Tasty Bytes

recipe analysis and popularity prediction

By Meiliang Pan

Table of Contents

Business Goals and Project Goals Overview

Business Goals

- Feature popular recipe to increase site traffic by at least 40%
- Predict which recipes will lead to high traffic(popular).
- Correctly predict high traffic(popular) recipes 80% of time.

Project Goals

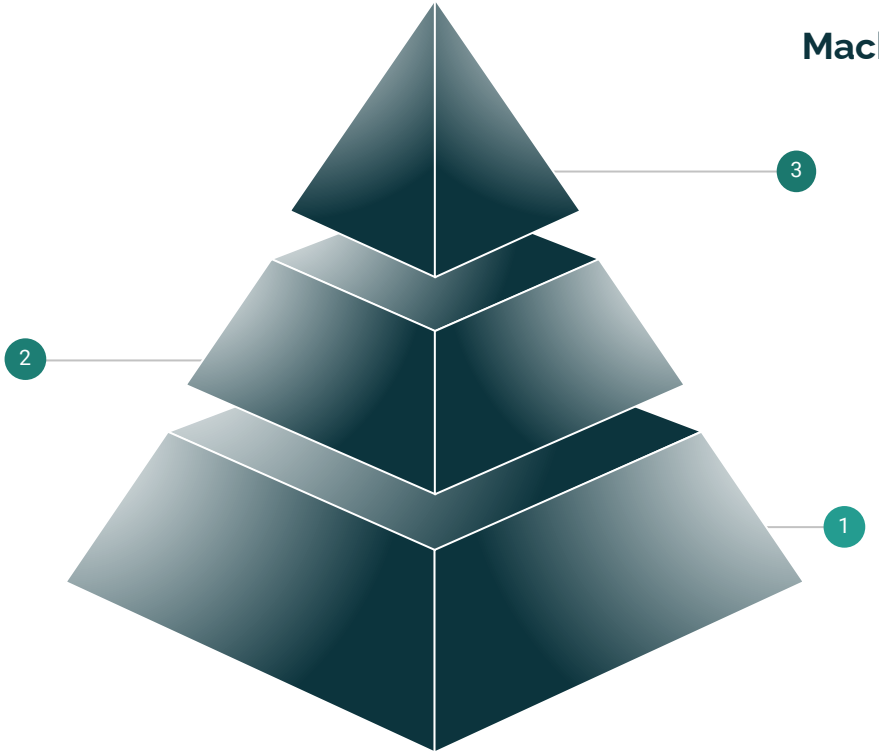
- Gain Insights of characteristics differentiate popular (high-traffic) recipes from unpopular recipes.
- Build Machine Learning models that can predict popular recipes with correctness 80% of the time.

Analysis Summary

Machine Learning Development

Picking Models:
- Logistic Regression
- Decision Tree Classifier

Train/Tune Model
Evaluate/choose best model base on business criteria.



Exploratory Analysis

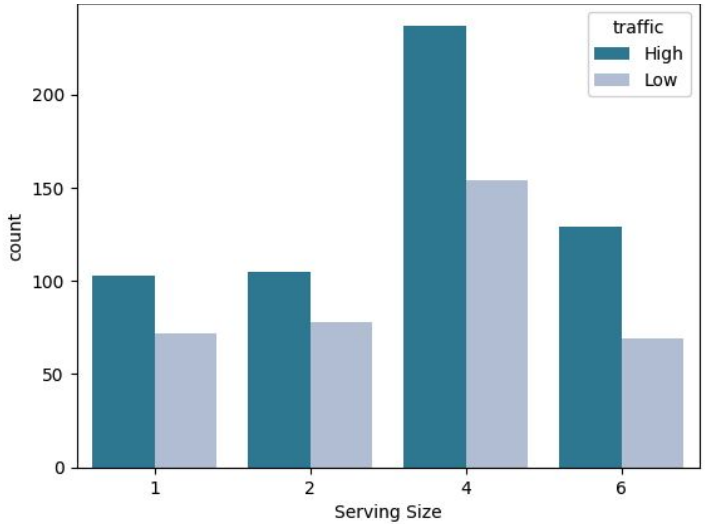
Explore variables in the dataset to gain insights of their roles in determining recipe popularity.

