Applied Data Science Capstone Project for IBM/Coursera

The Battle of the Neighborhoods
(Canary Islands Edition)

Author:

Erich Zehl

Date: August 8, 2020

Table of Contents

1. Introduction

1.1 Background

The Canary Islands are a Spanish territory composed of a chain of volcanic islands located off the coast of Morocco in the Atlantic Ocean. The islands are a very popular holiday destination and arguably one of the best places to retire due to its great weather, excellent healthcare system and affordability among others.

1.2 Business Problem and Interest

This project explores and compares different characteristics of the different islands of the Canary Islands (*Tenerife*, *Fuerteventura*, *Gran Canaria*, *Lanzarote*, *La Palma*, *La Gomera* and *El Hierro*) from data obtained from several different sources with the goal of helping any current or future retirees who might be considering relocating to the Canary Islands determine the best location to relocate on the islands.

The project generates a shortlist of promising suitable properties based on the following preferences or criteria: good climate (i.e. low rainfall, lots of sun, etc.), closeness to amenities and essential services (i.e. stores, entertainment and leisure options, pharmacies, etc.), and preferences regarding real estate (i.e. number of bathrooms, number of bedrooms, price, etc.). The retiree can then make use of the generated suggestions to make a more informed decision.

2. Data

To solve the problem mentioned above, data on weather, availability of amenities and essential services, as well as real estate listings was, gathered and analyzed in order to offer a recommendation of the best locations on the Canary Islands.

Yearly weather data for every main island was scraped from a common weather website (https://www.climatestotravel.com/climate/canary-islands) and the main focus was on the following parameters:

- Average rainfall
- Average daily sunshine hours
- Average monthly minimum temperature
- Average monthly maximum temperature
- Average monthly sea temperature

The data on availability of amenities and essential services was obtained through the FourSquare API.

Finally, current real estate listings and prices were scraped from a Spanish real estate website (https://www.tucasa.com/) and filtered based on a given set of requirements, such as:

- Maximum price
- Minimum number of bedrooms
- Minimum number of bathrooms
- Size

3. Methodology

In order to determine the best locations to retire on the Canary Islands in the form of a shortlist of promising suitable properties, two main steps were taken. The first step was to use weather data available on the internet to compare the seven main islands among each other in order to determine which of them has the "best climate". Of course, this is subjective, so some weather preferences were assumed. The second step was to examine the real estate market on that given island and narrow it to the expected shortlist of promising properties based also on assumed preferences with regards to nearby amenities, number of bathrooms, number of bedrooms, price, and so on.

3.1 Determining the Island with the Best Climate

To determine which island has the "best climate", yearly weather data from every island was scraped from the internet (https://www.climatestotravel.com/climate/canary-islands/). The data scraped included minimum average temperatures, maximum average temperatures, average sea temperatures, amount of precipitation, number of rainy days, and average daily hours of sun for each month of the year. All the weather data of every island was then combined into a single table to facilitate analysis.

The gathered data was plotted using *boxplots* from the Seaborn library to visually compare the weather characteristics of each island.

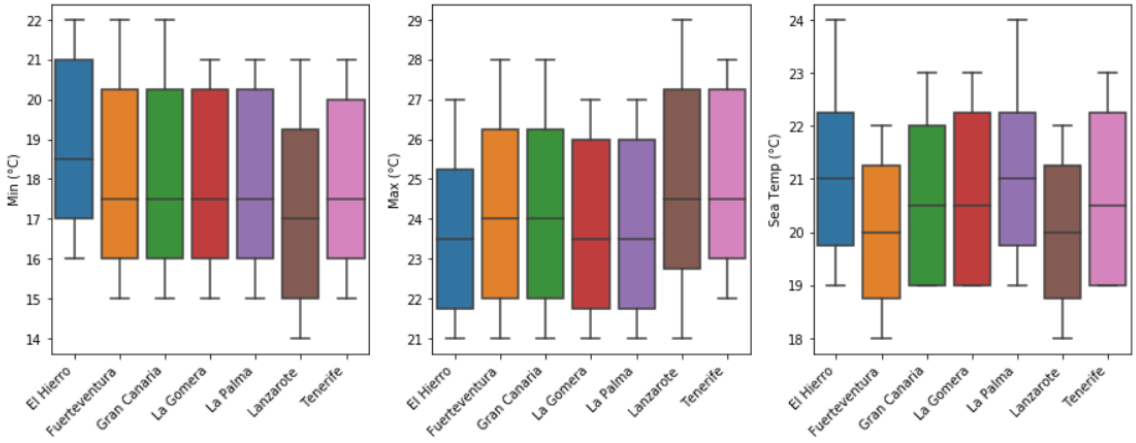

Figure 1. Monthly minimum average temperatures, monthly maximum average temperatures, and monthly average sea temperatures for the seven main islands of the Canary Islands.

