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 [178]:
```
q3c_df = ice_cream_shops.sort_values('timestamp').groupby('bid').agg('first')
q3c_df.head()
```

Out[178]:

bid	iid	date	score	type	timestamp	Missing Score	name	address	city	state	postal_code	latitude	longitude	phone_number	postal5	lowercase_name
31	31_20180615	06/15/2018 12:00:00 AM	96	Routine - Unscheduled	2018-06-15	False	Norman's Ice Cream and Freezes	2801 Leavenworth St	San Francisco	CA	94133	37.807155	-122.419004	-9999	94133	
758	758_20171212	12/12/2017 12:00:00 AM	94	Routine - Unscheduled	2017-12-12	False	BAKERY/ICE CREAM/STOREROOM	2 New Montgomery St	San Francisco	CA	94105	37.788286	-122.401543	-9999	94105	
4671	4671_20170117	01/17/2017 12:00:00 AM	98	Routine - Unscheduled	2017-01-17	False	MARCO POLO ITALIAN ICE CREAM	1447 TARAVAL St	San Francisco	CA	94116	37.742728	-122.481566	-9999	94116	
5032	5032_20170627	06/27/2017 12:00:00 AM	94	Routine - Unscheduled	2017-06-27	False	MITCHELLS ICE CREAM	688 SAN JOSE Ave	San Francisco	CA	94110	37.744122	-122.422722	14155642300	94110	
5524	5524_20161011	10/11/2016 12:00:00 AM	98	Routine - Unscheduled	2016-10-11	False	AT&T Park - Coffee and Ice Cream (4A+4B)	24 WILLIE MAYS PLAZA	San Francisco	CA	94107	-9999.000000	-9999.000000	-9999	94107	

```
bid
31              norman's ice cream and freezes
758                 bakery/ice cream/storeroom
4671              marco polo italian ice cream
5032                       mitchells ice cream
5524  at&t park - coffee and ice cream (4a+4b)
```

Each row in ice_cream_at_least_3 represents a single inspection from ice cream shops that had at least 3 inspections. On the other hand, each row in q3c_df represents the earliest inspection for each ice cream shop.

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

```
Out[182]:                 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.

                       city
          3704  San Francisco
          3972  San Francisco
          4182  San Francisco
          4186  San Francisco
          559   San Francisco
```

.groupby group all the unique businesses and .idxmax() get the index of the highest score among each of the ice cream shop group. .loc selects all the rows returned by .idxmax().

This line of code returns data frame that contains the inspection record with the highest score for each unique ice cream shop.

In [183]:
```
# 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
```