AI-Lab-numpy-pandas-matplotlib

NUMAN SHAFI

October 2024

Case Study 1: Weather Data Analysis

You are given a dataset containing daily weather information for a city over the last year. The dataset contains columns such as *Date*, *Temperature*, *Humidity*, *Wind Speed*, and *Weather Condition* (Sunny, Rainy, Cloudy, etc.).

1. **Data Generation Task:** Using Pandas and NumPy, generate a dummy dataset with 365 rows (representing days of the year). The *Temperature* column should contain random values between 10°C and 40°C, *Humidity* values between 30% and 90%, and *Wind Speed* between 0 and 20 km/h. The *Weather Condition* should randomly be assigned one of the three values: 'Sunny', 'Rainy', or 'Cloudy'.
2. **NumPy Task:** Convert the *Temperature* column into a NumPy array. Calculate the mean, median, and standard deviation of the temperature values.
3. **Pandas Task:** Filter the data to find all the days when the temperature was above 30°C and it was Sunny. How many such days are there?
4. **Pandas Task:** Group the dataset by *Weather Condition* and calculate the average *Humidity* for each weather condition.
5. **Matplotlib Task:** Plot a line graph showing the temperature variation over the year.
6. **Matplotlib Task:** Create a bar plot that shows the number of days for each weather condition (Sunny, Rainy, Cloudy, etc.).

Case Study 2: Sales Data Analysis

You are analyzing a dataset for an online store that includes information such as *Order ID*, *Product*, *Price*, *Quantity*, and *Date of Purchase*.

1. **Data Generation Task:** Generate a dummy dataset with 500 rows. Use random *Product* names (choose from 10 different products), random *Price*

values between $10 and $1000, and *Quantity* values between 1 and 20. Assign random dates for *Date of Purchase* within the past year.

2. **NumPy Task:** Create a NumPy array containing the *Price* and *Quantity* columns. Compute the total sales for each order (Price * Quantity) using NumPy.

3. **Pandas Task:** Filter the data to show only orders where the total sales value is greater than $100.

4. **Pandas Task:** Group the data by *Product* and calculate the total quantity sold for each product.

5. **Matplotlib Task:** Create a scatter plot to visualize the relationship between *Price* and *Quantity* of products sold.

6. **Matplotlib Task:** Create a histogram to show the distribution of total sales values.

Case Study 3: Employee Salary Analysis

You have a dataset that contains information about employees in a company, including columns such as *Employee ID*, *Name*, *Department*, *Salary*, and *Years of Experience*.

1. **Data Generation Task:** Generate a dataset with 300 rows. For *Employee ID*, use sequential integers, for *Name*, use random names from a list of 20 names. Assign *Department* randomly from 5 different departments. Generate random *Salary* values between $30,000 and $120,000 and *Years of Experience* between 1 and 25.

2. **NumPy Task:** Create a NumPy array with the *Salary* column. Calculate the average, maximum, and minimum salary.

3. **Pandas Task:** Filter the employees who have more than 5 years of experience and earn a salary above the average.

4. **Pandas Task:** Group the dataset by *Department* and calculate the mean salary for each department.

5. **Matplotlib Task:** Plot a bar chart showing the average salary in each department.

6. **Matplotlib Task:** Create a line plot showing the salary distribution for employees with increasing years of experience.

Case Study 4: Exam Score Analysis

You are analyzing the performance of students in an exam. The dataset contains *Student ID*, *Name*, *Subject*, *Score*, and *Total Marks*.

1. **Data Generation Task:** Generate a dummy dataset with 200 rows. Use random *Student ID*, *Name*, *Subject* (choose from 5 subjects), and assign random *Score* values between 0 and 100, with *Total Marks* fixed at 100 for all subjects.
2. **NumPy Task:** Using the *Score* column, calculate the overall mean, median, and standard deviation of the exam scores.
3. **Pandas Task:** Filter the students who scored above 80% in the exam. How many students achieved this score?
4. **Pandas Task:** Group the data by *Subject* and calculate the average score for each subject.
5. **Matplotlib Task:** Plot a histogram showing the distribution of scores across all students.
6. **Matplotlib Task:** Create a bar chart to compare the average scores across different subjects.

Case Study 5: Stock Market Analysis

You are working with stock market data containing columns like *Date*, *Open Price*, *Close Price*, *Volume Traded*, and *Company*.

1. **Data Generation Task:** Generate a dataset with 1000 rows. The *Date* column should contain random dates for the past two years, and the *Company* column should contain random selections from 5 companies. Assign random *Open Price* and *Close Price* values between \$50 and \$500, and *Volume Traded* values between 1000 and 1,000,000.
2. **NumPy Task:** Convert the *Close Price* column into a NumPy array. Calculate the daily percentage change in stock prices.
3. **Pandas Task:** Filter the data to find all the days when the stock price increased by more than 2% compared to the previous day.
4. **Pandas Task:** Group the data by *Company* and calculate the total *Volume Traded* for each company.
5. **Matplotlib Task:** Plot a line graph showing the trend of the *Close Price* over time for a particular company.
6. **Matplotlib Task:** Create a bar chart to compare the average percentage change in *Close Price* for different companies.