

## Exercises

Using the CSV files in the [exercises/](#) folder and what we have learned so far in this module, complete the following exercises:

1. With the [earthquakes.csv](#) file, select all the earthquakes in Japan with a magnitude of 4.9 or greater using the mb magnitude type.
2. Create bins for each full number of earthquake magnitude (for instance, the first bin is  $(0, 1]$ , the second is  $(1, 2]$ , and so on) with the ml magnitude type and count how many are in each bin.
3. Using the [faang.csv](#) file, group by the ticker and resample to monthly frequency. Make the following aggregations:
  - a) Mean of the opening price
  - b) Maximum of the high price
  - c) Minimum of the low price
  - d) Mean of the closing price
  - e) Sum of the volume traded
4. Build a crosstab with the earthquake data between the tsunami column and the magType column. Rather than showing the frequency count, show the maximum magnitude that was observed for each combination. Put the magnitude type along the columns.
5. Calculate the rolling 60-day aggregations of the OHLC data by ticker for the FAANG data. Use the same aggregations as exercise 3.
6. Create a pivot table of the FAANG data that compares the stocks. Put the ticker in the rows and show the averages of the OHLC and volume traded data.
7. Calculate the Z-scores for each numeric column of Amazon's data (ticker is AMZN) in Q4 2018 using `apply()`.

## 8. Add event descriptions:

a) Create a dataframe with the following three columns: ticker, date, and event. The columns should have the following values:

i) ticker: 'FB'

ii) date: ['2018-07-25', '2018-03-19', '2018-03-20']

iii) event: ['Disappointing user growth announced after close.', 'Cambridge Analytica story', 'FTC investigation']

b) Set the index to ['date', 'ticker'].

c) Merge this data with the FAANG data using an outer join.

9. Use the transform() method on the FAANG data to represent all the values in terms of the first date in the data. To do so, divide all the values for each ticker by the values for the first date in the data for that ticker. This is referred to as an index, and the data for the first date is the base ([https://ec.europa.eu/eurostat/statistics-explained/index.php/Beginners:Statistical\\_concept\\_-\\_Index\\_and\\_base\\_year](https://ec.europa.eu/eurostat/statistics-explained/index.php/Beginners:Statistical_concept_-_Index_and_base_year)). When data is in this format, we can easily see growth over time. Hint: transform() can take a function name.

10. The European Centre for Disease Prevention and Control (ECDC) provides an open dataset on COVID-19 cases called daily number of new reported cases of COVID-19 by country worldwide (<https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographicdistribution-covid-19-cases-worldwide>). This dataset is updated daily, but we will use a snapshot that contains data through September 18, 2020. Complete the following tasks to practice the skills you've learned up to this point:

a) Prepare the data:

i) Read in the data in the covid19\_cases.csv file.

ii) Create a date column by parsing the dateRep column into a datetime.

iii) Set the date column as the index.

iv) Use the replace() method to update all occurrences of United\_States\_of\_America and United\_Kingdom to USA and UK, respectively.

v) Sort the index.

- b) For the five countries with the most cases (cumulative), find the day with the largest number of cases.
- c) Find the 7-day average change in COVID-19 cases for the last week in the data for the five countries with the most cases.
- d) Find the first date that each country other than China had cases.
- e) Rank the countries by cumulative cases using percentiles.