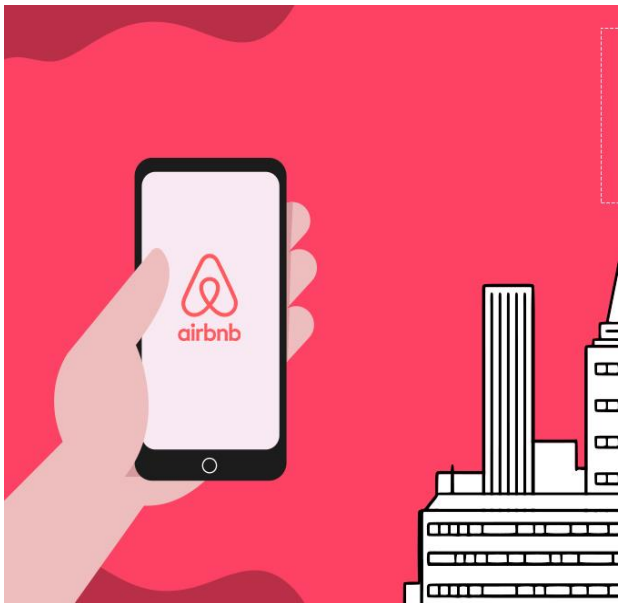


AIRBNB IN SICILY

Data Cleaning and Financial Estimates

DATA SCIENCE – UNICT
ACADEMIC YEAR 2023/2024

Giuseppe Leonardi
1000065630

BIG DATA ANALYTICS
PROFESSOR: MORANA GIOVANNI

Università di Catania

Contents

Introduction

Tableau is a powerful data visualization tool that consists of two main components: Tableau Prep and Tableau Desktop.

1. **Tableau Prep**: allows users to clean, shape, and combine data for analysis. It provides a visual and intuitive interface for performing various data cleaning and transformation tasks, such as filtering, aggregating, and joining datasets.

2. **Tableau Desktop**: is a data visualization tool that allows users to create interactive and insightful visualizations from their data. It offers a wide range of visualization options and advanced features for data analysis, such as calculated fields, parameters, and trend analysis.

Overall, Tableau is a versatile tool that empowers users to transform data into actionable insights through interactive and visually appealing visualizations.

In this report, I will use these tools to clean and analyze data regarding the trends in housing costs and rents in Sicily using the Airbnb application. Below, a comprehensive description of the initial data available to us will be provided.

Airbnb Dataset

The available info, essential for our analysis, are all contained in four Excel files: **CitiesInSicily**, **CitiesInSicily_BUY_RENT**, **airbnbPrice**, and **HouseInfo**. Lets' extract a conceptual model from the dataset:

- *CitiesInSicily* consists of two columns, one containing the names of the cities and the other with their corresponding provinces. This establishes a geographical dimension where each city is associated with a province.
- *CitiesInSicily_BUY_RENT* contains the prices of sale and rent per square meter in euros for each city in Sicily. This allows us to analyze real estate market trends within Sicily's cities.
- *airbnbPrice* focuses on short-term rental prices, listing the price per night for different houses available on Airbnb.
- *HouseInfo*, provides additional details about these Airbnb houses, such as the number of bedrooms and bathrooms, and the city in which each house is located.

Logical Design

The logical design translates this conceptual model into a structured format with specific tables and relationships. The CitiesInSicily table serves as the foundation, where **City** column is the primary key, ensuring each city is unique. The CitiesInSicily_BUY_RENT links to CitiesInSicily via the **City** column, which is a foreign key here. This linkage allows us to reference real estate prices directly back to their respective cities.

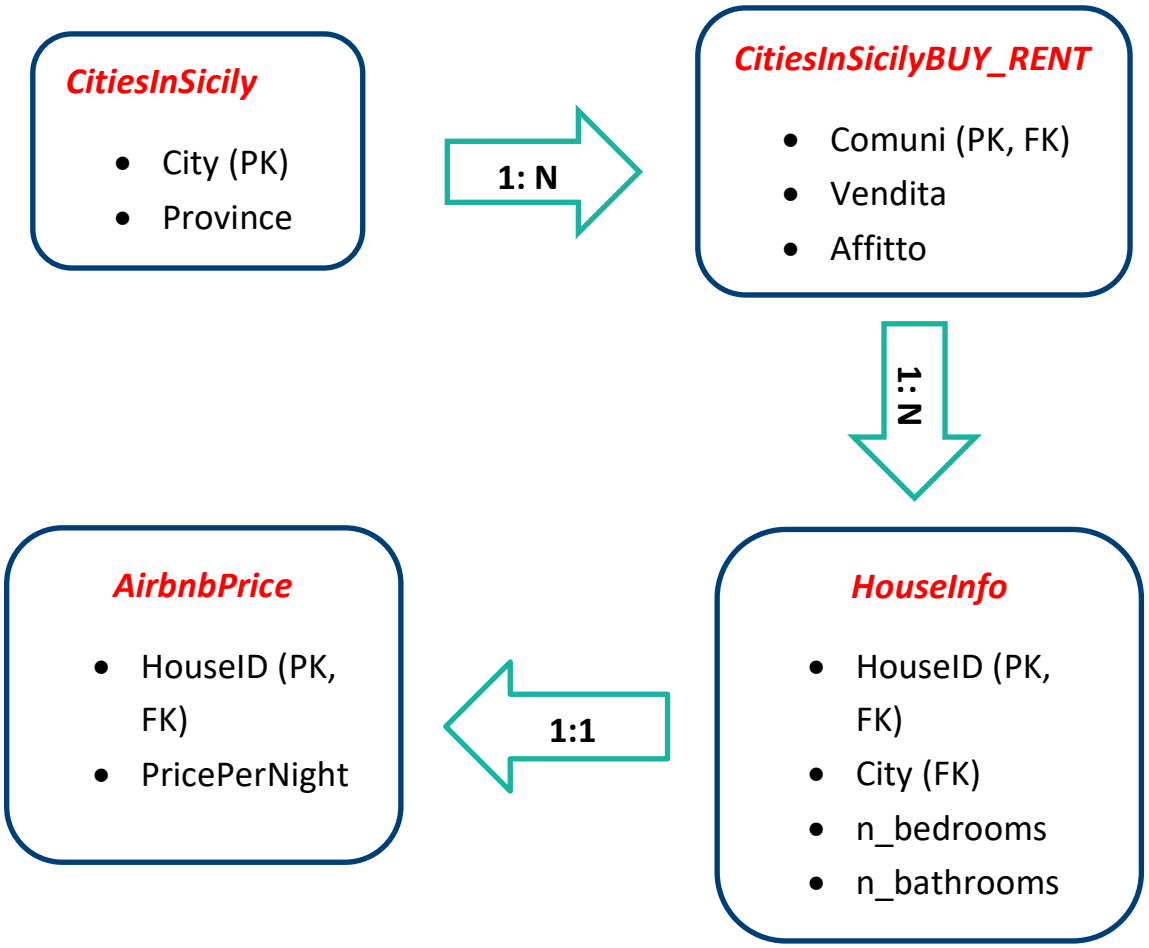
Next, airbnbPrice lists Airbnb listings with a unique identifier for each house (**HouseID**) and the nightly price (**PricePerNight**). Complementing this, the HouseInfo table provides more detailed information about each house listed on Airbnb. It includes the **HouseID** (which is also the primary key and a foreign key linking to airbnbPrice), the city where the house is located (**City**), and attributes like the number of bedrooms (**NumberOfBedrooms**) and bathrooms (**NumberOfBathrooms**). The city column in HouseInfo is a foreign key linking back to the CitiesInSicily table, ensuring that each house is associated with a specific city.

