PREDICTING HIGH TRAFFIC RECIPES

A DATA-DRIVEN APPROACH TO ENHANCING WEBSITE TRAFFIC

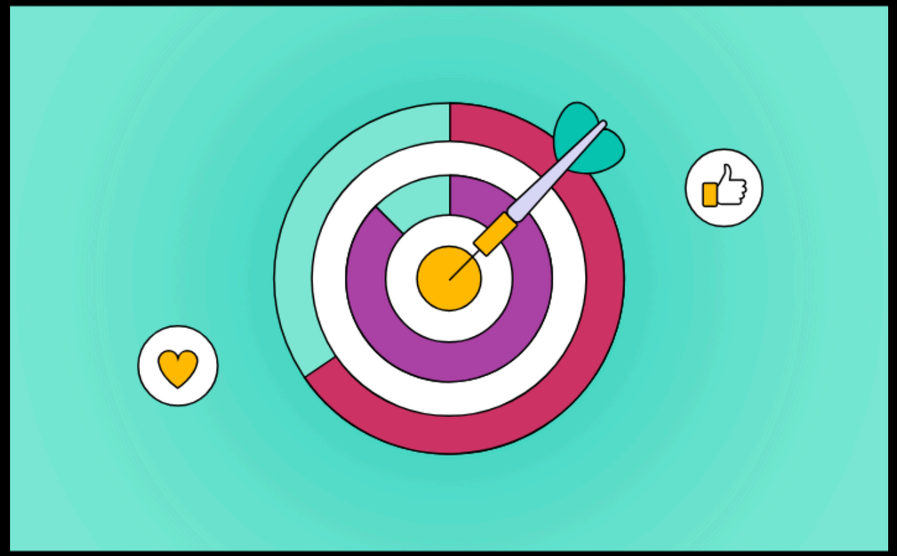
Bob Wesonga

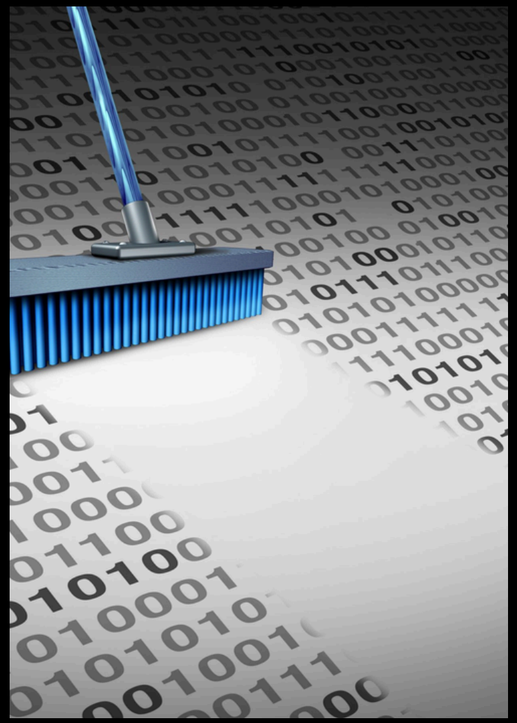
Project Overview

- **Objective**: Predict recipes that drive high traffic to the website.
- **Scope**: Analyze data, develop models, and provide actionable recommendations.
- **Outcome**: Enhance recipe selection to optimize website traffic.

Business Goals

- **Primary Goal**: Identify recipes leading to high traffic.
- **Secondary Goal**: Achieve 80% accuracy in predicting high traffic recipes.

