1. Without looking at the lecture notes or other material, explain the 68-95-99.7 rule in writing. Only look at the lecture notes if you get stuck.

**The 68-95-99.7 rule states that most of the data will be within 3 standard deviation. Most (68%) will be within one std dev, 95% will be within 2, and 99.7% will be within 3.**

1. What are two everyday examples of data that you would expect to have a bell-shaped distribution? Don’t rush over this problem.

**Two everyday examples I expect to be bell shaped are:**

* **The amount of money I spend at the grocery store in a given week.**
* **My electricity usage is another good one!**

1. What are two everyday examples of data that you would not expect to have a bell-shaped distribution? Don’t rush over this problem.

**Two examples of non-bell shaped would be:**

* **Real estate pricing**
* **Salaries**

1. Without looking at the lecture notes or other material, explain what CDF stands for and what it is. Only look at the lecture notes after you give it a try.

**CDF stands for Cumulative Distribution Function. It is used to find the probability of something happening for a value that is <= X.**

1. What is the maximum value a CDF can take?

**1.0**

1. In Python, plot the PDF of a normal distribution with a mean 5 and standard deviation of 0.5. Use scipy.stats.norm.pdf() to get values from the PDF, and matplotlib's plot() function to plot values of the PDF.
2. Sample 1000 values from the normal distribution we saw in the previous problem. Produce a histogram of your values. You can use matplotlib or seaborn. Run your code a few times to see how the histogram varies each time.
3. Suppose two points are exactly 10 meters apart. However, when we measure the distance between the points we get 10 + X, where X is a continuous random variable with a normal distribution with mean 0 and variance 0.04. The random variable X represents the measurement error, which may be positive or negative, but on average is 0.
4. plot the PDF of observed measurements using Python. This is similar to what was done in the earlier problem where we plotted a normal distribution. Plot the actual PDF, not a histogram based on sampled data.
5. what range of measurement values would you expect 68% of the data to fall within?
6. what is the probability that the observed measurement is within 0.4 meters of the true distance? Solve the problem in two ways: once analytically, and once through simulation. Hint: use the idea of part b for the analytical part.
7. If you still have time: in one of the problems above we plotted a histogram of 1000 samples from a normal distribution. Figure out how you could plot the histogram and then superimpose on the histogram the PDF of the normal distribution. Hint: it's easy to do this with Seaborn. With matplotlib, first plot the histogram, then plot values of the PDF. Use 'density=True' when plotting the histogram.
8. If you still have time, start reading “50 Years of Data Science” by David Donoho, available on the web. Write a 3-page technical review of the paper.