Relationships

The relationships between these tables are key to maintaining the integrity and usability of the dataset:

- CitiesInSicily table has a one-to-many relationship with CitiesInSicily_BUY_RENT, meaning each city can have multiple real estate price entries but each entry corresponds to a single city.
- CitiesInSicily has a one-to-many relationship with HouseInfo, as each city can host multiple Airbnb listings.
- The airbnbPrice and HouseInfo tables have a one-to-one relationship, where each house listing in airbnbPrice corresponds to a single entry in HouseInfo.

The best way to make this information more comprehensible is to summarize them through a graphical representation.



ETL Operations

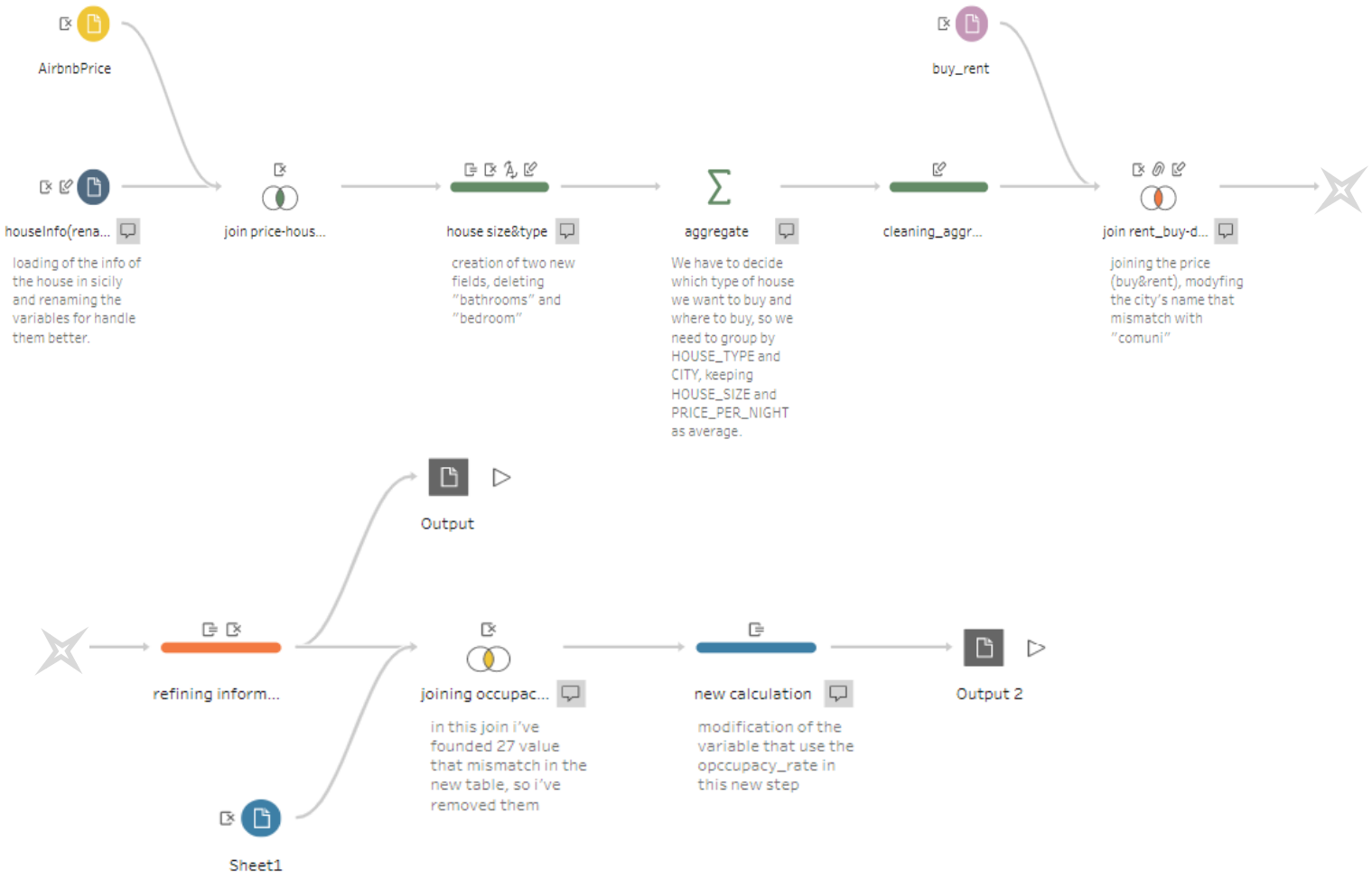
ETL stands for ***Extraction***, ***Transformation*** and ***Loading***. This is a set of fundamental procedures in data management. Here is a detailed explanation of each phase:

1. ***Extraction*****:** In this phase, data is extracted from one or more data sources. Data sources can include databases, text files, spreadsheets, web APIs, and so on. The goal is to gather raw data from these sources and bring it into the data management system.

2. ***Transformation*****:** After extraction, the raw data undergoes transformation into a format suitable for analysis, loading, or other purposes. This phase may involve various activities, such as data cleaning (removing duplicate or erroneous data), data normalization (bringing data to a common standard), data enrichment (adding current information to existing data), and data transformation (e.g., calculations or aggregations).

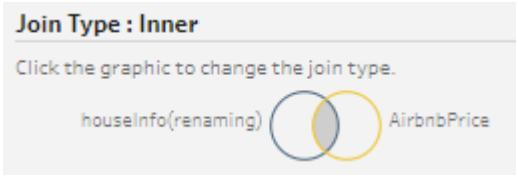
3. ***Loading*****:** Once transformed, the data is loaded into the target system, which could be a data warehouse, operational database, or another type of data management system. Data loading can be done in batch mode, where data is loaded periodically, or in real-time mode, where data is loaded continuously as it arrives.

