TP1: Introduction to Python and Jupyter Notebook

1. Install Python (the latest version).
2. Check the option (Add Python 3.10 to PATH).

3. Check if Python has been installed correctly through the command prompt (cmd) using the command: >python

```
C:\Users\difna>python
Python 3.10.7 (tags/v3.10.7:6cc6b13, Sep  5 2022, 14:08:36) [MSC v.1933 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

4. Exit the text editor and install Jupyter by following these lines of code:

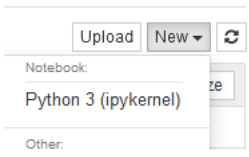
```
>>> quit()

C:\Users\difna>python -m pip install jupyter
Collecting jupyter
```

5. To verify if the installation is successful, type the following command in your console

```
:\Users\difna>jupyter notebook
```

6. Create a new notebook.



7. Rename the file, add cells of text type and code type.

8. Install the following libraries (in the command console): scipy, numpy, matplotlib, pandas, sklearn: pip install library_name.

9. Vectors:

- Create an array containing numbers from 1 to 9.
- Reshape this 1D array into a 2D array and display its dimensions.
- Display the first row and the first column of the reshaped array.
- Perform a matrix product between two 2D arrays.

10. Graphs

- Given the two lists X, Y: X = [-1, 0, 1, 2], Y = [3, 2, 4, 1], where X represents the list of abscissas and Y represents the list of ordinates, create a graph.
- Gradually insert the following commands:

```
plt.grid()
plt.axis([-2,3,0,5])
plt.xlabel("mes abscisses")
plt.ylabel("mes ordonnées")
plt.title("courbe d'équation $y =f(x)$")
```

- Replace the line plt.plot(X, Y) with plt.plot(X, Y, 'r.')

11. Datasets and Kaggle

Visit the Kaggle website (https://www.kaggle.com), a well-known platform
for datasets and data science competitions. The task is to:
1. Create an account (if one does not already exist) on Kaggle.
2. Explore the available datasets and select one of interest.
3. Download the chosen dataset.
4. Describe the content of the dataset by addressing the following points:
• What is the subject of the dataset?
• How many variables (columns) and observations (rows) are present?
• What is the nature of the variables (e.g., numeric, categorical)?What is the potential goal of analyzing this dataset (e.g., regression, classification)?

12. Pandas Dataframe

1. Create a random dataset of employee salaries with two features (Years and Salary) and 40 samples.
   - Years of experience: randomly generated integers between 1 and 10.
   - Salary: calculated based on the years of experience. The salary is computed as follow: Base salary is 2000 +200 for every year + a random integer number from 1 to 400.
2. Store the created dataset in a CSV file.
3. Import the dataset from the CSV file and load into in a Pandas dataframe.
4. Plot the salary of employees according to their years of experience. What do you observe?
5. Predict the salary of an employee with 10 years of experience.