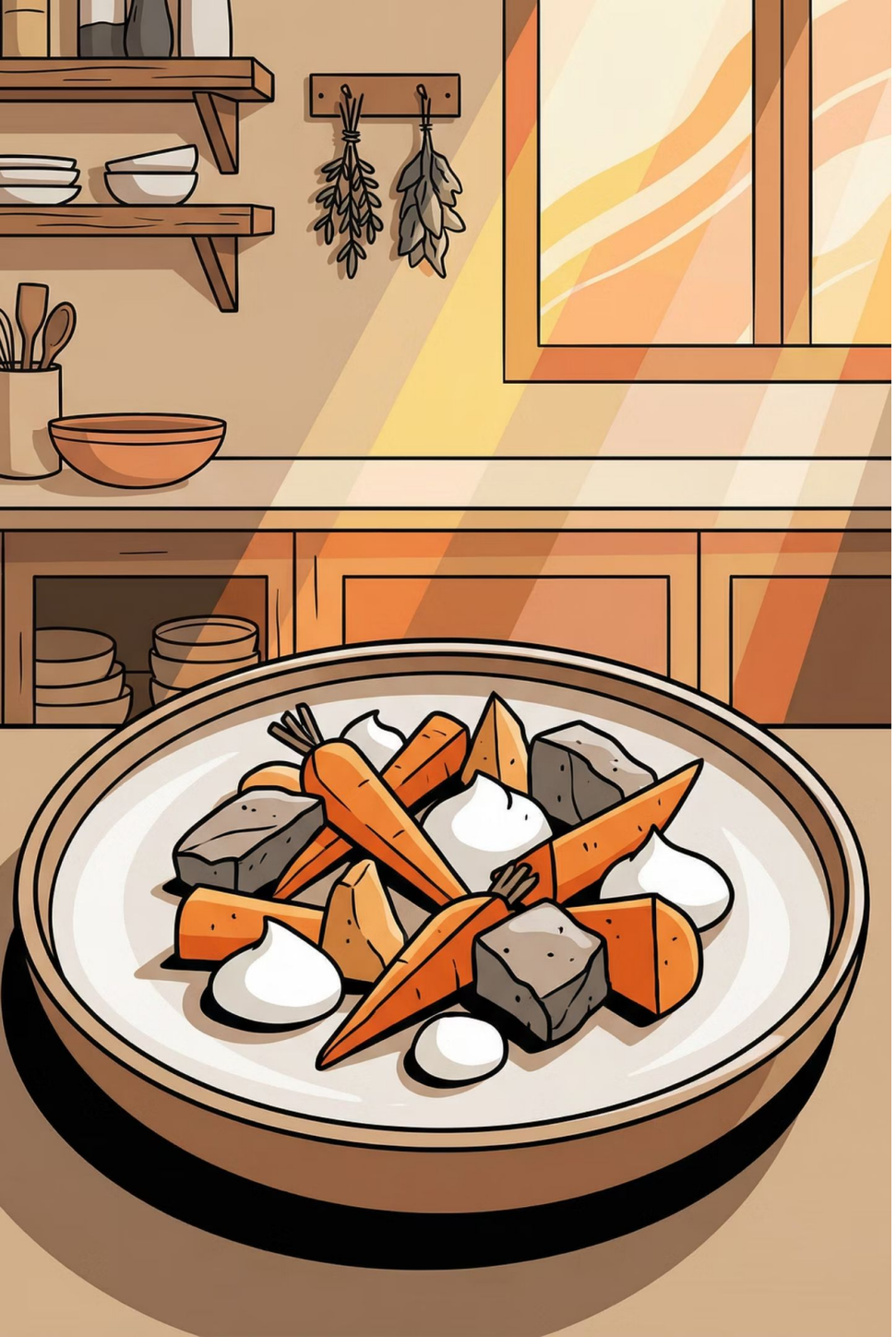


Recipe Site Traffic: Predicting Popularity

A data-driven approach to identifying high-traffic recipes for Tasty Bytes, optimizing user engagement through predictive modeling and nutritional analysis.

