

HOMEWORK

Answer the following questions using the wine review dataset from previous class.

1. Set the index of the dataframe to country
2. Using the index and any country of your choice, show the reviews for those country
Sample code `df.loc[df.country == 'Italy']`

can you glean any insight from this? Perhaps the percentage of wines from the selected country.
3. Supposing wines are reviewed on a 80-to-100 point scale, and wines with greater than 90 points are average, can you check what percentage of the total wines are above average? What about the percentage in the country you selected from question 2?
4. Check the pandas 'isin' operator and use it to perform any operation of your choice on the dataset
5. Suppose we'll buy any wine that's made in Italy or which is rated above average. Can you select or show these wines using the or (|) operator?
6. Read about the notnull operator and use it to check for wines whose prices are not null
7. Please share something you learnt outside of class on thursday

In [1]:

```
import pandas as pd
```

In [2]:

```
wine_review = pd.read_csv('Wine_review.csv')
```

In [3]:

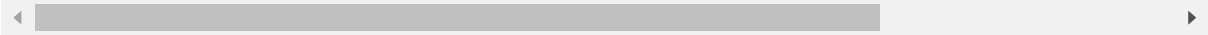
wine_review

Out[3]:

Unnamed: 0		country	description	designation	points	price	province	region_1
0	0	US	This tremendous 100% varietal wine hails from ...	Martha's Vineyard	96	235.0	California	Napa Valley
1	1	Spain	Ripe aromas of fig, blackberry and cassis are ...	Carodorum Selección Especial Reserva	96	110.0	Northern Spain	Toro
2	2	US	Mac Watson honors the memory of a wine once ma...	Special Selected Late Harvest	96	90.0	California	Knights Valley
3	3	US	This spent 20 months in 30% new French oak, an...	Reserve	96	65.0	Oregon	Willamette Valley
4	4	France	This is the top wine from La Bégude, named aft...	La Brûlade	95	66.0	Provence	Bandol
...
150925	150925	Italy	Many people feel Fiano represents southern Ita...	NaN	91	20.0	Southern Italy	Fiano di Avellino
150926	150926	France	Offers an intriguing nose with ginger, lime an...	Cuvée Prestige	91	27.0	Champagne	Champagne
150927	150927	Italy	This classic example comes from a cru vineyard...	Terre di Dora	91	20.0	Southern Italy	Fiano di Avellino
150928	150928	France	A perfect salmon shade, with scents of peaches...	Grand Brut Rosé	90	52.0	Champagne	Champagne

Unnamed: 0		country	description	designation	points	price	province	region_1
150929	150929	Italy	More Pinot Grigios should taste like this. A r...	NaN	90	15.0	Northeastern Italy	Alto Adige

150930 rows × 11 columns



In [9]:

```
## question1
wine_review.country
```

Out[9]:

```
0      US
1    Spain
2      US
3      US
4    France
...
150925  Italy
150926  France
150927  Italy
150928  France
150929  Italy
Name: country, Length: 150930, dtype: object
```

In [10]:

```
##question2
#insight from the dataset, shows that region_2 is missing data values

italian_wine = wine_review[wine_review.country == 'Italy']
italian_wine
```

Out[10]:

	Unnamed: 0	country	description	designation	points	price	province	region
10	10	Italy	Elegance, complexity and structure come togeth...	Ronco della Chiesa	95	80.0	Northeastern Italy	Co
32	32	Italy	Underbrush, scorched earth, menthol and plum s...	Vigna Piaggia	90	NaN	Tuscany	Brunello Montalc
35	35	Italy	Forest floor, tilled soil, mature berry and a ...	Riserva	90	135.0	Tuscany	Brunello Montalc
37	37	Italy	Aromas of forest floor, violet, red berry and ...	NaN	90	29.0	Tuscany	Vino Nobile Montepulcia
38	38	Italy	This has a charming nose that boasts rose, vio...	NaN	90	23.0	Tuscany	Chia Class
...
150920	150920	Italy	Rich and mature aromas of smoke, earth and her...	Brut Riserva	91	19.0	Northeastern Italy	Tre
150922	150922	Italy	Made by 30-ish Roberta Borghese high above Man...	Superiore	91	NaN	Northeastern Italy	Colli Orien del Fi
150925	150925	Italy	Many people feel Fiano represents southern Ita...	NaN	91	20.0	Southern Italy	Fiano Avell
150927	150927	Italy	This classic example comes from a cru vineyard...	Terre di Dora	91	20.0	Southern Italy	Fiano Avell

