HELP Dataset Glossary (Round 2)

This dataset is a simplified snapshot of HELP's catastrophe-exposed student housing portfolio.

- Each **row** represents one **coverage line** for one insured.
- Each **student_id** can therefore appear several times (once per coverage).

Identifier and Demographics

Student_id

- Unique identifier for each insured; treat this as the **policy or risk ID**.

Name

- Name label for the insured, useful for readability only, not required for analysis.

Class

- Freshman, Sophomore, Junior, Senior.
- Use this as a **risk grouping variable** (for example, different customer segments).

Study

- Business, Humanities, Science, Other.
- Another **categorical segmentation variable**, similar to property or business type.

Gpa

- Numeric attribute that varies by insured.
- Use this as a **proxy risk score** or stability indicator.

Greek

- Greek or Non-greek.
- Represents a **subtype of residence** or special housing classification.

Gender

- Male/Female/Other. It acts as a **grouping variable** rather than a literal rating factor.

Location and Protection

Off_campus

- On campus or Off campus.
- Interpret as **location type**, analogous to on-premises versus off-premises risks.

Distance_to_

- Numeric distance measure for off-campus risks.
- Use this as a **geographic hazard proxy**, such as distance to key services or higher-risk zones.
- (You can create distance bands to approximate "regions".)

Sprinklered

- TRUE or FALSE.
- Indicates whether the residence has **sprinkler protection**.

Coverage and Claims

Coverage

- Coverage line for that row. Typical values include:
    - Additional Living Expense
    - Guest Medical
    - Liability
    - Personal Property
- Treat these as standard commercial property coverages (ALE, medical payments, liability, contents).

Claim_id

- Identifier of a claim on that coverage.
    - 0 = **no claim** for that coverage in the period
    - Non-zero = **a claim occurred**

Amount

- Loss amount for the claim on that coverage.
- Use this for **severity** and **loss cost** analysis.

Underwriting and Modeling

Risk_tier

- Integer tier such as 1, 2, or 3.
- Represents HELP's **internal underwriting or pricing tier**.

Holdout

- TRUE or FALSE.
- Indicates whether the observation belongs to a **validation holdout set**.
- Only needed if you choose to build a predictive model.

Note on repeated student_id values

If you see the same **student_id** multiple times, that insured simply has multiple coverages (for example, Liability, Personal Property, Additional Living Expense, Guest Medical). This structure mirrors how real commercial property portfolios are stored and allows you to analyze patterns by coverage type.