AIRBNB SICILY REPORT – STEP 4

Paolo Vivera

Università di Catania

Summary

Introduction

This report showcases several key findings in the Airbnb housing market, taken from several adjustments and analyses of a dataset containing Airbnb listings. The dataset was merged with other datasets, including average rent, province names, occupancy rates per city, and house prices per square meter.

The aim of this report is to gain insights into the Airbnb housing market in the various cities, to determine in which city and which house to invest in and to verify if the house listing on Airbnb is significantly more convenient than traditional long-term renting, if at all.

The first part of the report covers the details of the data used, the outputs generated, the operations made and the logical schema. The second part shows the visualizations for the top cities and for the New Airbnb Yearly Return on Investment (YROI) and the traditional Rent YROI. Airbnb YROI is higher than the Rent YROI for most of the biggest cities.

The Step 4 is resolved using 2 different outputs representing the same fact, one at the level of single houses and another at the level of the different type of houses per City.

The main idea of the final step is to let the user explore and gain quick insight into the big housing market in Sicily with a dashboard using the aggregated data while exploring at a deeper level the single houses returns and costs using a dashboard representing single houses. Additionally, a personal visualization of which is the optimal solution to maximize the investment while keeping low the costs is provided.

Datasets Details

The initial dataset comprised five main sources:

- **Cities in Sicily**, composed of 2 rows:
  - **Città** (City categorical variable) that represents the sicilian cities.
  - **Provincia** (Province categorical variable).

- **House Info**, the main dataset composed of:
  - **ID**, a numerical identifier of the house.
  - **City**, the categorical variable expressing the city where the house is located.
  - **Bedrooms_Ten**, the number of bedrooms the house has.
  - **Bathrooms**, the number of bathrooms the house has.

- **Airbnb Price**, a dataset composed of 2 rows:
  - The **ID** identifier of the house.
  - The **price**, that is the price the guest pays.

- **Cities Buy Rent**, which provides information about rents and sales in 3 rows:
  - The **Comuni** (City) categorical variable.
  - The **Vendita**($€/m^2$) integer that provides the average price of a house in the city in euros by square meter.
  - The **Affitto**($€/m^2$) float that provides the average monthly rent price of a house in the city in euros by square meter.

- **Occupancy Rate by City**, composed of:
  - **City** categorical variable.
  - **Occupancy Rate**, average ratio of how many days in a year a house listed in Airbnb in that city is successfully rented.

Final Output

The source datasets have been merged and heavily modified, and many fields have been created altogether, resulting in working datasets made mostly of computed features. The final datasets, along with the Tableau Prep can be found in the same directory of this report.

Two final outputs have been made, one composed of the aggregated values
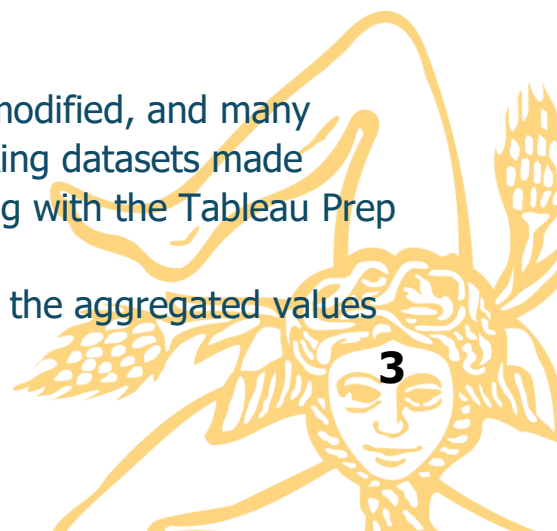
at a city and type of house level and another one composed of the individual houses, along with mean values shared by the type of house and city.

The Disaggregated dataset (Output-Vivera-Part2) is composed of

1 : CITY [string]: The City where the house is located.

2 : ID [integer]: the identifier of the house.

3 : HOUSE_TYPE [string]: The type of house, that contains the number of bedrooms and bathrooms the house has.
It is made by combining the variables Bathroom and Bedrooms together.

4 : HOUSE_SIZE [integer]: The size of the house computed by the number of bedrooms and bathrooms the house has.

IF[Bedrooms_Temp]==1 THEN 18+([Bedrooms_Temp]*16)+([Bathrooms]*9) ELSE 35+([Bedrooms_Temp]*16)+([Bathrooms]*9) END

5 : AVG HOUSE SIZE [integer]: The average size of the houses grouped by the city and type of house.

6 : HOUSE_PRICE [integer]: The house value, computed by the house size and the price per square meter.

7 : NUMBER_AIRBNB_HOUSE [integer]: The number of houses in the city by house type.

8 : ANNUAL_RENT [float]: the gross annual rent we would get in a year, obtained by multiplying the rent by square meters for the house price for 12 months.

[HOUSE_SIZE]*[RENT]*12

9: LOCAL ANNUAL TAXES [float]: The local taxes that the house owner must pay yearly, computed as 1.5% of the house price.

10: RENT YROI [float]: The Yearly Return on Investment of traditional rent, done by computing the net rent (annual rent – local taxes) and divided by the house price.

([ANNUAL_RENT]-[LOCAL_ANNUAL_TAXES])/[HOUSE_PRICE]

11: PRICE PER NIGHT [integer]: The price the host of the house charge for a single night.

12: OC [float]: The city average ratio of how many days in a year an house listed in Airbnb in that city is successfully rented, used for computing the Airbnb YROI in the tableau worksheet.

In addition, on Tableau Desktop, several variables and parameters have been added.

The parameters included are largely used for the creation of the new computed fields and are:

1: OC Var, an integer ranging from -5 to +5, influencing the Occupancy rate.
2: Chart Selection, a list of strings that allows to change the graph in the dashboard.
3: Initial investment, an integer that allows the user to choose how large the initial investment is and how much of the remaining cost will be taken as loan.
4: Interest Rate, a float ranging from 2.0 to 2.6, that determines with the Initial investment the annual interests the investor will pay for the selected houses.

The variables added are:

1: NEW ANNUAL AIR RENT [float]: estimate of the annual gross earnings made if the house is listed on AirBnb. The computed field

directly embeds the OC Var parameter. The formula is:

[Price Per Night]*365*([OC]+([OC Var]/100))

2: NEW ANNUAL SERVICE COSTS [float]: the estimated costs of listing the house on Airbnb as a percentage of the New Annual Air Rent. They are estimated by considering the second last student ID digit (8 in my case). The formula is:

[New Annual Air Rent]*(0.03+(8/300))

3: NEW AIR YROI [float]: the yearly net Return on Investment estimated for the house on Airbnb. It is computed by subtracting the Local Annual Taxes and Airbnb Service Costs from the New Annual Air Rent and then divided by the House Price. The formula is:

([New Annual Air Rent]-[New Air Service Costs]-[Local Annual Taxes])/[House Price]

4: HIGHEST N HOUSE CITY [integer]: keeps track of the constraint that no single city can have more than 4 houses.

5: COLOR MAX HOUSE, COLOR N HOUSE and COLOR PRICE HOUSE [string]: IF-ELSE statements for visual representation of constraints regarding the number of houses, number of houses in a city and the total price.

6: UNIQUE CHECKER and DUPLICATES REMOVER [string, Boolean]: It may happen that multiple houses in the same city can have the same house price and AirBnB YROI. Since we are interested in selecting only *one* house for a particular combination of city, house price and median YROI, I deploy unique checker to gather the information about House price and and duplicates remover to filter out multiple solution based on the lowest ID. Since a number of houses are removed, the number of houses is different when this filter is used.

