

4th of May, 2025

Presented By: Kofoworola Egbinola

OPTIMIZING RECIPE PROMOTION WITH ML

A Data-Driven Approach to Boosting User Engagement

Table of Contents

	pg.
I Project Objectives and Goals	3
II Dataset Overview	4
III EDA Highlights	5
IV Modeling Approach	6
V Model Comparison	7
VI Selected Model	8
VII Business Metric	9
VIII Conclusion & Strategic Recommendations	10

Objective:

To help Tasty Bytes promote recipes that are likely to become popular, using data-driven predictions to maximize homepage impact and user engagement.

Goals:

1

Predict which recipe will lead to high traffic.

2

Correctly predict high traffic recipes 80% of the time.

3

Define a business metric to track success post-deployment.

Dataset Sample:

recipe	calories	carbohydrate	sugar	protein	category	servings	high_traffic
1					Pork	6	High
2	35.48	38.56	0.66	0.92	Potato	4	High
3	914.28	42.68	3.09	2.88	Breakfast	1	null
4	97.03	30.56	38.63	0.02	Beverages	4	High
5	27.05	1.85	0.8	0.53	Beverages	4	null

Source: Tasty Bytes' Recipe database containing metadata, nutritional information, and traffic classification.

Data Validation & Cleaning:

Each feature was thoroughly inspected for data quality issues such as missing values, inconsistencies, and outliers, and appropriate cleaning steps were applied to ensure reliable analysis and modeling.