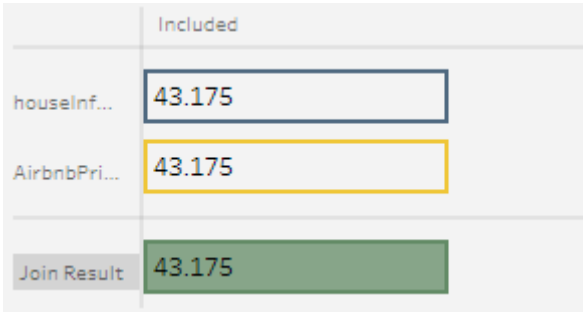
Now it is time for us to use Tableau Prep, to conduct these operations. Let us start with a general overview of the flow executed and then we concentrate on the most key steps of the process:



Let us now analyze more specifically the operations performed on the data. Recall that ETL operations follow each other and can be repeated whenever it is necessary to update the available information.

Let us start with the houseInfo table, for which I only modified the variable names to make them easier to call. Then, I performed an Inner Join with airbnbPrice, using the clause H.id = A.id, which found a perfect match.





So, we added a new column to HouseInfo, the price per night of each house. Now, in the third point of the flow (*house size&type*), I have created a Calculated Field called house_type, which unify the information contained in bedrooms and bathrooms in one single column, so we can delete the two columns just recently mentioned. Then, it is important to add another Calculated Field, house_size, which provides an estimate of the house size based on the following formula:

```
if [bedrooms] == 1 THEN 18+
([bedrooms]*16)+([bathrooms]*9) ELSE
35+([bedrooms]*16)+([bathrooms]*9)
END
```

The logic behind the formula is:

1. If the house has one bedroom, we use 18 as starting point to estimate the house size.
2. If the house has more than one bedroom, we use 35 as baseline.

Now, we can proceed with the Aggregation of our data. Aggregation is the process of summarizing data by combining multiple values into a single value. This is typically done by performing mathematical operations on a set of data, such as summing up numbers, calculating averages, counting occurrences, finding minimum or maximum values, and more.

In our case, I have grouped our data by city and by house_type, averaging house_size, price_per_night and summing the number of rows. These are the first rows of our dataset after the described operations:

Abc city	Abc house_type	# avg_house_size	# avg_price_per_night	# n_airbnb_house
Calatafimi-Segesta	4BR_2BT	117	38	1
Sperlinga	2BR_1.5BT	80,5	90	1
Butera	7BR_7BT	210	1.176	1
Ficarra	2BR_2BT	85	30	1
Acireale	2BR_2.5BT	89,5	127	2
Al Terme	1BR_1.5BT	47,5	70	1

As we can see, this table presents for each city and for each type of house the average house size and price per night. For example, Acireale has two houses with 2 bedrooms and 2.5 bathrooms. For these two houses, the mean of the sizes is 89,5 m^2 and have an average price per night of 127€.

Now it is time to join this new table with ***CitiesInSicilyBUY_RENT***, with the constrains of City = Comuni. This inner join finds some value that mismatch: this depends on the fact that the Column City do not have letters with accents (*Cefal = Cefalù*). So, I have proceeded to correct the Cities adding the final letters. Now, all the row of the table coming from the aggregation found a match, but 27 rows of second table do not have a match. In this case, we simply do not consider them because it means we do not have data on houses regarding those 27 excluded municipalities.

	Included	Excluded
cleaning_...	3.798	0
buy_rent	363	27
Join Result	3.798	

One the most important moment of this process is now: I have calculated some Field to have various measures concerning average house prices by type and by city, their performance, taxes paid on them, earnings from rent, and earnings from Airbnb rentals.

Calculated Field
annual_rent
[avg_house_size]*[rent]*12

Calculated Field
house_price
[avg_house_size]*[price]

Calculated Field
local_annual_taxes
[house_price]*0.015

Calculated Field
annual_airbnb_rent
[avg_price_per_night]*365*(0.35+0)

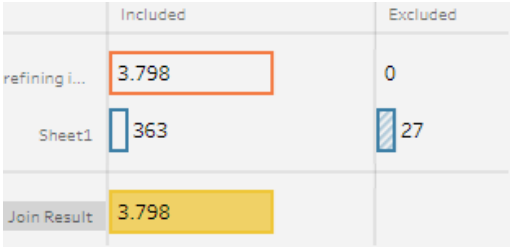
Calculated Field
airbnb_service_cost
[annual_airbnb_rent]*(0.03+(3/300))

Calculated Field
rent_yroi
([annual_rent]-[local_annual_taxes])/[house_price]

Calculated Field
airbnb_yroi
([annual_airbnb_rent]-[airbnb_service_cost]-[local_annual_taxes])/[house_price]

- annual_rent: calculated by multiplying the average house size with the monthly rent and then by 12 months.
- house_price: Calculates the total price of the house by multiplying the average house size with the price per square unit.
- local_annual_taxes: Determines the annual taxes on the property, calculated as a percentage (0.015) of the house price.
- annual_airbnb_rent: Computes the annual income from Airbnb multiplying the average price per night with 365 days and an occupancy factor (0.35+0), assuming 35% occupancy.
- airbnb_service_cost: Calculates the service cost for Airbnb rentals, considering a percentage fee (0.03) and a flat fee per booking (3/300) on the annual Airbnb rental income.
- rent_yroi: Computes the yield in a year on real estate investment based on traditional renting, subtracting taxes from rental income and then dividing by the house price.
- airbnb_yroi: Calculates the yield in a year on real estate investment based on Airbnb renting, accounting for service costs and taxes, then dividing by the house price.

Here, the first output was created, an Excel sheet containing all data obtained and processed up to now. Later, I became aware of necessary information that I did not have before: the occupancy rate for each city. So, the next step will be to perform a join with this data, present in the file ***Sheet1***. We need to join the clause City = city, where, for the same reason explained at the second join, we have 27 values mismatched in Sheet1, of which we simply do not care.



After having this information, it is important to modify the Calculated Field of the occupancy rate estimated by us with the real occupancy rate. Now that we have the clean and ready data, we can create the output that will transfered to Tableau Desktop to continue our analysis.

Type	Field Name
#	ID
Abc	city
Abc	house_type
#	avg_house_size
#	avg_price_per_night
#	n_airbnb_house
#	annual_rent
#	house_price
#	local_annual_taxes
#	annual_airbnb_rent
#	airbnb_service_cost
#	rent_yroi
#	airbnb_yroi
#	Occupancy Rate

Each variable of the output was explained during the process, but an important thing must be noticed: here, the variables do not represent only one house, but the mean of all houses of a specific type in a specific city. So, each row, identified by ID, represents aggregates measurements.

Data Visualizations

Let us start with a very general exploratory data analysis to get a broad overview, and then delve deeper according to our needs. In these two tables I have reported, respectively, the top ten City and the top five house type for number of houses. In this case choice of a gradient colour helps to evidence the difference between City or Type of house.

City	
Palermo	4.720
Siracusa	3.083
Catania	2.751
Noto	1.564
Castellammare del Golfo	1.459
Ragusa	1.156
San Vito Lo Capo	1.086
Lipari	1.037
Marsala	1.021
Taormina	958

Here the evidence is that Palermo stands out significantly from Siracusa, Siracusa less from Catania, which is third, while all the other cities hover around 1000 units.

