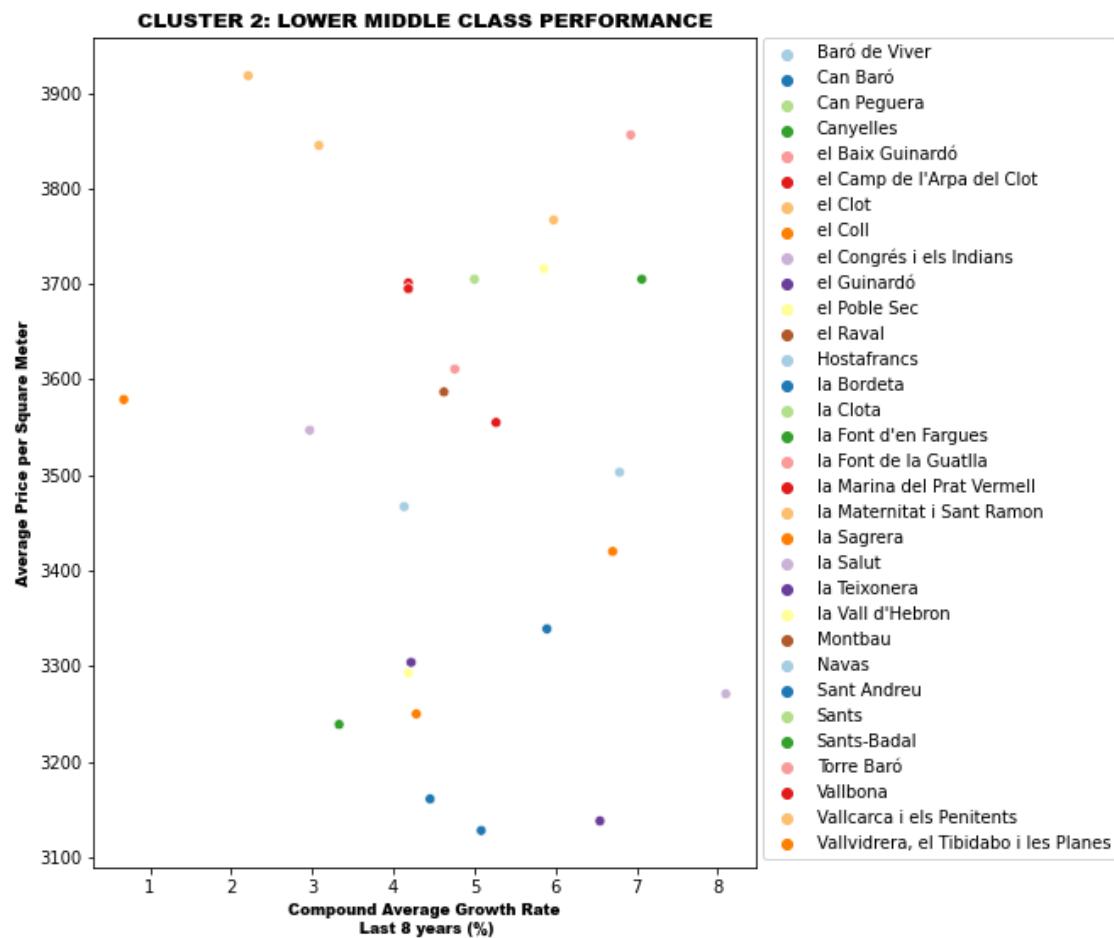
With the exception of el Raval which is overrun by vacation rentals, most of the neighborhoods in this cluster are composed of the following profile: middle class family that enjoy "Vida de Barrio", which means spending time with their friends and families within their own neighborhood, eating and drinking at nearby restaurants and bars, playing in the parks and plazas nearby. The venues listed are a reflection of the lifestyle.