Unnamed: 0

	country	description	designation	points	price	province	region
150929	Italy	More Pinot Grigios should taste like this. A r...	NaN	90	15.0	Northeastern Italy	Alto Ad

23478 rows × 11 columns

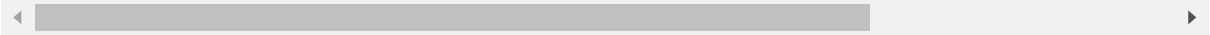
In [11]:

```
italian_wine = wine_review.loc[wine_review.country == 'Italy']
italian_wine
```

Out[11]:

	Unnamed: 0	country	description	designation	points	price	province	region_1
10	10	Italy	Elegance, complexity and structure come togeth...	Ronco della Chiesa	95	80.0	Northeastern Italy	Collio
32	32	Italy	Underbrush, scorched earth, menthol and plum s...	Vigna Piaggia	90	NaN	Tuscany	Brunello di Montalcino
35	35	Italy	Forest floor, tilled soil, mature berry and a ...	Riserva	90	135.0	Tuscany	Brunello di Montalcino
37	37	Italy	Aromas of forest floor, violet, red berry and ...	NaN	90	29.0	Tuscany	Vino Nobile di Montepulciano
38	38	Italy	This has a charming nose that boasts rose, vio...	NaN	90	23.0	Tuscany	Chianti Classico
...
150920	150920	Italy	Rich and mature aromas of smoke, earth and her...	Brut Riserva	91	19.0	Northeastern Italy	Trento
150922	150922	Italy	Made by 30-ish Roberta Borghese high above Man...	Superiore	91	NaN	Northeastern Italy	Colli Orientali del Friuli
150925	150925	Italy	Many people feel Fiano represents southern Ita...	NaN	91	20.0	Southern Italy	Fiano di Avellino
150927	150927	Italy	This classic example comes from a cru vineyard...	Terre di Dora	91	20.0	Southern Italy	Fiano di Avellino
150929	150929	Italy	More Pinot Grigios should taste like this. A r...	NaN	90	15.0	Northeastern Italy	Alto Adige

23478 rows × 11 columns



In [14]:

```
#question3(a)

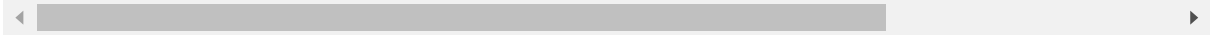
best_wines = wine_review.loc[(wine_review.points>90)]
best_wines
```

Out[14]:

Unnamed: 0		country	description	designation	points	price	province	region_1
0	0	US	This tremendous 100% varietal wine hails from ...	Martha's Vineyard	96	235.0	California	Napa Valley
1	1	Spain	Ripe aromas of fig, blackberry and cassis are ...	Carodorum Selección Especial Reserva	96	110.0	Northern Spain	Toro
2	2	US	Mac Watson honors the memory of a wine once ma...	Special Selected Late Harvest	96	90.0	California	Knights Valley
3	3	US	This spent 20 months in 30% new French oak, an...	Reserve	96	65.0	Oregon	Willamette Valley V
4	4	France	This is the top wine from La Bégude, named aft...	La Brûlade	95	66.0	Provence	Bandol
...
150923	150923	France	Rich and toasty, with tiny bubbles. The bouque...	Demi-Sec	91	30.0	Champagne	Champagne
150924	150924	France	Really fine for a low-acid vintage, there's an...	Diamant Bleu	91	70.0	Champagne	Champagne
150925	150925	Italy	Many people feel Fiano represents southern Ita...	NaN	91	20.0	Southern Italy	Fiano di Avellino
150926	150926	France	Offers an intriguing nose with ginger, lime an...	Cuvée Prestige	91	27.0	Champagne	Champagne

Unnamed: 0		country	description	designation	points	price	province	region_1
150927	150927	Italy	This classic example comes from a cru vineyard...	Terre di Dora	91	20.0	Southern Italy	Fiano di Avellino

