This notebook is an exercise in the Pandas course. You can reference the tutorial at this link.

Introduction

In these exercises we'll apply groupwise analysis to our dataset.

Run the code cell below to load the data before running the exercises.

In [ ]:

```
import pandas as pd

reviews = pd.read_csv("../input/wine-reviews/winemag-data-130k-v2.csv",
index_col=0)
#pd.set_option("display.max_rows", 5)

from learntools.core import binder; binder.bind(globals())
from learntools.pandas.grouping_and_sorting import *
print("Setup complete.")
```

Exercises

1.

Who are the most common wine reviewers in the dataset? Create a `Series` whose index is the `taster_twitter_handle` category from the dataset, and whose values count how many reviews each person wrote.

In [ ]:

```
# Your code here
reviews_written = reviews.groupby('taster_twitter_handle').taster_twitter_handle.count()
# Check your answer
q1.check()
```

In [ ]:

```
#q1.hint()
#q1.solution()
```

2.

What is the best wine I can buy for a given amount of money? Create a `Series` whose index is wine prices and whose values is the maximum number of points a wine costing that much was given in a review. Sort the values by price, ascending (so that `4.0` dollars is at the top and `3300.0` dollars is at the bottom).

In [ ]:

```
best_rating_per_price = reviews.groupby('price')['points'].max().sort_index()


# Check your answer
q2.check()
```

In [ ]:

```
#q2.hint()
#q2.solution()
```

3.

What are the minimum and maximum prices for each `variety` of wine? Create a `DataFrame` whose index is the `variety` category from the dataset and whose values are the `min` and `max` values thereof.

In [ ]:

```
price_extremes = reviews.groupby("variety").price.agg([min,max])
```

```
# Check your answer
q3.check()
```

In [ ]:
```
#q3.hint()
#q3.solution()
```

4.

What are the most expensive wine varieties? Create a variable `sorted_varieties` containing a copy of the dataframe from the previous question where varieties are sorted in descending order based on minimum price, then on maximum price (to break ties).

In [ ]:
```
sorted_varieties = price_extremes.sort_values(by=['min', 'max'], ascending=False)
# Check your answer
q4.check()
```

In [ ]:
```
#q4.hint()
#q4.solution()
```

5.

Create a `Series` whose index is reviewers and whose values is the average review score given out by that reviewer. Hint: you will need the `taster_name` and `points` columns.

In [ ]:
```
reviewer_mean_ratings = reviews.groupby('taster_name').points.mean()
# Check your answer
q5.check()
```

In [ ]:
```
#q5.hint()
#q5.solution()
```

Are there significant differences in the average scores assigned by the various reviewers? Run the cell below to use the `describe()` method to see a summary of the range of values.

In [ ]:
```
reviewer_mean_ratings.describe()
```

6.

What combination of countries and varieties are most common? Create a `Series` whose index is a `MultiIndex` of `{country, variety}` pairs. For example, a pinot noir produced in the US should map to `{"US", "Pinot Noir"}`. Sort the values in the `Series` in descending order based on wine count.

In [ ]:
```
country_variety_counts = reviews.groupby(['country', 'variety']).size()
.sort_values(ascending=False)

# Check your answer
q6.check()
```

In [ ]:
```
#q6.hint()
#q6.solution()
```

Keep going

Move on to the **data types and missing data**.

Have questions or comments? Visit the Learn Discussion forum to chat with other Learners.