



Predicting Recipe Popularity for Homepage Optimization

Using Data

Project Context

What is the current problem?

Recipes for the homepage are currently chosen based on personal preference.

Impact

Homepage traffic directly affects user engagement and subscription rates.

Goal

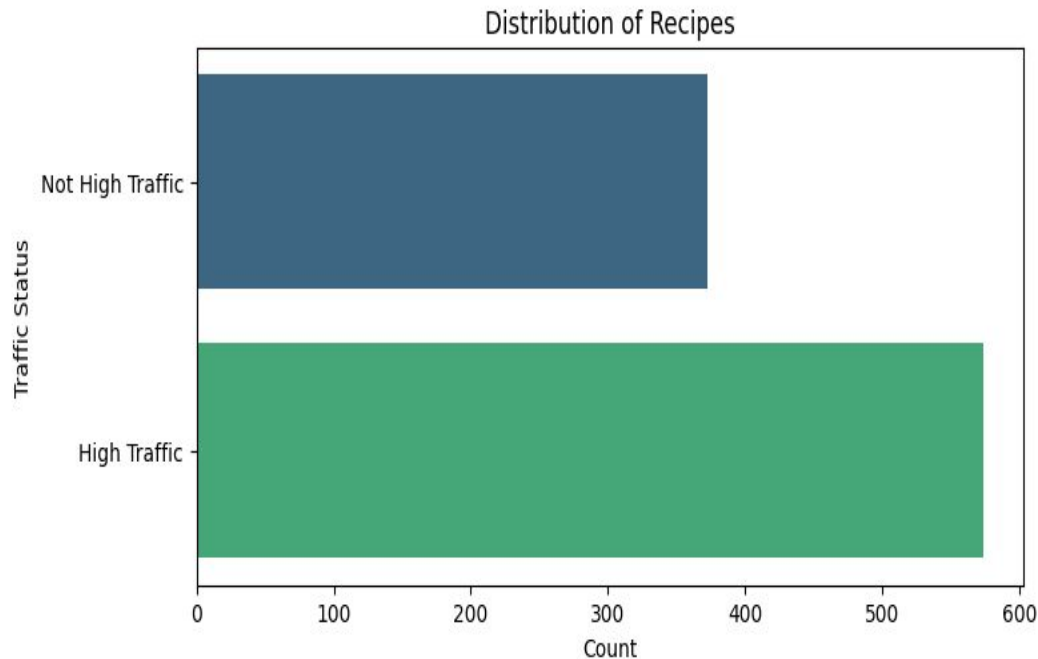
Develop a data-driven approach to select recipes that maximize traffic.

Project Objectives

1. Predict which recipes are likely to generate high website traffic.
2. Achieve 80% accuracy in predicting high-traffic recipes.
3. Increase user engagement and drive subscription growth.

Using Data

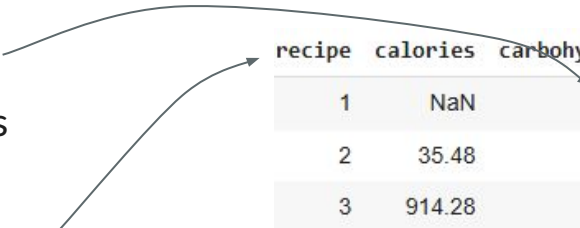
- Data includes 947 recipes
- Recipe attributes are used to predict whether or not a recipe performs high or low website traffic.
- This is a binary classification task.



Data Cleaning and Validation

1. Drop missing values

2. Drop unnecessary column



recipe	calories	carbohydrate	sugar	protein	category	servings	high_traffic
1	NaN	NaN	NaN	NaN	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	NaN
4	97.03	30.56	38.63	0.02	Beverages	4	High
5	27.05	1.85	0.80	0.53	Beverages	4	NaN