32237 rows × 11 columns



In [15]:

```
#question 3(b)

italian_best_wines = italian_wine.loc[italian_wine.points >90]
italian_best_wines
```

Out[15]:

	Unnamed: 0	country	description	designation	points	price	province	region_1
10	10	Italy	Elegance, complexity and structure come togeth...	Ronco della Chiesa	95	80.0	Northeastern Italy	Collio
72	72	Italy	This offers aromas of red rose, wild berry, da...	Bussia Riserva	91	NaN	Piedmont	Barolo
79	79	Italy	Underbrush, smoke, mocha and dark fruit aromas...	Boscato	91	75.0	Piedmont	Barolo
82	82	Italy	Berry, baking spice, dried iris, mint and a hi...	Palliano Riserva	91	NaN	Piedmont	Roero
83	83	Italy	This hearty red opens with aromas of leafy und...	del Comune di Serralunga d'Alba	91	59.0	Piedmont	Barolo
...
150776	150776	Italy	This very internationally-styled blend of Sang...	Summus	91	63.0	Tuscany	Sant'Antimo
150920	150920	Italy	Rich and mature aromas of smoke, earth and her...	Brut Riserva	91	19.0	Northeastern Italy	Trento
150922	150922	Italy	Made by 30-ish Roberta Borghese high above Man...	Superiore	91	NaN	Northeastern Italy	Colli Orientali del Friuli
150925	150925	Italy	Many people feel Fiano represents southern Ita...	NaN	91	20.0	Southern Italy	Fiano di Avellino
150927	150927	Italy	This classic example comes from a cru vineyard...	Terre di Dora	91	20.0	Southern Italy	Fiano di Avellino

5202 rows × 11 columns



In [16]:

```
#question 4
```

```
best_wines = wine_review.loc[  
    (wine_review.country.isin(['France', 'US']))  
    &(wine_review.points>=91)  
]
```


In [17]:

```
best_wines
```

Out[17]:

Unnamed: 0	country	description	designation	points	price	province	region_1
0	US	This tremendous 100% varietal wine hails from ...	Martha's Vineyard	96	235.0	California	Napa Valley
2	US	Mac Watson honors the memory of a wine once ma...	Special Selected Late Harvest	96	90.0	California	Knights Valley
3	US	This spent 20 months in 30% new French oak, an...	Reserve	96	65.0	Oregon	Willamette Valley V
4	France	This is the top wine from La Bégude, named aft...	La Brûlade	95	66.0	Provence	Bandol
8	US	This re-named vineyard was formerly bottled as...	Silice	95	65.0	Oregon	Chehalem Mountains V
...
150919	France	This classy example opens with a very floral n...	Cuvée President	91	37.0	Champagne	Champagne
150921	France	Shows some older notes: a bouquet of toasted w...	Blanc de Blancs Brut Mosaïque	91	38.0	Champagne	Champagne
150923	France	Rich and toasty, with tiny bubbles. The bouque...	Demi-Sec	91	30.0	Champagne	Champagne
150924	France	Really fine for a low-acid vintage, there's an...	Diamant Bleu	91	70.0	Champagne	Champagne
150926	France	Offers an intriguing nose with ginger, lime an...	Cuvée Prestige	91	27.0	Champagne	Champagne

20981 rows × 11 columns



In [19]:

#question 5

```
buy_best_wines = wine_review.loc[
    (wine_review.country.isin(['Italy']))
    | (wine_review.points > 90)
]
```

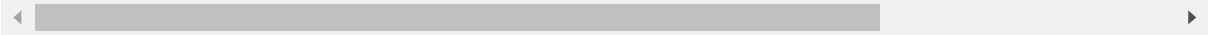
In [20]:

```
buy_best_wines
```

Out[20]:

