■ MOOC Dropout Prediction — Visualization Insights (OULAD Dataset)

Figure 1 — Outcome Distribution

Distribution of student outcomes (Pass / Fail / Withdrawn) in the MOOC.

• Shows overall retention vs dropout at a glance.

• Large Withdrawn slice → high dropout prevalence.

Insight: Prioritize retention strategies (onboarding, progress nudges) where Withdrawn proportion is largest.

Figure 2 — Dropout by Gender

Final outcomes grouped by gender.

• Compares pass/withdraw rates between genders.

• Any imbalance suggests gender-specific engagement differences.

Insight: Consider targeted support or tailored communication for the gender group with higher withdrawals.

Figure 3 — Dropout by Age Band

Dropout and completion patterns across age bands.

• Identifies which age groups withdraw most.

• Younger cohorts often show higher withdrawal; older cohorts may persist.

Insight: Design age-specific interventions (micro-modules for younger learners; flexible pacing for working adults).

Figure 4 — Studied Credits vs Final Result

Studied credits (engagement) distribution by final result (box plot).

• Low studied_credits cluster with Withdrawn/Fail groups.

• Higher credits correlate with better outcomes.

Insight: Use early-credit thresholds to flag at-risk learners for proactive outreach.

Figure 5 — Dropout by Region

Regional comparison of dropout vs completion (top 10 regions).

• Geographic disparities in retention are visible.

• Regions with high withdrawal rates may have access/support issues.

Insight: Investigate region-specific causes (connectivity, time zones, language) and localize interventions.

Figure 6 — Correlation Heatmap

Correlation matrix of numeric features (engagement, attempts, scores).

• Highlights strong positive/negative relationships among predictors.

• Identifies candidate features for modeling (e.g., studied_credits ↔ score).

Insight: Prioritize highly correlated features for model building and feature engineering.

Figure 7 — Feature Importance (Model Insight)

Top features driving dropout prediction (linear model coefficients).

• Lists strongest predictors and their relative influence.

• Direction (sign) indicates whether the feature increases or reduces dropout risk.

Insight: Target interventions on top positive predictors (those that increase dropout risk).

Figure 8 — Actual vs Predicted

Actual vs predicted outcomes — evaluates model accuracy.

• Points near diagonal = accurate predictions; large scatter = errors.

• Reveals bias or variance issues visually.
Insight: Use this diagnostic to refine the model (add features, reduce bias/variance, or use non-linear models).