Please submit the following to reflect your work:
[1] A fully-run ipnb file
[2] An html export of the fully-run ipnb file in [1]
[3] A document showing your answers. You can use or replicate Answer_Template.docx.

Gaps
The file **data.csv** contains historical market information on the SP500 index. Note that 1 basis point (bp) = 0.01%.

The return (in bp) on day t is calculated as

$$\mathrm{R}_t = \Big(\frac{\mathrm{Close}_t}{\mathrm{Close}_{t-1}} - 1\Big) \times 10000$$

Volatility (in bp/day) is estimated by the (sample) standard deviation of daily returns, based on the assumption that daily returns are independent of each other and are identically distributed.

There are random gaps between a trading day and the one before it. The gap (in bp) on day t is

$$\mathrm{Gap}_t = \Big(\frac{\mathrm{Open}_t}{\mathrm{Close}_{t-1}} - 1\Big) \times 10000$$

The size of a gap is $|\mathrm{Gap}_t|$. Therefore, $\mathrm{Gap}_t = \pm 10$ represents a gap of size 10bp on day t .

Large gaps, those larger than 75bp in size for example, have become more frequent in recent years. Gaps of large sizes are of a particular concern to traders, such as delta hedgers.

Let's investigate.

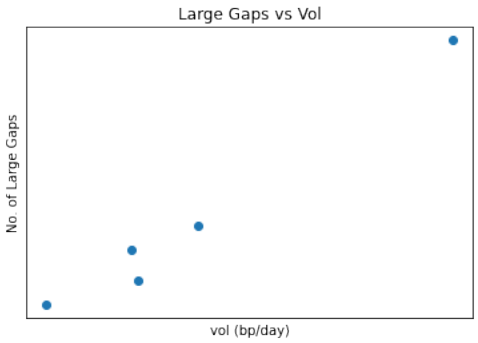
Please code in python and let your code answer the following questions.

[a] Please let your code output the following table (by printing a dataframe, for example.)
* A 75bpGap occurrs on day t when $|\mathrm{Gap}_t| > 75$bp.
* TDays stands for trading days.
* Each year's volatilty (in bp/day), is estiamated from daily returns within that year.

Year	No. of TDays	volatility	No. of 75bpGaps
2021			
2020			
2019			
2018			
2017			

Please see the back of this page for more questions.

[*b*] Please let your code procude a scatter plot based on data from the table in [*a*], in a similar format as follows. **I have hidden the axis tickmarks and labels but you should show them.**



Let's assume that the 75bpGaps are random draws from Poisson distributions.

A naive Poisson model with the same mean for each year would predict **22.67** 75bpGaps for 2022 (which has an estimated 252 trading days according to the NYSE.)

However, [*b*] reveals a potential connection between the frequency of large gaps and volatility. Let's attempt to build a better model by formulating a Poisson regression that allows for year-specific means. We use volatility as the only explanatory variable for now.

[*c*] Please fit this model with data from the table in [*a*] and let your code output the following table (by printing a dataframe, for example.)

Year	No. of TDays	No. of 75bpGaps	fitted No. of 75bpGaps
2021			
2020			
2019			
2018			
2017			

[*d*] For the year 2022, we estimate that there are 252 trading days and project a volatility of 121.5 bp/day.

[*d*1] Let your code report how many 75bpGap-trading-days your fitted model in [*c*] predicts for 2022.

[*d*2] Let your code report the pseudo-R^2 of the fitted model.

[*d*3] Let your code report the estimated prediction RMSE (root mean squared error) (in TDays) based on a modified LOOCV (leave one out cross validation). The modified LOOCV never leaves out the observation for 2020 (the year of the pandemic), in order to make each training set more representative of the vol space.