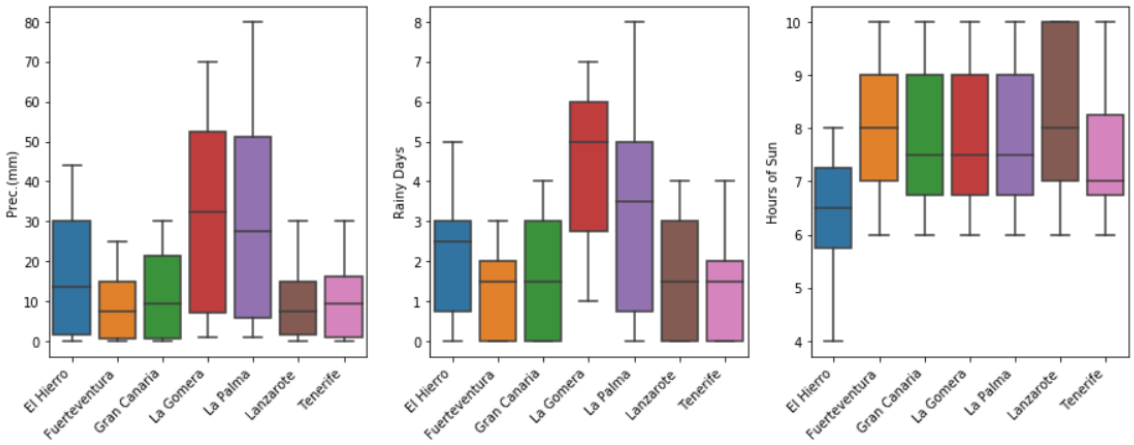

Figure 2. Monthly average rainfall in mm, average number of rainy days per month, and average daily hours of sunshine per month for the seven main islands of the Canary Islands.

From the temperature plots (Figure 1), we can see that all islands have very similar temperatures ranges, with *Lanzarote* having a slightly colder winter and a warmer summer, as well as lower sea temperatures. On the other side of the scale, we have *El Hierro*, which has a warmer winter, a not too warm summer and higher sea temperatures. But, as already stated, the difference is not very significant.

The rainfall and sunshine plots (Figure 2), however, offer a much clearer picture. We can clearly see, for example, that *El Hierro* has the least amount of sunshine hours per day. We can also see that *Fuerteventura*, *Lanzarote* and *Tenerife* have the least amount of rainfall, with *Fuerteventura* having the least amount of rainy days per month.

In summary, if we are looking for the island with the least amount of rain, the most sunshine, and relatively good temperatures, ***Fuerteventura*** has a slight edge over the other islands, making it the "photo-finish" winner of the "Battle of the Islands".

3.2 Examining the Real Estate Market

3.2.1 Data Pre-processing and Cleaning

After having determined that *Fuerteventura* is the island with the “best climate” (according to assumed preferences), the next step was to examine the real estate market in the island.

To do that, relevant data for property listings on the island was scraped from a popular Spanish real estate website (https://www.tucasa.com/compra-venta/viviendas/fuerteventura/). The data included the city, town or location, address or street, property size, number of rooms, number of bathrooms, and price.

The acquired data was put into a table and was further pre-processed and cleaned. Properties with empty data (except for street) or duplicate listings were removed. Numerical data was modified from type string to float or integer. Then, the index was reset. Finally, the price per square meter was calculated and added to the table. The resulting table is shown in Figure 3.

	Location	Street	Size	Rooms	Bathrooms	Price	Price/m2
0	Puerto del Rosario	CANDRAY 10	125.0	3.0	2.0	165000	1320.00
1	Corralejo		50.0	2.0	1.0	29900	598.00
2	Antigua		1000.0	3.0	3.0	275000	275.00
3	Corralejo		53.0	1.0	1.0	115900	2186.79
4	Tuineje		170.0	4.0	2.0	368000	2164.71
5	Puerto del Rosario	VIRGEN DE LA PEÑA	78.0	3.0	1.0	66000	846.15
6	Antigua		150.0	5.0	2.0	147500	983.33
7	Puerto del Rosario	Calle PITERA	96.0	3.0	2.0	82800	862.50
8	Puerto del Rosario		118.0	3.0	2.0	212000	1796.61
9	Puerto del Rosario		125.0	3.0	1.0	165000	1320.00

Figure 3. First ten rows of the pre-processed and cleaned table of property listings in Fuerteventura.

3.2.2 Exploratory Data Analysis

The data in the property listings table was plotted using *countplots* from the Seaborn library to get a better idea of its contents. First, to see the number of listings for each town or location and then to see the distribution of listings according to the number of rooms (as shown in Figure 4 and Figure 5 respectively).

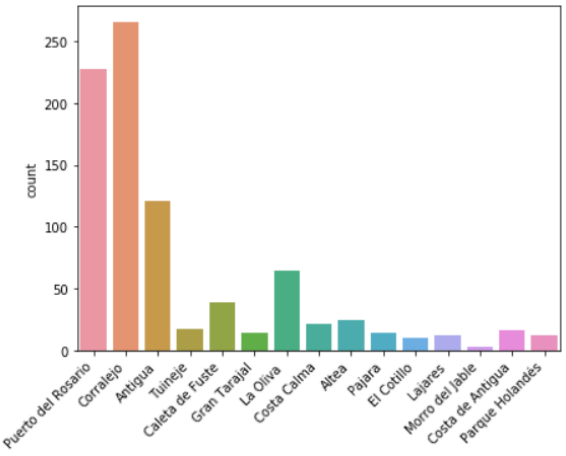

Figure 4. Number of property listings in each town or location in Fuerteventura.

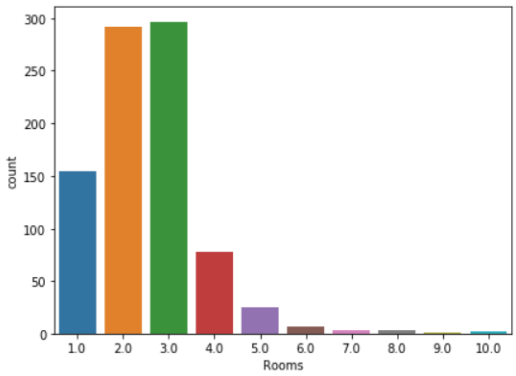

Figure 5. Distribution of property listings in Fuerteventura according to the number of rooms.

