

CravesByIb

Ai-powered Meal Recommendation & Automation Platform



Day 3 Task

Title: Recipe Feature Engineering Pipeline

Context (Give This to Students)

CravesByIB is an AI-powered food recommendation and meal-planning system. Before any AI models or GenAI logic can be applied, the system requires a **clean, engineered feature layer** derived from raw food data.

Today, you will build the **data backbone** of CravesByIB.

Objective

Using the provided dataset, engineer a new dataset called:

recipe_features.csv

This file will be used in later weeks for:

- Cost prediction
 - Nutrition analysis
 - AI-driven meal recommendations
-

What You Are Given

You are provided **three core datasets**:

1. `ingredients.csv`
2. `recipes.csv`
3. `recipe_ingredients.csv`

These represent raw, real-world data — **not ML-ready data**.

Your Task

Write Python code (script or notebook) that:

1. Loads the three datasets
 2. Joins them correctly
 3. Engineers meaningful numerical features per recipe
 4. Outputs a clean, ML-ready dataset:
`recipe_features.csv`
-

Required Output: **`recipe_features.csv`**

Your final dataset **must contain exactly these columns**:

`recipe_id`
`total_cost`
`total_calories`
`total_protein_g`
`total_carbs_g`
`total_fat_g`
`ingredient_count`
`difficulty_score`
`prep_time_minutes`

Feature Engineering Rules

1. Cost & Nutrition Aggregation

For each recipe:

- **total_cost**
→ Sum of: `quantity_required × avg_cost_per_unit`
 - **total_calories**
→ Sum of: `quantity_required × calories_per_unit`
 - **total_protein_g**
→ Sum of: `quantity_required × protein_g_per_unit`
 - **total_carbs_g**
→ Sum of: `quantity_required × carbs_g_per_unit`
 - **total_fat_g**
→ Sum of: `quantity_required × fat_g_per_unit`
-

2. Ingredient Count

- **ingredient_count**
→ Number of unique ingredients used in the recipe
-

3. Difficulty Encoding

Convert difficulty to a numerical score:

`easy → 1`
`medium → 2`
`hard → 3`

Store this as `difficulty_score`.

4. Data Quality Requirements

- All numeric columns must be floats or integers
 - No missing values in the final dataset
 - One row per recipe
-

Deliverables

Students must submit:

1. Python file or notebook
2. Generated `recipe_features.csv`
3. A short README answering:
 - What assumptions you made
 - One data issue you encountered
 - One feature you believe is most important, and why