

## Exercises\_wk6

Before you start, first import the NumPy (as `np`) and pandas (as `pd`) libraries. In addition, load the 'tips' dataset (via the Seaborn library) by using the following two lines of code:

```
import seaborn as sns
tips = sns.load_dataset('tips')
```

```
In [1]: import numpy as np
import pandas as pd

import seaborn as sns
tips = sns.load_dataset('tips')
```

**Q1:** Use attributes to respectively explore the dataset (in separate cells) in terms of its: (1) columns, (2) rows, (3) body (i.e., values), and (4) shape (i.e., the number of rows and columns).

**Q2:** Use methods to respectively explore the dataset (in separate cells) in terms of: (1) general information about the dataset, (2) its first five rows, (3) its first five columns, and (4) descriptive statistics of its numerical variables.

**Q3:** Use indexing and/or slicing to respectively select (in separate cells) the following parts of the dataset: (1) the 'total\_bill' column (as a DataFrame and only show the first five rows), (2) the 'total\_bill' and 'tip' columns (again only show the first five rows), (3) all columns for the first three rows, (4) the 'total\_bill' column for the first three rows (again as a DataFrame), and (5) the 'total\_bill' and 'tip' columns for the first three rows.

**Q4:** Use the **loc** (based on index label (row/column name)) attribute to respectively select (in separate cells) the following parts of the dataset: (1) all columns for the first four rows, (2) all columns for the last row, and (3) all columns for the second, fourth and sixth rows.

**Q5:** Use the **iloc** (based on row/column index (row/column number)) attribute to respectively select (in separate cells) the following parts of the dataset: (1) all columns for the first four rows, (2) all columns for the last row, and (3) all columns for the second, fourth and sixth rows.

**Q6:** Use the **loc** (based on index label (row/column name)) attribute to respectively select (in separate cells) the following parts of the dataset: (1) the 'total\_bill' and 'tip' columns for all rows (only show the first five rows), (2) the 'total\_bill' and 'tip' columns for the second, fourth and sixth rows, and (3) the value of 'total\_bill' on the second row.

**Q7:** Use the `iloc` (based on row/column index (row/column number) attribute to respectively select (in separate cells) the following parts of the dataset: (1) the 'total\_bill' and 'tip' columns for all rows (only show the first five rows), (2) the 'total\_bill' and 'tip' columns for the second, fourth and sixth rows, and (3) the value of 'total\_bill' on the second row.

**Q8:** Use conditionals to respectively select (in separate cells) the following parts of the dataset: (1) the 'total\_bill' and 'tips' columns when 'total\_bill' > 40, and (2) the 'total\_bill' and 'tips' columns when 'sex' = female and 'total\_bill' > 40.

**Q9:** Use the `groupby` method to perform the following calculations (in separate cells) on subsets of the dataset: (1) the average 'total\_bill' and 'tip' amounts during lunches and dinners, (2) the average 'total\_bill' and 'tip' amounts during lunches and dinners for males and females, and (3) the average and summed 'total\_bill' and 'tip' amounts during lunches and dinners.

**Q10:** Perform the following operations (in separate cells) on the dataset: (1) create a new column ('percentage') by dividing the 'tip' column by the 'total\_bill' column and then calculate the average of the new 'percentage' column (round the result to three digits), and (2) use the `groupby` method to calculate the average 'percentage' amounts during lunches and dinners for males and females.