From the previous figures, we can see that *Puerto del Rosario* (the capital city of *Fuerteventura*) and *Corralejo* have the highest number of properties in offer. We can also see that most properties offered have 2 or 3 rooms.

The data was then plotted using *jointplot* from the Seaborn library to better visualize the distribution of the properties with regard to price and size. Then, the data was plotted again using *regplot* to determine the correlation between price and property size after removing any potential outliers by limiting the range of the plot, as shown in Figure 6.

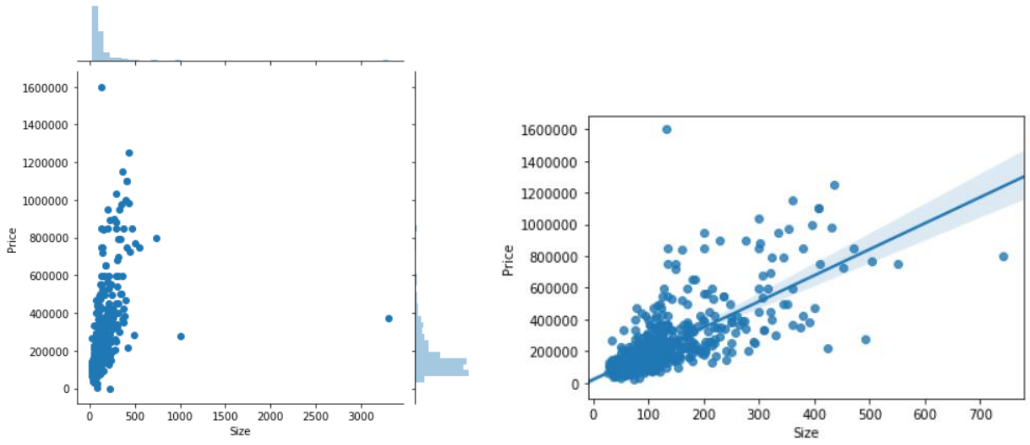

Figure 6. Distribution of property listings in Fuerteventura with regards to price and property size (left), and correlation between price and property size of property listings in Fuerteventura (right).

The second plot from Figure 6 shows a clear linear correlation between property size and price. Furthermore, to examine whether the number of rooms and number of bathrooms has an impact on price and price per square meter, the data was plotted using *boxplots* from the Seaborn library.

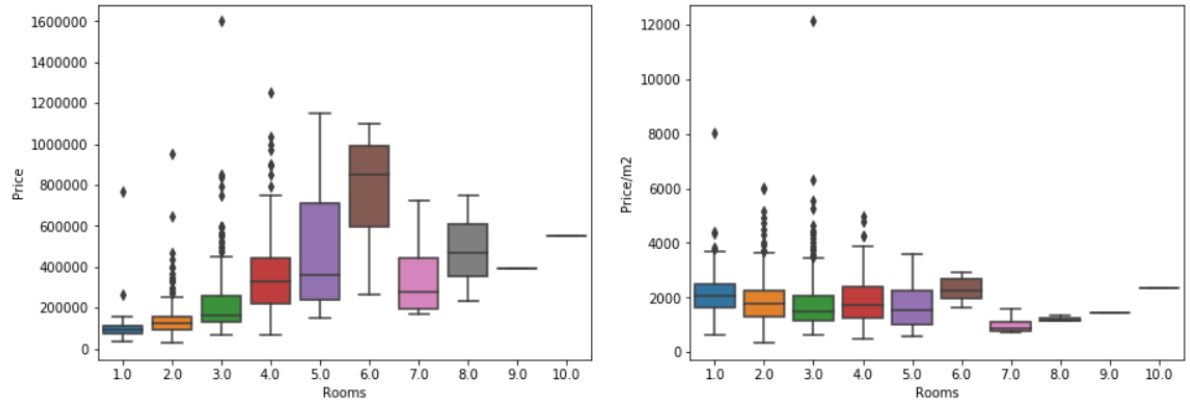

Figure 7. Distribution of prices and price per square meter according to the number of rooms of property listings.

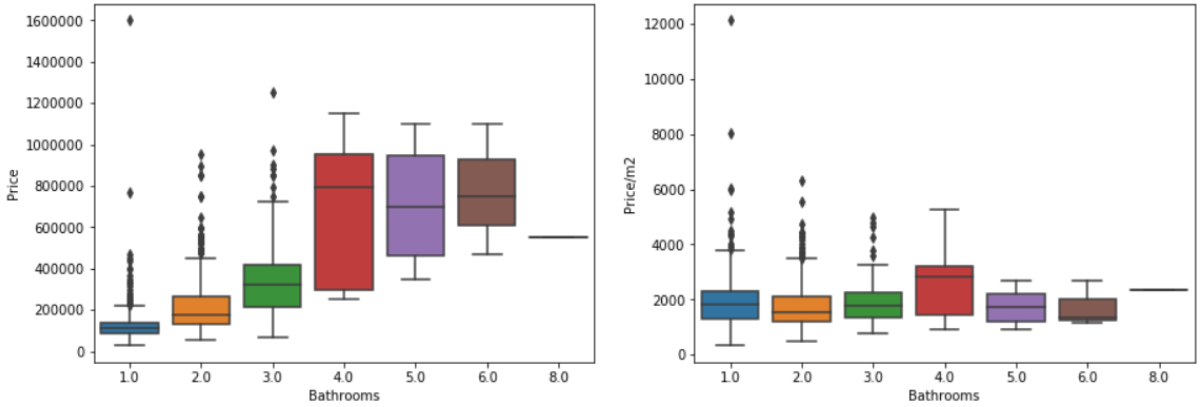

Figure 8. Distribution of prices and price per square meter according to the number of bathrooms of property listings.