House..	
1BR_1BT	16.160
2BR_1BT	10.458
2BR_2BT	4.020
3BR_2BT	3.000
3BR_1BT	1.461

Here, we notice a strong presence of small houses, with a clear predominance of those with one bedroom and one bathroom; every other type is significantly less common.

Now we can continue the analysis using our KPI. KPIs are metrics used to evaluate the success and performance of specific activities or investments. In this case, airbnb_yroi and rent_yroi (as explained before), measures the profitability and effectiveness of investing in a property for Airbnb rentals and traditional rentals, making them a valuable KPIs for assessing the financial performance of such investments. For the Airbnb index, it could be useful to consider the variation of the occupancy rate. To add this variation, I have created a parameter (oc_rate_var%) and added it in the formula of the Airbnb_yroi.

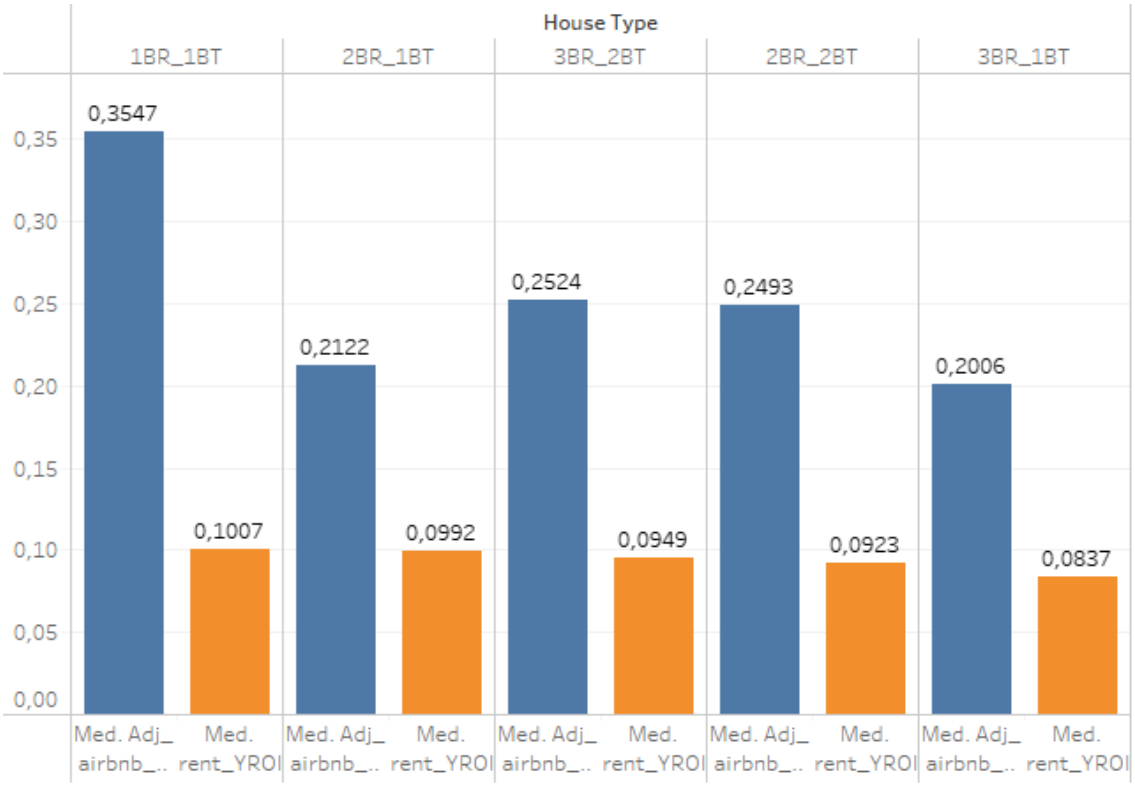
(([airbnb_yroi]*[house_price])+([airbnb_service_cost]+[local_annual_taxes]))*((100+[oc_rate_var_%])/100)) / [house_price]

Adding this term, we obtain the adj_airbnb_yroi, that directly consider the variation of occupation rate. So, it could be interesting to see the average behavior of this indexes in the cities (with the constraints that cities must have more than 300 houses):

City	
Mazara del Vallo	0,1785
Ispica	0,1160
Mascali	0,1007
Santa Croce Camerina	0,0995
Scicli	0,0980
Alcamo	0,0971
San Vito Lo Capo	0,0874
Ragusa	0,0854
Messina	0,0725
Siracusa	0,0722

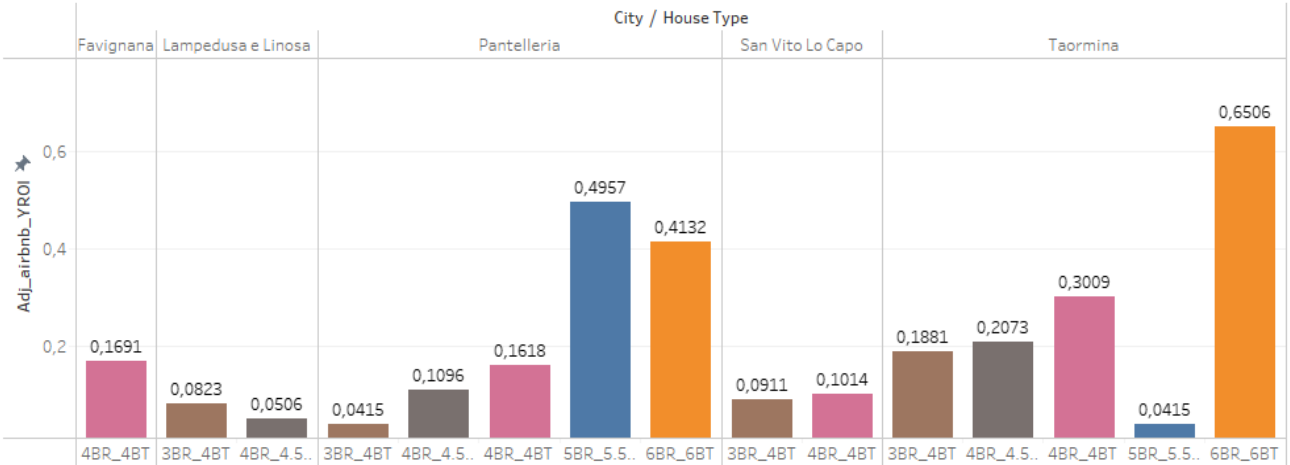
City	
Castelvetrano	0,5012
Ragusa	0,4543
Augusta	0,4016
Modica	0,3795
Ispica	0,3780
Mazara del Vallo	0,3704
Sciacca	0,3480
Marsala	0,3458
Siracusa	0,3347
Santa Croce Camerina	0,3021

We notice that the adj_airbnb_yroi (second table) contains significantly higher values compared to rent_yroi, also considering the less value of the oc_rate_var%, clearly suggesting that renting through Airbnb is more profitable than traditional channels. Additionally, we observe that the cities, with a few exceptions, are not the same for the two indices, indicating that it is better to use Airbnb in certain cities rather than others.



