Recipe Site Traffic Report

Presented by

Olayinka Julius Awoniyi

Outlines

- Overview of the project and business goals

- Summary

- Key findings, metric to monitor and current estimation

- Recommendations to the business

Overview of the Project and Business Goals

Project Goal: Predict high-traffic recipes for Tasty Bytes using nutritional and recipe data.

Business Goals: Improve website traffic by identifying and promoting high-traffic recipes, with a target of increasing accuracy in high-traffic predictions to 80%. The business aims to maximize the conversion rate of recipe views.

Summary of Work Undertaken

Data Validation:

Processed a dataset of 947 recipes with 8 features (e.g., calories, sugar, protein). Missing values were imputed using medians based on recipe category and servings.

Created binary target variable ('High' or 'Low' traffic) after cleaning categorical and numerical data.

Exploratory Data Analysis:

Key Insights: The data is right-skewed in terms of calorie distribution, and certain categories like beverages and breakfast dominate the recipe count.

Correlation Analysis: Weak correlations were observed between calories, carbohydrates, protein, and sugar in relation to recipe traffic.

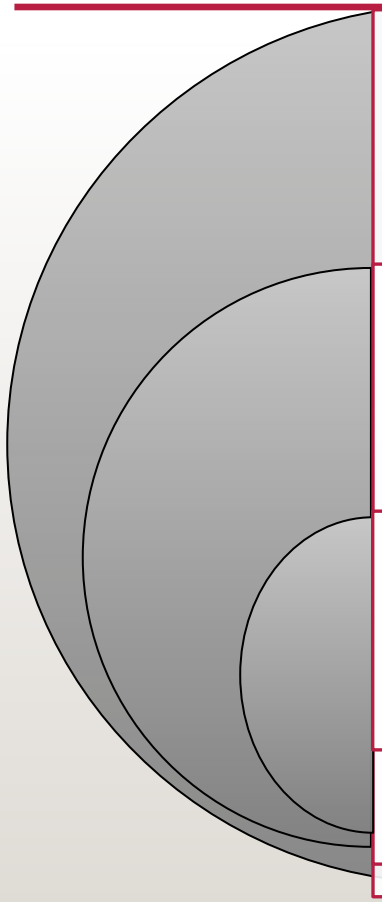
Model Development:

Baseline Model: Logistic Regression for binary classification.

Comparison Model: Random Forest Classifier, known for handling non-linear relationships, was used to improve prediction accuracy.

Feature Engineering: Applied one-hot encoding to categorical variables and normalized numeric data.

Key Findings



Metric to Monitor:

Definition: (Number of High Traffic Recipes Viewed/Total Number of Recipes Viewed)×100

Initial Conversion Rate for High Traffic Recipes: 60.61%

Key Findings

Current Estimation and Results:

Logistic Regression Model:

Accuracy: 72%

ROC AUC Score: 0.81

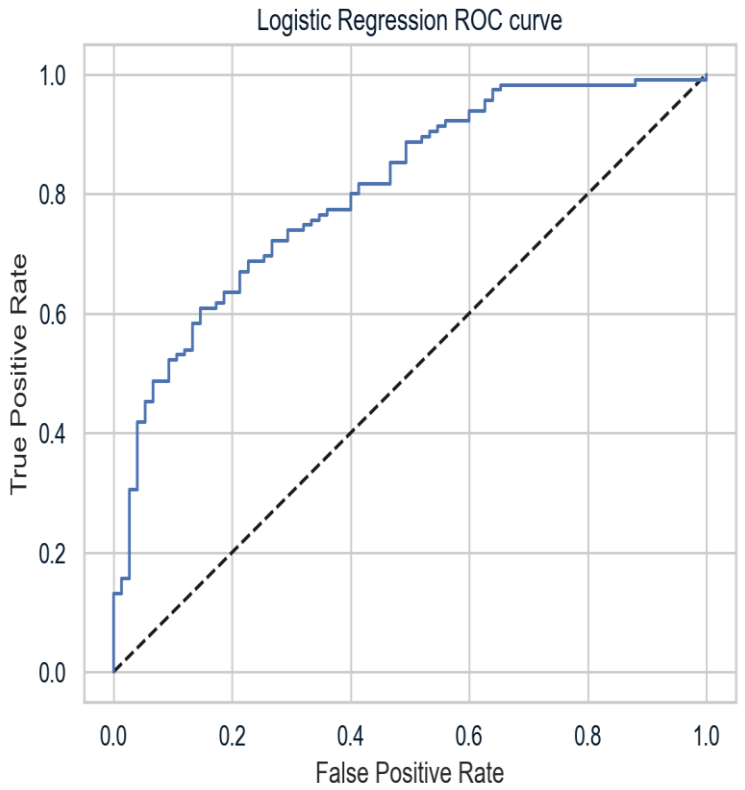
Random Forest Classifier:

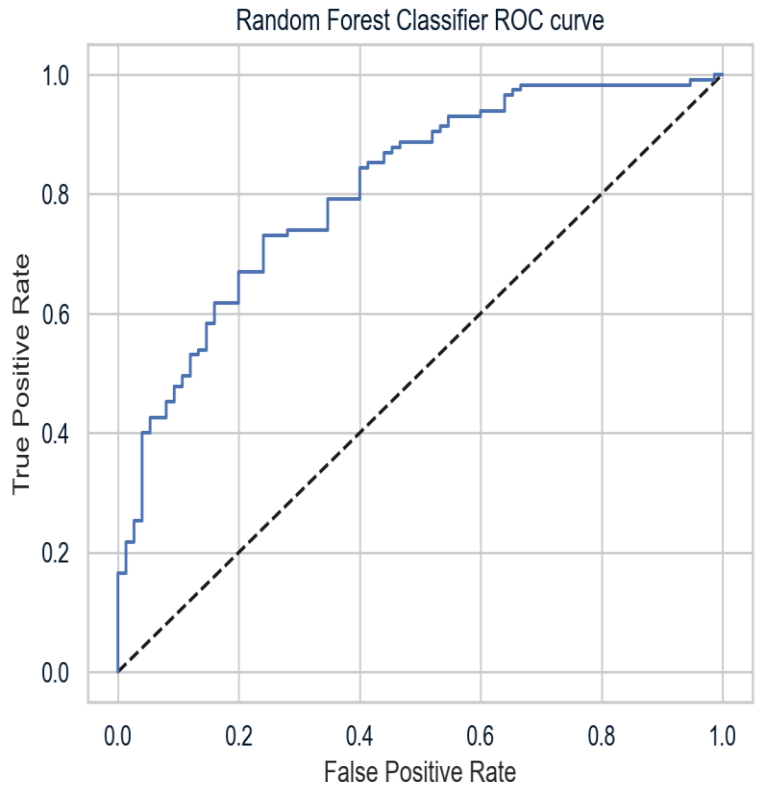
Accuracy: 74%

ROC AUC Score: 0.81

Most Significant Features: Beverages category, breakfast category, protein content.

The ROC curve for Logistic Regression is below 1, while the ROC curve for the Random Forest Classifier is above 1, indicating that the latter performs better than random guessing.





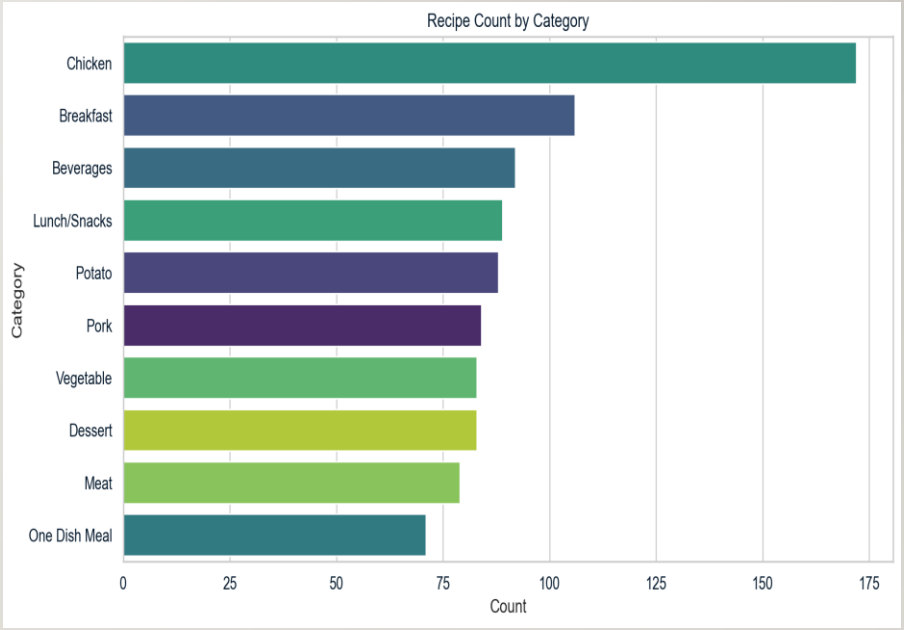
Recommendations to the Business

Emphasize High-Impact Categories: Focus marketing on recipes within the beverages, breakfast, and chicken categories, which show high potential for driving traffic.

Weekly Tracking of Revenue and Time: Implement a system to track revenue generation and time spent weekly for each sales method. This will provide ongoing insights into which strategies are most efficient.

Monthly Monitoring: Implement monthly reports tracking high-traffic recipe predictions and actual user interactions

Targeting 80% Accuracy: Achieving this is feasible with continuous data improvements and monitoring.



Thank you