The data, as shown in Figure 7 and Figure 8, indicates that the number of rooms and bathrooms has an influence on the overall price of the property up to 6 rooms and up to 4 bathrooms respectively, but not necessarily on the price per square meter.

Finally, the data was plotted again using *boxplot* to determine whether the location on the island has an influence on the price and price per square meter.

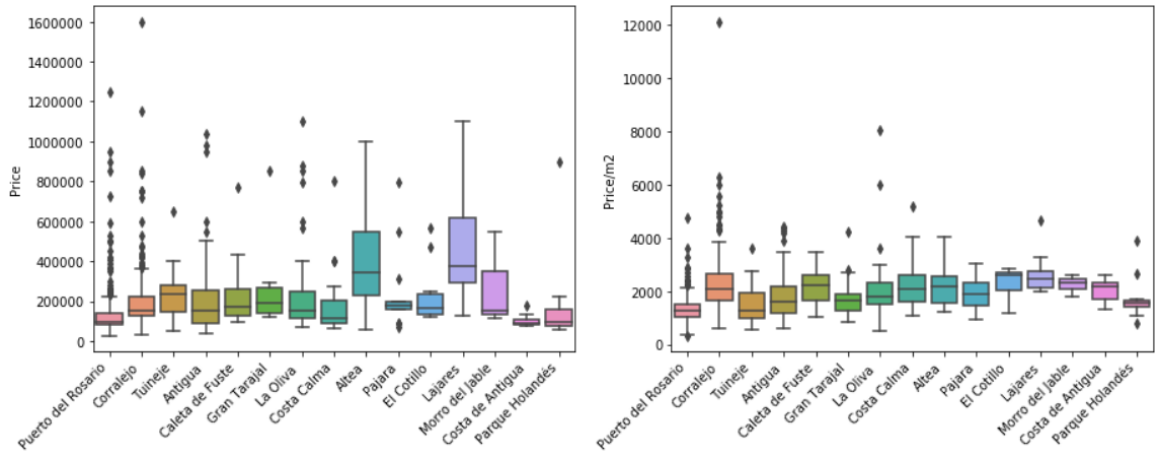

Figure 9. Distribution of prices and price per square meter according to town or location in Fuerteventura.

The data shown in Figure 9 appears to show that the properties in *Altea* and *Lajares* have consequently higher-priced units, with *Puerto del Rosario* and *Corralejo*, in particular, having lots of outliers. So, in general, location by itself seems to have a relative limited influence on price and price per square meter.

3.2.3 Limiting the Data According to Preferences

A retiree (or anyone else for that matter) has certain preferences and limitations when it comes to buying real estate. With that in mind, the list of properties was reduced based on a set of assumed preferences and limitations. The following assumptions were made:

- Maximum budget of 180,000 EUR
- Minimum 3 rooms
- Minimum 2 bathrooms
- Property size of at least 120 square meters

These boundary conditions limited the list of potential properties to 40.

3.2.4 Analysis of Nearby Amenities and Services Using the FourSquare API

An analysis of the amenities and services in close proximity to the reduced list of the properties for sale in Fuerteventura was conducted to try to find the most promising ones. In a first step, the geographical coordinates of all properties with a valid address from the previously reduced list of properties was obtained using the *geopy* library and added to the table, as shown in Figure 10. With this step, the number of properties on the list was further reduced to 20.

	ID	Location	Street	Size	Rooms	Bathrooms	Price	Price/m2	Latitude	Longitude
0	1	Puerto del Rosario	CANDRAY 10	125.0	3.0	2.0	165000	1320.00	28.499342	-13.859905
1	2	Gran Tarajal	Calle MAJUGA	122.0	3.0	3.0	165000	1352.46	28.210684	-14.030876
2	3	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	126.0	3.0	2.0	155000	1230.16	28.730723	-13.865191
3	4	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	129.0	3.0	2.0	150000	1162.79	28.730723	-13.865191
4	5	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	124.0	3.0	2.0	150000	1209.68	28.730723	-13.865191

Figure 10. First five rows of the reduced property listings in Fuerteventura with additional geographical coordinates.

Then, Folium was used to create a map with all the listed properties in the reduced list to get an overview of their locations (Figure 11).