Data Inspection/Cleaning

- Check data for common data constraints (i.e missingness).
- Cleaning data accordingly.
- Make sure data is analysis ready.

Key Findings

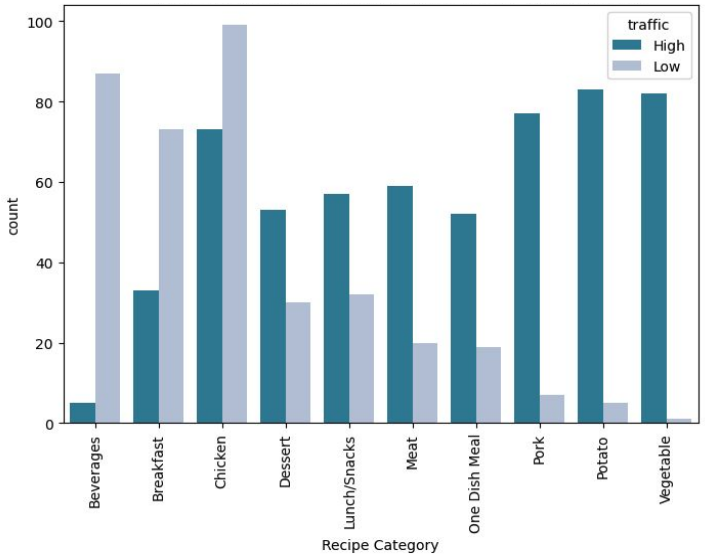
Count of each serving size in popular and unpopular recipes



From this count plot we can see the counts of popular (high traffic) recipes are higher in each serving size. This indicating that serving size has no effect on determining a recipe's popularity.

Key Findings

Counts of popular and unpopular recipes in each category



From this count plot we can see the counts of popular(high traffic) and unpopular recipes in each category.

- Category: Potato, Pork and Vegetables have the highest counts of popular recipes.
- Category: Beverages, Breakfast, and Chicken have the highest counts of unpopular recipes.

Key Findings

Nutritional Value Comparison Between Popular and Unpopular Recipes

	P	U	% Diff
Calories(cal)	328.16	282.57	14.92
Carbohydrates(g)	23.77	19.20	21.27
Sugar(g)	4.58	5.23	13.36
Protein(g)	12.47	9.62	25.80

*Values per serving
H= high traffic recipes
L = low traffic recipes

This tables shows the medium value of each nutritional variable in both popular and unpopular recipe per serving. As well as percent difference between them.

Popular recipe has higher calories, carbohydrates, protein and lower sugar.

The are more difference in carbohydrate and protein between the 2 recipe types.

Key Findings

Machine Learning Model Evaluation Scores

	Accuracy	Precision(P*)	Recall(P*)
Logistic Regression	0.779	0.81	0.83
Decision Tree Classifier	0.775	0.80	0.83

*P = popular = high traffic

KPI Compare Model Performance:

Precision: At 0.81, logistic regression model is correct 81% of the time with its predictions.

Recall: At 0.83, logistic regression model identified 83% of total high traffic recipes.

Accuracy: At 0.779, logistic regression model got 77.9% of prediction (high/low traffic) correct.

Logistic Regression is a better model, and scores indicate it met the business goals.

Business Metrics

For Tasty Bytes, **website traffic rate** is very important.
More traffic means more subscriptions. I would recommend use **website traffic as a metric.**

The product team noticed a **40%** increase previously in website traffic when display a popular recipe on the homepage.

Therefore after implementing the ML model and recommendations provided in this analysis, the website traffic rate goes up to 40% or higher, it's a good indicator that the business is on the right track to achieving its goals.

Recommendations

- Create New Recipe based on the following criteria:
  - **Categories** :
    - Pork / Potato /Vegetables /Lunch/Snacks/ One Meal /Dish Dessert/Chicken
  - **Nutrition values**:
    - Calories: ~328.155 cal Carbohydrates: ~23.77g Sugar: ~4.58g Protein: ~12.47g
  - Drop Beverage/Breakfast Category, add new category: dinner, vegan etc.
- Add new features: cooking time, cost per serving, difficulty level
- Implementing ML model, use predicted popular recipes on the homepage and monitor site traffic.
- Collect more data following changes for more in depth analysis