

Numpy & Matplotlib exercises

In the following exercises, no dataset is given. Generate the data as you please (random, predefined) or load it from some dataset the suits you. At the bottom, a list of data resources is provided in case you want

Numpy

1. Create a NumPy array with 10 random integers and print the array to the console.
2. Find the maximum value in a NumPy array with 50 random floats.
3. Use NumPy to generate a 3x3 identity matrix and print it to the console.
4. Calculate the mean and standard deviation of a NumPy array with 100 random integers.
5. Create a 2D NumPy array with 5 rows and 5 columns, and then extract the third row and the fourth column.
6. Use NumPy to create a 3x3 matrix with values ranging from 1 to 9, and then print the matrix to the console.
7. Calculate the dot product of two NumPy arrays of length 5.
8. Reshape a 1D NumPy array with 10 elements into a 2D array with 2 rows and 5 columns.
9. Use NumPy to calculate the eigenvalues and eigenvectors of a 3x3 matrix.
10. Create a new NumPy array that is the result of concatenating two existing arrays with different shapes.

Matplotlib

1. Create a line plot with 50 data points that shows the relationship between two variables, x and y.
2. Generate a scatter plot with 100 data points that shows the relationship between two variables, x and y.
3. Create a bar chart that shows the frequency of occurrence for different categories.
4. Generate a histogram that shows the distribution of a numerical variable.
5. Create a pie chart that shows the distribution of a categorical variable.
6. Use Matplotlib to create a stacked bar chart that compares the values of multiple variables across different categories.
7. Generate a box plot that shows the distribution of a numerical variable, including the median, quartiles, and outliers.
8. Create a heatmap that shows the relationship between two variables using a color

scale.

9. Use Matplotlib to create a time series plot that shows the trend in a variable over time.
10. Create a 3D scatter plot that shows the relationship between three variables, x, y, and z.

Ancillary datasets

- OutWorldInData: This is a collection of datasets covering many demographic and natural events worldwide, most of the datasets provide historical (time-series) and geographic information. You can access the datasets at <https://ourworldindata.org/>.
- UCI Machine Learning Repository: This repository contains a large collection of datasets, including the Iris dataset, Wine dataset, and Breast Cancer Wisconsin dataset. You can access the datasets at <https://archive.ics.uci.edu/ml/index.php>.
- Kaggle: This is a popular platform for data science competitions, and many of the datasets used in these competitions are available for download. You can access the datasets at <https://www.kaggle.com/datasets>.
- Scikit-learn: This is a popular machine learning library in Python, and it contains some built-in datasets that you can use for practice. You can access these datasets using the `load_` functions provided in the library. More information on these datasets can be found at <https://scikit-learn.org/stable/datasets/index.html>.
- Google Dataset Search: This is a search engine that allows you to find datasets from a variety of sources, including academic institutions, government agencies, and other organizations. You can access the search engine at <https://datasetsearch.research.google.com/>.