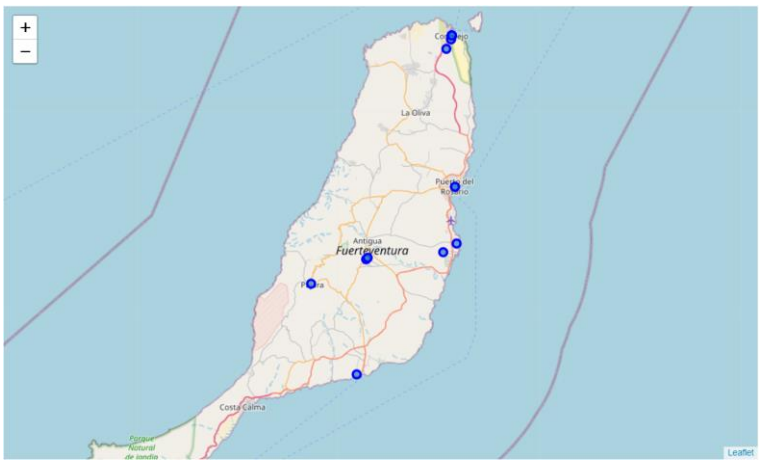

Figure 10. Map of preliminary selection of potentially suitable properties in Fuerteventura.

Using the FourSquare API, all venues within 500 meters of every property were obtained and added to a table. The resulting table contained a total of 241 venues with 36 unique categories.

	ID	Street	Location	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	1	CANDRAY 10	Puerto del Rosario	28.499342	-13.859905	Bar Terraza Los Paraguitas	28.498085	-13.858829	Tapas Restaurant
1	1	CANDRAY 10	Puerto del Rosario	28.499342	-13.859905	Cafeteria Son Son	28.500108	-13.863545	Café
2	1	CANDRAY 10	Puerto del Rosario	28.499342	-13.859905	Casa Toño	28.497128	-13.860780	Gastropub
3	1	CANDRAY 10	Puerto del Rosario	28.499342	-13.859905	La Saranda	28.497784	-13.860796	Breakfast Spot
4	1	CANDRAY 10	Puerto del Rosario	28.499342	-13.859905	Ciao Mare	28.495938	-13.860893	Pizza Place

Figure 10. First five rows of the table listing all venues within 500 m of every property in the reduced list.

To be able to better analyze the data, one-hot encoding was used and then the data was grouped by property ID to later take the mean of the frequency of occurrence of each category for each property. Finally, a new table with the 10 most common venues for each property was created with the obtained data, as shown in Figure 11.

	ID	Street	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	1	CANDRAY 10, Puerto del Rosario	Tapas Restaurant	Breakfast Spot	Coffee Shop	Clothing Store	Café	Pizza Place	Fast Food Restaurant	Gastropub	Bistro	Burger Joint
1	10	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
2	11	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
3	12	Calle PARDELA, Corralejo	Hotel	Business Service	Golf Course	Restaurant	Water Park	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café
4	13	Calle Atalaya Park, Antigua	Restaurant	Water Park	Business Service	Dessert Shop	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café	Burger Joint
5	14	Calle Real, Pajara	Plaza	Bed & Breakfast	Snack Place	Restaurant	Burger Joint	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café
6	15	Betoveen 110, Puerto del Rosario	Tapas Restaurant	Breakfast Spot	Coffee Shop	Clothing Store	Café	Pizza Place	Fast Food Restaurant	Gastropub	Bistro	Burger Joint
7	16	Calle Artemy, Caleta de Fuste	Resort	Water Park	Business Service	Dessert Shop	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café	Burger Joint
8	17	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
9	18	el roque, Corralejo	Café	Board Shop	Water Park	Dessert Shop	Diner	Burger Joint	Golf Course	Hostel	Hotel	Hotel Bar
10	19	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
11	2	Calle MAJUGA, Gran Tarajal	Restaurant	Water Park	Business Service	Dessert Shop	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café	Burger Joint
12	20	Maestro Leonardo, Antigua	Hotel	Bar	Spa	Grocery Store	Restaurant	Resort	Chinese Restaurant	Water Park	Business Service	Cocktail Bar
13	3	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
14	4	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
15	5	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
16	6	Calle Isla de La Graciosa, Antigua	Restaurant	Water Park	Business Service	Dessert Shop	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café	Burger Joint
17	7	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
18	8	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
19	9	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N, Co...	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant

Figure 11. Table listing the ten most common venues within 500 m of every property in the reduced list.

Also, to get a better idea of what venues are located where, the data was used to create a map using *Folium*, with blue markers depicting the properties and the green markers depicting the different venues. Figure 12 shows a zoomed-in image of the map to show the town of Corralejo as an example.

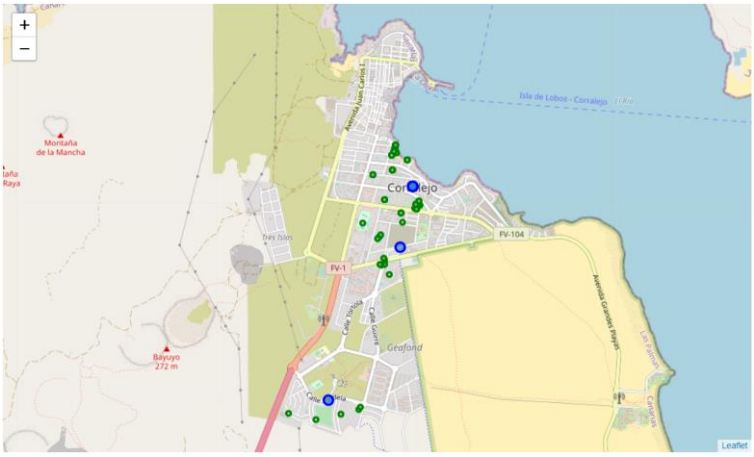

Figure 12. Map of potentially suitable properties in Corralejo, Fuerteventura (blue) and venues located within 500 m.