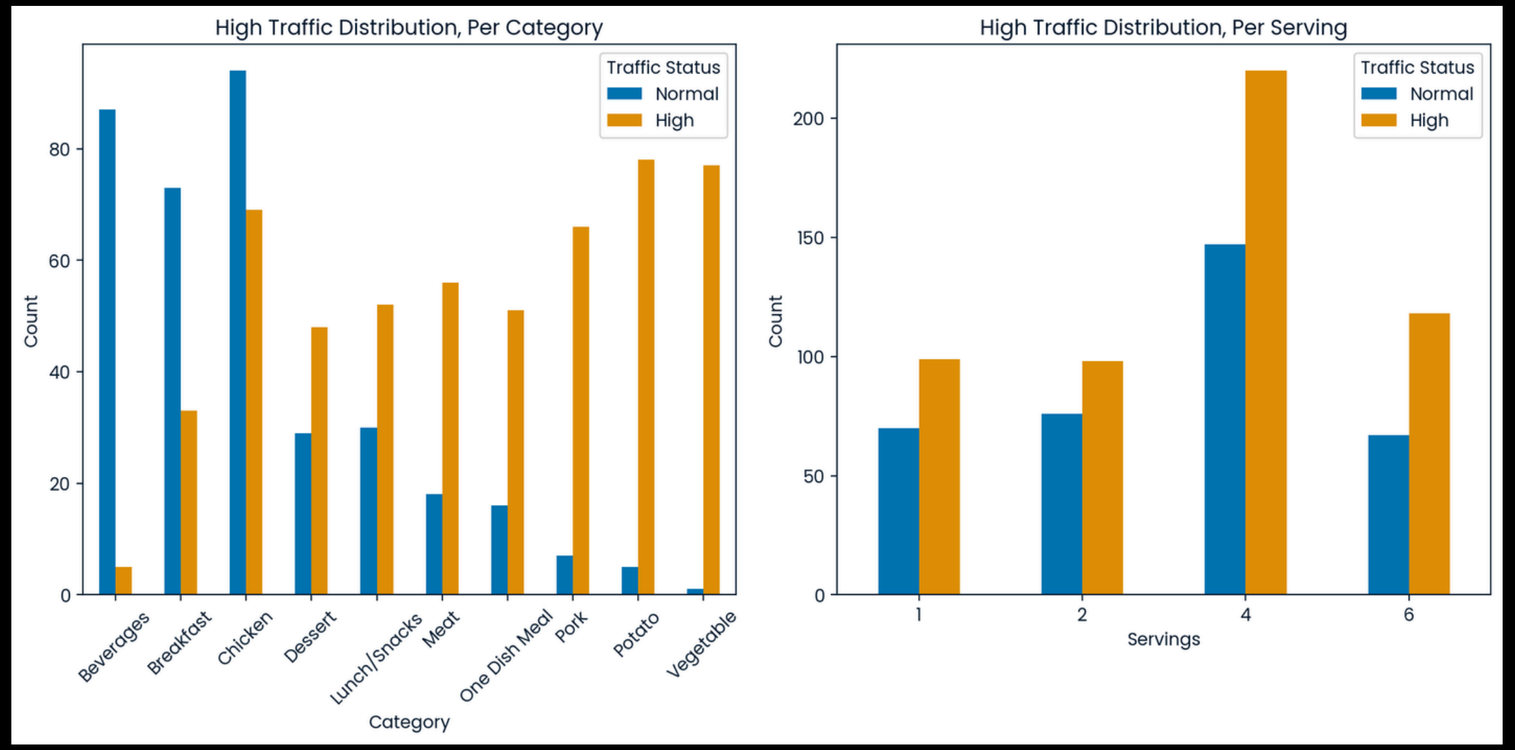




Data Preparation and Cleaning

- Steps Taken:
  - Handled missing values and outliers.
  - Applied log transformation to skewed numeric features.
- Outcome: Clean dataset ready for modeling.