Also for the type of house (filtered with types that have more than 100 units) we can observe the same tendency of the two indexes for the cities, with value more stable but so much smaller for the rent_yroi.

Now, we can continue our analysis with a bar graph comparing the adj_airbnb_yroi across different cities and house types:



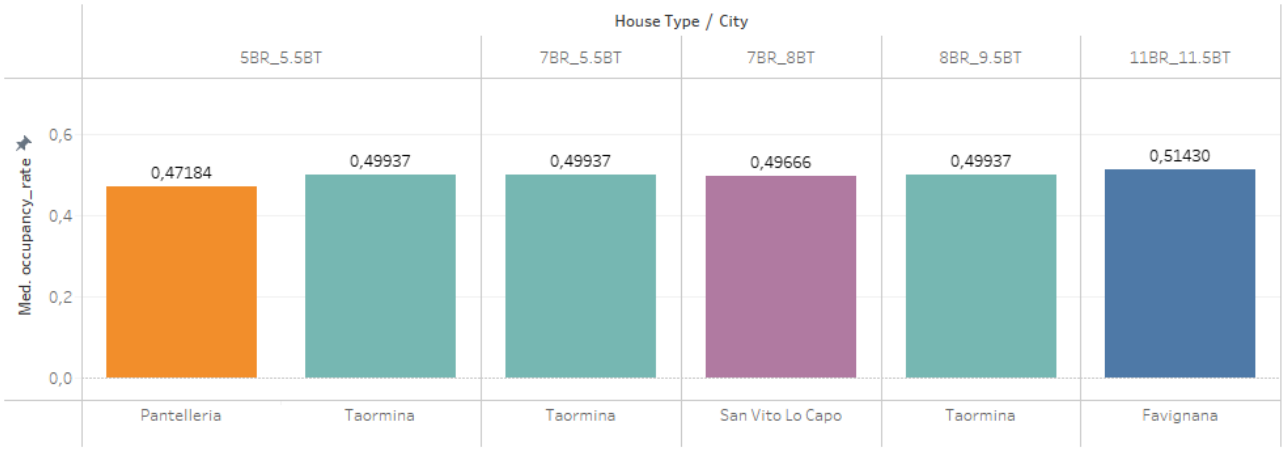
Given that the cities shown are the top five in terms of occupancy rate, each bar represents the adj_airbnb_yroi for a specific house type within a city. From this plot, we can have some conclusions:

1. Properties with more bedrooms and bathrooms tend to have higher adj_airbnb_yroi, especially noticeable in cities like Pantelleria and Taormina.
2. The ROI varies significantly between different cities and house types, indicating the importance of location and property type in determining profitability on Airbnb.
3. Investors should consider larger properties in cities like Pantelleria and Taormina to maximize returns, as these cities not only have high occupancy rates but also show high returns for larger properties.

There is a clear trend that larger properties tend to have better returns in cities with high demand. Therefore, investing in properties with more bedrooms and bathrooms could be a winning strategy.

On the other side, we can compare the median occupancy rates for different house types across several cities. These house types are among the top five in terms of the ratio between the adj_Airbnb_yroi and rent_yroi (***adj_index_yroi***), a new Calculated Field.

The adj_index_yroi provides a comparative measure of the profitability of renting out a property on Airbnb versus traditional long-term rentals: a ratio greater than 1 indicates that the Airbnb rental strategy is more profitable than the long-term rental strategy. Understanding the median occupancy rate in each city of these highly profitable properties provides deeper insight into their performance.



Each bar represents the median occupancy rate for a specific house type within a city. We can observe something important in this case:

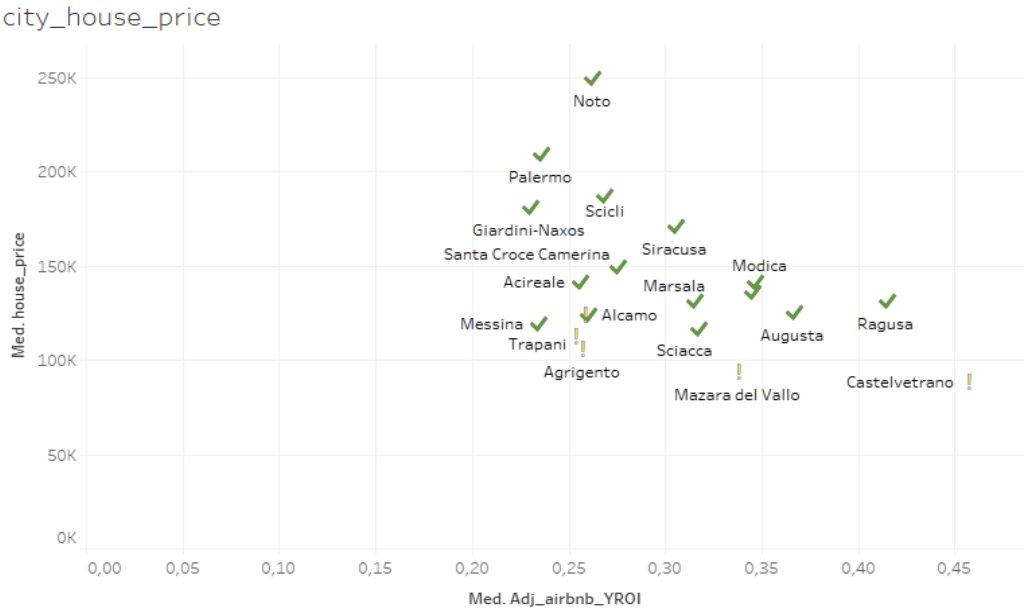
1. Taormina shows consistently high occupancy rates across different property types (5BR, 7BR, and 8BR), suggesting a strong demand for larger properties in this city.
2. Favignana's large properties (11BR_11.5BT) have the highest occupancy rate, indicating a strong demand for large properties.
3. Pantelleria's 5BR_5.5BT properties have a slightly lower occupancy rate compared to similar-sized properties in Taormina and San Vito Lo Capo.

Investors might find higher returns in cities like Taormina and Favignana where large properties have high occupancy rates. Understanding the median occupancy rate helps investors to catch the potential revenue from rentals in these cities. The difference in occupancy rates across different cities and property sizes highlights the varying demand levels and potentially different rental market dynamics.

Investment Strategy

Now that we have the data and graphical visualizations, the goal is to find a combination of houses (where each 'house' represents the average of houses of that type and in that city) that can maximize profit (winning strategy). To achieve this goal, given certain constraints that must be respected, we need to further explore the data to identify patterns or highlight the most profitable houses.

