

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are both tilted at an angle.

Tasty Bytes

Predicting Popular Recipes

Fernando Barros Magariños



Business Goals

- Today, **product** is in charge of **manually selecting** the recipes included in the home page.
No clear criteria used for this selection.
- Selecting Popular recipes in the home page have a positive impact (**up to 40% more traffic**)
- **Goals:**
 - Predict which recipes will lead to high traffic (**accuracy**)
 - Correctly predict high traffic recipes 80% of the time? (**precision**)



Data

	recipe	calories	carbohydrate	sugar	protein	category	servings	high_traffic
0	1	NaN	NaN	NaN	NaN	Pork	6	High
1	2	35.48	38.56	0.66	0.92	Potato	4	High
2	3	914.28	42.68	3.09	2.88	Breakfast	1	NaN
3	4	97.03	30.56	38.63	0.02	Beverages	4	High
4	5	27.05	1.85	0.80	0.53	Beverages	4	NaN

- 947 recipes
- Selection of Features:
 - Recipe id
 - Calories, carbohydrates, sugar, protein, category, servings (6 explanatory variables)
 - High_traffic (target)

Key Findings

- Other explanatory variables available not used in this first analysis

Example Recipe

This is an example of how a recipe may appear on the website, we haven't included all of the steps but you should get an idea of what visitors to the site see.

Tomato Soup

Servings: 4

Time to make: 2 hours

Category: Lunch/Snack

Cost per serving: \$

Nutritional Information (per serving)	
Calories	123
Carbohydrate	13g
Sugar	1g
Protein	4g

Ingredients:

- Tomatoes
- Onion
- Carrot
- Vegetable Stock

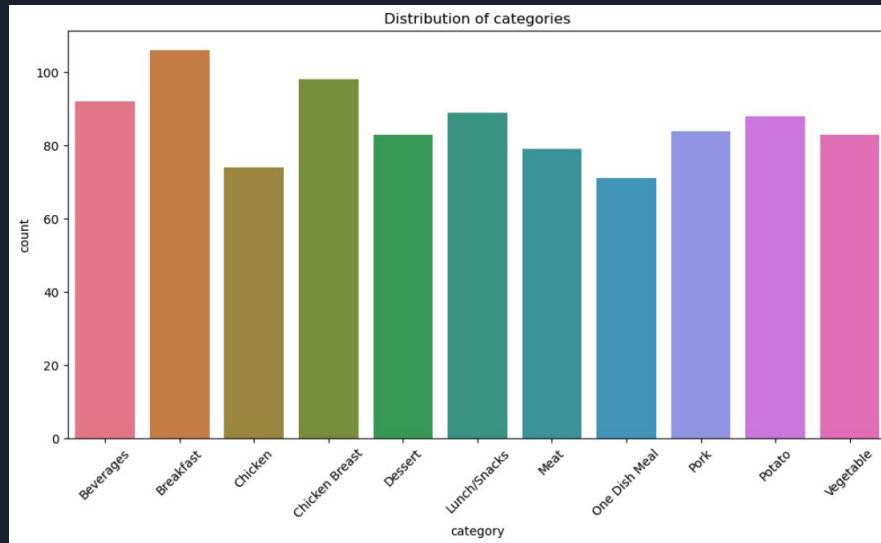
Method:

1. Cut the tomatoes into quarters...

- Other information like the cost, time to make or ingredients (vegan, veggie...) could be of help
- This information seems to be already available in the platform

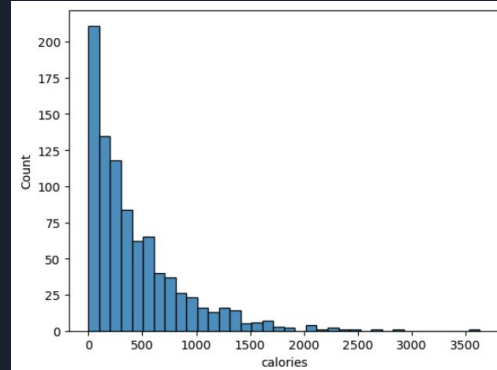
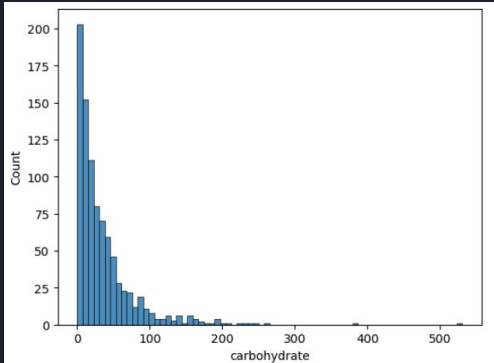
Key Findings II