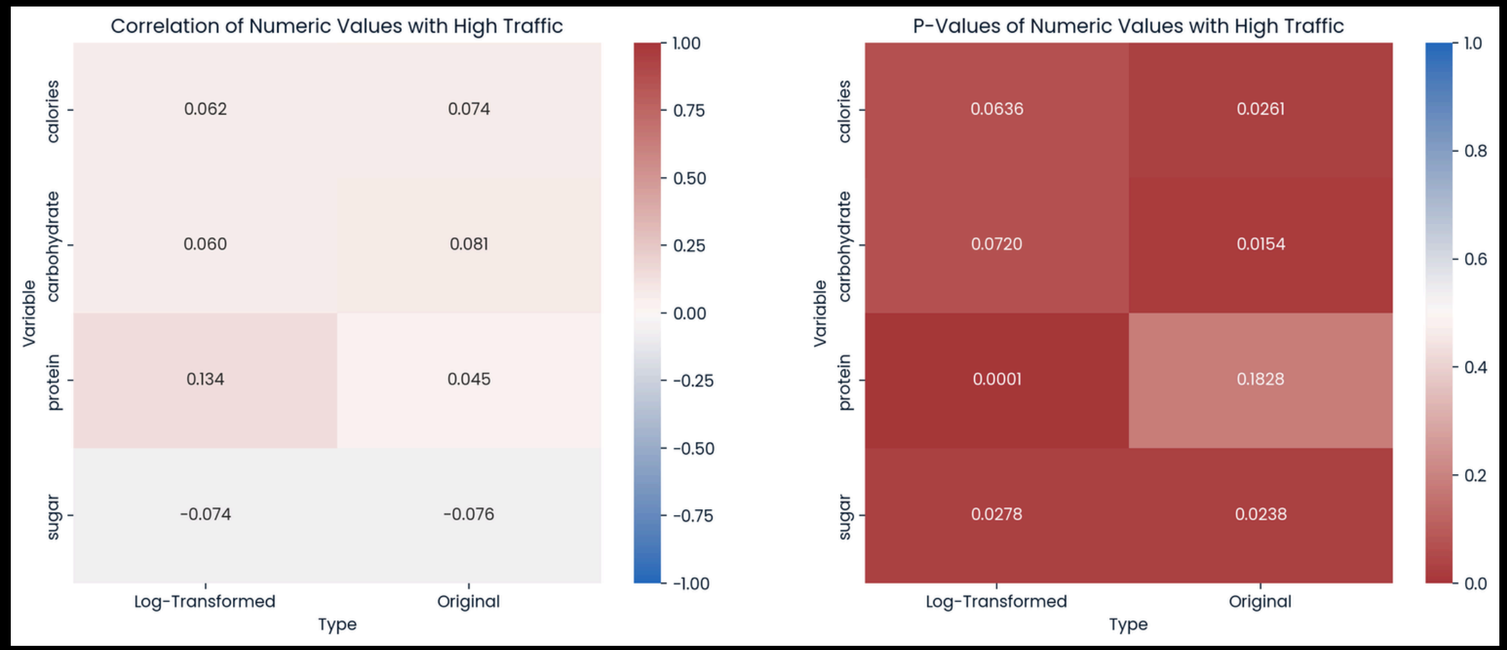
Exploratory Analysis



Single Variable Analysis:

- Distribution of Target Variable: High vs. normal traffic.

Feature Relationships



Two Variables Analysis:

- Correlation between nutritional values and high traffic.
- Impact of categorical features on high traffic.

- Findings: Categories show significant relationships; servings have minimal impact.

Model Development

- Problem Type: Binary classification.
- Models Used:
  - Logistic Regression: Linear approach.
  - Random Forest Classifier: Ensemble approach for complex interactions.

- Preprocessing Steps:
  - Normalization, log transformation, and one-hot encoding.

Model Evaluation

- Logistic Regression:
  - Precision: 81%
  - Accuracy: 77%
- Random Forest Classifier:
  - Precision: 75%
  - Accuracy: 75%
- Comparison: Logistic Regression is preferred due to higher precision.

Key Metrics and Recommendations

Key Metric:
Precision (81% with Logistic Regression).

Recommendations:

- Deploy the Logistic Regression Model.
- Monitor Performance: Regular updates and reviews.
- Enhance Data Collection: Additional features for improved predictions.

Next Steps

Implementation:

- Deploy the selected model into production.
- Establish monitoring and feedback mechanisms.

Future Improvements:

- Incorporate additional data for enhanced predictions.
- Regularly review and update the model to adapt to new trends.