A good starting point could be to investigate the relation between adj_airbnb_yroi and the house_price for each city, also considering the occupancy rate of that city.



The provided scatter plot visualizes the relationship between the medium house price (y-axis) and the medium adjusted Airbnb yield rate (x-axis) for various cities. The points are marked with distinct colors and symbols based on the occupancy rate:

- Green check marks (✓) indicate cities where the occupancy rate is greater than 30%.
- Yellow exclamation marks (!) indicate cities where the occupancy rate is less than 30%.

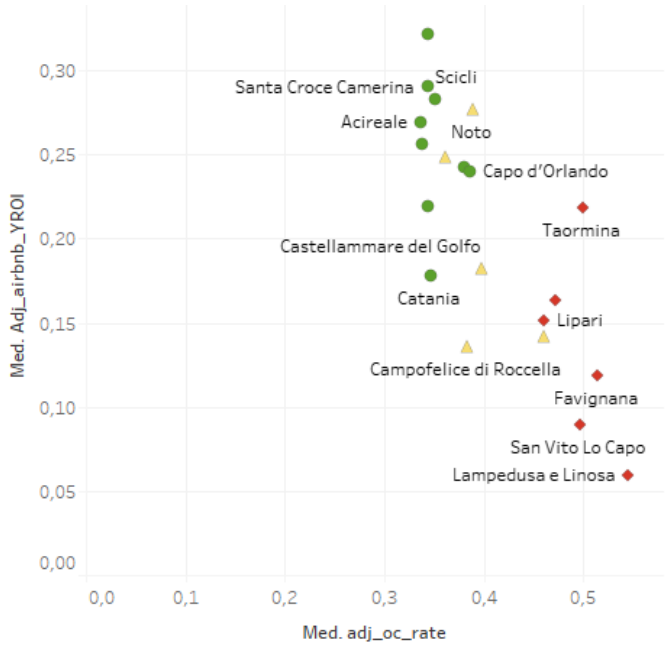
As we can see, Cities like *Ragusa* and *Augusta*, positioned towards the right of the x-axis with a median adjusted Airbnb yield rate above 0.35, have high profitability from Airbnb rentals and a high occupancy rate. *Noto* and *Palermo* have relatively high medium house prices and an occupancy rate above 30%, making them potentially attractive despite the higher initial investment. Cities such as *Sciacca* and *Marsala* have moderate house prices (around 150K) but a high ROI, which might indicate

a good balance between investment and return. *Castelvetrano* and other city are marked with a yellow exclamation mark, indicating an occupancy rate below 30%. Even though it has a high adjusted Airbnb yield rate the low occupancy rate could be a risk factor, potentially leading to unstable income streams.

In conclusion, to maximize profit while considering occupancy rates:

- Focus on cities with high adjusted Airbnb yield rates and high occupancy rates.
- Consider the balance between house prices and potential rental income.
- Be careful of cities with low occupancy rates despite high yields, as this could indicate volatility in income.

The following scatter plot depicts the relationship between two variables for various locations, with the colors of the markers representing different ranges of house prices.



- **Red:** House prices are greater than 300,000.
- **Yellow:** House prices between 200,000 and 300,000.
- **Green:** House prices are under 100,000.

We can observe that locations like *Lampedusa e Linosa*, *San Vito Lo Capo*, and *Favignana* (red markers) have high occupancy rates but relatively low Airbnb YROI. These places are expensive but may not yield high returns on Airbnb investments.

Locations such as *Catania*, *Castellammare del Golfo*, and *Capo d'Orlando* (green markers) have moderate occupancy rates and higher Airbnb YROI. These places are less expensive but offer better returns on Airbnb investments. Locations like *Scicli* and *Santa Croce Camerina* have a relatively high

median Airbnb YROI and moderate occupancy rates. These places are affordable and offer high returns on Airbnb investments.

Locations like *Taormina* and Lipari (red markers) have moderate to high occupancy rates but moderate YROI. These places are expensive but do not necessarily yield high returns.

The plot shows that there is a trade-off between house prices, occupancy rates, and Airbnb YROI across various locations. Cheaper locations (green markers) tend to offer higher returns on Airbnb investments, while more expensive locations (red markers) might not yield as high returns despite potentially higher occupancy rates.

To effectively invest in Airbnb properties in Sicily, it is essential to adopt a strategic approach. After the analysis, I have tried to outline a strategy to identify, evaluate, and select properties that align with our investment goals. We will divide the rest of the analysis in two scenarios: one where we have all the money with us and another where we need to get a loan to do our investment.

First Scenario

So, as we said, in this scenario we have all the money available and the choice of where to invest is based on several factors presented in the following paragraph.

Considerations

The key factors that we need to monitor and pay attention to are:

1. Budget & city constrains

With a total budget of €1.2 million, we need to spend at least €700,000 to purchase between 7 and 10 houses. These houses must be spread across at least four different cities, with no more than three houses in any one city.

2. High-Yield Locations

Focus on cities with high adj_airbnb_yroi and stable occupancy rates.

3. Property Type

Larger properties with more bedrooms and bathrooms typically yield higher returns.

4. Occupancy Rates

Occupancy rate is a critical factor in ensuring consistent rental income. Target cities with occupancy rates above 30%.

5. Balancing Cost and Yield

Utilize a balance between the initial investment and potential returns. Leverage data insights to find properties where the purchase price aligns with the projected yield.

Calculations

We also need to create some metrics (calculated fields) to obtain an accurate measure of the house's profitability. First, I have started calculating the day of occupation per house multiplying the adj_oc_rate by 365, so we have the precise day of occupation during a year when the occupation rate varies.

Then, we can consider the avg_price_per_night to have an estimation of the profitability:

```
[earnXyear]=[avg_price_per_night]*[n_days_occupied]
```

But this is not the net earn for each house. To have it, we need to subtract the local annual taxes and the taxes related to Airbnb.

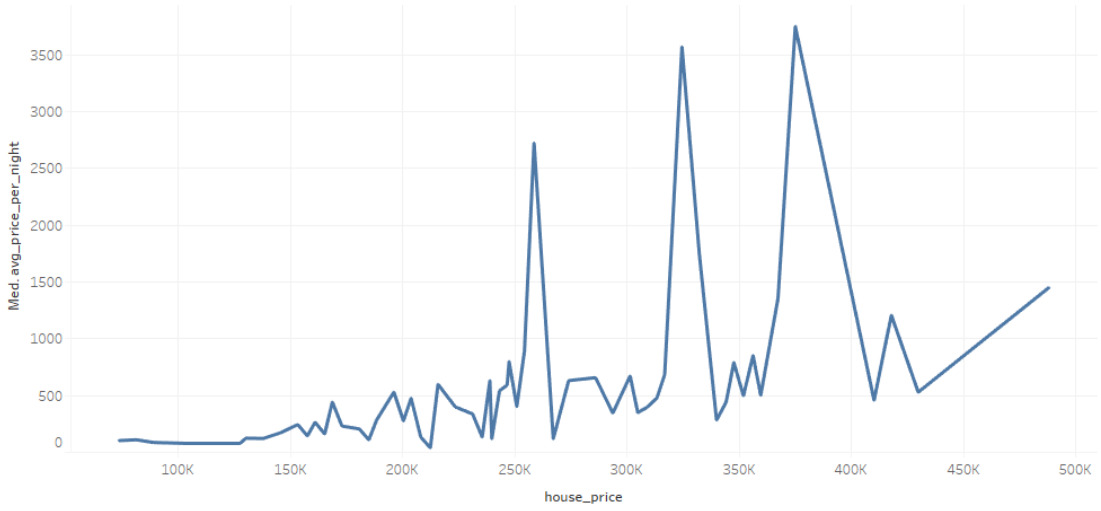
```
[net_earnXyear]= [earnXyear]-[airbnb_service_cost]-[local_annual_taxes]
```

Choice

Considering all the information obtained so far regarding the occupancy rate, adj_airbnb_yroi, costs, and earnings for each type of house and for each city, I have arrived at the following decision to maximize profit.

City	ID	House Type	house_price	Adj_airbnb_YROI	avg_price_night	days_occupied	net_earnXyear
Capo d'Orlando	2755	4BR_3.5BT	219.631,50 €	1,548	2.661,00 €	141	357.631,77 €
Letojanni	493	1BR_2.5BT	110.966,00 €	1,985	1.724,00 €	150	248.673,52 €
Partinico	656	1BR_4BT	50.120,00 €	2,549	1.000,00 €	108	101.943,50 €
	3416	5BR_7BT	127.448,00 €	3,658	3.649,00 €	108	372.823,42 €
Pollina	288	1BR_1BT	47.085,00 €	0,158	58,13 €	121	6.013,93 €
Ribera	2387	3BR_4.5BT	73.976,50 €	0,996	577,00 €	104	55.845,89 €
	2589	4BR_2BT	70.083,00 €	1,823	1.000,00 €	104	97.658,52 €
Santa Marina Salina	324	1BR_1BT	194.188,00 €	0,103	156,61 €	237	33.441,96 €
	548	1BR_2BT	234.832,00 €	0,018	34,00 €	237	4.370,28 €
Totale complessivo			1.128.330,00 €	*	*	*	1.278.402,78 €

As we can see from the table, all the constrains about the budget and the cities were respected but we notice something: the data does not seem to reflect the reality. For example, I report the distribution of houses in Noto, considering the house price and the price per night:



In this case, one would expect a sort of trend, namely that as the selling price of the house increases, the average price per night for that house would also increase. However, as we can see from the graph, the data do not show any pattern; rather, they seem to be randomly distributed. This means that the data do not reflect the real correlation between these two entities. And this happen for all the city we are analyzing.

Going in deep in this first choice, I've noted that: by comparing house prices and the average cost per night, we observe some outliers. For example, looking at ID 2589, we see that the cost per night is not consistent with the house price, which appears **undervalued** in this case. In my opinion, this bias is due to the aggregations performed in Prep. Remember that, taking ID 2589 as an example again, it represents the average of all houses with four bedrooms and two bathrooms located in Ribera. There might be a particular house in this average (for which a significantly higher price per night can be charged compared to the others) that is skewing our measurements (all based on basic assumptions, without considering other influences such as the area of the city and seasonality). Therefore, this combination, as attractive and profitable as it may seem, is not realistic at all. We need a way to discriminate based on the average cost per night, to obtain more accurate estimates of our profits.

I have therefore conducted some research on the real estate market and short-term rental apps to get a realistic picture of the market situation and to choose houses with prices consistent with their earnings, while still respecting the various constraints. So, I filtered the IDs based on the price per night, making three attempts: at most 800€, 400€, and 200€. The addition of the filter allowed me to obtain increasingly coherent and sensible results.

But why did I choose to filter by average price per night? For some simple reasons: it was the only feature that exhibited significant variability, thus it could better capture the true essence of the data. In contrast, the selling prices of the houses do not have a wide range and do not truly reflect the data. Furthermore, given our constraints, the house prices we can afford are only those that ensure a certain threshold of profit (without exorbitant amounts). Profit is directly linked to the price per night, so filtering by this variable seems to be a completely coherent and sensible choice, useful for achieving a performance-driven yet realistic strategy.

To better understand this choice, I will provide the selection of houses made each time the filter was adjusted, with an occupation rate stable, without any fluctuation.

choiche800

City	ID	House Type	house_price	Adj_airbnb_YROI	avg_price_night	days_occupied	net_earnXyear
Avola	3393	5BR_6BT	181.844,00 €	0,554	788,00 €	120	87.859,62 €
Caronia	3711	7BR_8BT	186.150,00 €	0,534	778,00 €	112	80.655,97 €
Castelmola	3042	4BR_5BT	175.824,00 €	0,556	765,00 €	125	89.084,47 €
Ispica	2440	3BR_5.5BT	142.172,50 €	0,717	798,00 €	120	89.522,73 €
Mazara del Vallo	2991	4BR_4BT	103.815,00 €	0,969	787,75 €	110	80.763,21 €
Ragusa	3407	5BR_7.5BT	172.827,50 €	0,569	770,00 €	116	82.542,00 €
Siracusa	3132	5BR_3.5BT	181.806,50 €	0,538	765,00 €	126	89.515,92 €
Totale complessivo			1.144.439,50 €	*	*	*	599.943,92 €

For the first filter, I set the average price per night to be less than €800. This resulted in a selection of seven houses. The houses selected under this filter exhibited substantial variability in terms of room types, ranging from 3 bedrooms and 3.5 bathrooms to 7 bedrooms and 8 bathrooms.

choiche400

City	ID	House Type	house_price	Adj_airbnb_YROI	avg_price_night	days_occupied	net_earnXyear
Casteldaccia	2751	4BR_3.5BT	113.274,00 €	0,192	170,00 €	113	16.639,31 €
Lipari	1920	3BR_2.5BT	261.534,50 €	0,187	383,33 €	168	58.464,80 €
Menfi	2577	4BR_2.5BT	147.865,50 €	0,337	390,00 €	125	44.489,03 €
Modica	2780	4BR_3BT	123.228,00 €	0,399	385,13 €	117	41.139,30 €
	2966	4BR_4BT	132.030,00 €	0,373	385,56 €	117	41.055,33 €
Ragusa	2638	4BR_2BT	110.799,00 €	0,461	399,62 €	116	42.521,16 €
Terrasini	1904	3BR_2.5BT	165.318,50 €	0,305	395,25 €	137	49.568,36 €
Vittoria	2217	3BR_3BT	82.500,00 €	0,601	388,33 €	109	39.092,27 €
Totale complessivo			1.136.549,50 €	*	*	*	332.969,55 €

