TASTY BYTES Site traffic boost

- Raise site traffic with better recipes' selection powered by machine learning
- Metrics to follow, improvements in data collection and next steps



Data validation

Exploratory Data Analysis

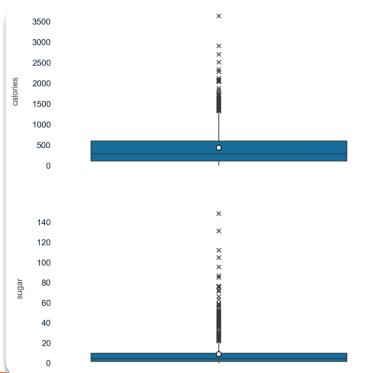
Null values, unique values and outliers

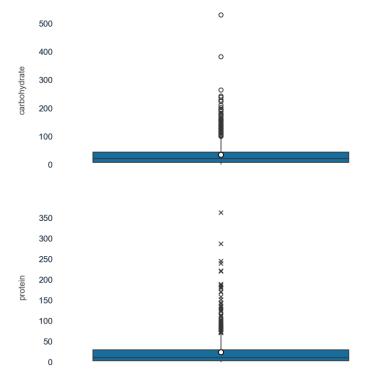
Dataset size = 947 registers

How much data is missing? In percentage

recipe 0.000000 calories 5.491024 carbohydrate 5.491024 5.491024 sugar protein 5.491024 category 0.000000 servings 0.000000 high_traffic 39.387540

