Project Instructions — Stage 08: Exploratory Data Analysis (EDA)

Today's Project Contribution: Today you'll complete a piece of your full data project. This task aligns with the **Exploratory Data Analysis** stage, where you will: • Understand the dataset's structure, distributions, and relationships between variables. • Add to your existing project repo (or update prior files).

By the end of this assignment, your project should include the elements listed below.

Deliverable Options Required:

- Notebook containing visualizations (e.g., histograms, scatter plots, boxplots, time series charts showing data patterns and outliers).
- Statistical summaries (e.g., mean, median, standard deviation, min/max, missing value counts, skewness describing data).
- Save EDA notebooks in `/notebooks/` folder to maintain a clean project layout.

Optional Formats (Choose One or more):

- Correlation matrix (show variables relationship).
- Additional distribution plots (show variable behavior).

How This Fits Into Your Final Project

Your work today builds toward a complete, end-to-end project. The EDA notebook becomes a **reference** for feature engineering and modeling decisions.

Before next class: • Save your files in the appropriate folders (`/notebooks/`). • Commit and push your changes to your GitHub repo. • Review any assumptions, risks, or notes — these will carry across your stages.

Explicit Chain *In your homework, you produced a clean EDA notebook with insights. Now, you will adapt those insights to propose feature ideas and justify preprocessing steps in the next stage.* '''