When the filter was adjusted to an average price per night of less than €400, the selection criteria led to a different set of seven houses, where the total net earnings per year was significantly lower than the first selections. The variability in house types was also reduced, with houses ranging from 3 bedrooms and 2.5 bathrooms to 4 bedrooms and 4 bathrooms. This suggests that mid-range properties tend to offer a more consistent level of accommodation.

choiche200

City	ID	House Type	house_price	Adj_airbnb_YROI	avg_price_night	days_occupied	net_earnXyear
Avola	1581	3BR_1.5BT	103.834,00 €	0,244	198,00 €	120	21.204,27 €
	3292	5BR_4BT	162.476,00 €	0,156	199,00 €	120	20.439,60 €
Casteldaccia	1259	2BR_2BT	73.780,00 €	0,344	198,40 €	113	20.295,32 €
Castellamm..	2143	3BR_2BT	182.810,00 €	0,137	195,61 €	145	24.635,67 €
Ispica	1419	2BR_2BT	91.205,00 €	0,280	199,66 €	120	21.564,08 €
	3107	5BR_2BT	142.709,00 €	0,177	198,00 €	120	20.600,91 €
Marsala	2743	4BR_3.5BT	123.714,00 €	0,202	196,00 €	116	19.821,54 €
Mascali	2796	4BR_3BT	154.728,00 €	0,163	197,40 €	125	21.393,98 €
Menfi	2776	4BR_3BT	153.342,00 €	0,162	195,00 €	125	21.053,37 €
Totale complessivo			1.188.598,00 €	*	*	*	191.008,74 €

Finally, with the filter set to an average price per night of less than €200, eight houses were selected. The total net earnings per year was the lowest among the three filters. As the filter criteria become more stringent (lower price per night), the total net earnings per year decrease significantly but the total house prices remain relatively similar across all filters.

And that's the point! The scenario that seems most realistic among the ones analyzed is the last one, where the cost of the houses reflects the earnings based on the price per night.

So, if I was a dreamer, I would choose the first combination of houses that could make me a millionaire in just one year. However, being a data analyst *(or at least trying to be one)*, I choose the last combination of houses, which doesn't ensure high earnings but provides ***real earnings***.

On the next page, the map of Sicily allows us to better understand the cities where the houses are located. We can see that they are all situated along the coast and are spread out from one another.

This choice allows us to cover the most touristic areas (the coastal ones) while also covering different geographical markets, as these coastal areas are literally at opposite ends.

Second Scenario

Now, let's add a new layer of realism to our analysis: since we do not have an amount of €1.2M, it is necessary to request a loan from the bank to make this investment. The constraints remain the same as before, but in this case, we only have €200k available. However, we know the annual interest rate of the loan and we can control the other variables. This leads us to create a set of parameters:

- interest_rate_%, that range from 2% to 2.6% with a step of 0.1.
- year, that goes from 5 to 30 with a step of 5.
- loan_capital, that goes from 500k to 1M, with a step of 1k.

We need now to calculate how much the loan impact our choice of investment. To do this, we need to create some calculated field:

- monthly_interest: ([interest_rate_%]/100)/12
- n_payments_loan: [years] * 12
- monthly_payment: [Loan_capital] * ([Monthly_interest] * (1 + [monthly_interest]) ^ [n_payments_loan]) / ((1 + [monthly_interest]) ^ [n_payments_loan] - 1)
- annual_loan_payment: [monthly_payment] * 12

Through these calculated fields and parameters that directly influence the fields, it is possible to establish different scenarios for our loan and compare its annual cost with the earnings from our investment. Calculated fields were also created to verify if the assumptions behind the loan were correct, such as the choice of a fixed-rate amortization (French method). These fields include, for example, total_cost_loan and total_interest.

So, starting again our analysis, the only thing that we know is the loan_capital. Given the amount spent, we set it equal to one million euros. This is the resulting table when we set the parameters in a certain way:

price_loan

loan_capital	1.000.000,00 €
interest_rate_%	2,3
years	15
monthly_payment	6.574,15 €
annual_loan_payment	78.889,85 €
total_cost_loan	1.183.347,72 €
total_interest	183.347,72 €

In this case, with an interest rate of 2.3% over 15 years, the monthly payment amounts to €6,574.15, leading to an annual payment of €78,889.85. Over the entire loan period, the total cost amounts to €1,183,347.72, with €183,347.72 paid in interest.

Now, given the annual loan payment, it is crucial to compare this cost with the potential earnings from the investment. The goal is to ensure that the net earnings from the properties not only cover the loan payments but also provide a reasonable profit. By incorporating these loan parameters into our analysis, we can simulate different financial scenarios and make informed decisions based on realistic financial conditions. To reach this goal, I have created another two calculated field: one is the difference between the summation of the earnings of each house and the annual payment of the loan.

Another is the coverage_ratio: it is calculated as the ratio between the net earnings per year and the annual payment:

- When the coverage ratio is greater than 1, it indicates that the net earnings from the investment are more than enough to cover the annual loan payments. This means the investment is generating a surplus after paying off the loan costs, leading to a profitable scenario.
- A coverage ratio equal to 1 means that the net earnings are just sufficient to cover the annual loan payments. There is no surplus.
- When the coverage ratio is less than 1, it signifies that the net earnings are not sufficient to cover the annual loan payments. This results in a shortfall, indicating that the investment is not generating enough income to pay off the debt, leading to losses.

loan200

oc_rate_var_%	-5,00
interest_rate_%	2,60
years	10
Total Earnings	180.112,81 €
Annual Payment	113.670,39 €
Earn after Loan	66.442,43 €
coverage_ratio	1,58

Given these parameters (first three row), we examine the total earnings, annual loan payment, earnings after loan payments, and the coverage ratio. the net earnings amount to €66,442.43. This is the profit retained after servicing the debt, indicating a positive cash flow and the potential for reinvestment or savings. The coverage ratio is greater than 1, which signifies that the earnings are more than sufficient to cover the loan payments. Specifically, for every euro paid towards the loan, there is an additional €0.58 available as profit.

Conclusion

This analysis demonstrates that, even with the worst prevision for the occupancy and the maximum interest rate expected, the investment properties are performing well. The net earnings after loan payments and the favorable coverage ratio underline the viability and attractiveness of this investment scenario.

But we need to pay attention. By modifying various parameters, we notice that if the loan term is five years, we obtain a coverage ratio of less than 1, even when lowering the interest rate and increasing the occupancy rate. This is due to the fact that, although the interests to pay decrease, the principal to be repaid is distributed over fewer monthly installments, resulting in higher installment amounts, which makes us unable to repay our debt.

Therefore, in this strategy, we need to focus only on the loan term. With a minimum term of 10 years, we will be able to earn a net gain from the loan even under negative projections (maximum interest rate and minimum employment rate).

GIUSEPPE LEONARDI