dtype: float64



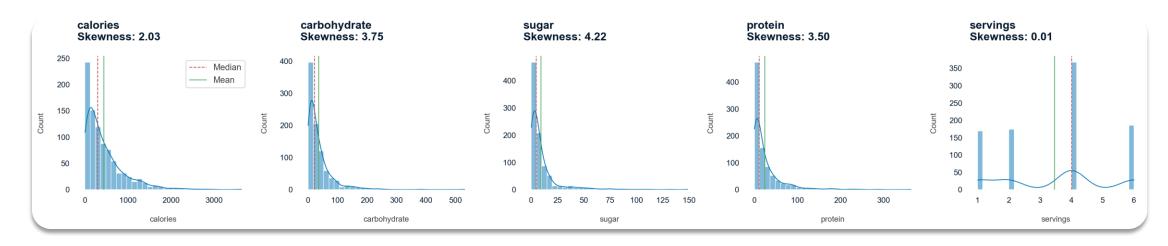


Missing values from the target

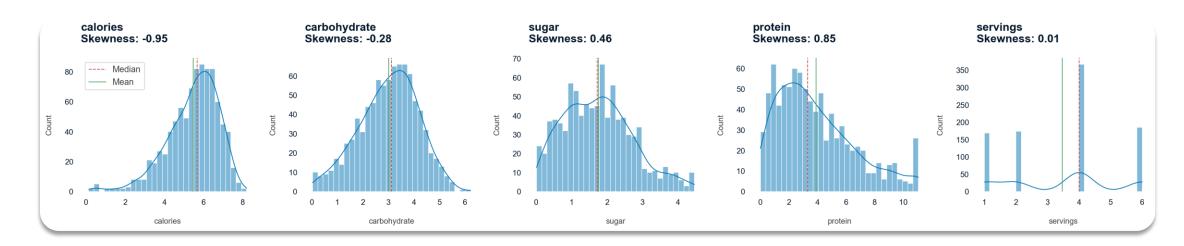
Data validation



Skewness



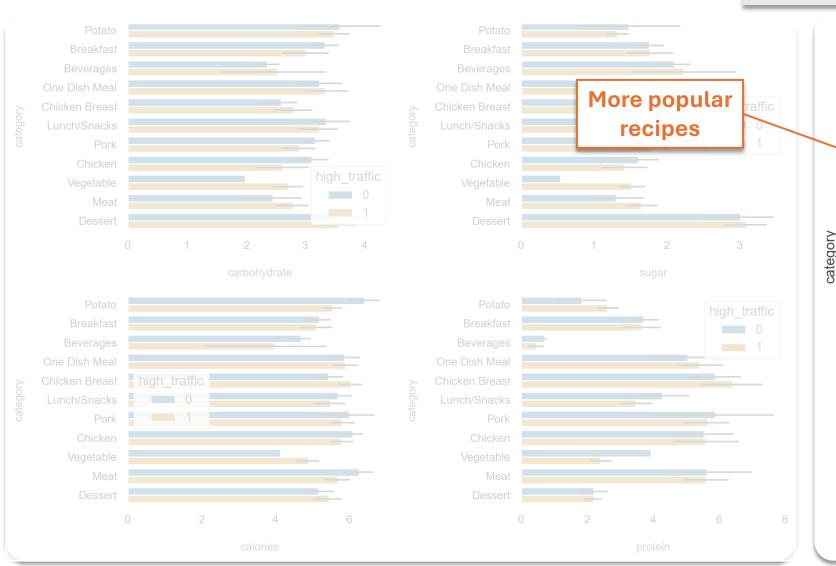
Log and sqrt transformations

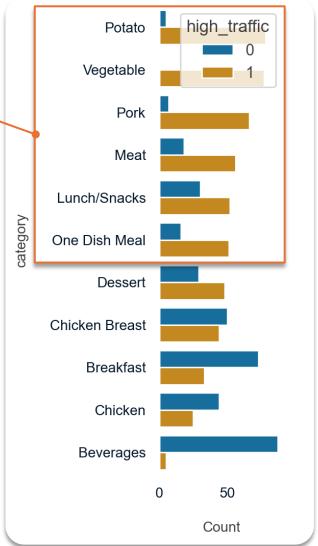






Patterns and correlations

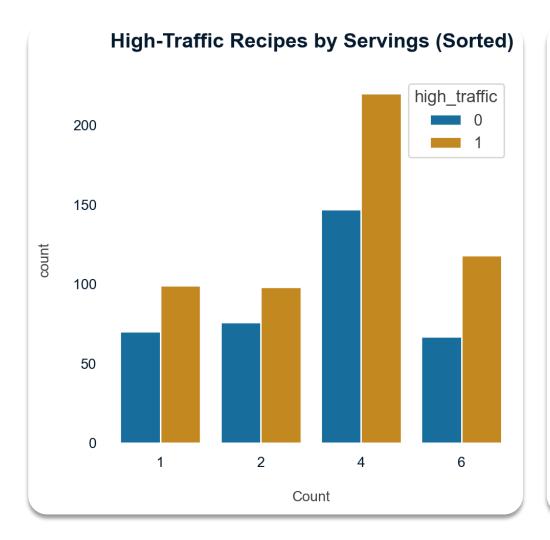


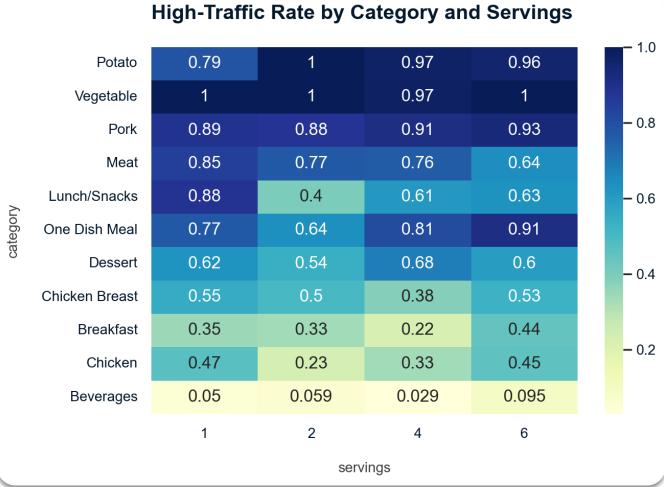


Feature engineering

Exploratory Data Analysis

Patterns and correlations

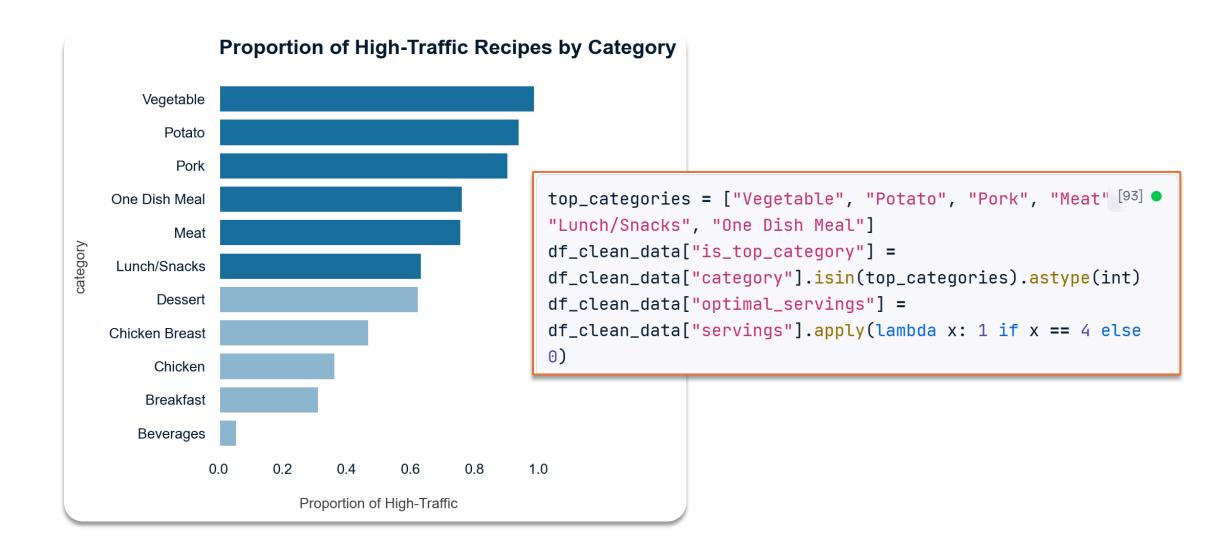




Feature engineering

Exploratory Data Analysis

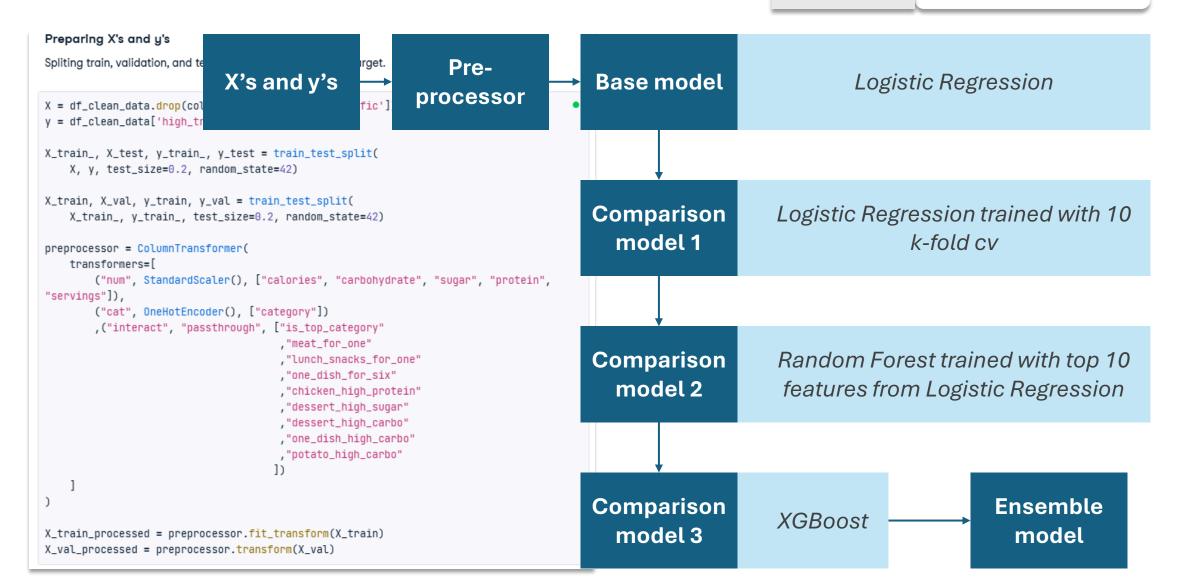
Patterns and correlations



Model selection

Modeling **Ensemble**

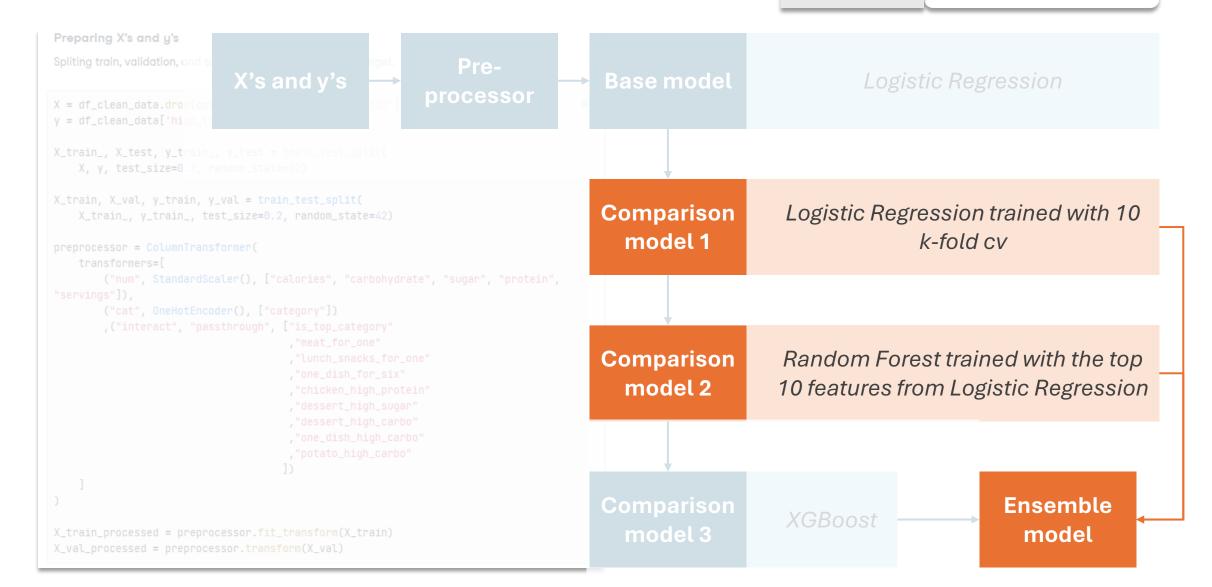
Logistic Regression and Random Forest



Model selection

Modeling **Ensemble**

Logistic Regression and Random Forest



Model selection

```
from sklearn.ensemble import VotingClassifier
ensemble = VotingClassifier(
    estimators=[("logreq", clf_tuned), ("rf", rf_pruned)],
    voting="hard" # Uses probabilities