TABLE 18: CLUSTER 2 - STATISTICAL ANALYSIS OF CATEGORICAL VARIABLES

	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	10th Most Common Venue
count	32	32	32	32	32	32	32	32	32	32	32
unique	32	15	20	21	18	17	25	24	25	25	23
top	Navas	Spanish Restaurant	Spanish Restaurant	Restaurant	Plaza	Grocery Store	Café	Restaurant	Breakfast Spot	Other Great Outdoors	Gym / Fitness Center
freq	1	6	7	5	4	4	2	4	3	3	3

As we can see in the scatter plot below, this cluster, obviously due to its size, is the most diverse cluster. In comparison to the other clusters, the data points, which represents neighborhoods, are significantly distanced, without a clear visible cluster within. That being said, this cluster is still the highest performing cluster in regard to the CAGR, and has the second lowest average price per sqm, making it the most attractive cluster to find a neighborhood to purchase a home.

FIGURE 14: SCATTER PLOT OF CLUSTER 2 NEIGHBORHOODS



4.5. Cluster 3 – High End

Within this cluster, there are 3 inner clusters that define its composition.

The first inner cluster, La Dreta de l'Eixample and l'Antiga Esquerra de l'Eixample are probably the best located neighborhoods in the city. They are located right in the middle of the city, but not in Ciutat Vella district which can be very noisy and crowded, but still surrounded by very nice, cool places nearby with very beautiful streets with a variety of retail and restaurants/bars offering to enjoy like Enric Granados and Rambla de Catalunya.

The outsiders of the cluster would be la Vila Olímpica de Poble Nou and Diagonal Mar, which are located at the beach side of the city. These neighborhoods represent the new and very modern areas of the city, mostly destined for beach side properties and offices.

The rest of the neighborhoods within this cluster are very similar and are close in proximity, representing very traditional, "old money", composed mostly of upper middle, high class families looking for a much more quiet city lifestyle but still want easy, quick access to the city center in case they ever want to.

TABLE 19: CLUSTER 3 - HIGH END NEIGHBORHOODS

	Neighborhood	Avg Price per m2	CAGR	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Diagonal Mar i el Front Marítim del Poblenou	6.237,0	4,06	Restaurant	Mediterranean Restaurant	Beach	Hotel	Beach Bar
1	l'Antiga Esquerra de l'Eixample	5.190,0	5,70	Hotel	Spanish Restaurant	Cocktail Bar	Mediterranean Restaurant	Japanese Restaurant
2	la Dreta de l'Eixample	5.395,0	3,31	Hotel	Boutique	Mediterranean Restaurant	Bakery	Tapas Restaurant
3	la Vila Olímpica del Poblenou	5.267,0	1,58	Hookah Bar	Café	Restaurant	Mediterranean Restaurant	Nightclub
4	Ies Tres Torres	5.749,0	3,53	Hotel	Spanish Restaurant	Train Station	Mexican Restaurant	Burger Joint
5	Pedralbes	6.460,0	3,99	Spanish Restaurant	Garden	Mediterranean Restaurant	Hot Dog Joint	Bar
6	Sant Gervasi - Galvany	5.326,0	4,56	Mediterranean Restaurant	Café	Japanese Restaurant	Sandwich Place	Italian Restaurant
7	Sant Gervasi - la Bonanova	5.115,0	4,08	Mediterranean Restaurant	Bakery	Seafood Restaurant	Japanese Restaurant	Bar
8	Sarrià	5.558,0	3,45	Bakery	Mediterranean Restaurant	Café	Pedestrian Plaza	Diner

As seen by the Average price per sqm, these are the most expensive neighborhoods to purchase a home in Barcelona. With an average price per sqm of 5.589€ and a CAGR of 3,8%, this is not the best cluster to purchase a home as an investment opportunity, considering the elevated price and low CAGR.