- Approximately same amount of recipes for each of the 11 categories



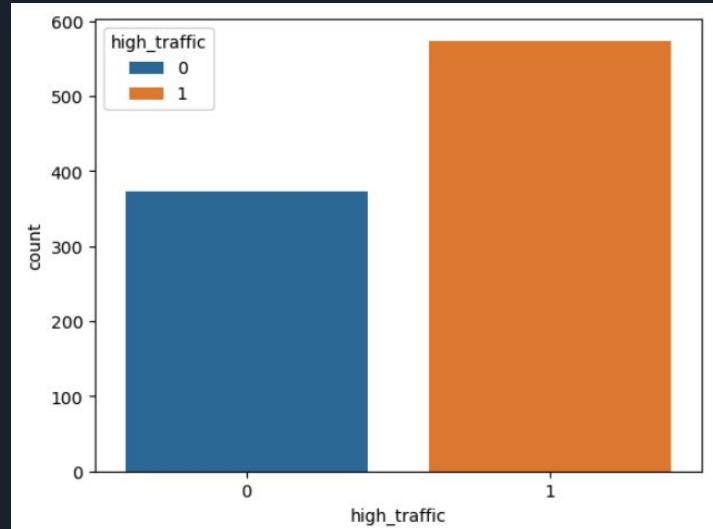
Key Findings III

- Calories, sugar, carbohydrate and protein present missing values.
- Their distributions have long tails



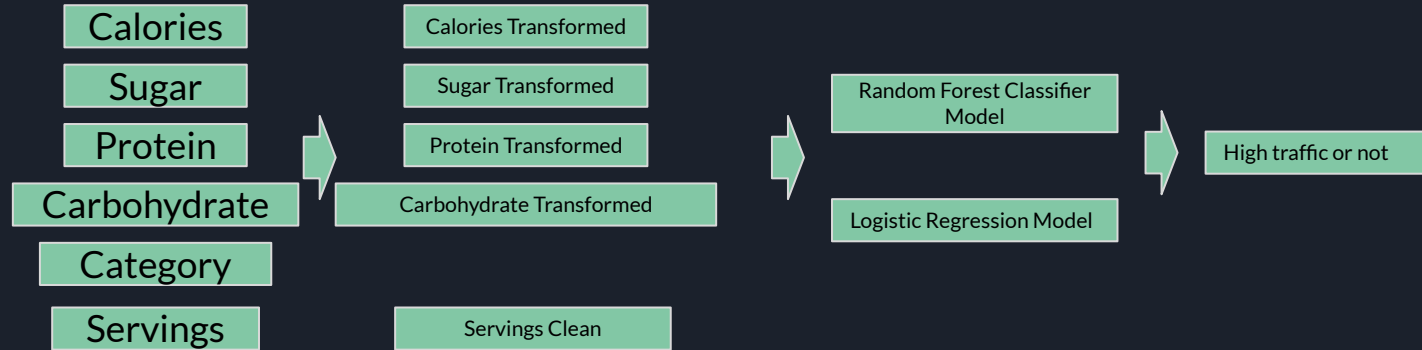
Key Findings IV

In the dataset, 60% high traffic and 40% not high traffic



Outcomes

We trained 2 models





Outcomes II

- Two metrics: accuracy and precision.

Metric	Accuracy	Precision
Description	The ratio of correct predictions to total predictions made	The ratio of true positives to all positive predictions.
Range	0-1 (1 is the best)	0-1 (1 is the best)



Outcomes III

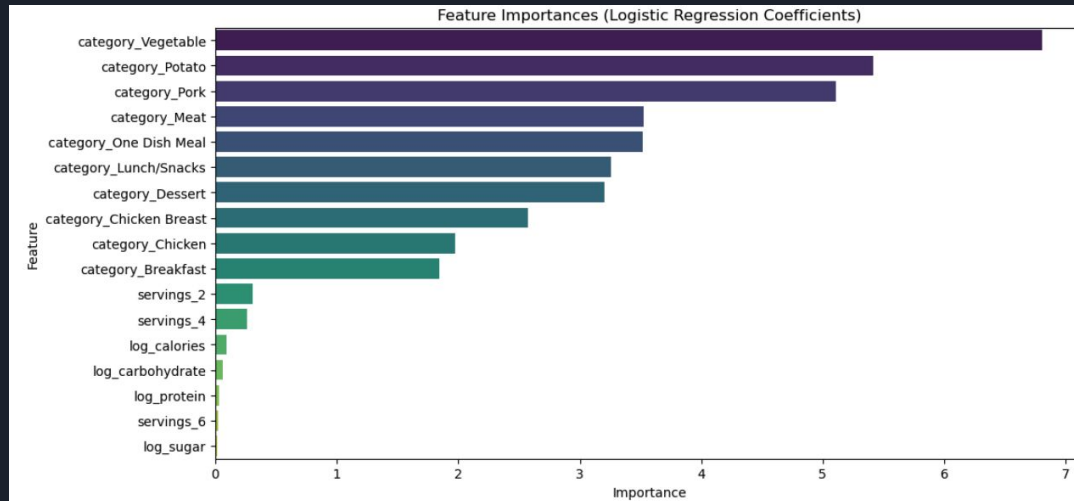
- Two metrics: accuracy and precision.

Metric	Accuracy	Precision
Logistic Regression	0.75	0.87
Random Forest	0.74	0.74

Outcomes IV

- The Logistic Regression Model gives more importance to

Category >> Servings >> Nutritional Information





Recommendation

- Identify and fix errors to improve accuracy after testing.
- Use the current model to assist the manual process.
- Automate the selection of the recipes for the home page.
- Continue collecting data.
- Assess the possibility of including more readily available features to retrain the model.
- Continuously monitor and improve the model.
 - New recipes.
 - Performance degradation (change in trends).
- Reach out to me if you have any question :)