
0.1 Question 3a

Consider the chained `pandas` statement below:

```
q3a_df = ins_named[ins_named["name"].str.lower().str.contains("taco")].groupby("bid").filter(lambda
sf: sf["score"].max() > 95).agg("count")
```

We can decompose this statement into three parts:

```
temp1 = ins_named[ins_named["name"].str.lower().str.contains("taco")]

temp2 = temp1.groupby("bid").filter(lambda sf: sf["score"].max() > 95)

q3a_df = temp2.agg("count")
```

For each line of code above, write one sentence describing what the line of code accomplishes. Feel free to create a cell to see what each line does. In total, you'll write three sentences.

Finally, write an example homework question whose answer is `q3a_df`.

- This example homework question should only be one sentence.

Note: While the first part of this question will be graded for correctness, the second part is a bit more open-ended. Answers that demonstrate correct understanding will receive full credit.

An example answer will look like the following: "`temp1` creates a … `temp2` transforms `temp1` by … Finally, `q3a_df` results in a `DataFrame` that … A question that is answered by this chain of operations is …"

```
In [29]: q3a_df = ins_named[ins_named["name"].str.lower().str.contains("taco")].groupby("bid").filter(la

         q3a_df
```

```
Out[29]: iid          20
         date         20
         score        20
         type         20
         timestamp    20
         bid          20
```

```
Missing Score    20
name             20
address          20
dtype: int64
```

temp1 filters ins_named to keep only rows where the name column contains “taco” (case-insensitive)

temp2 groups temp1 by bid (business ID) and filters out businesses where the highest inspection score is 95

Finally, q3a_df is a summary DataFrame that contains the count of inspections for businesses that passed the filtering criteria

Homework question: How many inspections were conducted for businesses with “taco” in their name, given that they have at least one inspection score greater than 95?

0.2 Question 3b

Consider `ins_named`, `temp1`, `temp2`, and `q3a_df` from the previous problem. What is the granularity of each `DataFrame`? Explain your answer in no more than four sentences.

Note: For more details on what the granularity of a `DataFrame` means, feel free to check the course notes!

ins_named has an inspection-level granularity, where each row represents a single inspection for a business.

temp1 maintains the same granularity but only includes businesses whose names contain "taco".

temp2 also retains inspection-level granularity but further filters businesses to include only those with at least one inspection score greater than 95.

q3a_df, if correctly counted by bid, would shift to business-level granularity, representing the number of businesses that meet the criteria.

0.3 Question 4e

Do you notice any trends? Are your results consistent with your prior knowledge about restaurants that receive high or low health inspection scores? Answer in the cell below.

This question is graded on effort, there is no one "correct" answer.

Looking at the results, there is a clear trend where restaurants with lower health inspection scores tend to have a higher proportion of recent reinspections. Conversely, restaurants with higher scores (95-100) have the lowest proportion of reinspections. This aligns with the expectation that establishments with lower scores are more likely to require follow-up inspections to ensure compliance with health regulations. It suggests that lower-scoring restaurants may have recurring issues that necessitate more frequent monitoring, whereas higher-scoring establishments maintain better compliance, reducing the need for reinspection.