
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 [19]: temp1 = ins_named[ins_named["name"].str.lower().str.contains("taco")]
         temp1.head(5)
```

Out[19]:

	iid	date	score	type	\
776	1726_20170314	03/14/2017 12:00:00 AM	92	Routine - Unscheduled	
777	1726_20180202	02/02/2018 12:00:00 AM	92	Routine - Unscheduled	
778	1726_20190611	06/11/2019 12:00:00 AM	92	Routine - Unscheduled	
885	18126_20190805	08/05/2019 12:00:00 AM	93	Routine - Unscheduled	
1387	2337_20161011	10/11/2016 12:00:00 AM	92	Routine - Unscheduled	

```
      timestamp    bid Missing Score               name          address
776  2017-03-14   1726            No      KFC/TACO BELL  4285 MISSION St
777  2018-02-02   1726            No      KFC/TACO BELL  4285 MISSION St
778  2019-06-11   1726            No      KFC/TACO BELL  4285 MISSION St
885  2019-08-05  18126            No    Tacos San Buena   768 Sansome St
1387 2016-10-11   2337            No  EL TACO LOCO INC.     3274 24th St
```

```
In [20]: temp2 = temp1.groupby("bid").filter(lambda sf: sf["score"].max() > 95)
         temp2.head(5)
```

```
Out[20]:                 iid                    date  score                   type  \
3118  37818_20171213  12/13/2017 12:00:00 AM     96  Routine - Unscheduled
4017   5136_20170710  07/10/2017 12:00:00 AM     94  Routine - Unscheduled
4018   5136_20180524  05/24/2018 12:00:00 AM     98  Routine - Unscheduled
4349    580_20170307  03/07/2017 12:00:00 AM     96  Routine - Unscheduled
4350    580_20170829  08/29/2017 12:00:00 AM     91  Routine - Unscheduled

      timestamp    bid Missing Score                 name           address
3118 2017-12-13  37818            No  Nick's Crispy Tacos  1500 Broadway St
4017 2017-07-10   5136            No            Taco Bell        7 Drumm St
4018 2018-05-24   5136            No            Taco Bell        7 Drumm St
4349 2017-03-07    580            No       TACO LOS ALTOS  737 CORTLAND Ave
4350 2017-08-29    580            No       TACO LOS ALTOS  737 CORTLAND Ave
```

```
In [21]: q3a_df = temp2.agg("count")
         q3a_df
```

```
Out[21]: iid              20
         date             20
         score            20
         type             20
         timestamp        20
         bid              20
         Missing Score    20
         name             20
         address          20
         dtype: int64
```

temp1 creates the dataframes and filter the business name that contain “taco” string in the name.

temp2 transforms temp1 and filter the score that is higher than 95 that means at least one inspection score should be more than 95

q3a_df results in dataframe that calculate the total number of rows for each column in temp2

A question that is answered by this chain of operations is "How many inspection records exist for businesses with 'taco' in their name that have received at least one inspection score higher 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 - The granularity is one inspection at a specific business. temp1 - The granularity remains one inspection at a specific business, but only for businesses whose names contain "taco". temp2 - The granularity is still one inspection at a specific business, but now only for businesses that have had at least one inspection score greater than 95. q3a_df - The granularity is column-level summary statistics.

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.

I think restaurants with lower initial inspection scores (70-74) have the highest proportion of follow-up inspections within 62 days, while those with higher scores (95-100) have the lowest. This aligns with expectations, as health departments prioritize reinspections for establishments with more violations.