

TASK:

Suppose you have a dataset with multiple variables that may be related to each other.

How can you use Python data visualization libraries to explore potential relationships and identify patterns between variables? Provide at least two creative ideas for data visualization techniques that can be used to explore relationships between variables in a dataset.

Exploring the relationships between variables in a dataset is essential for understanding underlying trends and making informed decisions. Python data visualization libraries such as Matplotlib, Seaborn, and Plotly provide powerful tools for effectively visualizing these relationships.

1. One creative idea for data visualization is to use a scatterplot or pair plot.

These techniques allow us to visualize pairwise relationships between several variables in a single graph. This helps to quickly identify patterns, correlations and potential outliers.

This study examines the factors influencing students' academic performance. The dataset consists of 10,000 student records, each record containing information on various predictors and a performance index.

Variables:

- Hours Studied: The total number of hours spent studying by each student.
- Previous Scores: The scores obtained by students in previous tests.
- Extracurricular Activities: Whether the student participates in extracurricular activities (Yes or No).
- Sleep Hours: The average number of hours of sleep the student had per day.
- Sample Question Papers Practiced: The number of sample question papers the student practiced.

1. line plot and box plot

- A line plot can be used to visualize the trend of several variables over time, or of another continuous variable. By plotting each variable on the same graph with different colored lines, we can observe how they change in relation to each other.
- A box plot can be used to compare the distribution of several variables or groups within a data set. By plotting the distribution of each variable side by side, we can visually compare their central tendency, variability and skewness.

The dataset provides information on poverty-level wages in the United States from 1973 to 2022.

It includes data on annual and hourly poverty-level wages, as well as wage shares for different income brackets.

The dataset is based on the Economic Policy Institute's State of Working America data library, which offers comprehensive economic data for analyzing labor market trends and patterns.

data set1: <https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression>

data set2: <https://www.kaggle.com/datasets/asaniczka/poverty-level-wages-in-the-usa-dataset-1973-2022>

References:

- Datacamp. (2023). *11 Data Visualization Techniques for Every Use-Case with Examples*. <https://www.datacamp.com/blog/data-visualization-techniques>
- Jack, D & Ilya, I. (2024). *Hands-On Data Visualization*. <https://handsondataviz.org/>
- Practical Data Science. (2021). *How to visualise correlations using pandas and Seaborn*. <https://practicaldatascience.co.uk/data-science/how-to-visualise-correlations-using-pandas-and-seaborn>
- Will, K.(2018). *Visualizing Data with Pairs Plots in Python*. Towardsdatascience. <https://towardsdatascience.com/visualizing-data-with-pair-plots-in-python-f228cf529166>