The aggregated dataset (Output-Vivera-Part2 Aggregated) is composed of:
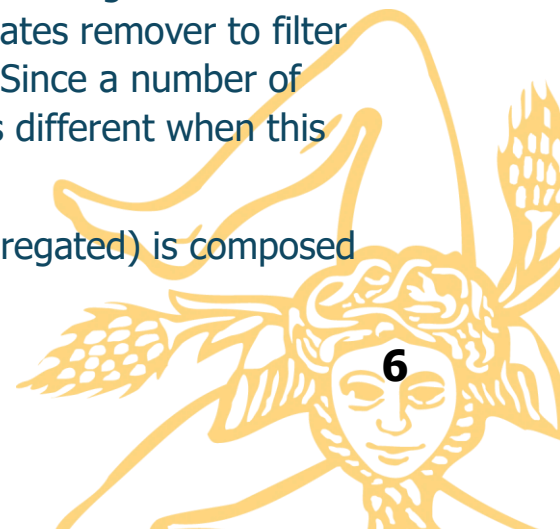
1: CITY [string]: The City where the house is located.

2: HOUSE TYPE [string]

3: HOUSE SIZE [integer]: The average size of the houses grouped by the city and type of house.

4: HOUSE_PRICE [integer]: The house value, computed by the house size and the price per square meter.

5: NUMBER_AIRBNB_HOUSE [integer]: The number of houses in the city by house type.

6: ANNUAL_RENT [float]: the average gross annual rent we would get in a year, obtained by multiplying the rent by square meter for the house price and then by 12 months.

7: LOCAL ANNUAL TAXES [float].

8: PRICE PER NIGHT [integer]: The price the host of the house charge for a single night.

9: OC [float] Average occupancy rate.

10: **Price** [integer]: average price for squared meter to buy an house.

11: **Rent** [float]: average price for squared meter to rent an house.

In addition, in the tableau desktop worksheet, the parameters added is the OC Var, while the variable added are:

1: Rent YROI [float]: The YROI of the rent computed by using the net (annual rent – local taxes) rent earnings and divide it by the house price.

2: NEW ANNUAL AIR RENT [float]: estimate of the annual gross earnings made if the house is listed on AirBnb. The computed field is made in the in order to embed directly into the field the parameter OC Var. The formula is:
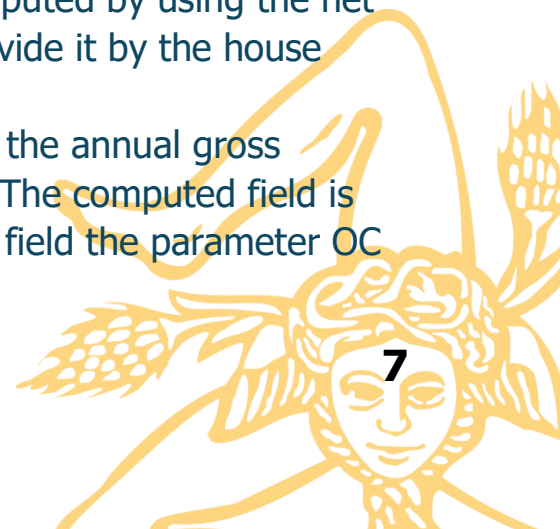
[Price Per Night]*365*([OC]+([OC Var]/100))

3: NEW ANNUAL SERVICE COSTS [float]: the estimated costs of listing the house on Airbnb as a percentage of the New Annual Air Rent. They are estimated by considering the second last student ID digit (8 in my case). The formula is:

[New Annual Air Rent]*(0.03+(8/300))

4: NEW AIR YROI [float]: the yearly net Return On Investment estimated for the house on Airbnb. It is computed by subtracting the Local Annual Taxes and Airbnb Service Costs from the New Annual Air Rent and then divided by the House Price. The formula is:

([New Annual Air Rent]-[New Air Service Costs]-[Local Annual Taxes])/[House Price]

5: AIR/RENT RATE [float]: a ratio that compares the AIR rate and the Rent rate. Higher than 1 means that the Airbnb YROI is higher than the rent YROI and vice versa.

**6: Index** [integer]: an index used for the additional plot 1.

Facts and Dimensions

The main fact is HouseInfo, which contains City and ID, the variables used for all joins, as well as the number of bedrooms and bathrooms of a variable.

The dimensions represented are:

1. **AirbnbPrice**, connected through the key-id

2. **Occupancy** Rate, through the key-City

3. **CitiesBuyRent**, through the key-City

4. **CitiesInSicily**, through the key-City

The Star schema is represented this way
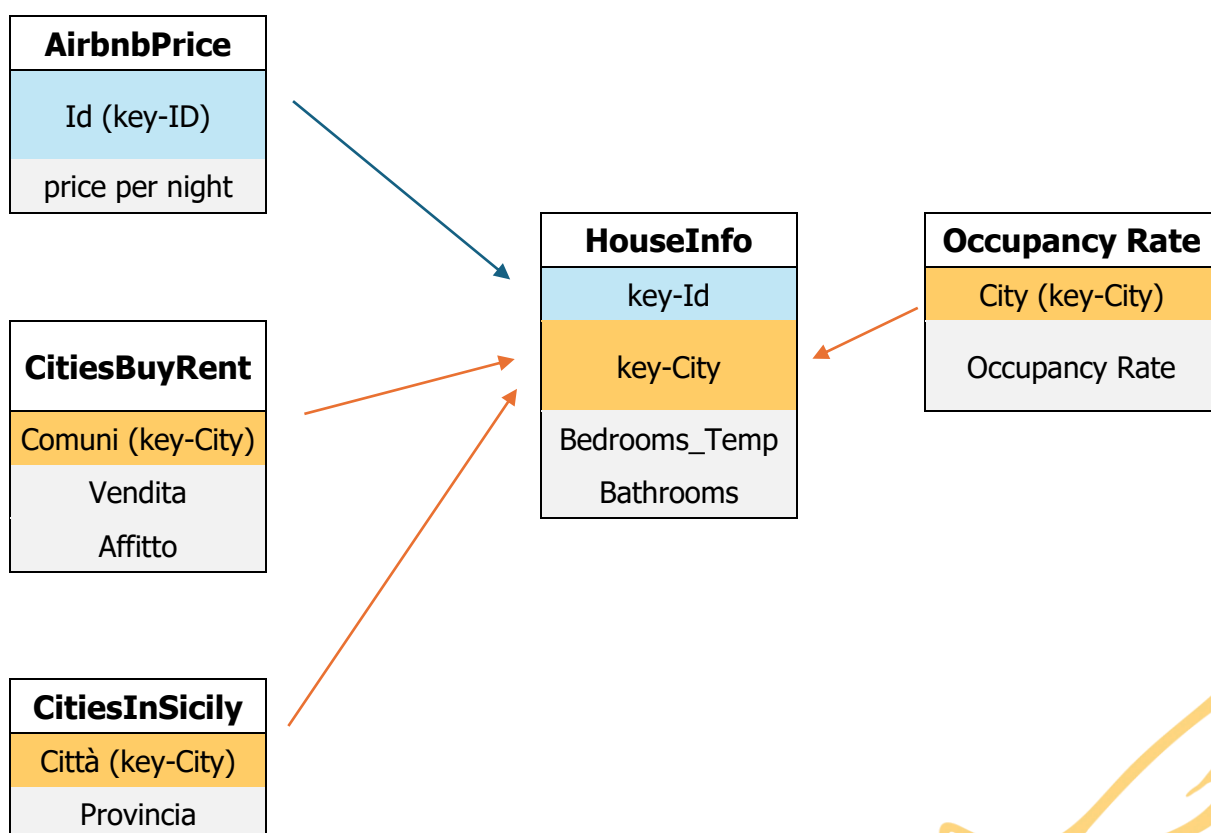
ETL Operations

