Data Science Fundamentals with Python

Worksheet 1

June 2025

We hope you had a good time watching the tutorials for the course and have learned a lot from them. Let's try out the following questions!

1 Question 1:

Create a Jupyter notebook or Google Colab notebook. Initialize and print a 2-dimensional NumPy array of shape (5, 5) filled with random* integers between 1 and 100. Perform the following tasks:

(a) Extract and print the middle element of the array using NumPy indexing.

(b) Calculate and print the mean of each row of the array.

(c) Create a new array that contains only the elements from the original array that are greater than the overall mean of the array.

(d) Write a Python function numpy_spiral_order(matrix) that takes a NumPy matrix and returns a list containing the elements visited in a spiral order. For example:

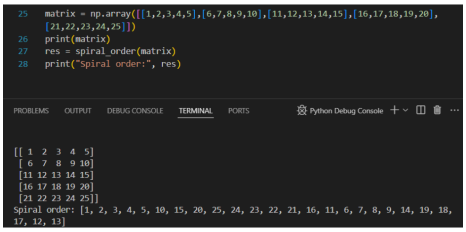

Figure 1: Output of numpy_spiral_order

Hints:

- Use np.random.randint() to create the initial array arr.
- Use NumPy functions like np.mean() and boolean indexing to perform calculations and extract elements.

2 Question 2:

This question involves the application of Python libraries on a dataset of video game sales. Here is the link to the dataset.

Create a Jupyter Notebook or a Google Colab Notebook, download the dataset in the drive link, and solve the following questions using the required Python libraries:

(a) Add a column of 'global_sales' showing the total sales of all the different regions to the data frame and sort (highest first) and print the DataFrame according to it.

(b) Display a plot of the total number of copies sold of each genre globally.

(c) Filter out only the games containing 'Grand Theft Auto' in their name and display the following as a DataFrame:

   i. their name
   ii. the platform they were released on
   iii. the year they were released in
   iv. the sum of sales in only Europe and Japan

(d) Display a pie chart of the total sales of all Grand Theft Auto games combined in North America, Europe, Japan, and other sales.