3.2.5 Clustering of Potentially Suitable Properties

The properties from the previous step were then clustered using the k-means clustering method to group them in previously unlabeled groups and try to find similarities among them. The best k-value for the data was determined using the Elbow Method by calculating the distortion for every value of k within the range 1-10.

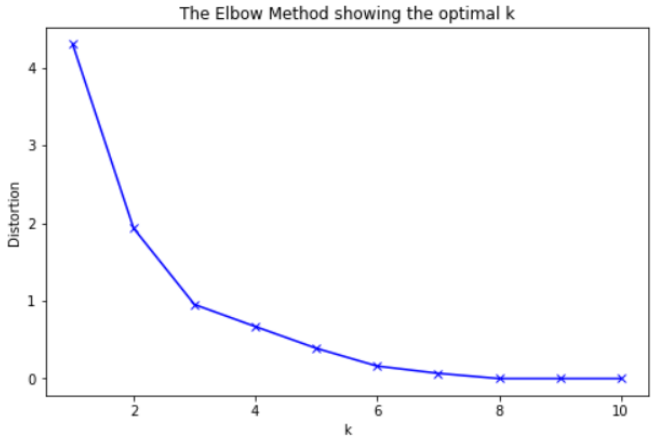

Figure 13. Elbow Method plot showing the distortion for every value of k within the range 1-10.

Figure 13 shows that the maximum change in slope appears to occur at a k-value of 3. So, the properties in were clustered into three groups before applying the k-clustering method. After the algorithm was run, Cluster 1 had one property, Cluster 2 had three properties, and Cluster 3 had sixteen properties. Figure 14 shows the geographical distribution of the properties according to their clusters.

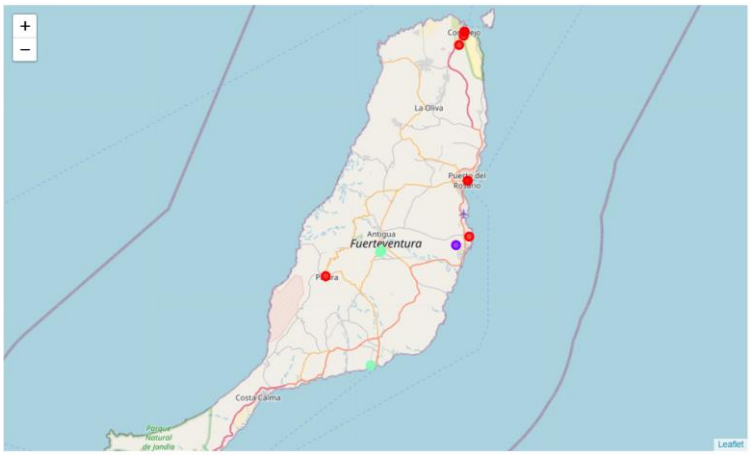

Figure 14. Map of property clusters after applying the k-means clustering method (Cluster 1 = purple, Cluster 2 = cyan, Cluster 3 = red).

The ten most common venues for each cluster were examined to determine the difference among the clusters. Figure 15, Figure 16 and Figure 17 show the respective clusters with their ten most common venues.

	ID	Location	Street	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
15	16	Caleta de Fuste	Calle Artemy	Resort	Water Park	Business Service	Dessert Shop	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café	Burger Joint

Figure 15. Cluster 1 and its ten most common venues.

	ID	Location	Street	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
1	2	Gran Tarajal	Calle MAJUGA	Restaurant	Water Park	Business Service	Dessert Shop	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café	Burger Joint
5	6	Antigua	Calle Isla de La Graciosa	Restaurant	Water Park	Business Service	Dessert Shop	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café	Burger Joint
12	13	Antigua	Calle Atalaya Park	Restaurant	Water Park	Business Service	Dessert Shop	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café	Burger Joint

Figure 16. Cluster 2 and its ten most common venues.

	ID	Location	Street	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	1	Puerto del Rosario	CANDRAY 10	Tapas Restaurant	Breakfast Spot	Coffee Shop	Clothing Store	Café	Pizza Place	Fast Food Restaurant	Gastropub	Bistro	Burger Joint
2	3	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
3	4	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
4	5	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
6	7	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
7	8	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
8	9	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
9	10	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
10	11	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
11	12	Corralejo	Calle PARDELA	Hotel	Business Service	Golf Course	Restaurant	Water Park	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café
13	14	Pajara	Calle Real	Plaza	Bed & Breakfast	Snack Place	Restaurant	Burger Joint	Coffee Shop	Cocktail Bar	Clothing Store	Chinese Restaurant	Café
14	15	Puerto del Rosario	Betoveen 110	Tapas Restaurant	Breakfast Spot	Coffee Shop	Clothing Store	Café	Pizza Place	Fast Food Restaurant	Gastropub	Bistro	Burger Joint
16	17	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
17	18	Corralejo	el roque	Café	Board Shop	Water Park	Dessert Shop	Diner	Burger Joint	Golf Course	Hostel	Hotel	Hotel Bar
18	19	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	Café	Hotel	Beach	Athletics & Sports	Bistro	Cocktail Bar	Hotel Bar	Pharmacy	Diner	Restaurant
19	20	Antigua	Maestro Leonardo	Hotel	Bar	Spa	Grocery Store	Restaurant	Resort	Chinese Restaurant	Water Park	Business Service	Cocktail Bar

Figure 17. Cluster 3 and its ten most common venues.

