Linear Regression Observation Sheet — Step-by-Step Fill Order

(Aligned to the One-Pager sections)

1. **Dataset (Section A)**
   - Dataset name, source, version/date, rows, columns
   - Target variable, train/test split, random seed
2. **Preprocessing (Section B)**
   - Standardize column names (Y/N)
   - Remove duplicates and record count
   - List columns with missing values
   - Choose imputation: numeric (mean/median), categorical (most_frequent)
   - Choose encoding: None / OHE / Ordinal
   - Choose scaling: None / Standard / MinMax
   - Outlier handling (None / IQR / Manual + which columns)
   - Feature selection or dropping (list and reason)
   - Leakage checks performed (Y/N + note)
3. **Feature Inventory (Section C) — quick pass**
   - For key features: type (num/cat), missing%, impute, encoding, scaling
   - Include? (Y/N) and short notes
4. **EDA — Univariate (Section C1)**
   - Record histogram observations: shape/skew, outliers, data quality notes
5. **EDA — Multivariate (Section C2)**
   - Correlations / scatter plots with target
   - Note features most related to the target
6. **Linear Regression — Setup (Section D, first row)**
   - List features you will start with (from Section C)
   - Brief comment on why these features
7. **Linear Regression — Metrics (Section D, second row)**
   - Record MAE, MSE, RMSE, R^2 (test), Train R^2, Train–Test gap
   - Short comment on fit quality
8. **Linear Regression — Interpretability (Section D, third row)**
   - Top coefficients: sign (+/−), magnitude
   - Plain-English interpretation in context
9. **Multicollinearity & Assumptions (Section E)**
   - Top VIFs (1–3)
   - Residual checks: Shapiro p, Durbin–Watson, Breusch–Pagan p
   - Overall: Pass / Needs Work; actions taken (drop/transform/etc.)
10. **Tuning F1 — Polynomial Features (Section F1)**
    - Choose strong feature; try degree range (e.g., 1–5)

- Best degree (CV) and best CV R^2; over/underfit note

11. **Tuning F2 — Encoding Impact (Section F2)**
    - Apply chosen encoding to full model
    - Record R^2 before/after; mark if it helped (Y/N) and why
12. **Tuning F3 — Regularization (Section F3)**
    - Try Ridge and/or Lasso with small α grid
    - Record type, alpha grid, best α, R^2 (test)
    - For Lasso: number of non-zero coefficients
13. **Tuning F4 — Cross-Validation / Learning Curve (Section F4)**
    - K-fold CV: R^2 (mean/std), RMSE (mean/std)
    - Bias/variance note from learning curve
14. **Final Summary & Reflection (Section G)**
    - Chosen final model (features + any poly/encoding/regularization)
    - Top 3 drivers (features)
    - Final test metrics (MAE/RMSE/R^2)
    - Limitations/ethics; next steps (data or features to add)

Tip: If you change preprocessing later (e.g., drop a feature), update Section C and add a note in Section B.