Unnamed: 0	country	description	designation	points	price	province	region_1
0	US	This tremendous 100% varietal wine hails from ...	Martha's Vineyard	96	235.0	California	Napa Valley
1	Spain	Ripe aromas of fig, blackberry and cassis are ...	Carodorum Selección Especial Reserva	96	110.0	Northern Spain	Toro
2	US	Mac Watson honors the memory of a wine once ma...	Special Selected Late Harvest	96	90.0	California	Knights Valley
3	US	This spent 20 months in 30% new French oak, an...	Reserve	96	65.0	Oregon	Willamette Valley
4	France	This is the top wine from La Bégude, named aft...	La Brûlade	95	66.0	Provence	Bandol
...
150924	France	Really fine for a low-acid vintage, there's an...	Diamant Bleu	91	70.0	Champagne	Champagne
150925	Italy	Many people feel Fiano represents southern Ita...	NaN	91	20.0	Southern Italy	Fiano di Avellino
150926	France	Offers an intriguing nose with ginger, lime an...	Cuvée Prestige	91	27.0	Champagne	Champagne
150927	Italy	This classic example comes from a cru vineyard...	Terre di Dora	91	20.0	Southern Italy	Fiano di Avellino
150929	Italy	More Pinot Grigios should taste like this. A r...	NaN	90	15.0	Northeastern Italy	Alto Adige

50513 rows × 11 columns



In [21]:

```
#checking for missing data value using notnull operator
wine_review.notnull()
```

Out[21]:

	Unnamed: 0	country	description	designation	points	price	province	region_1	region_2
0	True	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True	False
2	True	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True	False
...
150925	True	True	True	False	True	True	True	True	False
150926	True	True	True	True	True	True	True	True	False
150927	True	True	True	True	True	True	True	True	False
150928	True	True	True	True	True	True	True	True	False
150929	True	True	True	False	True	True	True	True	False

150930 rows × 11 columns



In [23]:

```
#question 6
```

```
wine_price = wine_review.price.notnull()  
wine_price
```

Out[23]:

```
0      True  
1      True  
2      True  
3      True  
4      True  
...  
150925  True  
150926  True  
150927  True  
150928  True  
150929  True  
Name: price, Length: 150930, dtype: bool
```

mini questions

1. what is the median of the points in the wine review dataset
2. what countries are represented in the dataset(no duplicates)
3. how often does each country appear in the dataset(hint: use value count)
4. create a variable 'centered price', it should contain the price after the mean value(of the price column) has been subtracted(hint: use df.price.mean() to get the mean value)
5. create a variable 'bargain_wine' with the title of the wine with the highest point -to-price ratio in the dataset (hint:use .idxmax())
6. we'd like to host these wine reviews in our website but a rating system ranging from 80 to 100 points is too hard to get
 - we'd like to transform them into simpler star ratings. a score of 95 and higher counts as 3 stars, a score of at least 85 but less than 95 is a 2 stars. Any other score is 1 star.
 - Also the canadian Vinters Association bought av lot of ads on the site, so any wines from Canada should automatically get stars regardless of the points.
 - create a series star_ratings with the number of stars corresponding to each review on the data set.

In [22]:

```
#question 1 using median operator
```

```
wine_review_point = wine_review.points.median()  
wine_review_point
```

Out[22]:

```
88.0
```


In [16]:

#question 2 using nunique operator

```
wine_review3 = wine_review.groupby('country').nunique()
wine_review3
```

Out[16]:

Unnamed: 0	description	designation	points	price	province	region_1	region_2
country							
Albania	2	1	0	1	1	1	0
Argentina	5631	3446	827	18	96	2	31
Australia	4957	3152	1144	20	118	7	83
Austria	3057	1958	934	18	97	28	0
Bosnia and Herzegovina	4	3	1	3	2	1	0
Brazil	25	14	10	7	8	6	0
Bulgaria	77	56	19	10	12	8	0
Canada	196	149	68	12	38	3	11
Chile	5816	3735	865	16	88	41	0
China	3	2	1	1	2	1	0
Croatia	89	70	40	11	24	13	0
Cyprus	31	15	8	7	10	6	0
Czech Republic	6	6	0	3	4	1	0
Egypt	3	3	2	2	0	1	0
England	9	9	7	3	8	1	0
France	21098	14453	4542	21	293	11	393
Georgia	43	30	22	11	14	3	0
Germany	2452	1529	826	18	151	12	0
Greece	884	541	261	13	45	62	0
Hungary	231	139	73	17	58	11	0
India	8	7	2	5	4	1	0
Israel	630	369	168	14	53	14	0
Italy	23478	14847	5593	21	192	10	371
Japan	2	1	1	1	1	1	0
Lebanon	37	27	11	9	17	2	0
Lithuania	8	4	4	2	1	1	0
Luxembourg	9	3	3	3	2	1	0
Macedonia	16	12	4	8	6	1	0
Mexico	63	54	35	13	24	4	0
Moldova	71	46	32	10	18	4	0