Cluster 3 appears to list the properties that have more diverse and more interesting venues nearby. That makes these properties the more appealing ones and, therefore, Cluster 3 would be the expected ***shortlist*** of potentially suitable properties (shown in Figure 18).

	ID	Location	Street	Size	Rooms	Bathrooms	Price	Price/m2	Latitude	Longitude
0	1	Puerto del Rosario	CANDRAY 10	125.0	3.0	2.0	165000	1320.00	28.499342	-13.859905
2	3	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	126.0	3.0	2.0	155000	1230.16	28.730723	-13.865191
3	4	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	129.0	3.0	2.0	150000	1162.79	28.730723	-13.865191
4	5	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	124.0	3.0	2.0	150000	1209.68	28.730723	-13.865191
6	7	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	120.0	3.0	2.0	126500	1054.17	28.730723	-13.865191
7	8	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	129.0	3.0	2.0	125500	972.87	28.730723	-13.865191
8	9	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	120.0	3.0	2.0	137500	1145.83	28.730723	-13.865191
9	10	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	122.0	3.0	2.0	134500	1102.46	28.730723	-13.865191
10	11	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	120.0	3.0	2.0	134500	1120.83	28.730723	-13.865191
11	12	Corralejo	Calle PARDELA	120.0	3.0	2.0	158700	1322.50	28.710112	-13.874482
13	14	Pajara	Calle Real	172.0	3.0	2.0	160103	930.83	28.350184	-14.110690
14	15	Puerto del Rosario	Betoveen 110	145.0	4.0	3.0	162000	1117.24	28.499342	-13.859905
16	17	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	120.0	3.0	2.0	127500	1062.50	28.730723	-13.865191
17	18	Corralejo	el roque	140.0	3.0	2.0	165000	1178.57	28.724865	-13.866551
18	19	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	122.0	3.0	2.0	129000	1057.38	28.730723	-13.865191
19	20	Antigua	Maestro Leonardo	120.0	3.0	2.0	149000	1241.67	28.411383	-13.856528

Figure 18. Shortlist of potentially suitable properties in Fuerteventura.

3.2.6 Further Clustering of Potentially Suitable Properties

As an additional analysis, the obtained *shortlist* was clustered into smaller clustered based on property data (i.e. number of rooms, bathrooms, size, price, etc.) to try to find similarities among them. The Elbow Method was used again to determine the optimum number of clusters.

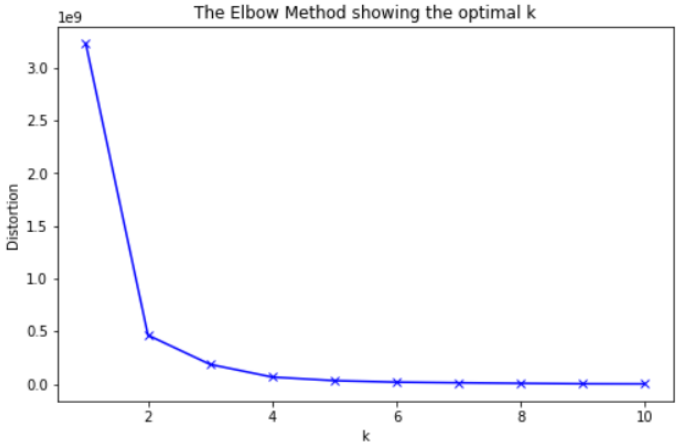

Figure 19. Elbow Method plot showing the distortion for every value of k within the range 1-10.

Figure 19 shows that the maximum change in slope occurs at a k-value of 2. So, the properties in the shortlist were clustered into two groups before applying the k-clustering method. After the algorithm was run, Cluster A had seven properties and Cluster B had nine properties. Figure 20 shows the geographical distribution of the properties according to their clusters.

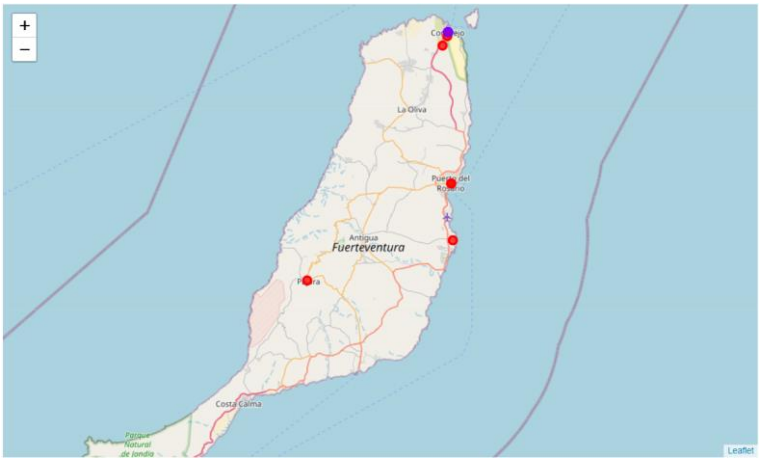

Figure 20. Map of property clusters after applying the k-means clustering method (Cluster A = purple, Cluster B = red).

The clusters were examined to determine the difference among them. Figure 21 and Figure 22 show the properties in each cluster and their characteristics.

