Homework — Stage 08: Exploratory Data Analysis (EDA)

Assignment

Create a comprehensive **EDA notebook** for your project dataset.

Requirements

- Statistical summaries: `.info()`, `.describe()`, missing counts
- Distributional plots: histograms/boxplots for key variables
- Bivariate visuals: scatter/line for at least 2 relationships
- (Optional) Correlation heatmap
- Clear markdown documenting **findings, risks, and assumptions**

Step-by-Step

1. Load your dataset (or adapt the synthetic generator from lecture to mimic your schema).
2. Profile numeric and categorical columns.
3. Plot at least 3 distributions and 2 relationships.
4. Note any skew, outliers, seasonality, or structure.
5. Conclude with **"Implications for next step"**: what you'll engineer or clean.

Grading Rubric (100 pts)

- (30) Completeness of numeric profiling
- (30) Quality and clarity of visuals
- (30) Insightfulness of interpretation (not just "what," but "so what")
- (10) Organization and readability

Example Submission Expectations

- A Jupyter notebook under `/notebooks/` named `eda_<team>.ipynb`
- Clean titles/labels; brief markdown per plot
- Final section: "Top 3 insights" + "Assumptions & risks"

Chain Statement

In the lecture, we learned how to profile and visualize data to surface patterns. Now, you will adapt those patterns to your own dataset and document what they imply for preprocessing and feature engineering.