COMP 333 — Lab Assignment 1

Motivation

The purpose of this assignment is for you to apply Python and pandas to a simple example.

It builds on Example 1 of Week 2 Lecture for restaurant tipping. It also builds on the labs for Weeks 2–4 on the use of Python and pandas.

Assignment

Create a Jupyter notebook using Python code and any of its libraries, but especially pandas, to analyse the data in the csv file for restaurant tipping, and to produce some plots similar to the wikipedia article.

Remember to add the new feature *"tip_rate"*.

You may need to do some self-learning of the plotting libraries for Python. Start with the cheat sheets for matplotlib and seaborn.

Deliverable

A notebook should indicate

- ▶ your identity (name and student number),
- ▶ the task (ie the course, the assignment number, and a short description of the task),
- ▶ the source of the input,
- ▶ a description of the input and the planned output.

For this assignment, your notebook must include the following:

- ▶ the above description and identification of the notebook;
- ▶ read in the csv file and print the first 10 rows of the dataframe;
- ▶ output the following plots:
    - ▶ a box plot for the total bill,
    - ▶ a scatter plot of tip versus total bill,
    - ▶ a scatter plot of *tip_rate* versus *party_size* also showing the line
      $tip\_rate = 0.18 - 0.1 \times party\_size$,
- ▶ an output of the above plots for the case of smokers; and
- ▶ a histogram of the sum of tip plus total bill where the bins cover 10 cent increments.

Your deliverable is the completed ipynb notebook showing all computation, output, and plots.