Task 1: Exploratory Data Analysis on a Dataset

Objective: Perform exploratory data analysis (EDA) on a given dataset to understand its structure and basic statistics.

Steps:

1. Load the Dataset: Use Pandas to load a CSV file into a DataFrame.

2. Data Cleaning: Check for missing values, duplicates, and data types. Handle missing values and duplicates appropriately.

3. Summary Statistics: Use Pandas and NumPy to compute summary statistics (mean, median, mode, standard deviation) for numerical columns.

4. Visualization: Use Seaborn to create visualizations such as histograms, box plots, and pair plots to explore the distribution and relationships between variables.

Example Dataset: Titanic dataset.

Task 2: Time Series Analysis

Objective: Analyze a time series dataset to identify trends, seasonality, and patterns.

Steps:

1. Load the Dataset: Use Pandas to load a time series dataset.

2. Date Handling: Ensure the date column is in datetime format and set it as the index.

3. Resampling and Aggregation: Use Pandas to resample the data (e.g., daily to monthly) and calculate aggregate metrics.

4. Trend Analysis: Use NumPy to calculate rolling means and standard deviations.

5. Visualization: Use Seaborn to create line plots, heatmaps, and seasonal plots.

Example Dataset: Stock prices.

Task 3: Data Manipulation and Feature Engineering

Objective: Perform data manipulation and create new features from an existing dataset.

Steps:

1. Load the Dataset: Use Pandas to load a dataset.

2. Data Transformation: Use Pandas and NumPy to transform columns (e.g., log transformation, normalization).

3. Feature Engineering: Create new features based on existing columns (e.g., extracting year/month from a date, creating interaction terms).

4. Grouping and Aggregation: Use Pandas to group data by a categorical variable and compute aggregated statistics.

5. Visualization: Use Seaborn to visualize the new features and their relationships with other variables.

Example Dataset: Housing prices,

Task 4: Statistical Analysis and Hypothesis Testing

Objective: Perform statistical analysis and hypothesis testing on a dataset to draw conclusions.

Steps:

1. Load the Dataset: Use Pandas to load a dataset.

2. Descriptive Statistics: Use Pandas and NumPy to compute descriptive statistics.

3. Hypothesis Testing: Perform t-tests, chi-square tests, and ANOVA using relevant libraries.

4. Correlation Analysis: Use Seaborn to create heatmaps and pair plots to visualize correlations between variables.

5. Visualization: Use Seaborn to create visualizations that support the statistical analysis (e.g., bar plots, violin plots).

Example Dataset: Customer satisfaction survey data,

Task 5: Machine Learning Preparation and Visualization

Objective: Prepare a dataset for machine learning and visualize key aspects of the data.

Steps:

1. Load the Dataset: Use Pandas to load a dataset.

2. Data Preprocessing: Handle missing values, encode categorical variables, and scale numerical features using Pandas and NumPy.

3. Feature Selection: Use correlation analysis and feature importance metrics to select relevant features.

4. Split the Data: Split the data into training and testing sets.

5. Visualization: Use Seaborn to visualize the feature distributions, relationships between features, and target variable distributions.

Example Dataset: loan default dataset.