Project focusing on data cleaning and preprocessing, as well as data exploration and visualization:

Project Title: Customer Churn Analysis

Project Overview:
In this project, you will work with a real-world dataset from a telecom company that includes customer information, such as demographics, usage patterns, and whether they churned (i.e., left the company) or not. The goal is to prepare the data for analysis, explore key patterns and trends, and create visualizations to gain insights into factors influencing customer churn.

Project Objectives:
1. Clean and preprocess the raw dataset to make it suitable for analysis.
2. Explore the dataset to understand the distribution of key variables.
3. Identify factors that may be associated with customer churn.
4. Create informative visualizations to communicate findings effectively.

Project Tasks:

Data Cleaning and Preprocessing (Days 1-2):
1. Load the raw dataset (e.g., CSV format) into a data analysis environment (Python with Pandas).
2. Examine the data for missing values, duplicates, and outliers.
3. Handle missing data through techniques like imputation (e.g., mean, median) or removal of rows/columns.
4. Address duplicates, if any, and remove unnecessary columns.
5. Encode categorical variables (e.g., one-hot encoding) for modeling purposes.
6. Normalize or scale numeric features if needed.

Data Exploration (Days 3-4):
7. Perform descriptive statistics to summarize the dataset (e.g., mean, median, standard deviation).
8. Explore the distribution of key variables, such as customer demographics (age, gender), usage patterns (monthly charges, total charges), and churn status.
9. Calculate and visualize correlations between variables (e.g., using a heatmap).
10. Conduct hypothesis tests (e.g., t-tests) to compare differences between churned and non-churned customers.
11. Create summary tables and statistics for various customer segments (e.g., by contract type, payment method).

Data Visualization (Days 5-6):

12. Generate insightful visualizations to communicate findings, including:
    - Bar plots and histograms to show distribution of categorical and numeric variables.
    - Box plots to identify outliers.
    - Scatter plots to explore relationships between numeric variables.
    - Pie charts to visualize the proportion of churned vs. non-churned customers.
    - Time series plots (if applicable) to observe trends over time.
13. Use Seaborn, Matplotlib, or other visualization libraries to enhance the aesthetics and clarity of plots.
14. Provide captions and labels for each visualization to make them self-explanatory.
15. Summarize key insights and findings from the visualizations.

Project Deliverables:

1. A clean and well-preprocessed dataset ready for analysis.
2. An exploratory data analysis (EDA) report or Jupyter Notebook documenting the data cleaning steps, exploratory analysis, and visualizations.
3. Interactive visualizations (if possible) that allow users to explore the data.
4. A presentation summarizing the project's findings, including insights into factors influencing customer churn.

Project Evaluation:

You will be evaluated based on the quality of your cleaned dataset, the depth of your exploratory analysis, and the effectiveness of your visualizations in communicating insights. Additionally, your ability to present and explain your findings will be assessed.

This project not only reinforces the skills of data cleaning, preprocessing, and exploratory data analysis but also provides valuable insights for business decision-making regarding customer churn mitigation strategies.