ACTU PS5841 Data Science in Finance & Insurance - Spring 2022 (Y. Wang)
Assignment - 2
Assigned 1/27/22, Due 2/3/22 (Thur)

Problem 1. Base Python, Data Wrangling, Probability Distribution
Let **base python** refer to the python distribution from www.python.org. Please use only base python to complete the following tasks and submit HW2.ipynb and HW2.html (which is an HTML export of the fully-run HW2.ipynb).

Chance LLC provides lottery enthusiasts with analytical insights on popular lottery games. For example, it estimates the distribution of the random gap between game draws on a specific number that is picked as one of the winning numbers.

Consider the number 9. If the number 9 is drawn in the n-th game, then in the $(n+3)$-th game but not in any interim games, it counts as a two-game gap (gap = 2).

The file, *Lottery_NY_Lotto_Winning_Numbers.csv*, contains winning numbers of the New York Lotto game since 2001.

Please code to output the raw distribution of the gap for the number 2, similar to the following:

number = 2

gap	count	percent
0	22	0.0957
1	26	0.1130
2	11	0.0478
3	17	0.0739
4	16	0.0696
5	11	0.0478
6	14	0.0609
7	14	0.0609
8	17	0.0739
9	9	0.0391
≥ 10	73	0.3174

Note 1 – The numbers in the table above are from a simulated game. They are close to but not identical to the answers to this assignment.
Note 2 – In your analysis, you should include the bonus numbers. That is, if the number 2 is a bonus number for a drawing date, it is also considered a winning number for your analysis.
Note 3 – One way to check the accuracy of your code output is to perform the same analysis in a spreadsheet.

You may find the following resources useful.

New York Lotto – information
https://nylottery.ny.gov/draw-game/?game=lotto

handling csv files
https://docs.python.org/3/library/csv.html

handling strings
https://docs.python.org/3/library/stdtypes.html#str