Assignment

For you stock, you have labeled each week as 'green" or "red" for years 1 and 2. Recall that for each week you (typically) have 5 daily returns. From these returns, for each of your week you compute the following two attributes ("feature values"):

1. μ: average of daily returns

2. σ: standard deviation ("volatility") of daily returns

Both μ and σ have the same units. For convenience, convert them both to percentages and round to two decimal points. The objects for our classifiers are weeks and for each week you have three values:

$$(\mu, \sigma, \text{Label})$$

In this assignment, you will visually compare your labels for both years and try to argue whether nearest neighbors approach could be a good (or bad) classifier. Specifically, your task is the following:

- for each year you will have around 50 weeks (52 trading weeks in a typical year). Plot these points on two separate charts (year 1 and year 2). For each chart, you have μ on the x-axis, σ on the y-axis. Each point has color. Add

week id to each point (hint: use plt.scatter function) and use parameter "s" (size) to make them highly visible on your chart.

- put both plots in a word file and describe what you see in each chart.

  1. Are there any "obvious" patterns? For example, for higher σ and μ are there more green points?
  2. are points of the same color located close to each other?
  3. do patterns repeat from year 1 to year 2?
  4. do you expect nearest-neighbor classifier trained in year 1 to do well in year 2?

For this assignment (aside from generating plots) do not use any Python code to answer the above questions. We want to get a "qualitative" assessment. Just tell us what you see.