Introduction

Run the following cell to load your data and some utility functions.

In [ ]:
```
import sys
from pathlib import Path
learntools_dir = Path().absolute().parents[1]
sys.path.append(str(learntools_dir))

import pandas as pd

reviews = pd.read_csv("../../pandas/datasets/winemag-data-130k-v2.csv", inde

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

Exercises

1.

What is the data type of the `points` column in the dataset?

In [ ]:
```
# Your code here
dtype = ____

# Check your answer
q1.check()
```

In [ ]:
```
#_COMMENT_IF(PROD)_
q1.hint()
#_COMMENT_IF(PROD)_
q1.solution()
```

2.

Create a Series from entries in the `points` column, but convert the entries to strings. Hint: strings are `str` in native Python.

In [ ]:
```
point_strings = ____

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

In [ ]:
```
#_COMMENT_IF(PROD)_
q2.hint()
#_COMMENT_IF(PROD)_
q2.solution()
```

3.

Sometimes the price column is null. How many reviews in the dataset are missing a price?

In [ ]:
```
n_missing_prices = ____

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

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

4.

What are the most common wine-producing regions? Create a Series counting the number of times each value occurs in the `region_1` field. This field is often missing data, so replace missing values with `Unknown`. Sort in descending order. Your output should look something like this:

```
Unknown                    21247
Napa Valley                 4480
                           ...
Bardolino Superiore            1
Primitivo del Tarantino        1
Name: region_1, Length: 1230, dtype: int64
```

In [ ]:
```
reviews_per_region = reviews.region_1.fillna('Unknown').value_counts().sort_

# Check your answer
q4.check()
```

In [ ]:
```
reviews_per_region
```

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

Keep going

Move on to **renaming and combining**.