Excellent — let's build you a **structured practice plan** that takes you from the basics of statistics all the way to advanced applications, using your dataset as the playground. Think of this as a checklist you can work through step by step, each stage reinforcing your analytical mindset.

📊 Structured Practice Plan for Your Dataset

Stage 1: Foundations — Descriptive Statistics

Goal: Get comfortable summarizing and describing data.

- Calculate **mean, median, mode** for exam scores, AI usage frequency, and study time.
- Compute **variance, standard deviation, range, IQR** for exam scores.
- Create **frequency tables** for categorical variables (e.g., gender, AI usage categories).
- Identify **skewness/kurtosis** of exam scores distribution.

👉 *Outcome:* You'll know how the dataset "looks" in terms of averages and spread.

Stage 2: Visualization — Seeing the Data

Goal: Translate numbers into pictures.

- Plot **histograms** of exam scores and study time.
- Draw **boxplots** comparing exam scores across AI usage groups.
- Create **scatterplots** of study time vs. exam scores.
- Use **bar charts** for categorical comparisons (e.g., engagement levels).

👉 *Outcome:* You'll visually spot trends, outliers, and group differences.

Stage 3: Relationships — Correlation & Association

Goal: Explore how variables relate to each other.

- Compute **Pearson correlation** between study time and exam scores.
- Use **Spearman correlation** for ordinal variables (e.g., engagement levels).
- Run a **Chi-Square test** for categorical associations (e.g., gender vs. AI usage frequency).

👉 *Outcome:* You'll see which variables move together and which are independent.

Stage 4: Inference — Hypothesis Testing

Goal: Move from description to testing ideas.

- Perform **t-tests**: Compare exam scores of frequent vs. infrequent AI users.
- Run **ANOVA**: Compare exam scores across multiple AI usage categories.
- Conduct **Chi-Square independence tests**: Is AI usage independent of demographics?
- Frame hypotheses (H_0 vs. H_1) and interpret p-values.

👉 *Outcome:* You'll learn to confirm or reject assumptions with evidence.

Stage 5: Prediction — Regression Analysis

Goal: Build models to predict outcomes.

- **Simple Linear Regression:** Predict exam score from study time.
- **Multiple Regression:** Predict exam score from study time, AI usage, engagement, demographics.
- **Logistic Regression:** Predict categorical outcomes (e.g., pass/fail).

👉 *Outcome:* You'll see how multiple factors combine to explain performance.

Stage 6: Advanced Exploration

Goal: Apply more complex statistical concepts.

- **Clustering:** Group students by AI usage patterns.
- **Trend Analysis:** Study if more AI usage consistently improves performance.
- **Interaction Effects:** Does AI usage benefit some demographics more than others?
- **Causal Inference (optional):** Explore whether AI usage *causes* improvement, not just correlates.

👉 *Outcome:* You'll practice higher-level thinking about causality and grouping.

Stage 7: Interpretation & Reporting

Goal: Turn analysis into insights.

- Summarize findings in plain language.
- Highlight practical implications (e.g., "Students who use AI tools daily improve scores by X%").
- Reflect on limitations (e.g., self-reported data, sample size).
- Suggest actionable recommendations for educators.

👉 *Outcome:* You'll learn to communicate results clearly — the most important skill in data science.

📝 **Suggested Workflow**

1. Start with **Stage 1 & 2** (Descriptive + Visualization).
2. Move to **Stage 3 & 4** (Relationships + Hypothesis Testing).
3. Progress to **Stage 5 & 6** (Regression + Advanced).
4. End with **Stage 7** (Interpretation).

Would you like me to **expand this into a week-by-week practice schedule** (e.g., Week 1: Descriptive stats, Week 2: Visualization, etc.) so you can pace yourself and build confidence gradually?