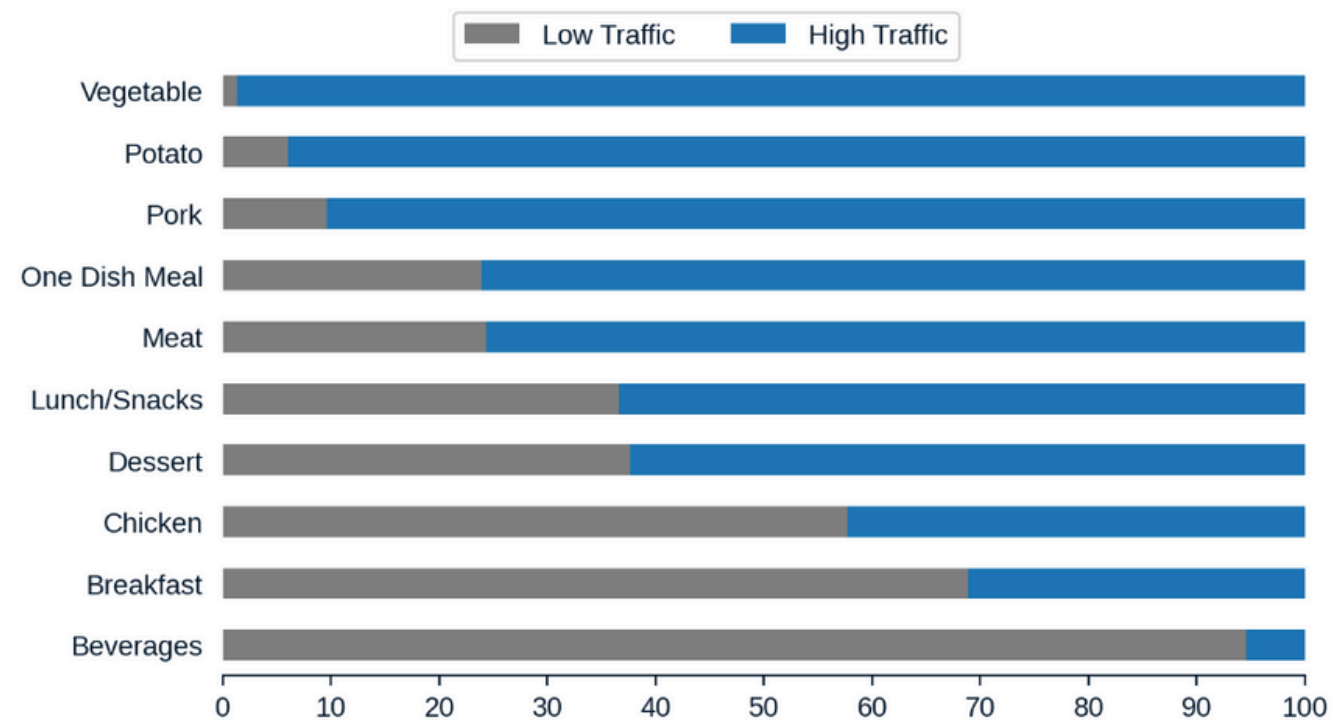
Traffic is Slightly Skewed

A higher proportion of recipes drive **high traffic**

60% 40%

Category Influence on Traffic

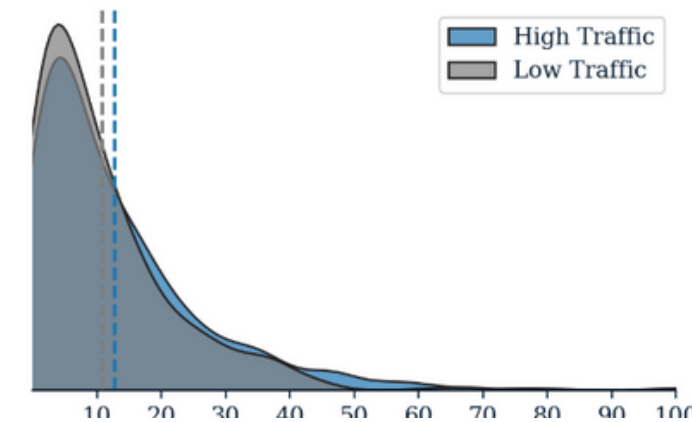
Certain categories (e.g., Potato, Vegetable) consistently drive high traffic, with a strong, statistically significant relationship.



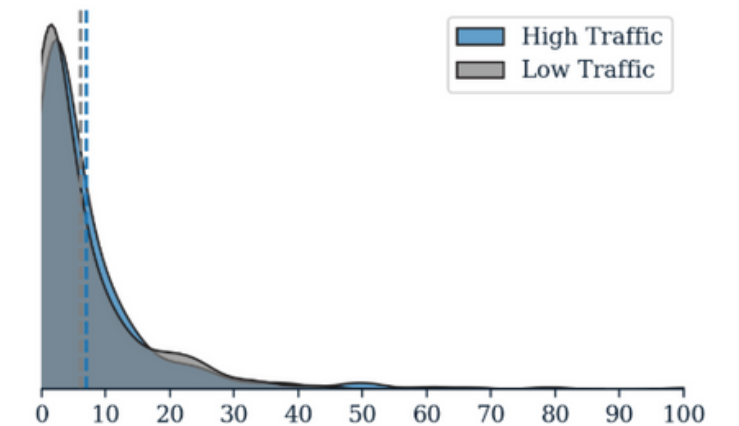
Nutritional Patterns

High-traffic recipes generally have higher nutritional values (excluding sugar), though correlations are not statistically significant.

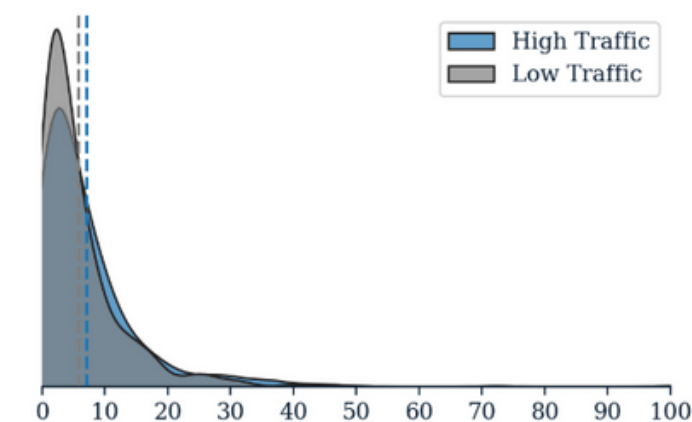
High-traffic recipes tend to have more **calories**



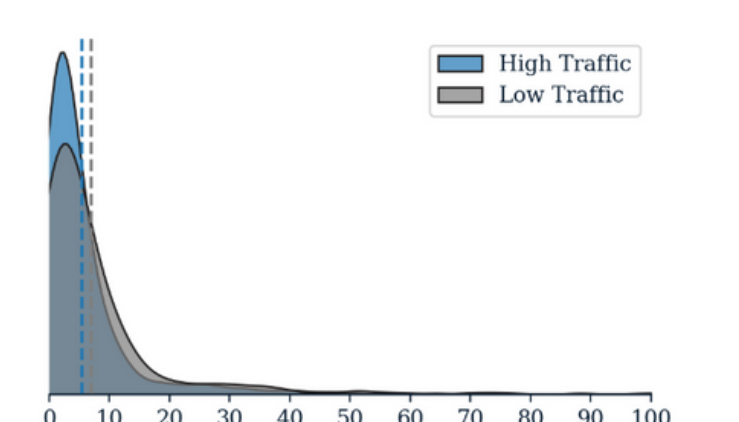
Slight **protein** differences exist between traffic groups



Carbs are slightly higher in **high-traffic** recipes



Low-traffic recipes tend to have more **sugar**



Goal:

Predict whether a recipe will drive traffic — with a focus on achieving at least 80% precision to ensure effective promotions.