Unnamed: 0

	description	designation	points	price	province	region_1	region_2
--	-------------	-------------	--------	-------	----------	----------	----------

country								
Montenegro	2	1	0	1	1	1	0	0
Morocco	12	11	5	7	8	4	0	0
New Zealand	3320	1863	416	15	71	27	0	0
Portugal	5322	3507	1570	19	129	45	0	0
Romania	139	106	61	11	20	17	0	0
Serbia	14	10	6	4	7	3	0	0
Slovakia	3	2	0	2	2	1	0	0
Slovenia	94	72	38	11	27	9	0	0
South Africa	2258	1481	449	16	73	36	0	0
South Korea	4	2	2	2	2	1	0	0
Spain	8268	5443	2158	19	172	8	82	0
Switzerland	4	4	3	3	4	4	0	0
Tunisia	2	2	2	2	0	1	0	0
Turkey	52	46	27	8	18	5	0	0
US	62397	40534	11241	21	149	24	265	18
US-France	1	1	0	1	1	1	0	0
Ukraine	5	5	5	4	1	1	0	0
Uruguay	92	52	36	11	27	9	0	0



In [14]:

```
#question 2 using value count

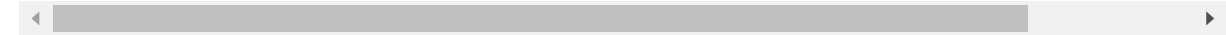
wine_review2 = wine_review.groupby(['country']).count()
wine_review2
```

Out[14]:

Unnamed: 0								
description designation points price province region_1 region_2								
country								
Albania	2	2	0	2	2	2	0	0
Argentina	5631	5631	3936	5631	5587	5631	5627	0
Australia	4957	4957	3347	4957	4894	4957	4957	0
Austria	3057	3057	2654	3057	2483	3057	0	0
Bosnia and Herzegovina	4	4	1	4	4	4	0	0
Brazil	25	25	18	25	25	25	0	0
Bulgaria	77	77	51	77	77	77	0	0
Canada	196	196	145	196	194	196	196	0
Chile	5816	5816	4917	5816	5766	5816	0	0
China	3	3	1	3	3	3	0	0
Croatia	89	89	60	89	83	89	0	0
Cyprus	31	31	25	31	31	31	0	0
Czech Republic	6	6	0	6	6	6	0	0
Egypt	3	3	2	3	0	3	0	0
England	9	9	9	9	8	9	0	0
France	21098	21098	14507	21098	14785	21098	21083	0
Georgia	43	43	38	43	43	43	0	0
Germany	2452	2452	2299	2452	2347	2452	0	0
Greece	884	884	605	884	872	884	0	0
Hungary	231	231	182	231	230	231	0	0
India	8	8	4	8	8	8	0	0
Israel	630	630	568	630	610	630	0	0
Italy	23478	23478	16890	23478	18784	23478	23478	0
Japan	2	2	2	2	2	2	0	0
Lebanon	37	37	29	37	37	37	0	0
Lithuania	8	8	8	8	8	8	0	0
Luxembourg	9	9	9	9	9	9	0	0
Macedonia	16	16	7	16	16	16	0	0
Mexico	63	63	46	63	63	63	0	0
Moldova	71	71	58	71	71	71	0	0

Unnamed: 0

	description	designation	points	price	province	region_1	region_2	
country								
Montenegro	2	2	0	2	2	2	0	0
Morocco	12	12	5	12	12	12	0	0
New Zealand	3320	3320	1585	3320	3070	3320	0	0
Portugal	5322	5322	4769	5322	4176	5322	0	0
Romania	139	139	118	139	139	139	0	0
Serbia	14	14	11	14	14	14	0	0
Slovakia	3	3	0	3	3	3	0	0
Slovenia	94	94	60	94	81	94	0	0
South Africa	2258	2258	1200	2258	2237	2258	0	0
South Korea	4	4	4	4	4	4	0	0
Spain	8268	8268	6554	8268	8160	8268	8268	0
Switzerland	4	4	3	4	4	4	0	0
Tunisia	2	2	2	2	0	2	0	0
Turkey	52	52	34	52	50	52	0	0
US	62397	62397	40345	62397	62139	62397	62261	60953
US-France	1	1	0	1	1	1	0	0
Ukraine	5	5	5	5	5	5	0	0
Uruguay	92	92	77	92	85	92	0	0