Project Workflow

My systematic approach ensures data integrity and model reliability through five critical phases.

Cleaning & Validation

Removing nulls, duplicates, etc. to ensure a clean dataset.

Exploratory Analysis

Identifying trends in servings, categories, and nutrients.

Model Development

Building Logistic Regression and Random Forest models.

Model Evaluation

Assessing precision and accuracy on test data.

Business Metric

Establishing the True High Traffic Recipe Index (THTRI).

Data Cleaning & Validation

Numerical Integrity

- Imputed rows with nulls in nutrient columns group-wise.
- Extracted numbers from mixed-type columns.
- Converted data types for optimal model processing.

Categorical Standardization

- Standardized recipe categories for consistency.
- Converted data types for optimal model processing.

Before

```
RangeIndex: 947 entries, 0 to 946
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   recipe        947 non-null    int64
 1   calories      895 non-null    float64
 2   carbohydrate  895 non-null    float64
 3   sugar         895 non-null    float64
 4   protein       895 non-null    float64
 5   category      947 non-null    object
 6   servings      947 non-null    object
 7   high_traffic  574 non-null    object
dtypes: float64(4), int64(1), object(3)
memory usage: 59.3+ KB
```

After

```
RangeIndex: 947 entries, 0 to 946
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   recipe        947 non-null    int64
 1   calories      947 non-null    float64
 2   carbohydrate  947 non-null    float64
 3   sugar         947 non-null    float64
 4   protein       947 non-null    float64
 5   category      947 non-null    category
 6   servings      947 non-null    int64
 7   high_traffic  947 non-null    uint8
dtypes: category(1), float64(4), int64(2), uint8(1)
memory usage: 46.7 KB
```

What Makes a Recipe Popular?

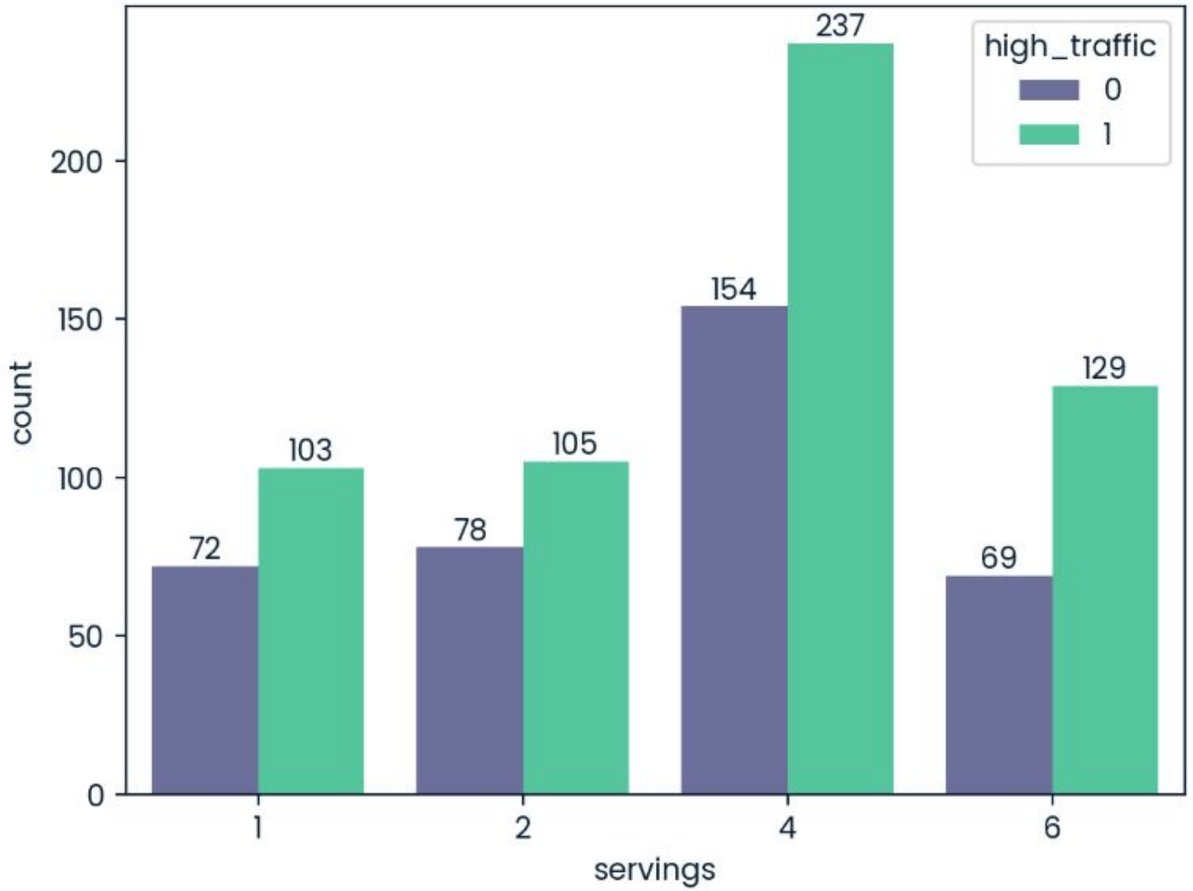
This analysis revealed distinct patterns in user preferences across:

1. Serving size
2. Category
3. Nutrition type.

Impact of Serving Size

The "Four" Rule

Meals with four servings are significantly more popular than any other serving size.



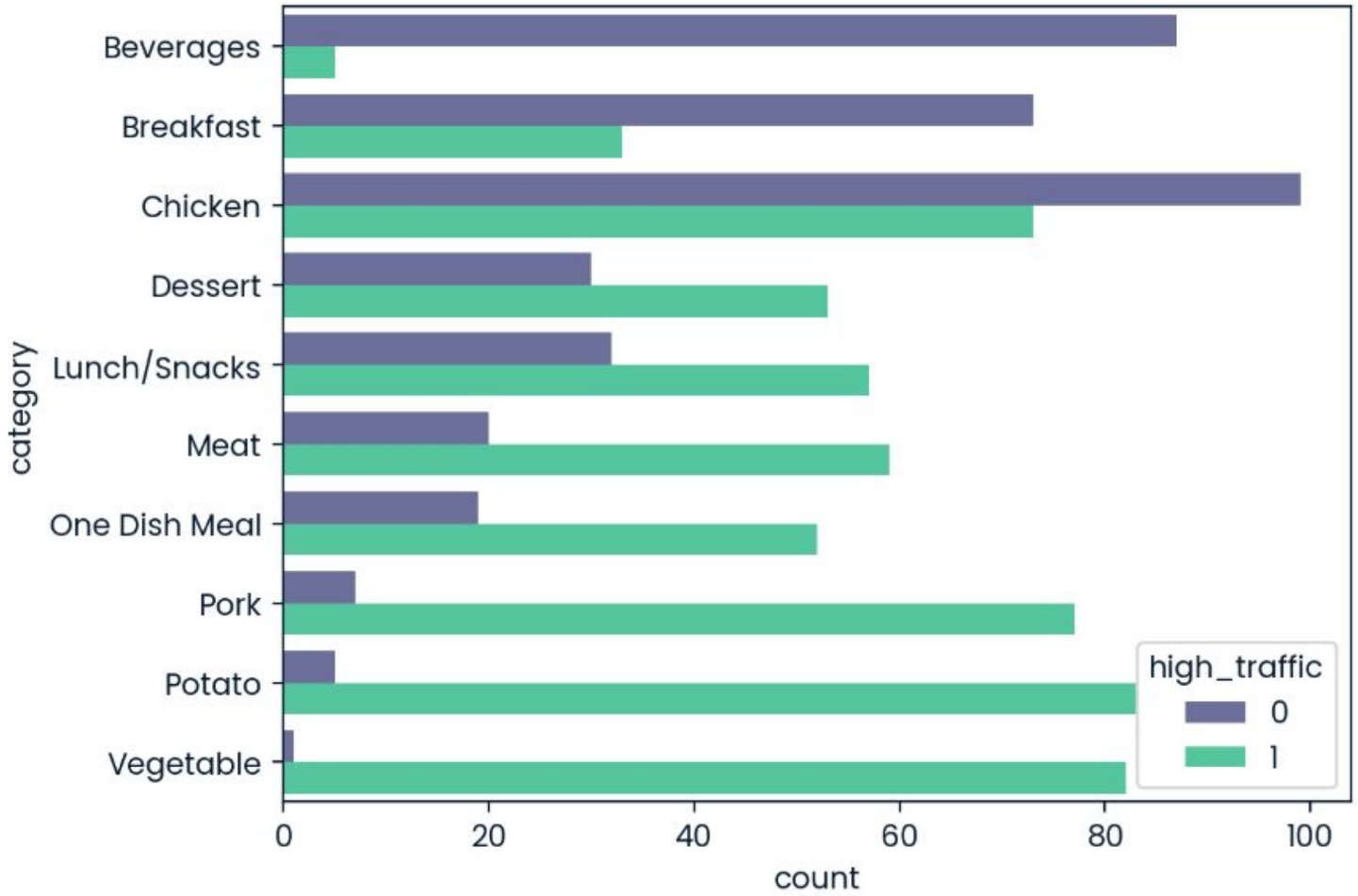
Categorical Influence

Top Categories

Potato, Pork, and Vegetable recipes consistently drive the highest traffic.

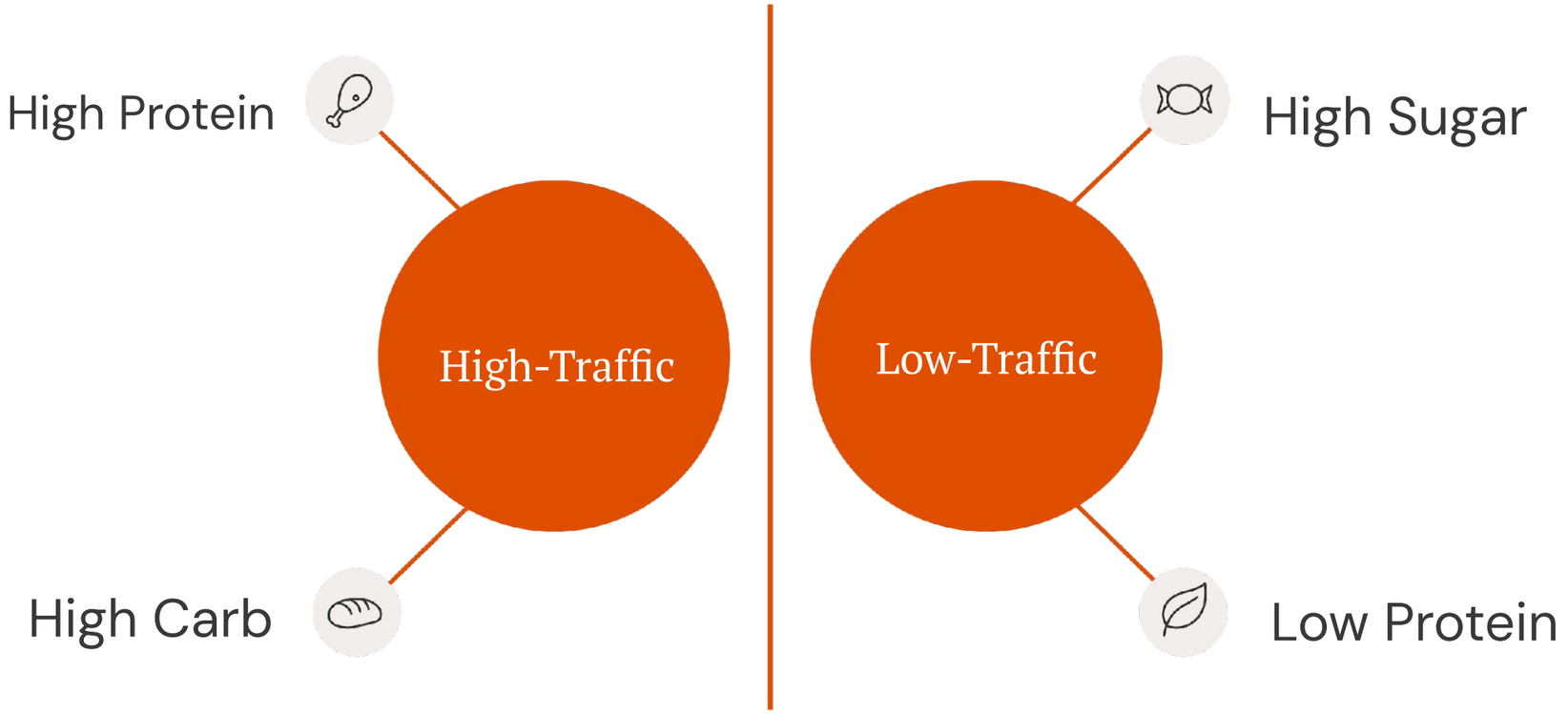
Low Performers

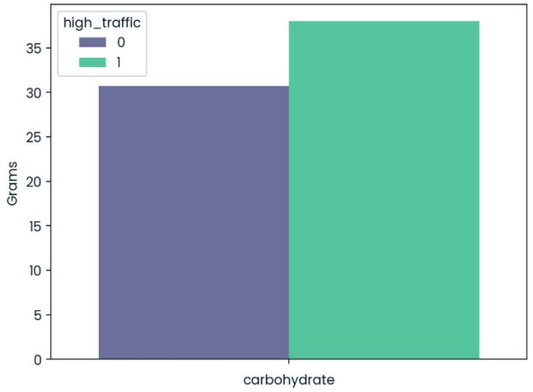
Beverages and Breakfast categories gain very little traction on the site.

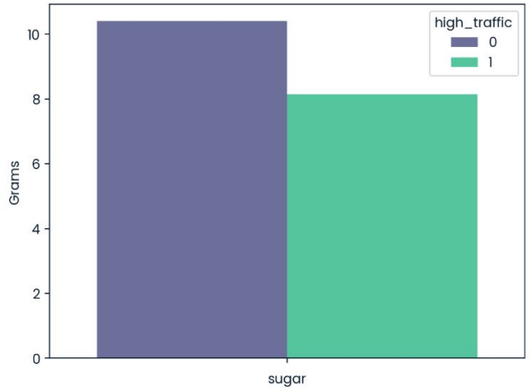


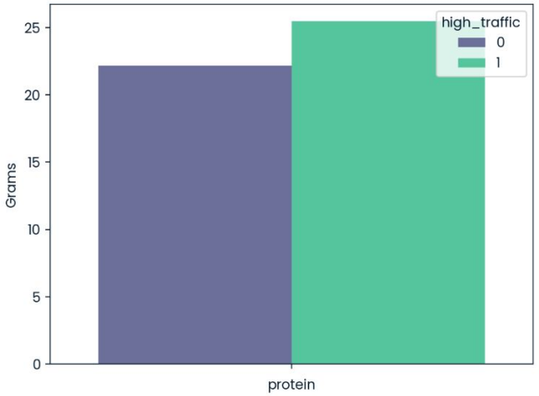
Nutritional Drivers of Traffic

Users show a clear preference for specific nutritional profiles when selecting recipes.

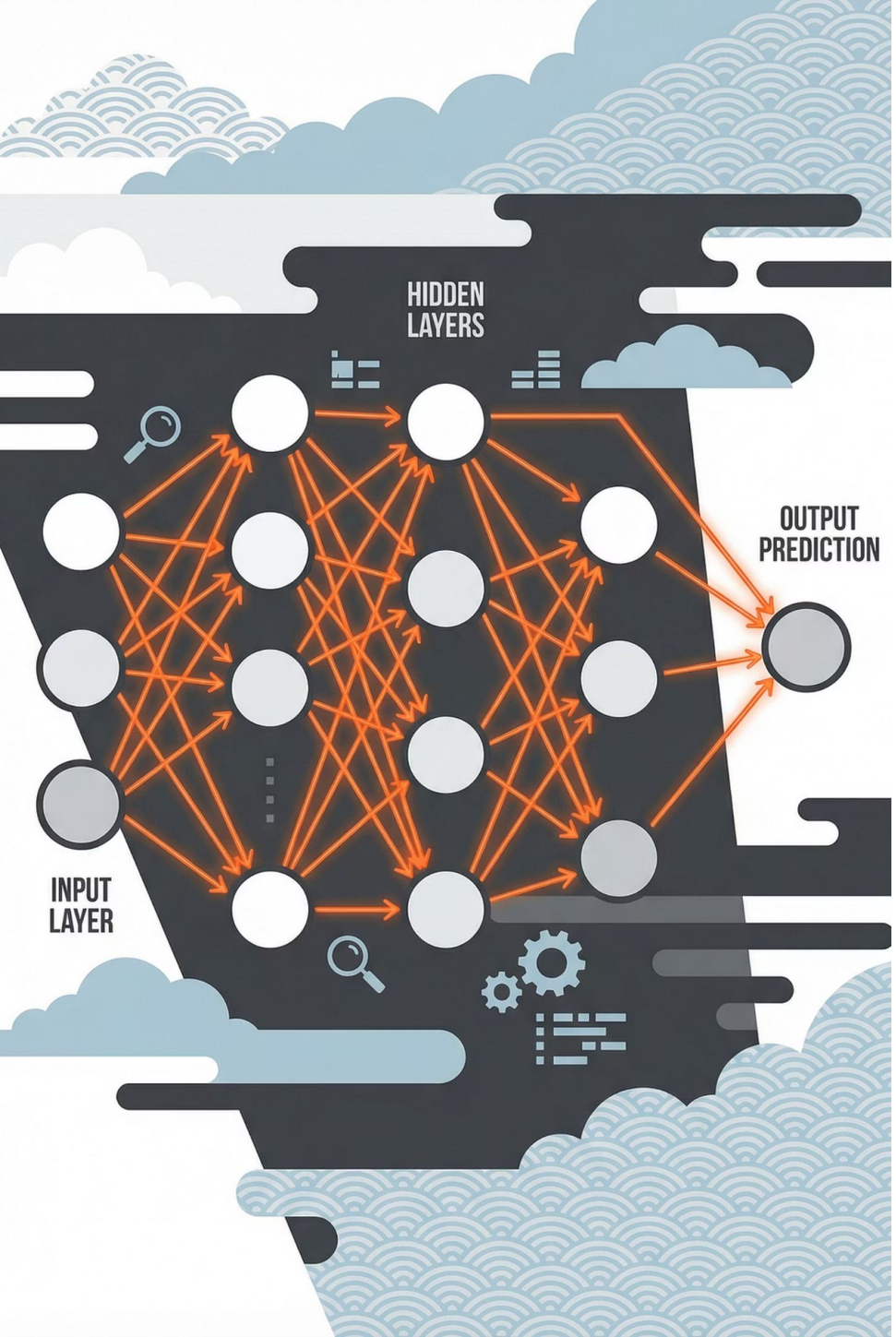








Insight: There is no strong association between sugar and calories, but popular meals consistently have on average less sugar across all serving amounts.



Model Development Strategy

Primary Goal: 80% Precision

I prioritize Precision over Recall because predicting a "dud" as high-traffic (False Positive) hurts the business more than missing a popular recipe.

Baseline: Logistic Regression

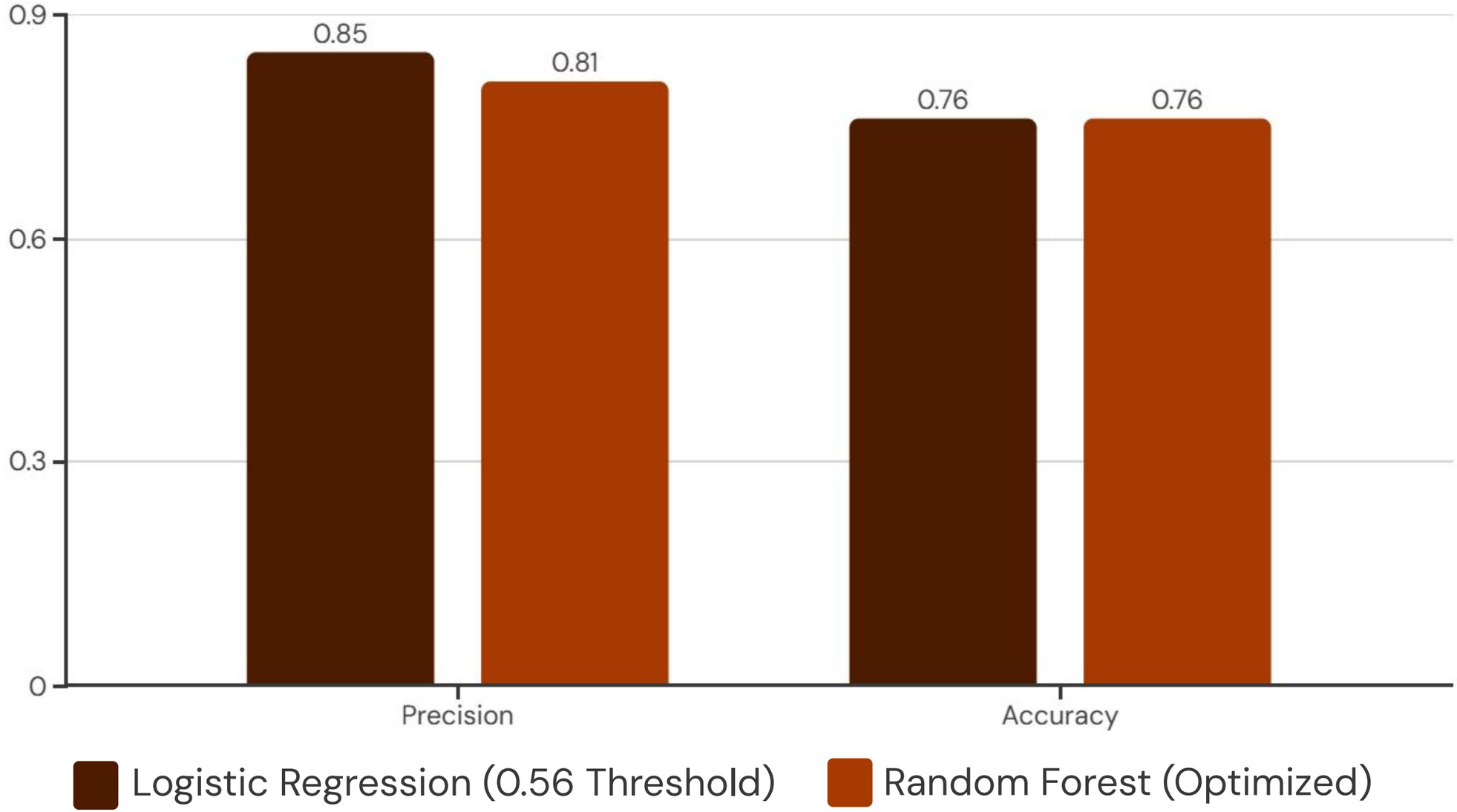
Chosen for its efficiency with limited datasets and clear interpretability.

Comparison: Random Forest

Tested to see if ensemble methods could capture non-linear relationships in this case better or not.

Model Evaluation Results

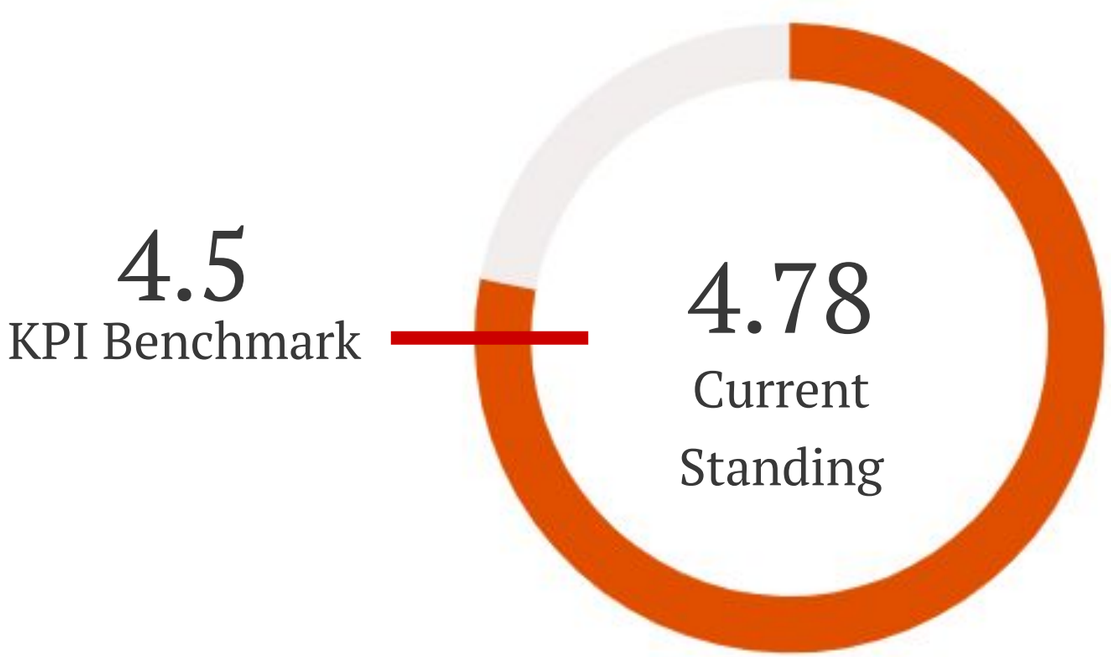
The Logistic Regression model outperformed Random Forest, especially after threshold optimization.



Cross-validated accuracy score of **0.76** confirms the model is well-generalized and not overfit.

The Business Metric: THTRI

To monitor success, we established the **True High Traffic Recipe Index (THTRI)**, a ratio of False Positives to True Positives.



I set the THTRI benchmark at 4.5 to minimize business risk and maximize traffic.

Final Recommendations

→ Promote Top Categories

Prioritize 4 servings' Potato, Pork, and Vegetable recipes for the homepage display.

→ Apply Nutritional Filters

Use the "High Protein/High Carb/Low Sugar" filter to select new recipes for promotion.

→ Deploy the Model

Integrate the Optimized Logistic Regression model with the THTRI KPI for real-time monitoring.