xercises

1. Exercise 1

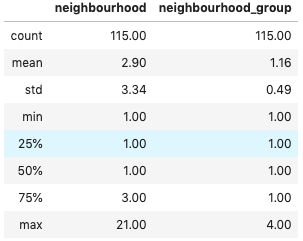
In these exercises, you will continue investigating AB\_NYC\_2019.csv listings in NYC.

We are interested in the following question: *Do owners of multiple listings spread their listings around the city, or are they focused in the same location?*

There are many ways to answer this. For now, we are going to use the groupby method with these steps:

* + For hosts that have 10 listings or more, use the groupby method with an aggregation, to calculate the number of unique neighborhoods and neighborhood groups per host.
  + Use the .describe() method on the grouped dataframe to present descriptive statistics regarding the number of listings per neighborhood and per neighborhood group

Your result should look like this:



1. Exercise 2

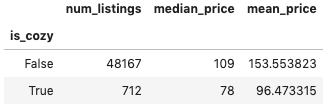
You suspect that listings that contain the word 'cozy' in the description are cheaper than other listings.

To check this assumption, follow the next steps:

* + Use the syntax df['col'].str.contains('string') to create a boolean series that checks if a specific column contains a specific string
  + Group the dataframe by the series you've just creates (lists that have 'cozy' in their name against the others)

Use .agg() to calculate the number of listings in each group, the median price and the mean price.

Your result should look like this:



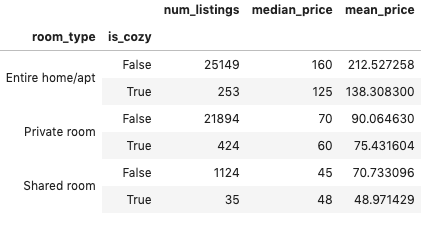
1. Exercise 3

We know that listing prices are also affected by room type. Maybe all of the "cozy" rooms belong to the same type?

Repeat the analysis from the previous exercise, but this time group the dataframe by both the boolean series (cozy/ not cozy) and the room type.

Does the 'cozy' effect still exist?

Your result should look like this:



1. Exercise 4

Do hosts tend to offer the same types of listings, or do they diversify the rooms types?

Use groupby to check the distribution of number of room types (1,2 or 3) for hosts that have at least 3 listings.

ראש הטופס