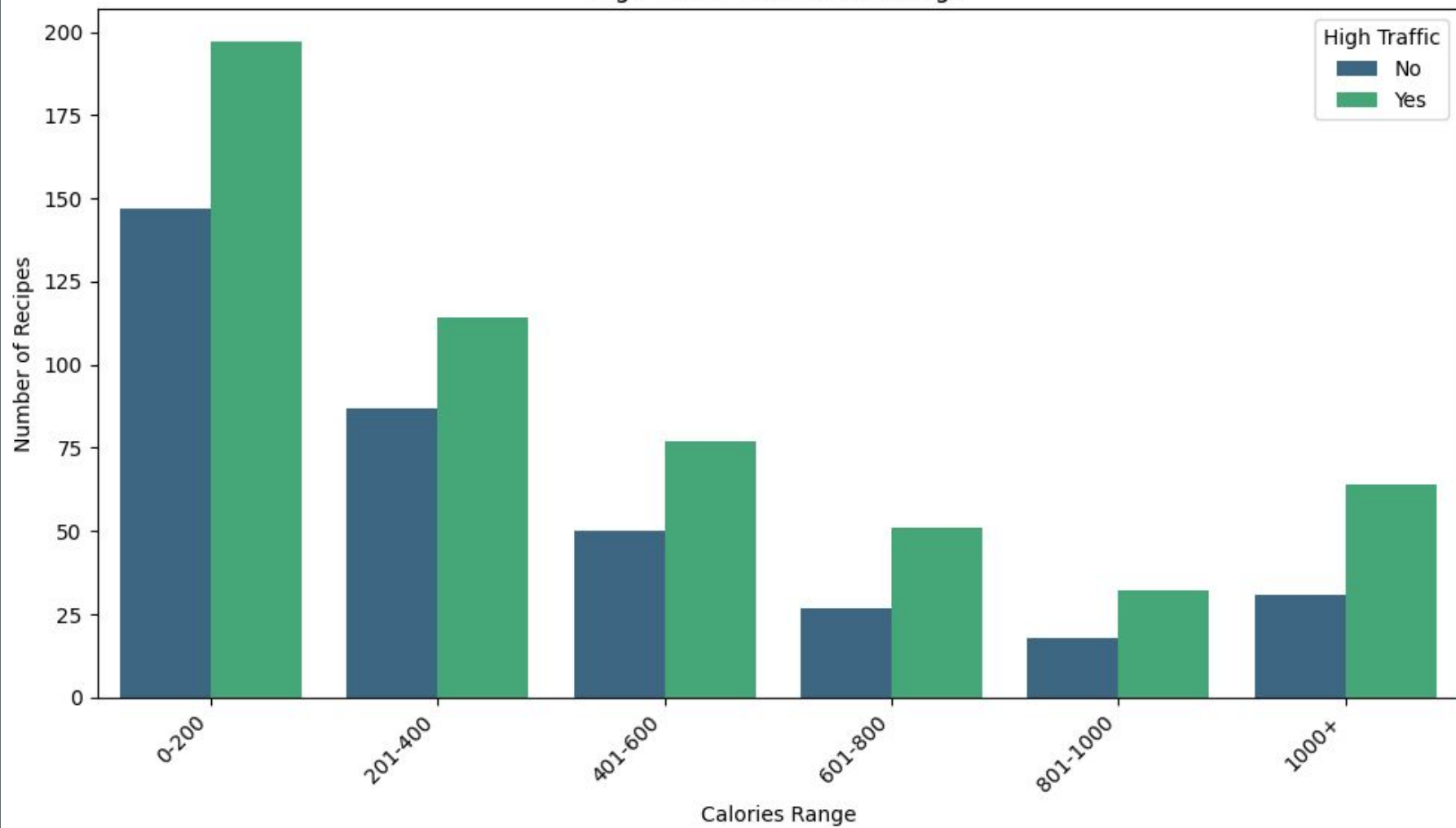
3. Data type conversions

4. Create new features from existing data

- **energy score** = (calories+carbohydrate) / servings
- **sugar to calorie ratio** = sugar / calories
- **protein density** = protein / servings
- **protein to carb. ratio** = protein/ carbohydrate

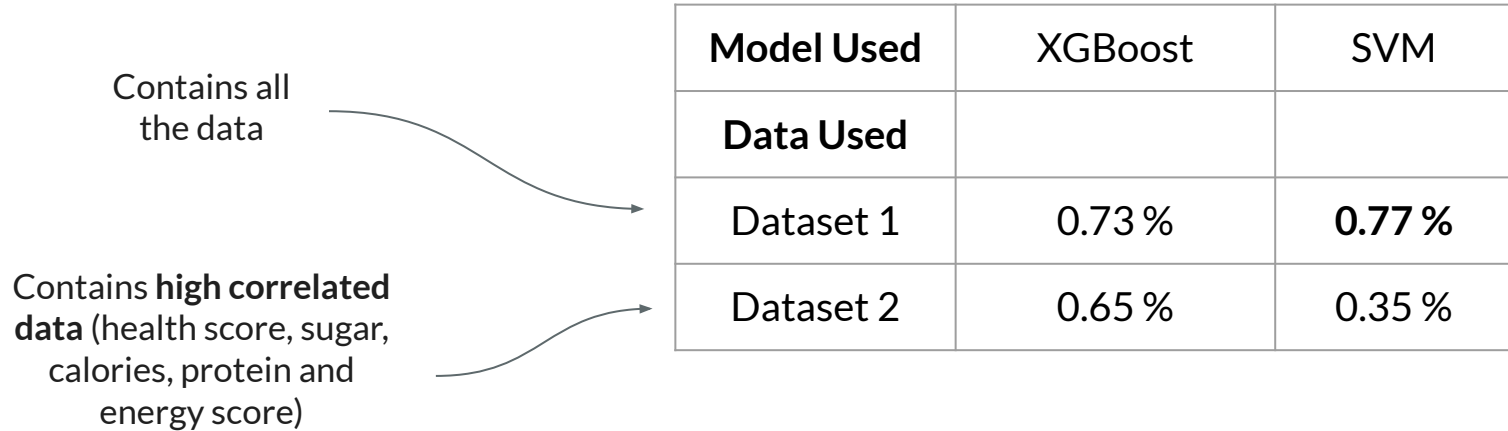
Old Categories	New Categories
Breakfast, Lunch, Snacks	Meal
Chicken, Meat, Pork	Protein
Beverages	Beverage
Potato, Vegetable	Vegetable
Dessert	Dessert

High Traffic vs. Calories Range



Model Development

The business should use **Precision for High Traffic Predictions**, a measure of how often the recipes we predict to be popular *actually* are popular.



Model Used	XGBoost	SVM
Data Used		
Dataset 1	0.73 %	0.77 %
Dataset 2	0.65 %	0.35 %

Contains all the data

Contains **high correlated data** (health score, sugar, calories, protein and energy score)

Recommendations

1. Gather additional data (user clicks, recipe ratings, social media engagement) to enhance predictive power.
2. Deploy the Support Vector Machine (SVM) model due to its superior performance.
3. Periodically retrain the model with new data.
4. Prioritize publishing recipes that are low in calories and high in protein or vegetables.

Thank you for your attention!