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 simplier 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 [3]:

```
import pandas as pd
```

In [4]:

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

In [5]:

```
wine_review
```

Out[5]:

	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 [6]:

```
#question 1 using median operator

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

Out[6]:

```
88.0
```

In [8]:

```
#question 2 using value count

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

Out[8]:

country	Unnamed: 0	description	designation	points	price	province	region_1	region_2
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

country	Unnamed: 0	description	designation	points	price	province	region_1	region_2
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 [9]:

```
#question 3

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

Out[9]:

```
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 [10]:

```
#question 4

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

Out[10]:

```
33.13148249353299
```

In [16]:

```
#question 4

centered_price =  wine_review.price - price_mean_value
centered_price
```

Out[16]:

```
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 [17]:

```
#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[17]:

```
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 [18]:

```
wine1 = ser2.idxmax()
wine1
```

Out[18]:

```
2145
```

In [19]:

```
#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[19]:

```
nan
```

In [20]:

```
#question 6

def stars(row):
    if row.country == 'Canada':
        return 3
    elif row.points >= 95:
        return 3
    elif row.points >= 85:
        return 2
    else:
        return 1

star_ratings = wine_review.apply(stars, axis = 'columns')
star_ratings
```

Out[20]:

```
0         3
1         3
2         3
3         3
4         3
         ..
150925    2
150926    2
150927    2
150928    2
150929    2
Length: 150930, dtype: int64
```

In [ ]:

```
wine_review.apply(star_ratings['Canada'])
```