DIY Master's in Data Science (AI Era) — Study Session 2 Checklist

Today's goal: Build your visual intuition and storytelling skills using matplotlib, plotly, and Markdown narration.

■ PREP (10–15 min)

- ■ Activate environment: conda activate ds
- ■ Launch Jupyter Lab → jupyter lab
- ■ Open notebook: ds-zero-to-one/notebooks/02_visual_exploration.ipynb
- ■ Load dataset: df = pd.read_csv('../data/processed/tips_cleaned.csv')

■ PART 1 — Quick Visual Checks (20–30 min)

- Plot a histogram: df['tip_pct'].hist(bins=20)
- Create boxplots by gender: df.boxplot(column='tip_pct', by='gender')
- Plot scatter: df.plot.scatter(x='bill_total_usd', y='tip_usd', alpha=0.7)
- Observe distributions and patterns in tipping behavior

■ PART 2 — Interactive Plotly Views (30–45 min)

- Import Plotly Express: import plotly.express as px
- Scatter with trendline: px.scatter(df, x='bill_total_usd', y='tip_usd', color='gender', trendline='ols')
- Grouped boxplot: px.box(df, x='weekday', y='tip_pct', color='is_smoker')
- Faceted scatterplots: px.scatter(df, x='bill_total_usd', y='tip_pct', facet_col='weekday', color='gender')
- Write Markdown notes below each plot describing observed trends

✍■ PART 3 — Story Cells & Markdown Narration (20 min)

- Add section header: ## Visual Insights & Mini-Story
- Summarize three strongest insights (3 sentences each)
- Use bullets, italics, and inline code for emphasis
- Example: - **Big spenders tip more**, but variance ↑ after $30.

■ PART 4 — Commit Your Work (10 min)

- git add notebooks/02_visual_exploration.ipynb
- git commit -m 'Session 2: visualization and storytelling practice'
- git push
- Confirm GitHub Actions → CI → check 'CI env OK'

■■■■ REFLECT (Optional 5 min)

- Which chart type best communicated the data story?
- What was your biggest visualization mistake?
- How would you explain your findings to a non-technical friend?

■ Tip: Pair visuals with brief Markdown notes — practice explaining what the chart shows *and why it matters*.