In [12]:

#question 3

```
wine_review2 = wine_review.groupby(['country']).size()  
wine_review2
```

Out[12]:

country	
Albania	2
Argentina	5631
Australia	4957
Austria	3057
Bosnia and Herzegovina	4
Brazil	25
Bulgaria	77
Canada	196
Chile	5816
China	3
Croatia	89
Cyprus	31
Czech Republic	6
Egypt	3
England	9
France	21098
Georgia	43
Germany	2452
Greece	884
Hungary	231
India	8
Israel	630
Italy	23478
Japan	2
Lebanon	37
Lithuania	8
Luxembourg	9
Macedonia	16
Mexico	63
Moldova	71
Montenegro	2
Morocco	12
New Zealand	3320
Portugal	5322
Romania	139
Serbia	14
Slovakia	3
Slovenia	94
South Africa	2258
South Korea	4
Spain	8268
Switzerland	4
Tunisia	2
Turkey	52
US	62397
US-France	1
Ukraine	5
Uruguay	92

dtype: int64

In [5]:

```
#question 4
```

```
price_mean_value = wine_review.price.mean()  
price_mean_value
```

Out[5]:

```
33.13148249353299
```

In [7]:

```
#question 4
```

```
centered_price = wine_review.price - wine_review.price.mean_value()  
centered_price
```

Out[7]:

```
0          201.868518  
1           76.868518  
2           56.868518  
3           31.868518  
4           32.868518  
...  
150925    -13.131482  
150926     -6.131482  
150927    -13.131482  
150928     18.868518  
150929    -18.131482  
Name: price, Length: 150930, dtype: float64
```

In []:

```
star_ratings = {'Star 3':[95, 96, 97, 98],  
                'Star 2':[85, 84, 83, 81],  
                'Star 1':[]}
```

In [54]:

```
#question 5  
#first convert from dataframe to series inorder  
#to use the .idxmax operator  
#here i converted only the specific column i was interested in  
  
ser2 = wine_review['points'].squeeze()  
ser2
```

Out[54]:

```
0      96  
1      96  
2      96  
3      96  
4      95  
..  
150925  91  
150926  91  
150927  91  
150928  90  
150929  90  
Name: points, Length: 150930, dtype: int64
```

In [55]:

```
#question 5 contd.  
#using the idxmax operator to get the max index value for points on the data set  
  
wine1 = ser2.idxmax()  
wine1
```

Out[55]:

```
2145
```

In [58]:

```
#using the maximum index for points to access the description of the wine  
#using the row and column index numbers  
  
bargain_wine = wine_review.iloc[2145, 3]  
bargain_wine
```

Out[58]:

```
nan
```

In [59]:

```
#checkig to see if bargain_wine description is true using price index
```

```
price_column = wine_review['price'].squeeze()  
price_column
```

Out[59]:

```
0      235.0  
1      110.0  
2       90.0  
3       65.0  
4       66.0  
...  
150925    20.0  
150926    27.0  
150927    20.0  
150928    52.0  
150929    15.0  
Name: price, Length: 150930, dtype: float64
```

In [60]:

```
price_index = price_column.idxmax()  
price_index
```

Out[60]:

```
34920
```

In [61]:

```
#wine description same as bargain_wine
```

```
price_desc = wine_review.iloc[34920, 3]  
price_desc
```

Out[61]:

```
nan
```


In []:

```
def star_ratings = wine_review

country = []
stars1 = [3]
stars2 = [2]
stars3 = [1]

for i in star_ratings:
    if i >=95:
        return(stars1)
    elif i >=85:
        return(stars2)
else:
    return(stars3)
for i in star_ratings:
    if i == country:
        country = 'canada'
        if country:
            return(stars1)
print(star_ratings)
```

```
def ratings(row):
    rate_stars = 1
    if row.points >= 95:
        rate_stars = 3
    elif rows.points >= 85:
        rate_stars = 2
    return rate_stars
star_ratings =
wine_review.apply(ratings,
axis = 'columns')
q5.check()
```

In []: