Denton Kunz, Sam Beilenson

INGREDIENT VERSATILITY: FINDING THE MOST FLEXIBLE FOODS IN RECIPE DATA

CONTEXT & MOTIVATION

- We love cooking, and we're fascinated by how some ingredients seem to show up *everywhere*.

- We wanted to explore which ingredients are the most useful in everyday cooking—like, what gives you the most bang for your buck?

- This could help:
  - Students or families shop smarter
  - Food banks recommend versatile staples
  - App developers suggest better recipes based on pantry items

ANALYTICAL QUESTION

Which foods (ingredients) are used in the highest number of unique recipes?

Follow Up Questions:

- How does ingredient versatility vary across tags or cuisines?

- What's the smallest group of ingredients that unlocks the most recipe options?

DATA PIPELINE PROCESS

CSV Dataset (Kaggle) , then → Python Script (CSV Parser) → SQL INSERT Statements → Postgres DB → SQL Queries & Visualization (Grafana)

DATA SOURCES

Main Dataset: Kaggle – *Food.com Recipes and Interactions!*

Supplementary APIs (not used):

- API Ninjas

- Spoonacular

- Edamam
  (These lacked ingredient detail or access flexibility)

INGESTION PROCESS

Format: CSV files (from Kaggle)

Tooling:

- Parsed in **Python** using csv and parquet
- Converted to SQL using generated **INSERT statements**
- Inserted into a **PostgreSQL** database

Tables Created:

- recipes – basic recipe metadata (ID, title, instructions, servings)
- ingredients – unique ingredient list with standardized names
- recipe_ingredients – join table linking recipes to their ingredients, with quantities and units.
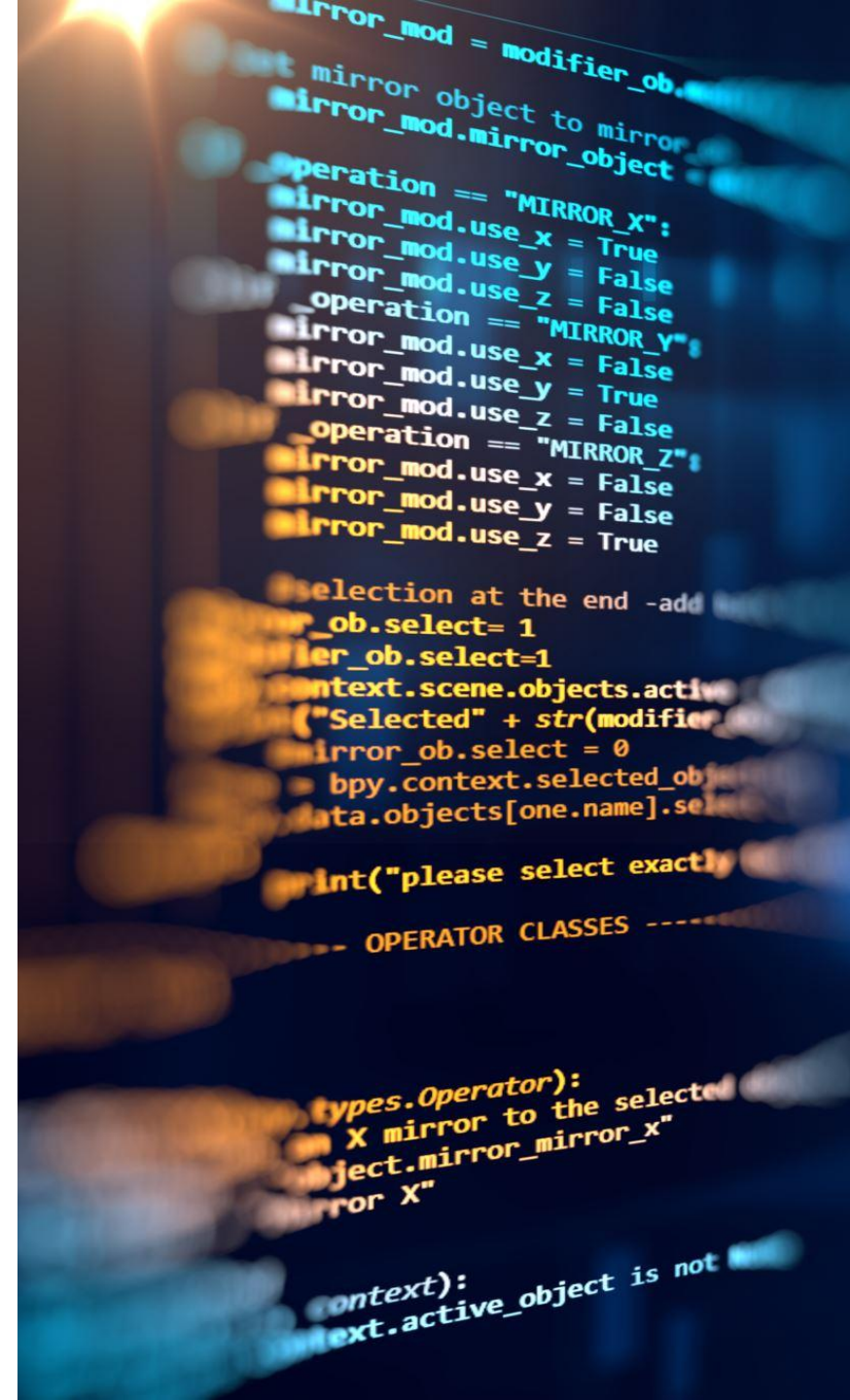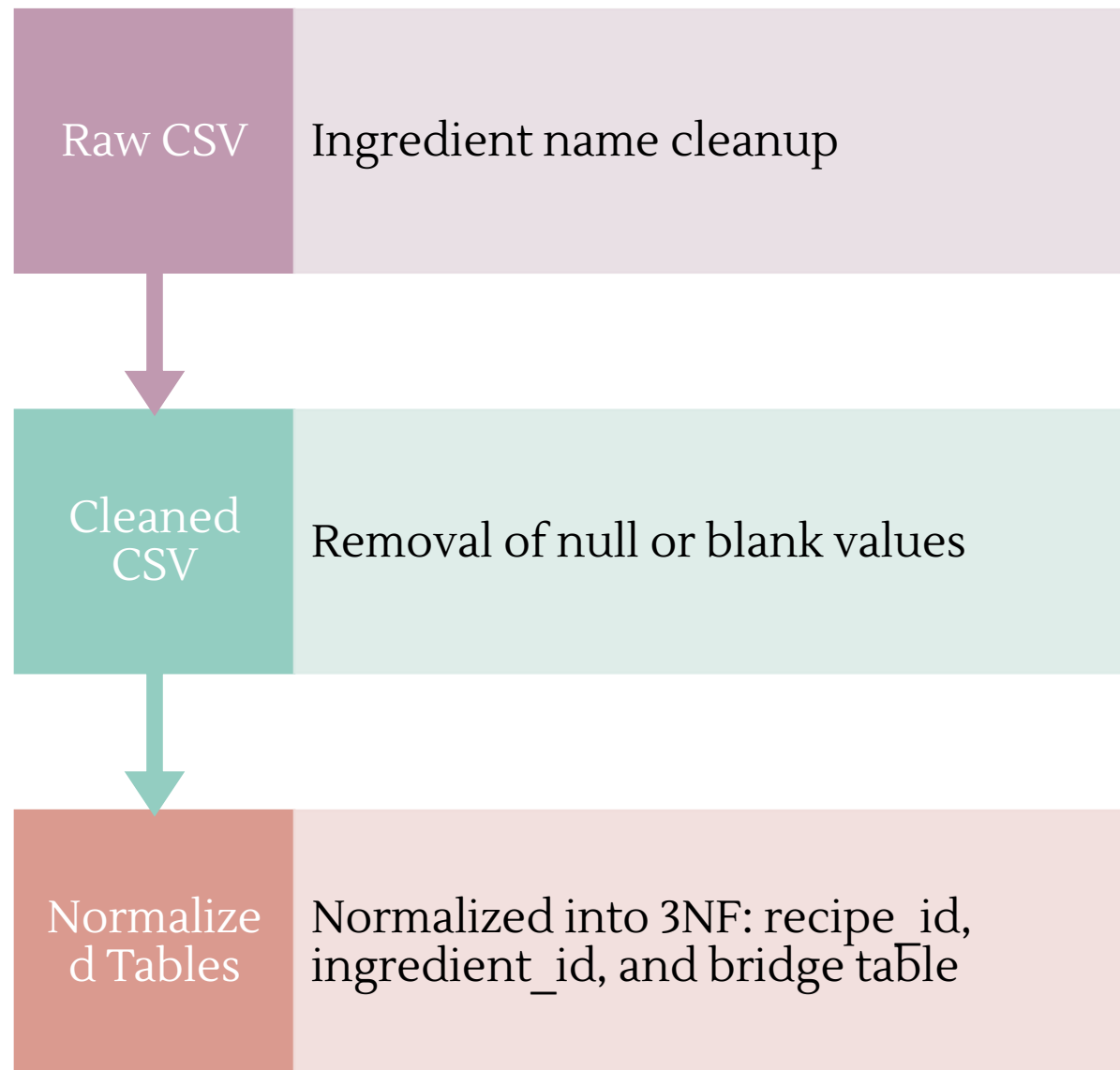
TRANSFORMATION OVERVIEW

Code snippet (Python):

```python
ingredients = ast.literal_eval(row["ingredients"])
for item in ingredients:

insert_sql("RECIPE_INGREDIENT", recipe_id, normalize(item))
```

Raw CSV	Ingredient name cleanup
Cleaned CSV	Removal of null or blank values
Normalized Tables	Normalized into 3NF: recipe_id, ingredient_id, and bridge table

ER DIAGRAM

Tables and Relationships:

- **recipes** (*recipe_id* PK, title, description, cook/prep times, nutrition info, etc.)

- **ingredients** (*ingredient_id* PK, name)

- **authors** (*author_id* PK, name)

- **recipe_ingredients** (*recipe_id*, *ingredient_id*)

  → Bridge table with composite PK

 → Many-to-Many relationship between recipes and ingredients

Normalization:

- Fully normalized to **3NF**

  → Reduces redundancy

  → Allows for flexible queries (e.g., top ingredients by nutrition, by category)

Foreign Key Constraints:

- recipes.author_id → authors.author_id

- recipe_ingredients.recipe_id → recipes.recipe_id

- recipe_ingredients.ingredient_id → ingredients.ingredient_id

Recipe Management Data Model
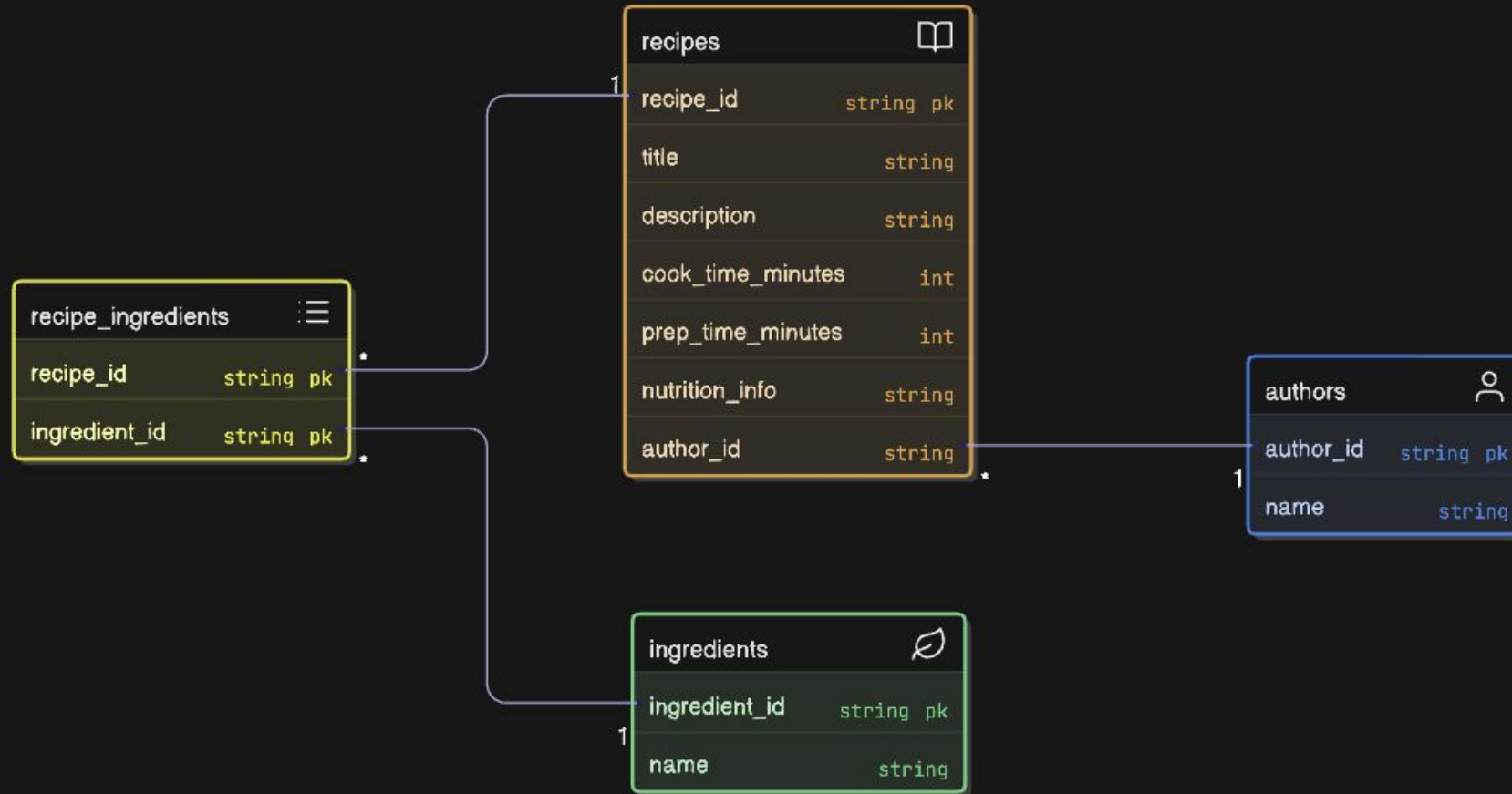
SERVING THE DATA*

POSTGRESQL AND GRAFANA
BOTH HOSTED ON **RAILWAY**

CONNECTED USING
**POSTGRESQL DATASOURCE
URL**

VISUALIZATIONS BUILT WITH
LIVE SQL QUERIES

PANELS SHOW INGREDIENT
TRENDS, NUTRITION STATS,
AND RECIPE COMPARISONS

RESULTS

Category	Top Ingredients
Most Common (All Recipes)	Salt, Butter, Sugar, Onion, Eggs, Water, Olive Oil, Flour, Milk, Garlic Cloves
Desserts	Sugar, Butter, Salt, Eggs, Flour, Baking Powder, Vanilla, Baking Soda, Milk, All-purpose Flour
Quick Recipes (< 15 mins)	Green Onions, Red Onion, Dijon Mustard, Dry Mustard, Onion Powder, Red Wine Vinegar, Kosher Salt, Lime Juice, etc.
Longer Recipes (< 4 hrs)	Ground Beef, Kosher Salt, Mozzarella Cheese, Tomato Paste, Tomato Sauce, Potatoes, Dried Oregano, Heavy Cream, etc.
Unhealthy Recipes	Soy Sauce, Cheddar Cheese, Chicken Broth, Worcestershire Sauce, Bacon, Ground Beef, Chili Powder, Cornstarch, etc.
Healthy Recipes	Carrot, Walnuts, Pecans, Potatoes, Ground Cumin, Nutmeg, Chili Powder, Cornstarch, Zucchini, Chicken Broth, etc.
Average Nutrition (All)	Calories: 484, Fat: 24.6g, Saturated Fat: 9.56g, Cholesterol: 86.5mg, Sodium: 767mg, Carbs: 49.1g, Fiber: 3.84g, Sugar: 21.9g, Protein: 17.5g

CUTOFF FOR HEALTHY/ UNHEALTHY

Nutrition Metric	Threshold for "Too Much"
Calories	> 600 kcal
Fat Content	> 30 grams
Saturated Fat Content	> 10 grams
Sodium Content	> 600 mg
Sugar Content	> 25 grams
Cholesterol Content	> 200 mg
Carbohydrate Content	> 60 grams

CONCLUSIONS

Basic ingredients dominate across all recipe types

- (e.g., salt, butter, sugar, flour appear frequently in every category, regardless of nutrition or prep time.)

Unhealthy recipes feature richer, fattier ingredients

- (e.g., cheddar cheese, bacon, ground beef, heavy sauces.)

Healthy recipes favor vegetables, nuts, and spices

- (e.g., carrots, walnuts, pecans, ground cumin.)

Faster recipes (under 15 minutes) use simple flavorings

- (e.g., onions, mustard, vinegar, quick seasonings.)

Longer recipes (over several hours) lean toward hearty ingredients

- (e.g., ground beef, potatoes, tomato sauce.)

Challenges and Future Work

Challenges Encountered:

- **Data Cleaning:**

  Needed extensive filtering to remove basic ingredients (e.g., salt, butter) that skewed results.

- **Normalization:**

  Converting raw CSVs with nested ingredient lists into normalized SQL tables required manual processing and scripting.

- **Database Connection:**

  Initial setup issues connecting Grafana to PostgreSQL via Railway (passwords, permissions, URL confusion).

Future Improvements:

- **Deeper Analysis:**

  Incorporate nutrition density (e.g., calories per gram) rather than raw nutrition totals.

- **Predictive Modeling:**

  Build models to predict recipe healthiness based on ingredient lists and preparation methods.