The commented Tableau Prep Flow is present with this report and details all changes. The main ETL operations made are:

1. **Inner Join** on ID that combines AirbnbPrice and HouseInfo.
2. **Left Join** on City with AirbnbPrice+HouseInfo and CitiesInSicily. This operation is not necessary but allows me to easily see city names discrepancies before the other joins.
3. **Aggregation** of the Number of Houses, House Size and Price per Night, **grouped by** City and House Type. The aggregations are combined with the original dataset with an **Inner Join** on City and house Type.
4. **Left Join** on City with CitiesInSicily_BUY_RENT to add rent and price per square meter.
5. In step Four, **Right Join** on City with OccupancyRateByCity.
6. For the aggregated data, we **Aggregate** Rent, Price, House price, Occupancy Rate, Price per Night, Number of Houses, House Size, Local Annual Taxes and Annual Rent. They are **grouped by** City and House Type again, with the fields ready to be exported.
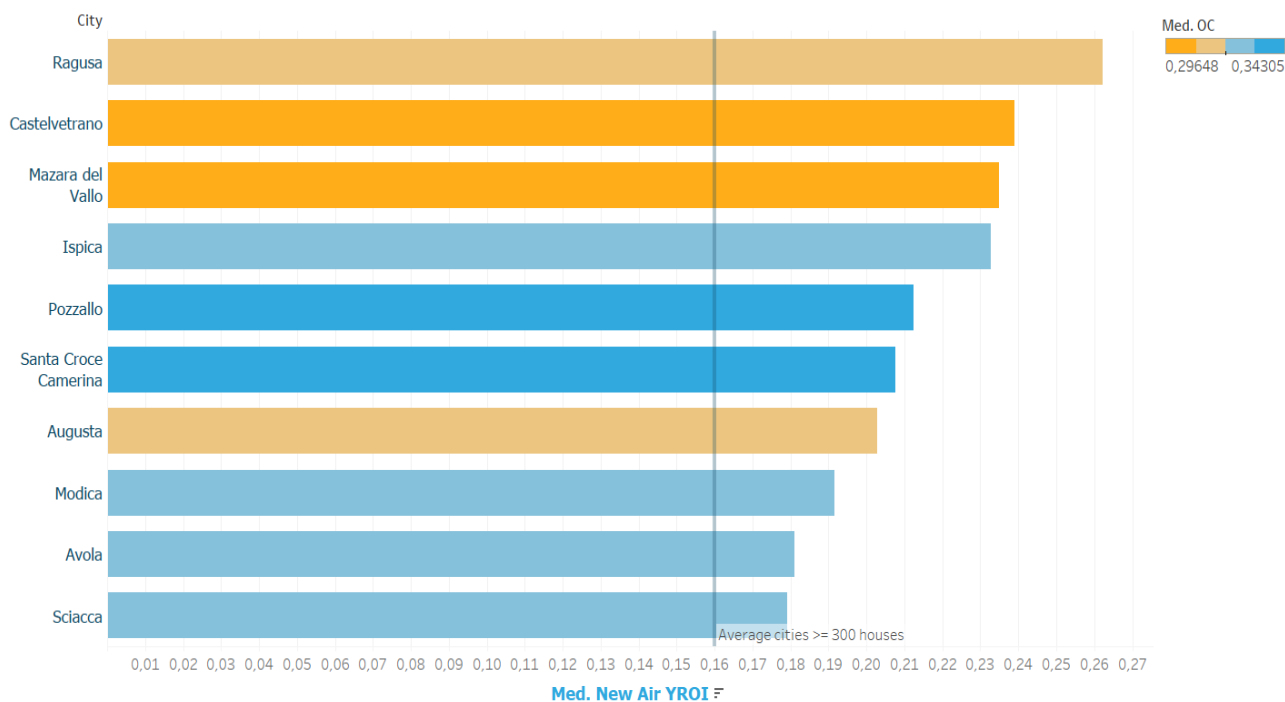
Remake of the First Part

The main points made on the first part of the report are remade by updating the Airbnb YROI with the new Occupancy rate data on the unaggregated Dataset.

Top 10 Cities for New Air YROI
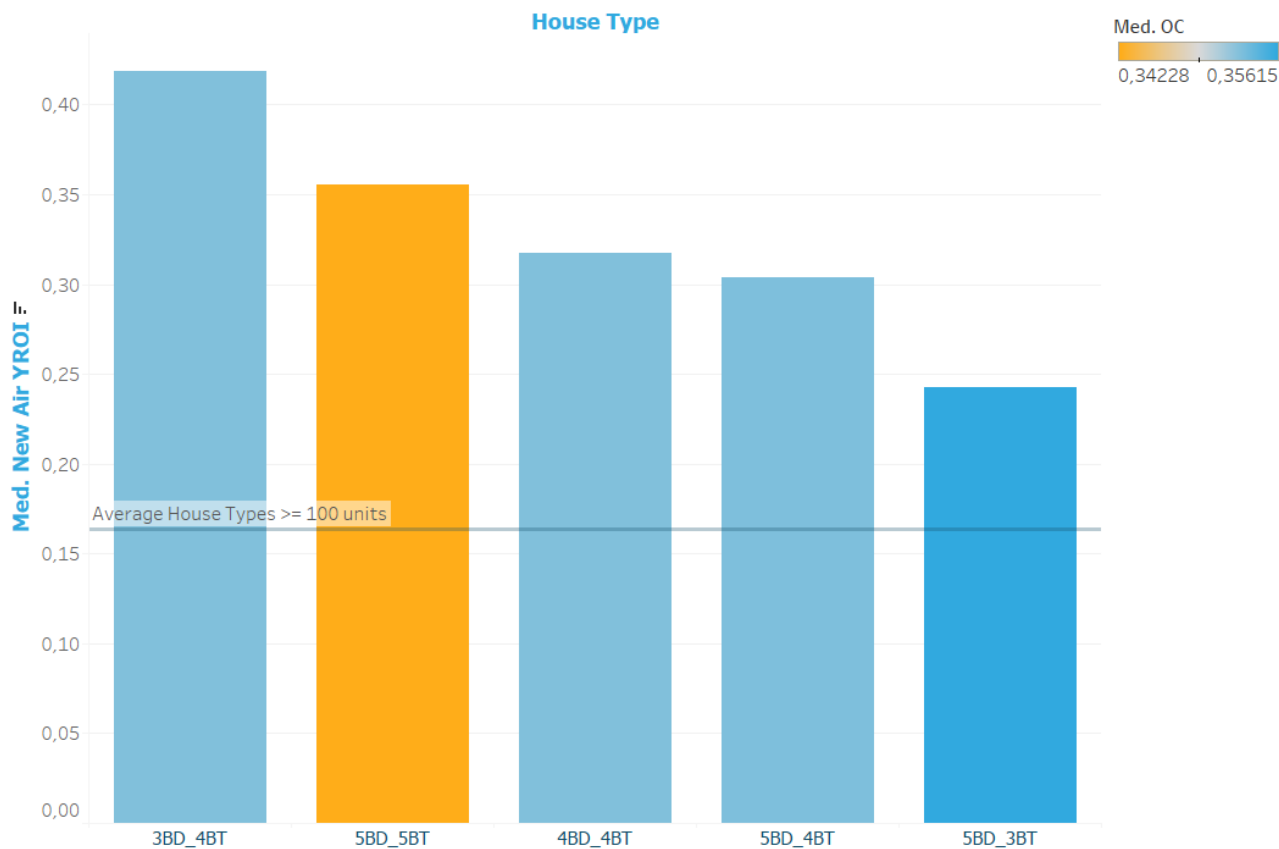### Cities with at least 300 houses

Media di New Air YROI per ogni City. Il colore mostra media di OC. La vista è filtrata su City, che mantiene 10 di 366 membri.

The first plot shows the top 10 cities with at least 300 houses based on the New Air YROI. Ragusa, Castelvetrano, Mazara del Vallo, Ispica, and Pozzallo are the top 5 cities, maintaining the same ranking as the initial report. Colour denotes the city's OC rate, and the line shows the average New Air YROI for all cities with at least 300 houses.

Media di New Air YROI per ogni House Type. Il colore mostra media di OC. La vista è filtrata su House Type, che mantiene 3BD_4BT, 4BD_4BT, 5BD_3BT, 5BD_4BT e 5BD_5BT.
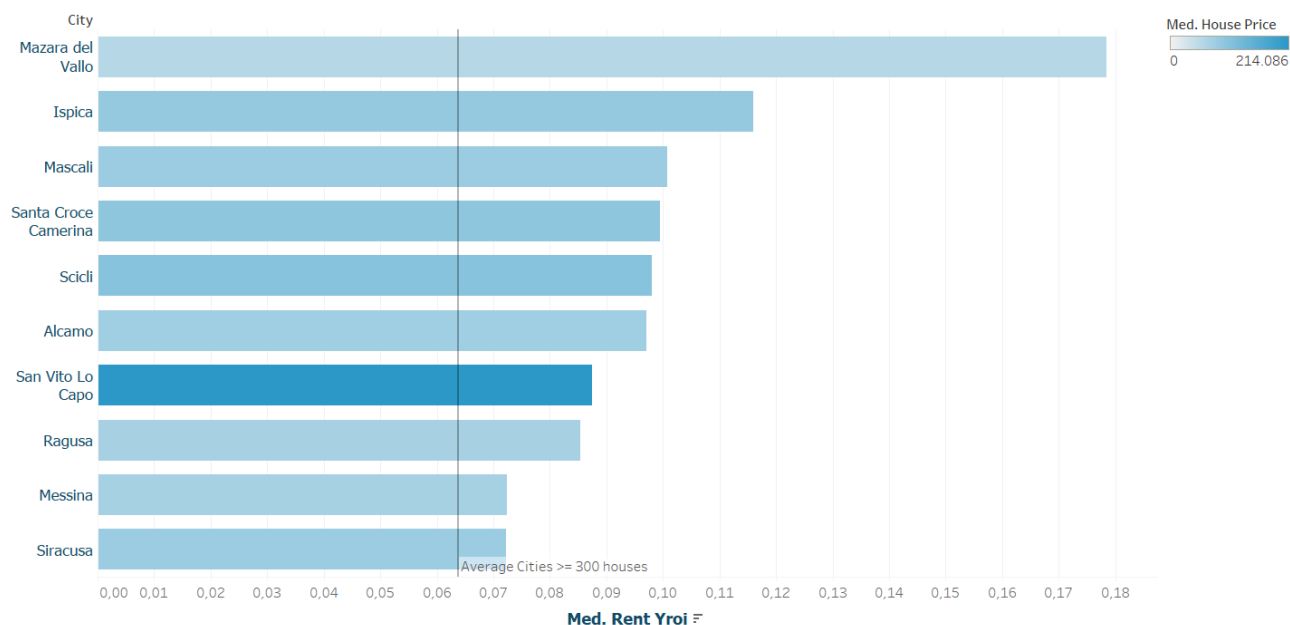
The top 5 type of houses plot is instead different compared to the first part, mainly due to the filtering of all type of houses with less than 100 units in the dataset. Like the previous plot, the color is used to show different levels of Occupancy Rate, while the line marks the average New Air YROI among house types with more than 100 units.

Media di Rent Yroi per ogni City. Il colore mostra media di House Price. La vista è filtrata su City, che mantiene 10 di 366 membri.

Comparing traditional Rent YROI reveals that some cities, like Mascali and Alcamo, appear in the ranking, changing significantly from the Airbnb ranking. The colour represents the average House Price (OC does not affect rent YROI), and the line shows the average New Air YROI for cities with at least 300 houses.

Top 5 Type of Houses by Rent YROI
Type of houses with at least 100 units

House Type

Med. House Price

50.000   158.578

Med. Rent Yroi

0,07 — Average House Types >= 100 units

| | 2BD_3BT | 4BD_2.5BT | 3BD_1.5BT | 4BD_2BT | 3BD_4BT |
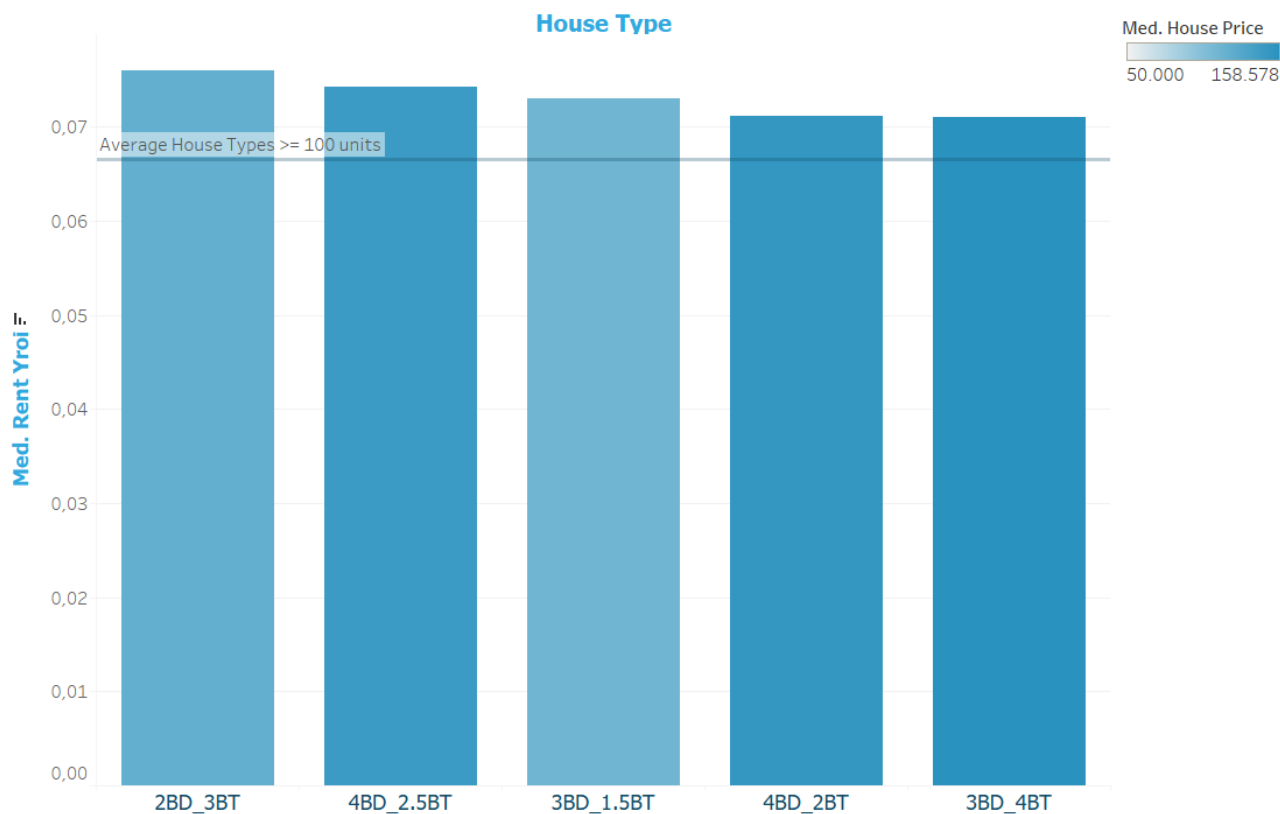
Media di Rent Yroi per ogni House Type. Il colore mostra media di House Price. La vista è filtrata su House Type, che mantiene 2BD_3BT, 3BD_1.5BT, 3BD_4BT, 4BD_2.5BT e 4BD_2BT.

The raking for the top 5 type of houses for the traditional Rent shows more profitability for smaller type of houses and less variance for the Rent YROI among all house types with at least 100 units, for the average house type Rent YROI is closer to the value of the type of houses' Rent YROI in the ranking.

Air vs Rent
For the top 10 cities By Numer of Houses
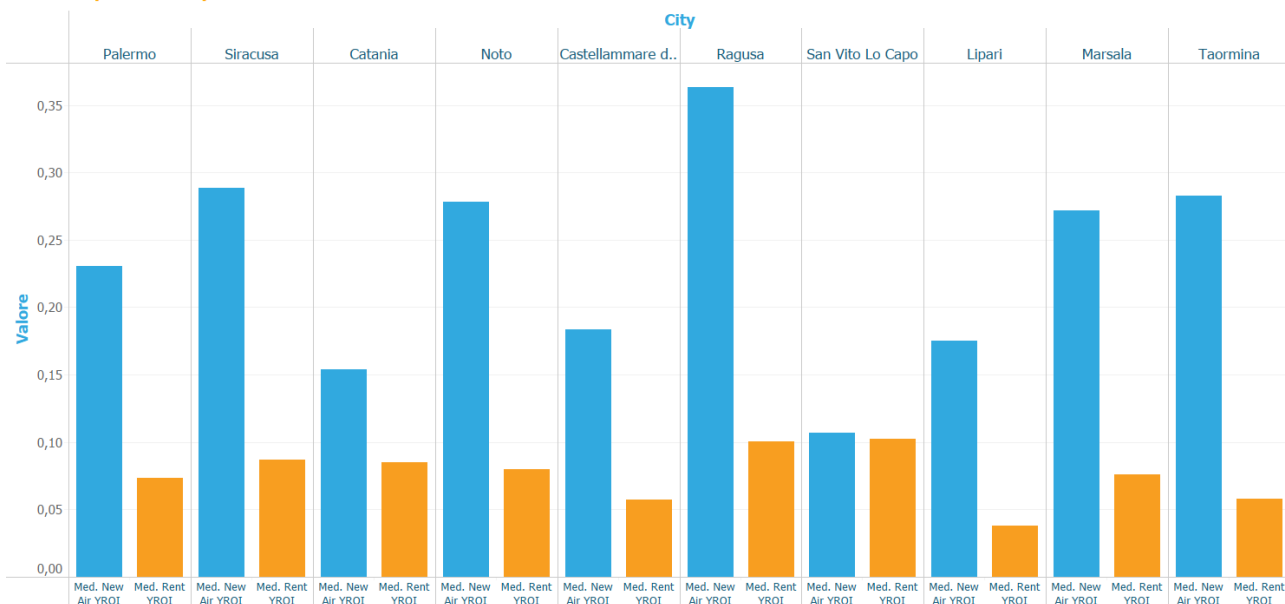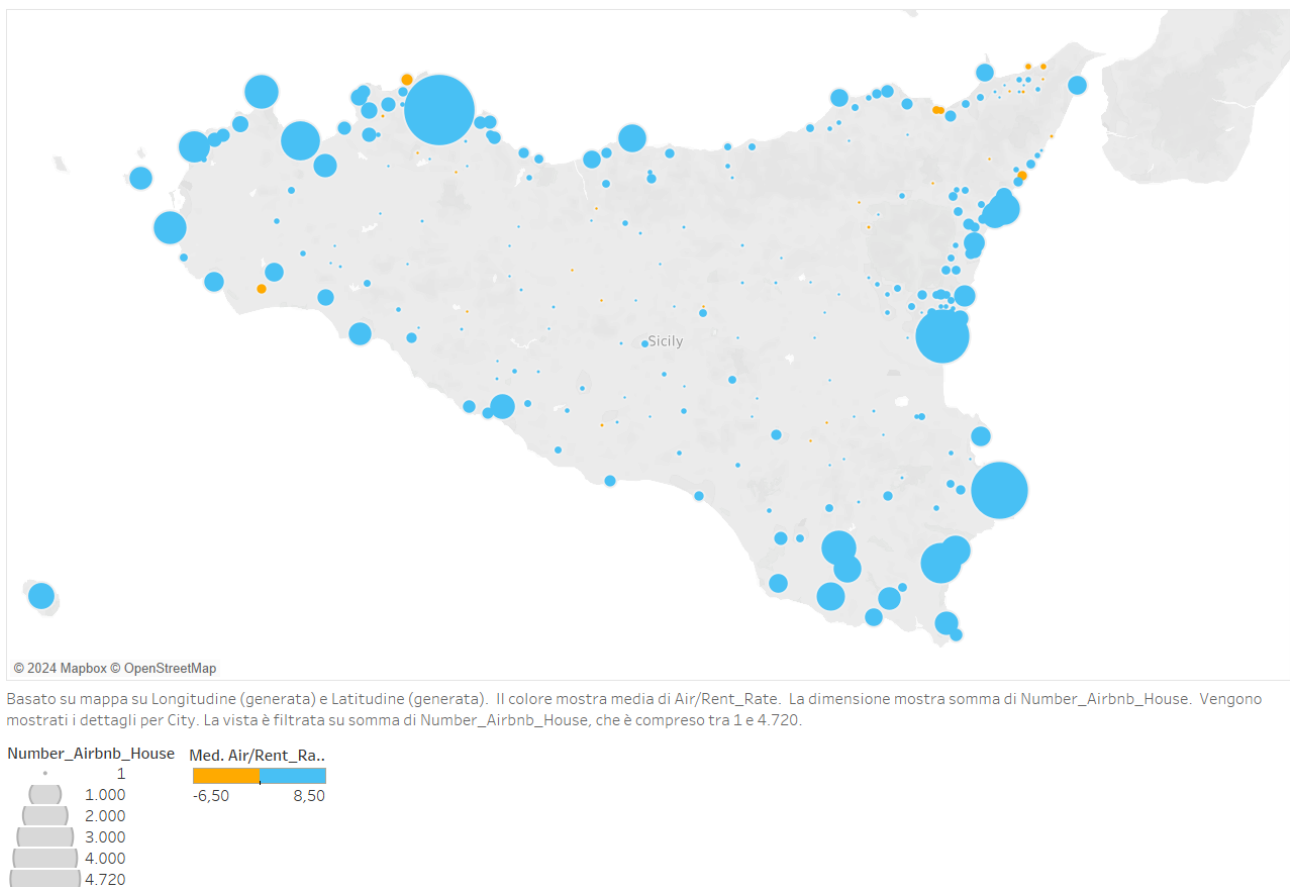
City

Med. New Air YROI e Med. Rent YROI per ogni City. Il colore mostra i dettagli relativi a Med. New Air YROI e Med. Rent YROI. La vista è filtrata su City, che mantiene 10 di 366 membri.

In the end, is it more profitable to rent or to list the house on Airbnb? Using the aggregated data set, I compare the Average New Air YROI against the Rent YROI for the 10 largest cities by the Number of Houses listed in Airbnb, with OC Var = 0 (modifiable in Tableau). For all these cities, Airbnb is by far the best choice in terms of return on investment. Only for San Vito Lo Capo the two YROIs have similar returns, and if we change the OC Var parameter to -5, we can see that in fact the Rent YROI is higher than the Airbnb one.

AIR vs RENT
For blue cities, Airbnb renting is more profitable.
For yellow cities, trantidional renting is more profitable.

Basato su mappa su Longitudine (generata) e Latitudine (generata). Il colore mostra media di Air/Rent_Rate. La dimensione mostra somma di Number_Airbnb_House. Vengono mostrati i dettagli per City. La vista è filtrata su somma di Number_Airbnb_House, che è compreso tra 1 e 4.720.

Number_Airbnb_House
- 1
- 1.000
- 2.000
- 3.000
- 4.000
- 4.720

Med. Air/Rent_Ra..
-6,50   8,50

Using the Air/Rent Rate computed field, I created a map of Sicilian cities. Blue dots represent cities with an Air/Rent rate higher than 1, that is, a city whose Average Airbnb YROI is higher than the Rent YROI, whilst yellow dots represent cities where renting is more profitable. The dimension of the dot represents the number of houses. As we can see, the map is dominated by blue dots, meaning that Airbnb is almost always the most profitable option. Even if we set the OC Var to -5, leading to lower Airbnb YROI, the blue dots remain the majority, with San Vito Lo Capo being the only significant yellow dot. On Tableau, is it possible to change the filter by the number of houses to exclude some cities and change the OC Var parameter.

Additional Examples

Top 5 types of Houses for the Top 5 Cities
Type of houses selected by New Air YROI and
cities selected on Occupancy Rate.
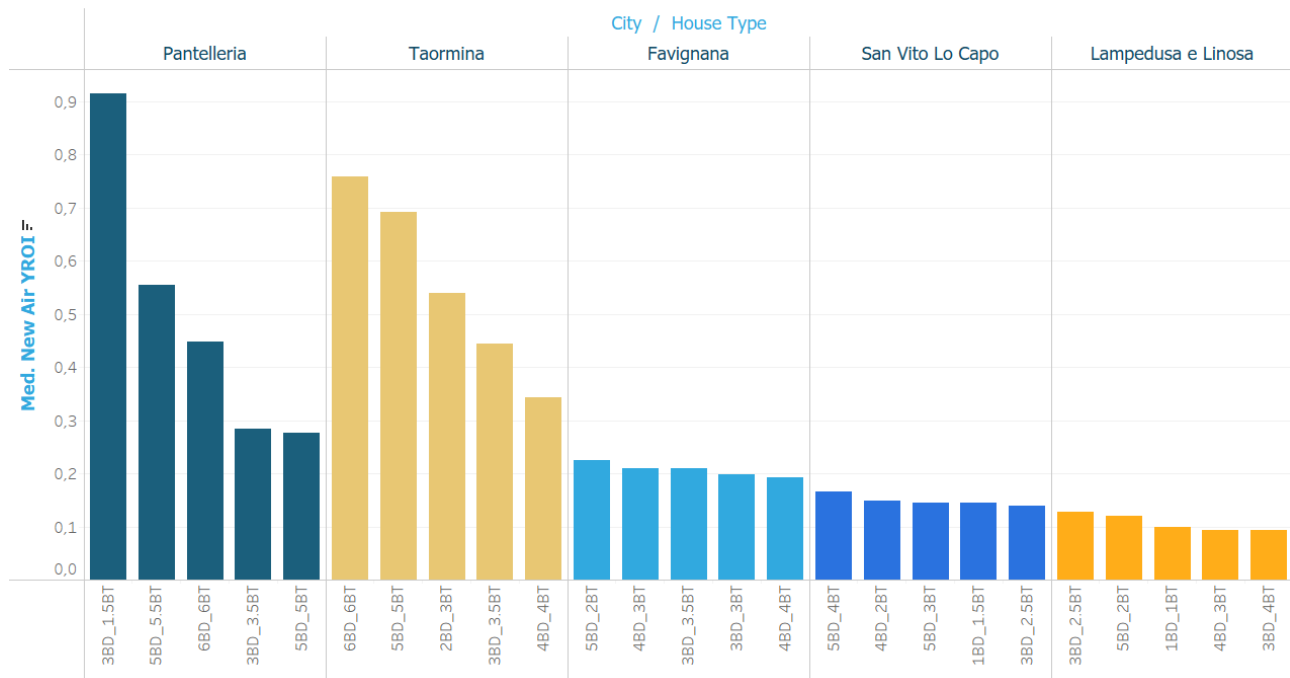
In the plot, the top 5 cities, selected by Occupancy Rate, are shown with the highest 5 types of houses ranked by the average New Air YROI. For the additional plots, first we filter the Cities by Number of houses (>=300) selecting the top 5 Cities by Occupancy rate and then we filter for Type of House (>=20). I employ the usage of an Index to rank the 5 highest houses for each city, regardless of their global ranking. Otherwise, if I used Type of House as a filter, we would have only 5 types of houses for all cities, leading to less than 25 types of houses in total.

Top 5 Cities for the top 5 type of houses.
Cities selected by Occupancy Rate and
type of houses selected for Air/Rent_Rate
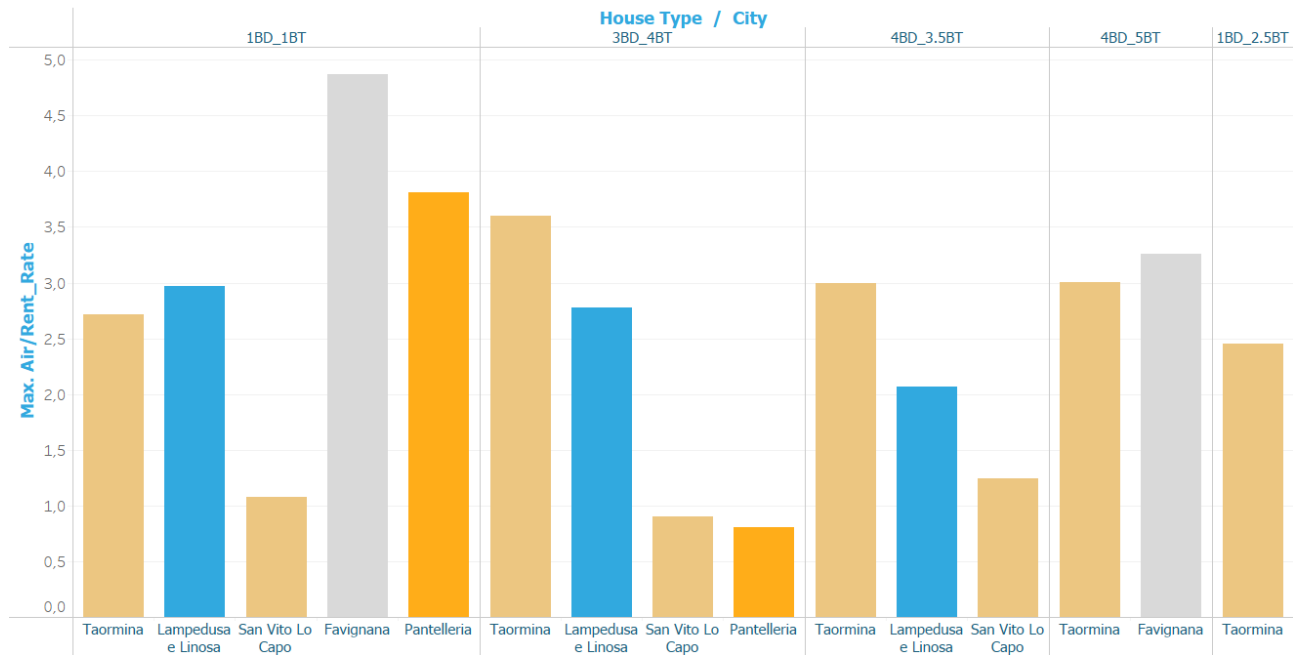
Massimo di Air/Rent_Rate per ogni City suddiviso per House Type. Il colore mostra media di Occupancy Rate. La vista viene filtrata su City e House Type. Il filtro City mantiene Favignana, Lampedusa e Linosa, Pantelleria, San Vito Lo Capo e Taormina. Il filtro House Type mantiene 1BD_1BT, 1BD_2.5BT, 3BD_4BT, 4BD_3.5BT e 4BD_5BT.

