

This notebook is an exercise in the [Pandas](#) course. You can reference the tutorial at [this link](#).

---

## Introduction

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

```
In [ ]: import pandas as pd

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

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

## Exercises

View the first several lines of your data by running the cell below:

```
In [ ]: reviews.head()
```

1.

`region_1` and `region_2` are pretty uninformative names for locale columns in the dataset. Create a copy of `reviews` with these columns renamed to `region` and `locale`, respectively.

```
In [ ]: # Your code here
# Your code here
renamed = reviews.rename(columns={"region_1": "region", "region_2": "local
e"})

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

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

## 2.

Set the index name in the dataset to `wines` .

```
In [ ]: reindexed = reviews.rename_axis('wines', axis='rows')

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

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

## 3.

The [Things on Reddit](#) dataset includes product links from a selection of top-ranked forums ("subreddits") on reddit.com. Run the cell below to load a dataframe of products mentioned on the `/r/gaming` subreddit and another dataframe for products mentioned on the `r//movies` subreddit.

```
In [ ]: gaming_products = pd.read_csv("../input/things-on-reddit/top-things/top
-things/reddits/g/gaming.csv")
```

```
gaming_products['subreddit'] = "r/gaming"
movie_products = pd.read_csv("../input/things-on-reddit/top-things/top-
things/reddits/m/movies.csv")
movie_products['subreddit'] = "r/movies"
```

Create a `DataFrame` of products mentioned on *either* subreddit.

```
In [ ]: combined_products = pd.concat([gaming_products, movie_products])

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

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

## 4.

The [Powerlifting Database](#) dataset on Kaggle includes one CSV table for powerlifting meets and a separate one for powerlifting competitors. Run the cell below to load these datasets into dataframes:

```
In [ ]: powerlifting_meets = pd.read_csv("../input/powerlifting-database/meets.
      csv")
powerlifting_competitors = pd.read_csv("../input/powerlifting-database/
openpowerlifting.csv")
```

Both tables include references to a `MeetID`, a unique key for each meet (competition) included in the database. Using this, generate a dataset combining the two tables into one.

```
In [ ]: powerlifting_combined = powerlifting_meets.set_index("MeetID").join(pow
erlifting_competitors.set_index("MeetID"))

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

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

## Congratulations!

You've finished the Pandas micro-course. Many data scientists feel efficiency with Pandas is the most useful and practical skill they have, because it allows you to progress quickly in any project you have.

If you'd like to apply your new skills to examining geospatial data, you're encouraged to check out our [Geospatial Analysis](#) micro-course.

You can also take advantage of your Pandas skills by entering a [Kaggle Competition](#) or by answering a question you find interesting using [Kaggle Datasets](#).

---

*Have questions or comments? Visit the [Learn Discussion forum](#) to chat with other Learners.*