



Checkpoint I: Project Proposal

Group: G23

Date: 2024/09/24

Problem Domain

"How was the real estate market in Portugal in 2021?"

As students who dream of one day moving out of our parents' house, the housing crisis in Portugal is a topic that concerns us. With this domain, we will be able to analyze the real estate market in Portugal in order to better understand trends and identify opportunities.

These visualizations aim to help users find the home that best suits their needs and possibilities, using information related to each house such as its price, location, number of bedrooms and bathrooms, or even the total area.

Task Abstraction

Example Questions

1. Which location has the highest average area in Portugal?

- Determine the average house area for a specific location
- **Question type:** comparison (comparing different average area values), geolocation
- **Concrete question:** "What is the average area of houses in Porto?"

2. What is the relationship between the number of rooms and bathrooms and the average price per square meter of houses?

- For a specific location, is it possible to obtain a correlation between the number of bedrooms and the average price per square meter?
- **Question type:** correlation, geolocation
- **Concrete question:** "In Beja, having 3 rooms instead of 2, will it significantly affect the price per square meter of the house?"

3. Is it more cost-effective to buy a new house or a renovated house?

- For a specific location, compare the prices of new houses with used houses to indicate if there is a relationship between the price and the condition (new or renovated) of these houses.
- **Question type:** comparison, correlation
- **Concrete question:** "For Bragança, how much more expensive is it to buy a new house instead of a renovated house?"

4. Is it more expensive to buy a house or to rent?

- For a specific location, compare the annual cost of buying houses with the annual cost of renting houses to determine if there is a relationship between the annual costs and the current condition (new or renovated) of these houses.
- **Question type:** comparison, correlation
- **Concrete question:** "For Lisbon, is it more expensive to buy a house or to rent a house for 40 years?"

5. What is the ranking of **districts** based on the average price per square meter for rental houses?
 - Determine how to rank a specific location based on the average price per square meter.
 - **Question type:** rank, geolocation
 - **Concrete question:** "Is the average price per square meter of rental housing more expensive in Lisbon or Faro?"
6. Where am I most likely to find my ideal home knowing that I need at least three bedrooms and two bathrooms, and I can spend a maximum of half a million euros?
 - Indicate the number of houses per location that meet the specified search criteria.
 - **Question type:** geolocation, proportion
 - **Concrete question:** "I need a house with 3 bedrooms and 2 bathrooms, and I can only spend a maximum of one and a half million euros. Are there more houses that meet these criteria in Braga or Aveiro?"

Data

Initial Dataset

Dataset:

[Portugal proprieties - rent, buy and vacation.](#)

The dataset used was obtained from the Kaggle website and is a static snapshot of the properties available in Portugal on March 21, 2021.

It contains information on properties available for sale, rent and holiday rentals. In total, the dataset includes 62658 items and 8 attributes:

1. Locations -> neighborhood, city, and state
2. Rooms -> how many rooms
3. Price -> price value in € (Euro)
4. Area -> area that is available in the house
5. Bathrooms -> how many bathrooms in the house
6. Condition -> house's condition, for example: New, Used, Renewed, and etc
7. AdsType -> determine the ads type, for example: to sell, to rent, or to vacation rent
8. ProprietyType -> House or Apartment for example

Selected/Derived Data

1. Annual price (rent has monthly price)
 - Description: For rental houses, the annual price will be calculated by multiplying the monthly price by 12.
 - Formula: $\text{Price} * 12$
2. Price per square meter
 - Description: The price per square meter will be calculated by dividing the property price by the available area of the house.
 - Formula: $\text{Price} / \text{Area}$
3. Simplified localization
 - Description: Information on municipalities and districts will be separated.

Data Processing

To ensure the integrity and relevance of the data for the analyses, the following data processing steps will be carried out:

1. Cleaning up null data
 - Description: Items with null values will be removed to ensure that the analysis is not affected by incomplete data.
2. Removal of vacation rental properties
 - Description: The vacation rental items will be removed from the dataset as they are not relevant to the analysis, only the purchase and long-term rental cases will be considered.
3. Removal of the 'ProprietyType' column
 - Description: The 'ProprietyType' column will be removed from the dataset after the data has been cleaned. As it is a single-value column, it will not contribute to the analyses.

Outliers were not removed because they represent the natural variability of the real estate market.

Data Abstraction

The dataset used will be a table made up of 9 columns (attributes), contains 36350 items that represent houses availables after data processing:

Attribute	Description	Attribute Type
Rooms	How many rooms in the house	Ratio
Price	Home purchase price (for homes for sale) and annual rental price (for home for rent)	Ratio
Area	Area that is available in the house	Ratio
Bathrooms	How many bathrooms in the house	Ratio
Condition	House's condition, for example: new, used, renewed, and etc.	Ordinal
AdsType	Determine the ads type, for example: to sell, to rent	Nominal
PricePerSquareMeter	The cost associated with one square meter of a available area in the house	Ratio
District	Indicate house's district	Nominal
Municipality	Indicates house's municipality	Nominal

Mapping (Data sample/Questions)

1. Which location has the highest average area in Portugal?

Area	District	Municipality
52.2	Faro	Tavira

2. What is the relationship between the number of rooms and bathrooms and the average price per square meter of houses?

Rooms	Bathrooms	PricePerSquareMeter	District	Municipality
4	4	163.265306122449	Aveiro	Ovar

3. Is it more cost-effective to buy a new house or a renovated house?

Price	Condition	District	Municipality
6600	Renovated	Beja	Odemira

4. Is it more expensive to buy a house or to rent?

Price	AdsType	District	Municipality
28800	Rent	Braga	Fraião e Lamações

5. What is the ranking of districts based on the average price per square meter for rental houses?

PricePerSquareMeter	District	Adstype	Municipality
271.698113207547	Lisboa	Rent	Oeiras

6. Where am I most likely to find my ideal home knowing that I need at least three bedrooms and two bathrooms, and I can spend a maximum of half a million euros?

Rooms	Bathrooms	Price	District	Municipality
3	3	245000	Aveiro	Oliveira de Azeméis