Lab 2

**This lab provides insight into the concept of probability and statistical measures that you could use to classify the data as spam or not spam
and also clean the data by removing the outliers using measures of central tendency and dispersion.
The datasets have been enclosed.

Part A

Use the **Lab2 dataset** provided.
1. Load the dataset and split it into test and train.
2. After that, train the model using Gaussian and Multinominal classifiers
3. Check which model performs better.
4. Use the trained model to perform some predictions on test data.

Part B

Use this **air bnb new york city dataset**
1. Remove outliers based on price per night for a given apartment/home.
2. You can demonstrate why using other techniques like mean/median/ percentile works.
3. The task is to come up with a clean dataset that does not have outliers showcasing all the possibilities.

Submission Guidelines:

1. How to find labs on eConestoga? Content>Course Tools>Assignments>Labs i(i = 1-2) Or Content>Evaluations> Labs 1-2
2. Submission of the others work in any form is an academic integrity.
3. How to complete the Labs/Assignments? Practice lecture pdf files and python coding files.
4. Files naming conventions: FiirstName_LastName_LabName_No
5. Emergency situations: Email coordinator or instructor 72 hours before the due dates for all labs/assignments/midterm/final exam.
6. Submit jupyter notebook python file(.ipynb) and html file on eConestoga: Course Tools>Assignments>Labs i(i = 1-2)
7. Resubmission, rewriting, rescheduling, and make-up of the assignments/Labs/Midterm/Final Exam are not allowed.
8. At the end of the semester, students will not be allowed to make-up missed labs/assignments/midterm/final exam in order to pass the course.
9. Jupyter notebook pyhton file(.ipynb) to html converter:
   1. https://www.runcell.dev/tool/ipynb-to-html
   2. https://www.vertopal.com/en/convert/ipynb-to-html
10. This lab should be submitted as a notebook and an HTML two weeks after the due date. Follow https://docs.github.com/en/pages/quickstart.

Scoring:

1. Scoring is based on the clarity and accuracy of the Labs/Assignments/Midterm/Final exam and on your ability to answer the questions.
2. Starts with a markup cell with your name and Lab/Assignment/Midterm/Final Exam name. Write the question numbers and statements before the code cells.
3. Contains all the setup and configuration steps necessary to run the code that follows.

Late deliveries:

1. 1 day late: 20% cumulative penalty(0 - 24 hours → 20% penalty).
2. 2 days late: 40% cumulative penalty(24 - 48 hours→ 40% penalty).
3. 3 days late: 100% cumulative penalty(48 - 72 hours → 100% penalty)