Project Title: Exploratory Data Analysis Using Python Libraries

Project Description:

In this project, you will perform exploratory data analysis (EDA) on a dataset of your choice using Python libraries such as NumPy, Pandas, Matplotlib, and Seaborn. EDA is a critical step in data analysis and involves cleaning, preprocessing, visualizing, and summarizing data to gain insights and discover patterns. Through this project, you will learn how to apply Python libraries to real-world datasets and develop the skills needed for data analysis.

Tasks:

1. Select a dataset of your choice (e.g., from Kaggle, UCI Machine Learning Repository, or other sources) and clean and preprocess the data.
2. Load the dataset into a Pandas DataFrame.
3. Perform exploratory data analysis to get an understanding of the data using NumPy and Pandas.
4. Visualize the data using Matplotlib and Seaborn to identify trends, patterns, and relationships among variables.
5. Perform data preprocessing tasks such as data cleaning, handling missing values, and data transformation.
6. Draw conclusions and insights from the data and communicate your findings through a report or a presentation.
7.

Deliverables:

A Jupyter Notebook containing the Python code and the results of the analyses.
A report or a presentation summarizing the findings and insights from the data.

Note: You are free to choose any dataset, and you are encouraged to be creative with your analysis. You may also choose to work in teams of two, and each team member should contribute to the analysis.

References:

1. Exploratory Data Analysis in Python: An Introduction - https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15
2. Step-by-Step Exploratory Data Analysis (EDA) using Python - https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/
3. Pandas documentation - https://pandas.pydata.org/docs/
4. NumPy documentation - https://numpy.org/doc/
5. Matplotlib documentation - https://matplotlib.org/3.3.3/contents.html
6. Seaborn documentation - https://seaborn.pydata.org/tutorial.html

Finding Datasets:

1. Kaggle - https://www.kaggle.com/datasets
2. UCI Machine Learning Repository - https://archive.ics.uci.edu/ml/index.php
3. Data.gov - https://www.data.gov/
4. Google Dataset Search - https://datasetsearch.research.google.com/