Selected Models:

- Logistic Regression (baseline, interpretable)
- Random Forest (ensemble, handles complexity & non-linearity)

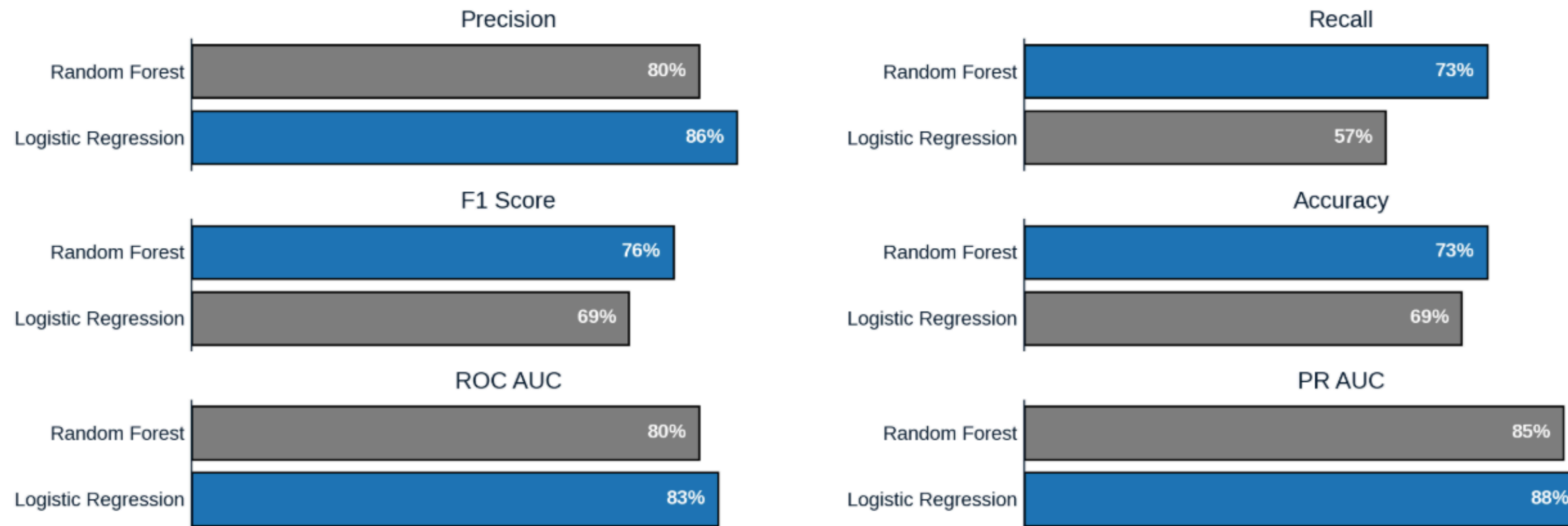
Target Variable:

Traffic Level (High vs. Low), derived from recipe pages views.

Evaluation Metrics:

- **Precision, Recall, F1-score** — to balance correct predictions and reduce false promotions
- **ROC-AUC and PR-AUC** — to assess overall classification performance

Key Metrics on Test Data



🎯 Both models met the 80% *precision target*, meaning most promoted recipes are likely to drive traffic.

⚠️ **Logistic Regression** had slightly higher **Precision**, but lower **Recall**, may miss too many high-potential recipes.

✅ **Random Forest** outperformed in **Recall** and **F1 Score**, making it more effective for capturing more truly high traffic recipes.

Chosen Model — Random Forest

Why Random Forest?

- Met the business target with 80% Precision — ensuring effective recipe promotions
- Achieved higher Recall (0.73) and F1 Score (0.76) — better at identifying truly popular recipes
- Strong overall performance

Strengths:

- Captures complex, non-linear relationships in the data
- Robust to overfitting and works well with mixed data types

Random Forest's Prediction Performance

	Predicted Unpopular	Predicted Popular
Actual Unpopular	<div>True Unpopular 53 29.6%</div>	<div>False Popular 19 10.6%</div>
Actual Popular	<div>False Unpopular 29 16.2%</div>	<div>True Popular 78 43.6%</div>

Business Metric — Promotion Hit Rate (PHR)

Definition:

The percentage of promoted recipes that actually drive traffic

→ Measures how effective the promotion strategy is

Formula:

$$\text{PHR} = \frac{\text{Number of promoted recipes that drove traffic}}{\text{Total number of promoted recipes}} \times 100$$

Why PHR Matters:

- Ensures homepage is used on recipes that truly attract traffic
- Directly aligns with the business goal of maximizing user engagement through popular content

Estimated Starting Value:

Based on the Random Forest model's 80% precision, PHR is estimated to start at ~80% if only predicted-popular recipes are promoted.

Conclusion

Random Forest outperformed Logistic Regression with stronger recall and F1-score, achieving the 80% precision target for confident promotions. Traffic is strongly influenced by recipe categories, with Potato, Vegetable, and Pork driving the highest engagement.

Actionable Recommendations for Tasty Bytes

- Adopt the Random Forest model for selecting recipes to promote.
- Use Promotion Hit Rate (PHR) as a core business KPI.
- Promote only recipes predicted as "high-traffic" to maximize engagement opportunities.
- Monitor PHR regularly and retrain the model when performance declines.
- Run A/B tests comparing model-guided promotions vs. current strategy.
- Enhance underperforming categories (e.g., Beverages) through better content, visuals, or SEO.

4th of May, 2025

Presented By: Kofoworola Egbinola

Thank You
For Listening