Med. Occupancy R..

0,47184   0,54501

While I could have kept the same approach for the second additional example, I think it was more interesting to show how we can leverage the colour to show both a city and the occupancy rate. The plot shows the Highest Air/Rent Rate for each of the top 5 Cities, ranked by occupancy rate and Type of House, ranked by Air/Rent rate.

In this case, we filter both Types of Houses for Air/Rent rate and Cities by Occupancy Rate directly. This results in fewer cities for each type of House, for the view must satisfy 2 global filters at the same time. Since each city has a unique, we can exploit this feature to represent with the colour both the city and the average Occupancy Rate.

Step 4 Overview

Step 4 of the assignment focuses on investing in houses, given some constraints:

- we must spend from 700K to 1.2M euros.

- we can buy (at least) 7 to (no more) 10 houses.

- we must buy houses at least in 4 different locations.

- we cannot buy more than 3 houses per location.

- we have **200k** available, and you must borrow the rest of the money with an interest rate ranging from [2 to 2.6] with 0.1 steps.

- Occupancy rate can vary from -5% to 5% to the expected one

This section of the report and the corresponding dashboard are designed to give the end user the freedom to explore the datasets at both aggregated and non-aggregated levels.

At the aggregated level, I focus on giving an overview of the market house in Sicily for the biggest cities (with at least 300 houses).

At the non-aggregated level, I deploy a dashboard that satisfies the constraints. Given the ETL transformations and computed fields using fixed dimensions for city and house type, it may happen that multiple houses in the same city can have the same house price and Airbnb YROI. Since we are interested in selecting only one house for a particular combination of city, house price and YROI, I filter the values by DUPLICATES REMOVER.

The dashboards made are the following:

1. The first Dashboard is made using aggregated data, designed to give the user the opportunity to quickly get insights into the Airbnb market, by focusing on the house price, the city and the type of house.

2. A dashboard showing the optimal solution, the one that maximizes the YROI and leads to the lowest total house price, making the most profitable investment.

3. The dashboard is made of non-aggregated data, that allows the user to choose individual houses or cities and interactively verifies if the constraints in the assignment are respected.

Aggregated Exploratory Dashboard

Before selecting individual houses, users might first explore the broader Airbnb house market in Sicily. This dashboard offers tools to understand the market, including house price distribution among cities and the average New Air YROI for each city. To avoid an overload of information and to keep the results relevant, I focus only on cities with at least 300 houses, maintaining the constraint made in the first part of the report. The result is the following one:
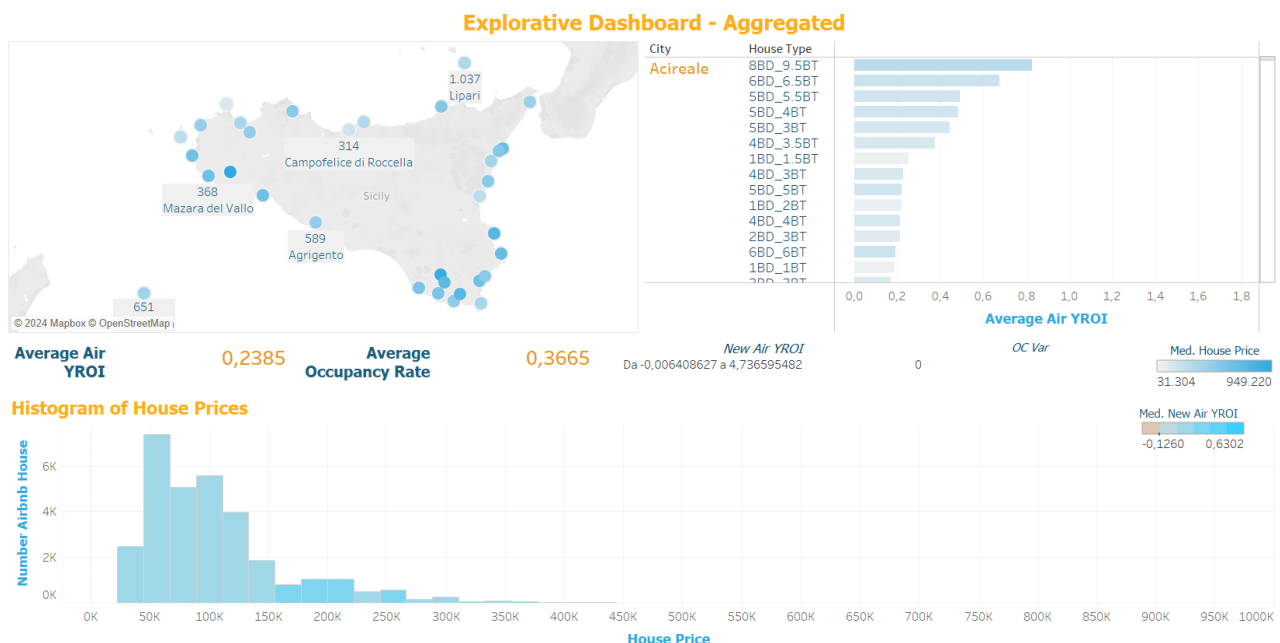

In the top-right corner, the map of Sicily shows the location of the towns in the region with more than 300 houses, along with the number of houses there are in each city and the colour corresponds to their Average New Air YROI.

The plot in the top right is a bar plot of the average New Air YROI for each type of house in each city, with the colour representing the house price. At the bottom, a histogram of the house price distribution is laid out, with the

colour representing the Average New Air YROI. At the center of the dashboard, 2 metrics are shown, Average Air YROI for all the values in the view and the average occupancy rate, along with the OC Var parameter and a filter to select cities and values depending on the New Air YROI. Each plot acts as a filter for the others. Selecting a bar in the histogram filters the cities and house types with the corresponding average house price range. Selecting a city on the map shows its house price distribution and average Airbnb YROI in the bar plot. This interactivity provides a comprehensive view of location, house price distribution, and average YROI for cities and house types.