TABLE 20: CLUSTER 3 - STATISTICAL ANALYSIS

	Avg Price per m2	CAGR
count	9	9
mean	5.589	3,80
std	474	1,10
min	5.115	1,58
25%	5.267	3,45
50%	5.395	3,99
75%	5.749	4,08
max	6.460	5,70

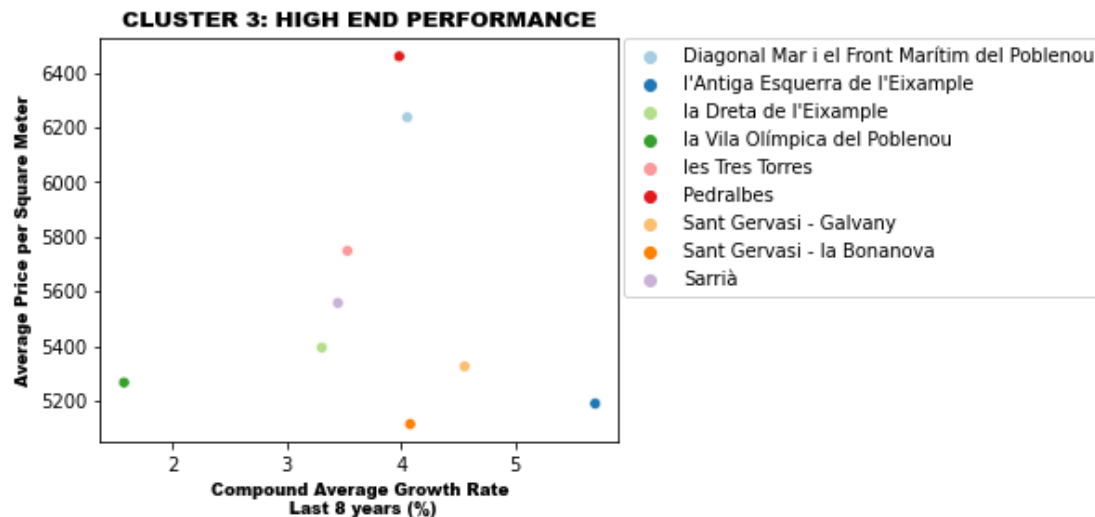
As seen below, the most common venues include more expensive types of restaurants like seafood and Japanese, which matches the profile of the people in this cluster. The two most common venue in this cluster are bakeries and Mediterranean restaurants, which go in line with the local habit of eating out most meals a day, mostly breakfast and lunch.

TABLE 21: CLUSTER 3 - STATISTICAL ANALYSIS OF CATEGORICAL VARIABLES

	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	10th Most Common Venue
count	9	9	9	9	9	9	9	9	9	9	9
unique	9	7	8	9	5	8	8	9	8	9	8
top	Sant Gervasi - la Bonanova	Bakery	Mediterranean Restaurant	Seafood Restaurant	Japanese Restaurant	Italian Restaurant	Sandwich Place	Bakery	Restaurant	Spanish Restaurant	Restaurant
freq	1	2	2	1	3	2	2	1	2	1	2

In the scatter plot below, it's important to highlight the best performing neighborhood within this cluster, which is l'Antiga Esquerra de l'Eixample, which has the second lowest price but the highest CAGR, positioning itself as the most promising neighborhood within the cluster.

FIGURE 15: SCATTER PLOT OF CLUSTER 3 NEIGHBORHOODS



5. Discussion

With the results presented above, there's a cluster for every type of home buyer. The profiles are as follows:

- For individuals or businesses looking to purchase a house to use as a vacation rental or short-term rental, Cluster 0 and Cluster 2 are a perfect match, especially Cluster 0 whose neighborhoods are close to the tourist attractions.
- Cluster 1 satisfies the need for a low-income family wishing to purchase a home within the city and with easy access to the city center.
- Individuals or businesses looking for investment opportunities in properties with relatively low cost and high compound annual growth rate would invest in Cluster 2.
- High-income families looking to purchase a home in very nice neighborhoods, with neighbors that share similar lifestyles would purchase a home within Cluster 3.
- Businesses looking to purchase property for their installations, wanting to be near the business hubs in the city would most likely purchase from Cluster 0 or Cluster 3.

