**📝 EDA Notes — Banking Fraud Project**

**Purpose**

This document summarizes the exploratory data analysis (EDA) workflow added in this PR. It captures the intent of the scripts and notebooks, and provides reviewers/assignees with guidance on interpreting the outputs.

**Added Files**

* scripts/eda\_banking\_project\_data\_ingestion.py
* notebooks/EDA\_Banking\_Project\_Data\_Ingestion.ipynb
* Generated outputs: plots, summary tables (CSV), and logs.

**Key EDA Steps**

1. **Data Ingestion**
   * Reads raw banking transaction data.
   * Handles missing values, dtypes, and basic cleaning.
   * Produces overview tables (overview.csv, missing\_values.csv, etc.).
2. **Fraud Distribution**
   * Computes fraud vs. non-fraud counts.
   * Saves bar plots and distribution charts.
3. **Transaction Patterns**
   * Boxplots of amounts by fraud label.
   * Visualizes categorical distributions (type of transaction, customer vs merchant).
4. **Missing Data**
   * Quantifies nulls per column.
   * Outputs CSV + plot for reviewer interpretation.
5. **Numeric Summaries**
   * describe\_numeric.csv with basic stats.
   * Outlier inspection via log/boxplots.

**Reviewer Guidance**

* Please **run the EDA scripts/notebooks** and confirm that:
  + Plots are generated correctly in plots/EDA/.
  + Summary tables are stored in tables/EDA/.
  + Fraud ratios align with expectations from the dataset.
* After verification, reviewers should add **written interpretations**:
  + What transaction types dominate fraud?
  + Are fraud transactions higher in amount than non-fraud?
  + Do missing values appear systematic (e.g., only in fraud cases)?
  + Any anomalies in distribution (long tails, heavy skew)?
* These interpretations will feed into the **final report** and support the **feature engineering strategy**.

**Next Steps**

* Once reviewers confirm EDA consistency, we will:
  1. Consolidate findings into docs/EDA\_Report.md.
  2. Link outputs into the modeling pipeline (src/capstone/features.py).