

---

## 0.1 Question 3c

In the cell below, run the following line of code: `q3c_df = ice_cream_shops.sort_values('timestamp').groupby('bid').`

Is the granularity of `ice_cream_at_least_3` the same as the granularity of `q3c_df`? In other words, what does a single row of `q3c_df` represent, and what does a single row in `ice_cream_at_least_3` represent? Explain the granularity of each `DataFrame`. Your answer does not need to be more than 2-3 lines, but you should be specific.

```
In [171]: q3c_df = ice_cream_shops.sort_values('timestamp').groupby('bid').agg('first')
          q3c_df.head()
```

```
Out[171]:
```

	iid	date	score	type
bid				
31	31_20180615	06/15/2018 12:00:00 AM	96	Routine - Unscheduled
758	758_20171212	12/12/2017 12:00:00 AM	94	Routine - Unscheduled
4671	4671_20170117	01/17/2017 12:00:00 AM	98	Routine - Unscheduled
5032	5032_20170627	06/27/2017 12:00:00 AM	94	Routine - Unscheduled
5524	5524_20161011	10/11/2016 12:00:00 AM	98	Routine - Unscheduled

  

	timestamp	Missing	Score	name
bid				
31	2018-06-15	False		Norman's Ice Cream and Freezes
758	2017-12-12	False		BAKERY/ICE CREAM/STOREROOM
4671	2017-01-17	False		MARCO POLO ITALIAN ICE CREAM
5032	2017-06-27	False		MITCHELLS ICE CREAM
5524	2016-10-11	False		AT&T Park - Coffee and Ice Cream (4A+4B)

  

	address	lowercase_name
bid		
31	2801 Leavenworth St	norman's ice cream and freezes
758	2 New Montgomery St	bakery/ice cream/storeroom
4671	1447 TARAVAL St	marco polo italian ice cream
5032	688 SAN JOSE Ave	mitchells ice cream
5524	24 WILLIE MAYS PLAZA	at&t park - coffee and ice cream (4a+4b)

The difference is that `ice_cream_at_least_3` gives us ice cream shops that have been around long enough to get inspected at least 3 times. `q3c_df` gives us the first of each group (in this case `bid`) which makes it easier for the reader to look at the data. But in both cases, it focuses on ice cream shops, and it showcases the type of inspection, their score, when it happened, and where it happened.



---

## 0.2 Question 3e

Finally, to examine different parts of a chained pandas statement, describe the purpose of each of the functions used (`.loc`, `.groupby`, `idxmax()`) in words.

Secondly, share what you think this line of code accomplishes. In other words, write a question that could be answered using this statement.

While the first part of this question will be graded for correctness, the second part of this question is a bit more open-ended. Answers demonstrating your understanding will get full credit.

```
In [230]: ice_cream_at_least_3.loc[ice_cream_at_least_3.groupby("bid")["score"].idxmax()].head()
```

```
Out [230]:
```

	iid	date	score	type	\
3704	4671_20171130	11/30/2017 12:00:00 AM	100	Routine - Unscheduled	
3972	5032_20190718	07/18/2019 12:00:00 AM	100	Routine - Unscheduled	
4182	5524_20170919	09/19/2017 12:00:00 AM	100	Routine - Unscheduled	
4186	5528_20170424	04/24/2017 12:00:00 AM	100	Routine - Unscheduled	
559	14743_20161103	11/03/2016 12:00:00 AM	100	Routine - Unscheduled	

  

	timestamp	bid	Missing	Score	\
3704	2017-11-30	4671	False		
3972	2019-07-18	5032	False		
4182	2017-09-19	5524	False		
4186	2017-04-24	5528	False		
559	2016-11-03	14743	False		

  

	name	address	\
3704	MARCO POLO ITALIAN ICE CREAM	1447 TARAVAL St	
3972	MITCHELLS ICE CREAM	688 SAN JOSE Ave	
4182	AT&T Park - Coffee and Ice Cream (4A+4B)	24 WILLIE MAYS PLAZA	
4186	AT&T - Juma Cart 1 - Ice Cream	24 WILLIE MAYS PLAZA	
559	Polly Ann Ice Cream	3138 Noriega St.	

  

	lowercase_name
3704	marco polo italian ice cream
3972	mitchells ice cream
4182	at&t park - coffee and ice cream (4a+4b)
4186	at&t - juma cart 1 - ice cream
559	polly ann ice cream

The difference between each of the functions is as follows: `loc` allows the user to access specific parts of columns or arrays; `groupby` splits the dataframe, performs a function, and then combines it; `idxmax` returns the first max over an axis (in this case score).

This code accomplishes showing the reader the only icecream shops that got a max score that were inspected at least 3 times.

```
In [ ]: # You may do some scratch work in this cell, however, only your written answer will be graded.  
        # Any outputs or dataframes you generate here will not be counted as part of your explanation.
```