# CYCLE 1: Burnout Busters → Clean Data Champions

5 Days · 5 Tasks · 1 Mini Challenge · 1 Main Submission

### Case File #001 – Data cleaning and eda (Day 1)

**Syllabus Topics:**

* Missing values, outlier handling
* Pandas for EDA & plots

**Tasks:**

* Handle missing values
* Fix outliers (e.g., >18 hr workdays)
* Visualize relationships: stress vs sleep, mental health vs burnout

Deliverable: Cleaned dataset + 2–3 visualizations with insights

### Case File #002 – Feature Forge + Regression (Day 2)

**Syllabus:**

* Encoding (one-hot, ordinal), scaling, normalization
* Feature selection (correlation, mutual info)
* Linear, Ridge, Lasso Regression
* Evaluation metrics - MSE, R² metrics

Focus: Feature Engineering

**Raw data isn’t enough. Forge better signals for stronger predictions.**

**Task:**

*Regression Mini task*

* Encode categorical columns (one-hot or ordinal as appropriate)
* Normalize numerical features (e.g., MinMaxScaler, StandardScaler)
* Select features using correlation/mutual information
* Create 2 interaction features (e.g., stress × work\_hours, Sleep vs Stress)
* Train Linear, Ridge, and Lasso regression using Stress\_level as a proxy target

**Deliverables:**

* Encoded + scaled data
* Feature selection logic
* Model training + MSE and R² scores
* Short summary: Which model performed best and why?

### Case File #003 – Classifier Arena (Day 3)

**Objective:**

Build classifiers to predict **burnout (0/1)** and evaluate them using proper metrics.

**Syllabus:**

* Logistic Regression
* Linear discriminant analysis
* Classification metrics: Accuracy, ROC-AUC

**Burnout prediction under pressure. Train, compare, and evaluate**

**Tasks:**

Train:

* Logistic Regression
* Linear Discriminant Analysis (LDA)

Evaluate with:

### Accuracy

### Confusion Matrix

### ROC-AUC (plot ROC curve)

### Deliverables

### Model code, performance metrics, ROC plot

### Summary: Which model performed better and why?

### Case File #004 – Tree-Based Models + k-NN + Feature Selection (Day 4)

**Objective:**

Train and compare advanced classifiers. Select top 3 important features and attempt a **3-feature model challenge**.

**Syllabus:**

* Decision Trees (structure, criteria)
* Random Forests (feature importance, hyperparams)
* k-NN
* Drop weak features using importance/Mutual info

Tasks:

* Train: Decision tree, random forest, knn
* Use correlation, mutual info, or Random Forest importances
* Drop weak features
* Retrain and compare results

Deliverable: Before vs After model comparison

### Case File #005 – 3-Feature Showdown (Day 5 – Mini Challenge) (Bonus Task)

Focus: Compact Burnout Prediction

**Minimal model. Maximum accuracy.**

Tasks:

* Select 3 features using feature importance + EDA
* Train a minimal model
* Justify your selection

Deliverable: 3 chosen features, model accuracy, short reasoning

Justification: Why these 3 features?