

TD Series 2: Python Language

Exercise 1:

Write a Python program that:

- Imports the numpy and pandas libraries.
- Creates a 1D NumPy array containing random annual growth rates (e.g., values between 1.0 and 5.0).
- Creates a Pandas DataFrame with 3 columns: **Country**: The name of the country, **GDP** (in billions of dollars): **The Gross Domestic Product of the country**, **Population** (in millions): The population of the country.
- Adds 5 rows of data.
- Displays the NumPy array and the DataFrame.

Exercise 2:

A dataset represented by the table below describes a set of products with their respective prices, quantities, and dates:

PRODUCT	PRICE	QUANTITY	DATE
A	10	5	2024-01-10
B	20	2	2024-01-15
A	10	3	2024-01-20
C	15	4	2024-02-01
B	20	1	2024-02-10
A	10	2	2024-02-15
C	15	3	2024-03-01
B	20	2	2024-03-05
A	10	4	2024-03-10
C	15	1	2024-03-15

Write a Python program that:

- Loads this dataset into a Pandas DataFrame.

- Displays the first few rows, the last few rows, and a random sample.
- Uses `df.info()` and `df.describe()` to explore the data.
- What do the `info()` and `describe()` functions display?
- Is the "Date" column in the correct format? If not, convert it.
- What is the total revenue?
- What is the revenue per product?
- What is the best-selling product in terms of quantity?
- Calculate the average revenue per month.
- Create a line graph showing the evolution of total revenue over the months.
- Add labels to the x-axis (Months) and the y-axis (Total Revenue).
- Include a title for the graph.

Additional Instructions:

- ✓ Use the appropriate Pandas functions to answer the questions (e.g., `groupby()`, `sum()`, `mean()`, `sort_values()`, etc.).
- ✓ Justify your choices of methods and operations.
- ✓ Interpret the results of your analysis.

Exercise 3: Using the file

`student_study_habits.csv`, represented by the dataset below write a Python program to :

HOURS_STUDIED	ATTENDANCE_RATE (%)	SLEEP_HOURS	TEST_SCORE	PASSED
3	80	6	60	0
5	90	7	75	1
2	60	5	45	0
4	85	8	70	1
6	95	7	82	1
1	50	4	40	0
7	98	8	90	1
3	70	5	55	0
5	88	6	78	1
2	65	5	48	0

1. Load the dataset.
2. Display the first 10 rows.
3. Display dataset information (`info()` and `describe()`).
4. Check for missing values.
5. How many students are included?
6. What is the average attendance rate?
7. Explore the correlation between hours studied and sleep hours.

Exercise 4:

Using the same dataset:

1. Predict `Test_Score` using `Hours_Studied`.
2. Split the dataset (80% for training, 20% for testing).
3. Train a Linear Regression model.
4. Predict on the test set.
5. Calculate and display the Mean Squared Error (MSE).
6. Visualize the regression results:
7. Scatter plot of actual points.
8. What does the slope of the regression line represent?
9. Interpret the model's predictive strength.

Exercise 5:

using `student_study_habits.csv`:

The inputs are : `Hours_Studied`,
`Attendance_Rate (%)`, `Sleep_Hours`.

Target: `Passed` (0 = Failed, 1 = Passed).

1. Split the data.
2. Train a Decision Tree Classifier (max depth = 3).
3. Predict and evaluate accuracy.
4. Visualize the Decision Tree.
5. Which features are most important for predicting the pass/fail status?