

# Data Toolkit

## Practice Questions



## People Dataset :-

<https://drive.google.com/file/d/13x8f8HNKieSRAzxTlzAojaYp8Up8cefk/view?usp=sharing>

**Demonstrate three different methods for creating identical 2D arrays in NumPy. Provide the code for each method and the final output after each method.**

- 2. Using the Numpy function, generate an array of 100 evenly spaced numbers between 1 and 10 and Reshape that 1D array into a 2D array.**
- 3. Explain the following terms:**
  - The difference in np.array, np.asarray and np.asanyarray.
  - The difference between Deep copy and shallow copy.
- 4. Generate a 3x3 array with random floating-point numbers between 5 and 20. Then, round each number in the array to 2 decimal places.**
- 5. Create a NumPy array with random integers between 1 and 10 of shape (5, 6). After creating the array perform the following operations:**
  - a) Extract all even integers from array.
  - b) Extract all odd integers from array.
- 6. Create a 3D NumPy array of shape (3, 3, 3) containing random integers between 1 and 10. Perform the following operations:**
  - a) Find the indices of the maximum values along each depth level (third axis).
  - b) Perform element-wise multiplication of between both array.
- 7. Clean and transform the 'Phone' column in the sample dataset to remove non-numeric characters and convert it to a numeric data type. Also display the table attributes and data types of each column.**
- 8. Perform the following tasks using people dataset:**
  - a) Read the 'data.csv' file using pandas, skipping the first 50 rows.
  - b) Only read the columns: 'Last Name', 'Gender', 'Email', 'Phone' and 'Salary' from the file.
  - c) Display the first 10 rows of the filtered dataset.
  - d) Extract the 'Salary' column as a Series and display its last 5 values.
- 9. Filter and select rows from the People\_Dataset, where the "Last Name" column contains the name 'Duke', 'Gender' column contains the word Female and 'Salary' should be less than 85000.**
- 10. Create a 7\*5 Dataframe in Pandas using a series generated from 35 random integers between 1 to 6?**

**11. Create two different Series, each of length 50, with the following criteria:**

- a) The first Series should contain random numbers ranging from 10 to 50.
- b) The second Series should contain random numbers ranging from 100 to 1000.
- c) Create a DataFrame by joining these Series by column, and, change the names of the columns to 'col1', 'col2', etc.

**12. Perform the following operations using people data set:**

- a) Delete the 'Email', 'Phone', and 'Date of birth' columns from the dataset.
- b) Delete the rows containing any missing values.
- d) Print the final output also.

**13. Create two NumPy arrays, x and y, each containing 100 random float values between 0 and 1. Perform the following tasks using Matplotlib and NumPy:**

- a) Create a scatter plot using x and y, setting the color of the points to red and the marker style to 'o'.
- b) Add a horizontal line at  $y = 0.5$  using a dashed line style and label it as ' $y = 0.5$ '.
- c) Add a vertical line at  $x = 0.5$  using a dotted line style and label it as ' $x = 0.5$ '.
- d) Label the x-axis as 'X-axis' and the y-axis as 'Y-axis'.
- e) Set the title of the plot as 'Advanced Scatter Plot of Random Values'.
- f) Display a legend for the scatter plot, the horizontal line, and the vertical line.

**14. Create a time-series dataset in a Pandas DataFrame with columns: 'Date', 'Temperature', 'Humidity' and Perform the following tasks using Matplotlib:**

- a) Plot the 'Temperature' and 'Humidity' on the same plot with different y-axes (left y-axis for 'Temperature' and right y-axis for 'Humidity').
- b) Label the x-axis as 'Date'.
- c) Set the title of the plot as 'Temperature and Humidity Over Time'.

**15. Create a NumPy array data containing 1000 samples from a normal distribution. Perform the following tasks using Matplotlib:**

- a) Plot a histogram of the data with 30 bins.
- b) Overlay a line plot representing the normal distribution's probability density function (PDF).
- c) Label the x-axis as 'Value' and the y-axis as 'Frequency/Probability'.
- d) Set the title of the plot as 'Histogram with PDF Overlay'.

**16. Set the title of the plot as 'Histogram with PDF Overlay'.**

**17. Create a Seaborn scatter plot of two random arrays, color points based on their position relative to the origin (quadrants), add a legend, label the axes, and set the title as 'Quadrant-wise Scatter Plot'.**

18. With Bokeh, plot a line chart of a sine wave function, add grid lines, label the axes, and set the title as 'Sine Wave Function'.
19. Using Bokeh, generate a bar chart of randomly generated categorical data, color bars based on their values, add hover tooltips to display exact values, label the axes, and set the title as 'Random Categorical Bar Chart'.
20. Using Plotly, create a basic line plot of a randomly generated dataset, label the axes, and set the title as 'Simple Line Plot'.
21. Using Plotly, create an interactive pie chart of randomly generated data, add labels and percentages, set the title as 'Interactive Pie Chart'.