    , weights=[2, 1]
ensemble.fit(X_train_processed_final, y_train_)
v_test_pred = ensemble.predict(X_test_processed)
print(classification_report(y_test, y_test_pred))
              precision
                          recall f1-score
                                              support
                   0.67
                             0.75
                                       0.71
                                                   73
                            0.75
                                       0.78
                   0.81
                                                  106
                                       0.75
                                                 179
    accuracy
                                      0.74
  macro avq
                   0.74
                             0.75
                                                 179
weighted avg
                   0.76
                             0.75
                                       0.75
                                                  179
```

Modeling **Ensemble**

Logistic Regression and Random Forest

Ensemble model

81% 75%

Precision Recall

An ensemble voting classifier combining **Logistic Regression** and **Random Forest**

Recommendations for the marketing team

Ensemble model

81% 75%

Precision Recall

An ensemble voting classifier combining **Logistic Regression** and **Random Forest**

The metric

Precision > 80%

Deploy ensemble model

 Build admin tool (Python + Streamlit/Flask) to predict "POPULAR"/"NOT POPULAR" w/ probability.

Monitor Performance

Track % high-traffic recipes weekly; retrain if precision
 <75%.

Hybrid human-model decision making

 Allow overrides from the product manager when model's confidence is between 60-80%.

Enhance data

- Retroactively tag missing high_traffic labels (40% gap).
- Collect user feedback (ratings, cooking time).

A/B test

 Run 2-week trial: model vs. manual picks. Measure traffic & subscriptions.

TASTY BYTES Site traffic boost

Thank you!



André Canal May 18th, 2025