

Quiz ADA Solutions

3rd Trimester

Ant nio Fidalgo Miguel Salema

March 8, 2024

Instructions

- You have 70 minutes to complete your quiz.
- This is an individual exam.
- All relevant R code and written answers must be answered in a .qmd file and then rendered to pdf. In case you cannot render to pdf, you may render it to html, but you will have a 5 points discount.
- Upload the full directory of your assignment to Moodle by the indicated deadline as a .zip file.
- Your code must run on the grader's computer.
- The detection of any form of plagiarism in your work will result in a “fail” grade.
- You must separate all your questions and sub-questions using headers in the .qmd document.
- You can access all class materials but internet access is forbidden.

Portugal's Housing Data

To answer this quiz, you must read the file `housing.csv` from Moodle. The dataset contains information about properties from adds in the website *Idealista*. The dataset was extracted in March 21, 2021. Each observation is a property. Table 1 contains the description of the variables. **Assume that you know nothing else about the Portuguese housing market and draw no conclusions that cannot be drawn from this data alone.**

Table 1: Description of the Variables

Variable	Description
Location	County (Freguesia), Municipality and District.
Rooms	Number of rooms.
Price	Price value in euros.
Area	Available area in squared meters.
Bathrooms	Number of bathrooms.
Condition	Properties's condition.
distrito	District.

Questions

- (20 points) Read the data into R and name the object `df`.
- (40 points) Copy the following sentence into your quarto and, using the dataset `df`, fill the “xxx” using inline R code: “The housing dataset has xxx observations”.
- (30 points) What was the average price of a property? (Hint: don't forget to ignore the NA values.)
- (30 points) What was the average price of a property that is smaller than 150 squared meters?
- (30 points) Add a column to `df` (overwrite it) that contains the number of rooms plus the number of bathrooms.
- (50 points) Print the a dataset showing the average price per number of rooms ordered in descending order by price. We did it and it looks like this:

```
# A tibble: 11 x 2
  Rooms av_price
  <dbl>   <dbl>
1     1 128735.
2     2 195493.
3     0 220740.
```

```
4      3 394375.  
5      4 602338.  
6      5 794254.  
7      6 912814.  
8      9 982444.  
9      8 1011928.  
10     7 1110069.  
11    10 1309975.
```

7. (Extra 20 points) Give a possible explanation to why 0 rooms is on average so much more expensive than 1 room. No code required.

Solutions

1)

```
library(tidyverse)  
  
df <- read_csv("data/housing.csv")
```

2)

The housing dataset has 61730 observations.

3)

```
df$Price %>% mean(na.rm = T)
```

```
[1] 489230.6
```

4)

```
df2 <- df %>% filter(Area < 150)  
df2$Price %>% mean(na.rm = T)
```

```
[1] 230560.2
```

5)

```
df <- df %>%  
  mutate(n_div = Rooms + Bathrooms)
```

6)

```
df %>%  
  group_by(Rooms) %>%  
  summarise(av_price = mean(Price,  
                           na.rm = T)) %>%  
  arrange(av_price)
```

A tibble: 11 x 2

	Rooms	av_price
	<dbl>	<dbl>
1	1	128735.
2	2	195493.
3	0	220740.
4	3	394375.
5	4	602338.
6	5	794254.
7	6	912814.
8	9	982444.
9	8	1011928.
10	7	1110069.
11	10	1309975.

7)

Properties with no rooms (T0) are only common in city centers while one bedroom properties exist also outside city centers. Thus, no bedroom apartments are on average more expensive because of this confounding effect with location.