

Due date: September 29, 2019 at 11:59pm

Individual assignment. No group work allowed.
Weight: 10% of the final grade.

Description

Write python code in a jupyter notebook that will accept a CSV file and summarize the data within it for each column. For each column in the CSV file you need to calculate the following:

1. the minimum value
2. the maximum value
3. the average
4. the standard deviation
5. the most common value in the column
6. histogram of the values

You need to format the results nicely in a table.

Items 1-4 can only be computed if all data within the column are numeric. If at least one item in a column is non-numeric, then report 'n/a' for that column.

Item 4 can only be computed if there are at least 2 numeric values. Report 'n/a' if less than 2 values are available.

Item 5 must be computed for every column. If there is a tie, pick one of the most common values.

Item 6 needs to be computed with your own code. You should use matplotlib to display the bar chart. Do not use matplotlib.pyplot.hist or any other function you did not write to compute the histogram.

A sample .ipynb file is provided on the assignment page. Please download this file and implement your assignment within this file.

Some sample CSV files are also on the assignment page. You should test your code on these.

Submit your .ipynb file to D2L for grading before the assignment is due. Please note all work submitted has to be your own work.

You can use the 'csv' and 'matplotlib' libraries in this assignment. Use of any other library is not allowed, without explicit permission from your instructor.