Optimal choice

Investors often aim to maximize YROI by selecting the best group of houses. Based on this assumption, a dedicated dashboard shows the best 10 houses that leads to the highest YROI.
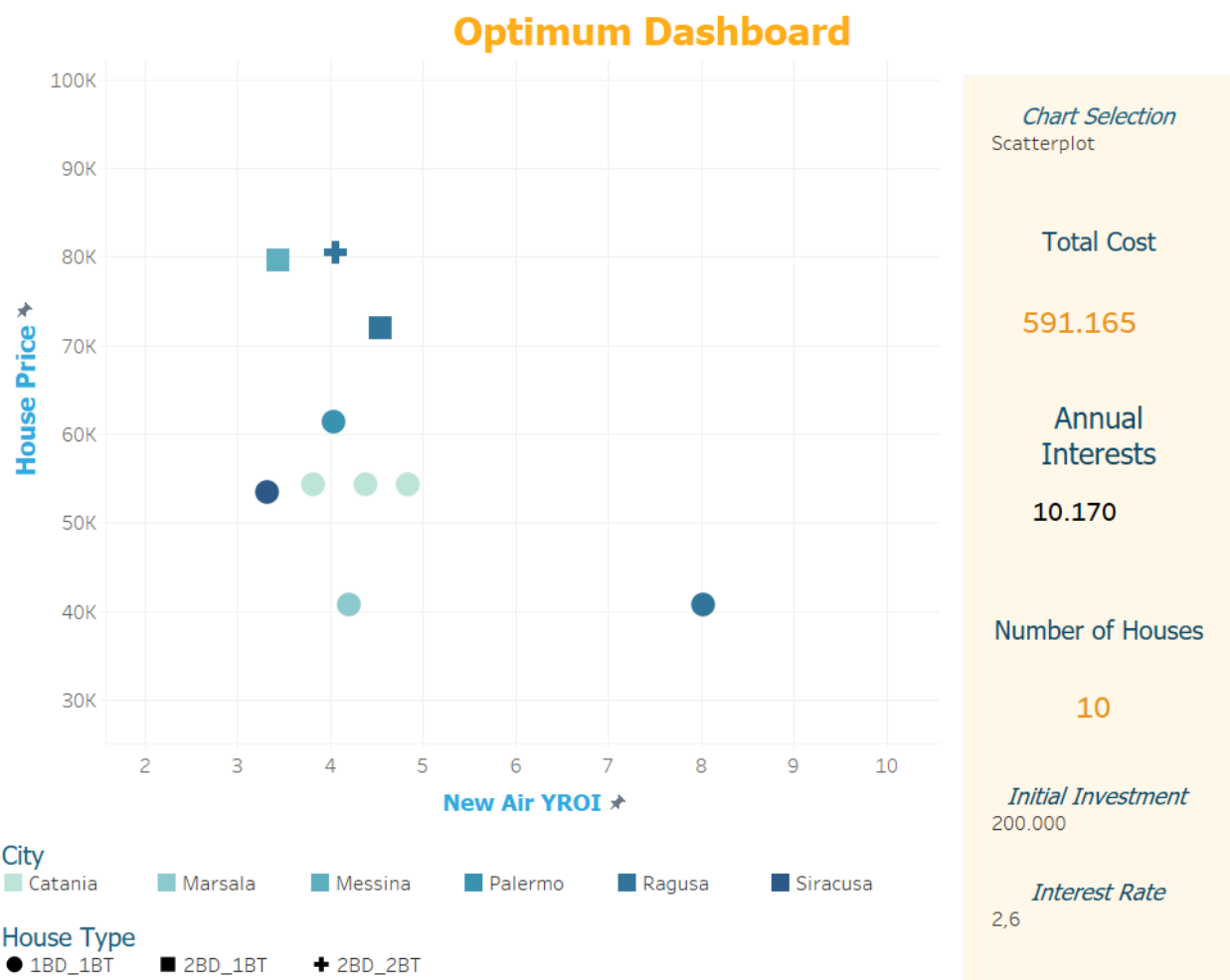
The Optimum Dashboard shows a scatterplot with the New Air YROI on the x-axis and the house price on the Y-axis, along with the legends. On the right side, the Total Cost and number of houses are shown. The Chart Selector allows to switch between the scatterplot and a map showing the location of the 10 houses, along with the count of houses in each city.

Initial investment and Interest Rate are two parameters that allow to compute the annual interests paid on the loan.

Negative interest indicates an initial investment higher than the total cost. Results are filtered using the duplicates remover to avoid duplicate houses with the same new Airbnb YROI, city, and house price. According to this analysis, the best houses include 3 from Ragusa and Catania, and 1 each from Palermo, Messina, Marsala, and Siracusa. The mean YROI for this selection is 4.61 with an OC Var of 0.

Exploratory Dashboard Step 4 – Individual houses.

The final dashboard meets all assignment requirements. It allows users to individually select houses by city or house type.

Explorative Dashboard - Individual houses
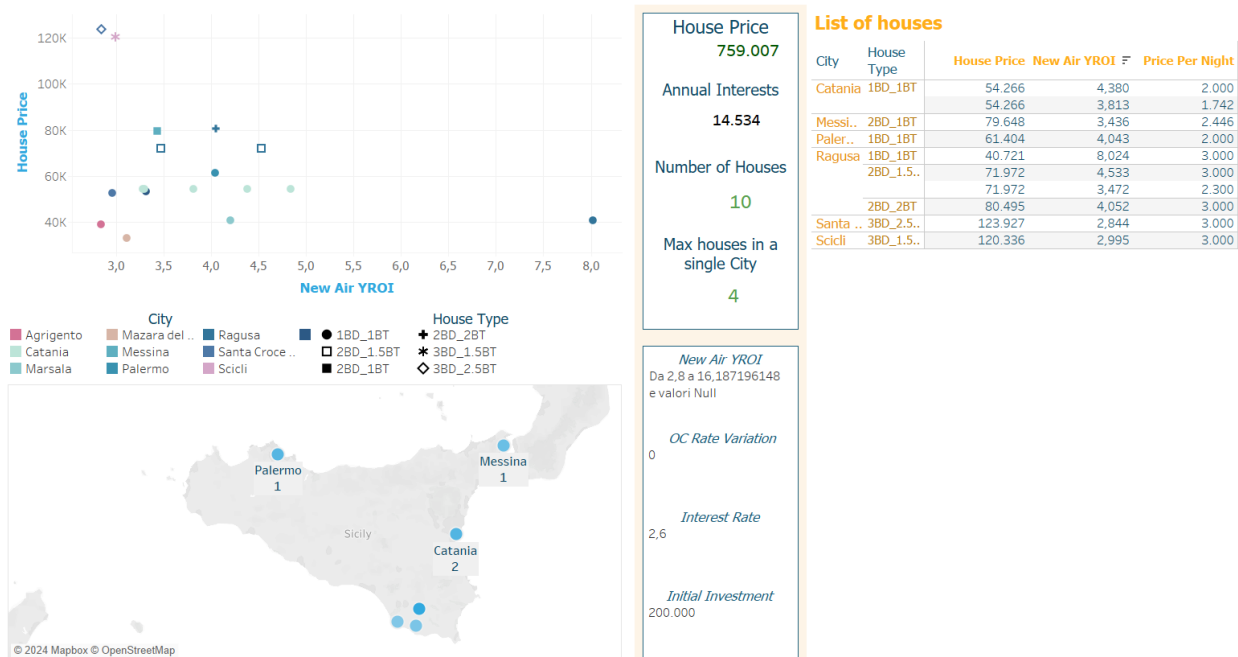
- On the left, the **scatterplot** shows all houses filtered by duplicates remover, with colour representing the city and shape representing the house type.

- The center column shows two boxes, one for the **key metrics** :
  - Total House Price.
  - Annual Interests on the loan.
  - The number of houses selected
  - Max number of houses in a single city.
  - The constraints are checked with a conditional formatting: if the constraints are not met, the metric is shown in red.
- The second box contains **parameters** and a **filter** to select houses based on their New Air YROI, the OC Var parameter to see how the YROI changes, and Interest Rate and Initial Investment to choose the initial cash put into the investment and see how the interest changes.
- The right **table** shows all filtered houses, along with their house type, house price and price per night.

Each plot can filter the others. Selecting a city on the map updates the scatterplot and table to show only houses in that city. Users can manually select houses from the table or scatterplot, depending on their preferred metrics.

Appendix

Color Usage

The choice of the colours in this report were made specifically to recall the Sicily in its main symbols:

Orange [#FFAD19] to represent the ever-present sun of the Sicily, its sandy beaches and the wheat, one of the main agricultural products of the island.

The Indigo Dye [#104B66] represents the deep blue color of the sea that surrounds the island. It's easily and used for writing

The Picture Blue [#31A9DF] represents the color of the blue skies and good weather, typical of an ordinary summer day in the region.

Also, the combination respects the needs of colorblind people by being well-differentiable.

Color Palette

True	Prot.	Deut.	Trit.