1. Read "Data Science: the impact of statistics" and “7 ways data scientists use statistics”
2. Read about Law of Total Probability (also known as Total Probability Rule)
3. Read about Bayes’ Rule
4. Watch Probability Walkthrough
5. Solve the following probability questions:
    a. You have two fair six-sided dices, and you roll both together. What is the probability that the sides they land on sum up to 8?
    b. You have a glass jar containing jelly beans; 6 of them are red, 2 of them are blue, 4 of them are green, and 1 of them is yellow. You pull out three jelly beans from the jar without looking, one after the other, without putting either back in. What is the probability that at least one of the beans you pulled out is blue?
    c. There are 200 students in an Introduction To Data Science course. 140 students use Jupyter and 60 students use Colab. Colab conducts A/B testing of a new feature they are developing, such that each Colab user is randomly assigned to either see the new feature or see the old version (i.e. the probability of a Colab user getting the new feature is the same as the probability of that user getting the old version), for the the testing duration. During that period of time, you randomly select a student out of all the students in the course. What is the probability that they have encountered the new Colab feature?
6. Watch Visualizing quantitative data walkthrough
7. Watch Visualizing qualitative data walkthrough
8. Download Airbnb_NYC_2019.csv
9. Start a notebook in python3 and name it ***Name_A2.ipynb*** where Name consists of your name the way you would want it on your certificate if you pass this course, with an underscore after each part of your name e.g. if I was a student I would name my notebook *Lavanya_Vijayan_A2.ipynb*; if Gloria was a student she would name her notebook *Gloria_Tumushabe_A2*.ipynb.
10. In the notebook write code to do the following tasks/answer the following questions:
    a. Select a subset of the Airbnb NYC 2019 dataset — the subset should only contain the data where the neighbourhood group is Manhattan.
        i. How many rows are in the subset?
        ii. How many columns are in the subset?
    b. Create a histogram to visualize the distribution of prices for the Manhattan listings. Write a sentence to describe what you see.
    c. Create a scatter plot to visualize the number of reviews over the price for the Manhattan listings. Write a sentence to describe what you see.
    d. Identify the “top 10” neighbourhoods from the Manhattan listings — the neighborhoods with the 10 highest counts i.e. number of listings. What are they, in order of first highest count to tenth highest count?
    e. Select the subset of the Manhattan listings data corresponding to the “top 10”.
        i. How many rows are in the subset?
        ii. How many columns are in the subset?
    f. Create a count plot to visualize how many Airbnb listings are in each of those “top 10” neighbourhoods in Manhattan in 2019. Here is an example of tackling

"crowded"/"overlapping" axis labels that may help you. Write a sentence to describe what you see.

11. Fill out this form. You will be asked to:
    a. Enter your answers to the questions for the probability portion of the assignment
    b. Describe your thought process/the steps you took to arrive at your answers to the probability questions
    c. Enter your answers to the questions for the coding portion of the assignment
    d. Attach your completed ipynb via file upload
    e. Enter any conceptual questions you may have at this point in time (Optional)