Genztechs

Data Science Internship Program

Assignment 2

Learn Matplotlib, Pandas and Numpy Libraries of Python and Implement the given tasks and project.

Resources for Matlplotlib:

Youtube Tutorial in English: https://youtu.be/vBCXsAd_swk?si=SAliX-6ylljoeT57

Youtube Tutorial: https://youtu.be/9GvnrQv138s?si=f2FxaEGl7BjMr1-1

Youtube Tutorial: https://youtu.be/3Xc3CA655Y4?si=VLlomerFKSF0KYFu

Website Tutorial: https://www.tutorialspoint.com/matplotlib/index.htm

WebsiteTutorial: https://www.geeksforgeeks.org/matplotlib-tutorial/

CheatSheets: https://matplotlib.org/cheatsheets/

Resources for Pandas:

Youtube Tutorial: https://youtu.be/RhEjmHeDNoA?si=YebXTNWdtq50tXm7

Youtube Tutorial: https://youtu.be/ZyhVh-qRZPA?si=che_S_xxJqEbkrZn

Youtube Tutorial: https://youtu.be/kQQaO5Cm5AI?si=UYcVOmIySRYz34FT

Youtube Tutorial in English: https://youtu.be/2uvysYbKdjM?si=aAFq4XRhQE8I5E98

WebsiteTutorial: https://www.w3schools.com/python/pandas/default.asp

WebsiteTutorial: https://www.geeksforgeeks.org/pandas-tutorial/

WebsiteTutorial: https://www.tutorialspoint.com/python_pandas/index.htm

WebsiteTutorial: https://www.kaggle.com/learn/pandas

Cheatsheet:

https://images.datacamp.com/image/upload/v1676302204/Marketing/Blog/Pandas_Cheat_Sheet.pdf

Resources for Numpy:

Youtube Tutorial in English: https://youtu.be/QUT1VHiLmmI?si=9RRnXaNqHe1Mesq7

Youtube Tutorial: https://youtu.be/Rbh1rieb3zc?si=4YFlywbTyGoQoCYv

Youtube Tutorial: https://youtu.be/ZaKzw9tULeM?si=cjQXOpItol89Q_1Q

Youtube Tutorial: https://youtu.be/awP79Yb3NaU?si=Qsum2pbV6akWx9rh

WebsiteTutorial: https://www.w3schools.com/python/numpy/default.asp

WebsiteTutorial: https://www.geeksforgeeks.org/numpy-tutorial/

WebsiteTutorial: https://www.tutorialspoint.com/numpy/index.htm

Cheatsheet:
https://images.datacamp.com/image/upload/v1676302459/Marketing/Blog/Numpy_Cheat_Sheet.pdf

Tasks:

1. Create a line plot for a mathematical function like y=x2y = x^2y=x2 for values of xxx ranging from -10 to 10. Add labels, a title, and a grid.

2. Visualize the population of five cities using a bar chart. Include different colors for each bar and add appropriate labels.

3. Generate random data points and create a scatter plot. Differentiate points based on their value (e.g., by color or size).

4. Display the percentage distribution of expenses (e.g., rent, groceries, entertainment) in a month using a pie chart with custom labels and a legend.

5. Create a figure with 2x2 subplots showcasing different types of plots (e.g., line, bar, scatter, and histogram).

6. Load a CSV file containing sales data. Perform operations like displaying the first 5 rows, column names, and statistical summaries.

7. Handle missing values in a dataset by either filling them with a specific value or dropping rows/columns.

8. Filter rows in a dataset where a particular column (e.g., "sales") is greater than a certain threshold.

9. Using a dataset, group the data by a categorical column (e.g., "region") and calculate the mean of numerical columns for each group.

10. Using Pandas and Matplotlib, create a time-series plot of stock prices from a dataset.

11. Create arrays of zeros, ones, and random numbers. Perform operations like reshaping, slicing, and indexing.

12. Perform matrix multiplication between two 2D arrays and find the determinant and inverse of a square matrix.

13. Generate a large array of random numbers and compute mean, median, standard deviation, and variance.

14. Perform operations like adding a scalar to an array and adding two arrays with different shapes using broadcasting.

15. Write a function using NumPy to calculate f(x) = sin(x2) + cos(x)f(x) = sin(x^2) + cos(x)f(x) = sin(x2) + cos(x) and apply it to an array of values.

Mini-Project: Sales Data Analysis and Visualization

Objective: Analyze sales data for a company and visualize trends.

1. Load a CSV file containing columns like `Date`, `Product`, `Region`, and `Sales`.
2. Handle missing values and remove any duplicate entries.
3. Calculate statistics like total sales, average sales, and sales variance.
4. Extract sales data for a specific product and region for further analysis.
    - Create a bar chart showing total sales per product.
    - Create a line chart showing sales trends over time.
    - Create a pie chart showing sales distribution across regions.
5. Save cleaned data and visualizations as separate files for reporting.