- Lab 1 : Data Analytics
- Lecturer : Pr. Youness Moukafih
- Lab Instructor: Pr. Berkani Safaa

Lab 1 – Introduction to Python

Objective:

The goal of this lab is to introduce you to the essential parts of the Python programming language and its data-oriented libraries, which will equip you to become an effective data analyst. Python provides powerful tools for data analysis, interactive computing, and data visualization.

In recent years, Python has gained popularity due to its versatile open-source libraries like Pandas, Matplotlib, and Scikit-learn, which make it a strong choice for performing a wide range of data analysis tasks. Moreover, Python's general-purpose capabilities also make it an excellent option for building data applications, beyond just analysis. This versatility is why Python has become a preferred language for many data analysts and data scientists.

As you progress through the labs, you will not only follow instructions but are also encouraged to explore beyond what's provided. You'll receive a cheat sheet and detailed instructions to guide you, but your curiosity and initiative will help you deepen your understanding of the tools.

Instructions:

1. Work in groups of 2 or 3 students. Your group will remain the same for all labs this semester.
2. The final report is due by 11:59 PM on the day of the lab.
3. Prepare one final lab report as a single file for your group.
4. The questions marked with (Analysis) require a written response. You must answer these in your report using complete sentences, and use the data or plots you've generated to support your reasoning.
5. The report must be uploaded to the CONNECT platform.
   IMPORTANT: Every member of the group must individually upload the exact same file to CONNECT to receive a grade.

- Lab 1 : Data Analytics
- Lecturer : Pr. Youness Moukafih
- Lab Instructor: Pr. Berkani Safaa

Exercise 1: Sleep tracker analysis:

You've been feeling tired lately, so you decide to track how much sleep you get every night for a week to see whether you're getting enough rest. Here is the data for your sleep over the last week:

First week	Second week
Monday : 6 hours of sleep Tuesday : 7 hours of sleep Wednesday : 8 hours of sleep Thursday : 5 hours of sleep Friday : 9 hours of sleep	Monday : 7 hours of sleep Tuesday : 6 hours of sleep Wednesday :8 hours of sleep Thursday : 7 hours of sleep Friday : 6 hours of sleep

Q1: Assign the number of hours you slept each day during the first week to the list *week1_sleep and* do the same for the second week in the list *week2_sleep*.

Q2: Create a variable *days* to store the days of the week (as strings).

Q3: Create a variable *daily_difference* that shows how your sleep changed each day between the two weeks (e.g., how much more or less sleep you got each day from Week 1 to Week 2).

Q4: a - Calculate the total amount of sleep you got during the first week and assign it to the variable *total_week1*.

b - Similarly, calculate the total sleep for the second week and assign it to *total_week2*.

c - Now calculate your overall average sleep per day for each week and store them in *average_week1* and *average_week2*.

Q5: Compare whether you slept more in the first week than in the second week. Assign the result of this comparison to the Boolean variable slept_more_first_week. Based on this, show a message saying which week you slept more.

Q6: a- Assign the number of hours you slept on Wednesday during the first week to the variable *sleep_wednesday_week1*.

b - Assign the sleep hours for Tuesday, Wednesday, and Thursday during the second week to the variable *midweek_sleep_week2*.

Q7: Check if you slept 8 hours or more on each day of the first week, and assign the result to the variable *sleep_enough_week1*. These are the days when you met or exceeded your sleep goal.

Q8: Create a list *sleep_successful_days_week1* that contains the hours you slept on the days where you met or exceeded the 8-hour goal in Week 1.

- Lab 1 : Data Analytics
- Lecturer : Pr. Youness Moukafih
- Lab Instructor: Pr. Berkani Safaa

Exercise 2: Exploring the students' performance dataset:

In this exercise, you will learn the fundamental steps of exploring a new dataset. The goal is to load the data, understand its structure, and use basic visualizations to find initial insights. First, download the ***StudentsPerformance.csv*** file from the CONNECT platform and place it in the same folder as your Notebook.

Q1: Import the *pandas, matplotlib.pyplot*, and *seaborn* libraries. Then, load the *StudentsPerformance* dataset into a pandas DataFrame named df.

Q2: Display the first 5 rows of the dataset using df.head().

- Based on the output, identify which columns are categorical and which are numerical? (Analysis)

Q3: Run both df.info() and df.describe() to get basic information about the dataset.

- Does this dataset have any missing values? (Analysis)
- What is the average math score for the students? (Analysis)
- Which subject has the most variation in scores? (Analysis)
- What does the 50% value represent, and how does it differ from the mean? (Analysis)

Q4: Select and display a random sample of 10 students from the dataset using df.sample(10).

Visualizations:

In the suggested code snippets, the visualizations provided are kept minimal to focus on the core concepts. However, there are many ways to improve these visualizations by customizing the style, labels, colors, and other parameters. Explore the documentation for the plotting libraries used, such as *Matplotlib* and *Seaborn*, to further optimize and enhance the code.

Q5: Plot histograms for numerical variables (math score, reading score, writing score).

Example:

```
df['math score'].hist(bins=10)
plt.show()
```

Q6: Visualize the distribution of math scores by gender using a boxplot.

```
sns.boxplot(x='gender', y='math score', data=df)
plt.show()
```

- What are the main information that you can extract from a boxplot? (Analysis)
- Based on the above plot, are there any outliers for math scores among the female students?? (Analysis)

Q7: Create a scatter plot to visualize the relationship between math score and reading score. Discuss if there is any correlation.

```
plt.scatter(df['math score'], df['reading score'])
plt.show()
```

- Is there an association between math score and reading score? (Analysis)

Q8: Calculate the correlation between the numerical variables (math score, reading score, writing score) and visualize the correlation matrix using a heatmap.

```
correlation_mx = df[['math score', 'reading score', 'writing score']].corr()
sns.heatmap(correlation_mx)
plt.show()
```

- Based on this heatmap, which two subjects have the strongest relationship? (Analysis)

End of the lab.