The results are fairly accurate when it comes to segmenting Barcelona's neighborhoods and identifying the similarities certain neighborhoods share, however, there are certain factors that could have been included in the K-means clustering algorithm, which could have improved not only the clustering process, but narrowing down the home buyer's decision process regarding which neighborhoods are ideal for purchasing a house. One factor that could be decisive is the crime stats per neighborhood, which definitely would be relevant for the home purchaser. Public

transport accessibility would normally be a key factor for purchasing a house, however, Barcelona has a superb public transport system, where you have guaranteed options for moving around the city, whether it be by bus, metro, train, or trans. Additionally, Barcelona has many cycling roads available, facilitating movement around the city through bicycles, electric scooters and other alternative means of transportation.

6. Conclusion

In conclusion, this Project fulfills its purpose in aiding people or businesses looking to invest in residential real estate in Barcelona, by helping them narrow down their search for the right investment, taking into account their budgets, investment objectives, possible returns on investment and last but not least, the type of neighborhood desired. From an investment standpoint, Cluster 2 has been identified as the best cluster to purchase a home in.

Utilizing data science methodology, this project provided a better understanding of the residential real estate market in Barcelona, as well as valuable insights regarding each neighborhood and its most common venues. The future of this project includes taking in account other factors such as crime rate, population density, and demographics, all of which would provide the home buyer with a more informed decision.

Bibliography

- PopulationStat*. (2020, July 9). Retrieved from PopulationStat: <https://populationstat.com/spain/barcelona>
- Idealista*. (2020, Noviembre). From Idealista: <https://www.idealista.com/sala-de-prensa/informes-precio-vivienda/alquiler/cataluna/barcelona-provincia/>
- Ajuntament de Barcelona. Departament d'Estadística i Difusió de Dades*. (2020). From Ajuntament de Barcelona. Departament d'Estadística i Difusió de Dades: <https://www.bcn.cat/estadistica/angles/dades/timm/ipreus/hab2mave/evo/t2mab.htm>
- O'Sullivan, F. (2018, June 6). *Barcelona Finds a Way to Control Its Airbnb Market*. From Bloomberg CityLab: <https://www.bloomberg.com/news/articles/2018-06-06/how-barcelona-is-limiting-airbnb-rentals>
- González, M. (2020, November 3). *Github*. From Github: <https://github.com/martgnz/bcn-geodata>
- Trevino, A. (2016, December 6). *Introduction to K-means Clustering*. From Oracle AI and Data Science Blog: <https://blogs.oracle.com/datascience/introduction-to-k-means-clustering>
- Transforming Barcelona: The future of Les Glòries*. (2021). From Sivoris: <https://sivoris.com/transforming-barcelona-the-future-of-les-glories/>
- Record number of the foreign population in Barcelona*. (n.d.). From www.barcelona.cat: https://www.barcelona.cat/internationalwelcome/en/noticia/record-number-of-the-foreign-population-in-barcelona_842387
- Explorer Publishing. (2020, December 28). *Study in Barcelona: universities and higher education institutions*. From Expatica: <https://www.expatica.com/es/education/higher-education/study-in-barcelona-102817/>
- BarcelonaYellow. (2020, October 26). *How many tourists visited Barcelona in 2019*. From BarcelonaYellow: <https://www.barcelonayellow.com/barcelona-faq/1035-how-many-tourists-visited-barcelona>
- Novoselteva, E. (2019, September 05). *BARCELONA LLEGA AL TOP 6 DE LOS HUBS DE INNOVACIÓN*. From Apium Hub: <https://apiumhub.com/es/tech-blog-barcelona/barcelona-hubs-de-innovacion/>
- Roberts, D. (2019, April 29). *Barcelona's remarkable history of rebirth and transformation*. From Vox: <https://www.vox.com/energy-and-environment/2019/4/8/18266760/barcelona-spain-urban-planning-history>