	ID	Location	Street	Size	Rooms	Bathrooms	Price	Price/m2
6	7	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	120.0	3.0	2.0	126500	1054.17
7	8	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	129.0	3.0	2.0	125500	972.87
8	9	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	120.0	3.0	2.0	137500	1145.83
9	10	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	122.0	3.0	2.0	134500	1102.46
10	11	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	120.0	3.0	2.0	134500	1120.83
16	17	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	120.0	3.0	2.0	127500	1062.50
18	19	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	122.0	3.0	2.0	129000	1057.38

Figure 21. Properties in Cluster A.

	ID	Location	Street	Size	Rooms	Bathrooms	Price	Price/m2
0	1	Puerto del Rosario	CANDRAY 10	125.0	3.0	2.0	165000	1320.00
2	3	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	126.0	3.0	2.0	155000	1230.16
3	4	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	129.0	3.0	2.0	150000	1162.79
4	5	Corralejo	Calle PARDELA URB.MIRADOR DE LAS DUNAS S/N	124.0	3.0	2.0	150000	1209.68
11	12	Corralejo	Calle PARDELA	120.0	3.0	2.0	158700	1322.50
13	14	Pajara	Calle Real	172.0	3.0	2.0	160103	930.83
14	15	Puerto del Rosario	Betoveen 110	145.0	4.0	3.0	162000	1117.24
17	18	Corralejo	el roque	140.0	3.0	2.0	165000	1178.57
19	20	Antigua	Maestro Leonardo	120.0	3.0	2.0	149000	1241.67

Figure 22. Properties in Cluster B.

The data from each cluster was plotted using *boxplots* from the Seaborn library to see how the two clusters compare to each other with regards to price, price per square meter and size (Figure 23).

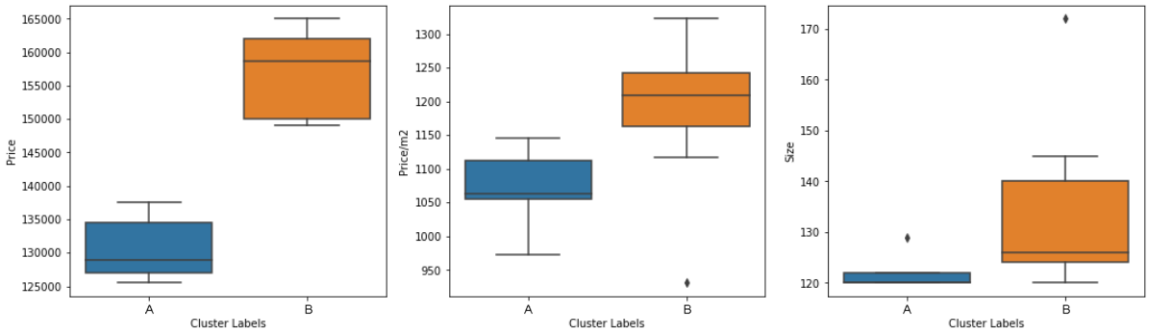

Figure 23. Distribution of prices, price per square meter and size for property listings in Clusters A and B.

A Closer inspection of the properties within the two clusters show that:

- Cluster A is made up solely of apartments that are in the same complex and are relatively smaller in size.
- Cluster B is made up of single-family homes with a small garden.

4. Results

The analysis of weather data from the Canary Islands determined that, if what is preferred is the least amount of rain, the most sunshine, and relative mild temperatures, ***Fuerteventura*** had a slight edge over the other islands, making it the island of choice for retirement.

The analysis of the real estate market in the island of *Fuerteventura* determined a couple of things. First, as was to be expected, there was a clear linear correlation between property size and price. The number of rooms and bathrooms also had a positive correlation to the properties' price, up to 6 rooms and up to 4 bathrooms respectively, but not necessarily on the price per square meter.

With regards to location on the island, *Altea* and *Lajares* appeared to have consequently higher-priced properties, while *Puerto del Rosario* and *Corralejo*, in particular, had lots of outliers. In general, location by itself did not have a significant influence on property price or price per square meter.

The final result, and goal of this project, was a shortlist of potentially suitable properties for retirement in *Fuerteventura*. The list included 16 properties chosen based on the variety, type and overall appeal of the amenities nearby. Additionally, this shortlist was further divided into two smaller lists; one composed of apartments within the same complex, and the other one composed of single-family homes with a small garden. These shortlists can then be presented to a retiree in search of a retirement property in *Fuerteventura*.

5. Discussion

Based on the results, there didn't seem to be enough data available on the Canary Islands in FourSquare, which might have limited the analysis carried out in this project to some extent. One possible suggestion to solve this issue would be to complement the data with data from another source, or use another source entirely.

Also, it is worth mentioning that the data analysis presented in this project is meant as an example only. Therefore, it would be recommended to carry out the analysis regularly in order to be up to date with the constantly changing real estate market, and similarly, the preferences or starting conditions for the analysis should also be modified in order to accommodate the personal preferences of the person carrying out the analysis.

6. Conclusion

This project determined the island of the Canary Islands with "best" weather based on certain subjective criteria (i.e. low rainfall, lots of sunshine, and relative mild temperatures). It then explored the real estate market on that island and narrowed down the options based on preferences regarding property characteristics (i.e. maximum price, minimum number of rooms, minimum number of bathrooms, and minimum property size) as well as the availability of amenities and services nearby.

As a result of such analysis, the project obtained a shortlist of top properties, which was then further divided into two smaller lists; one with apartments within a same complex and another one with single-family homes with a small garden. Based on this pre-selection, a potential retiree could plan his or her visits to the listed properties and make a final decision for buying a suitable home for retirement in the Canary Islands.