

Tasty Bytes

Recipe Site Traffic

The project review

Recipes on the Tasty Bytes has different bits of information.

The following data was provided by the product team:

- Category of the recipe (Lunch, Meat, etc., total: 10)
- Servings (1, 2, 4 or 6)
- Nutritional info: Calories, Sugar, Protein (per serving)

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....

Business aims

Traffic to the rest of the Tasty Bytes goes up by as **much as 40%** if a user picks a popular recipe.

Two goals are formulated:

- Predict **which recipes will be in high traffic**
- Predict high value of traffic with **80% probability**

Outline of the work

- validation of dataset with 947 recipes (data cleaning, missing values filling, etc.)
- data visualization (correlations between variables, most prominent predictors)
- model fitting (fit two models for binary classification)
- model selection and evaluation (calculating and comparing metrics)
- KPI definition and business advice

Modelling details

Solving binary classification problems with the help of

- Logistic regression
- Random forest

Similar performance, the former is slightly better and achieved all business goals

	Logistic Regression	Random forest
Precision	80 %	80%
Recall	80%	75 %
F1 Score	80 %	78 %
Accuracy	76 %	74 %

Key performance indicators

We opt to use **accuracy** as a KPI since the product teams wants to increase the quality of predictions to increase revenue.

Logistic regression has better KPI compared to the other model

	Logistic Regression	Random forest
Precision	80 %	80%
Recall	80%	75 %
F1 Score	80 %	78 %
Accuracy	76 %	74 %

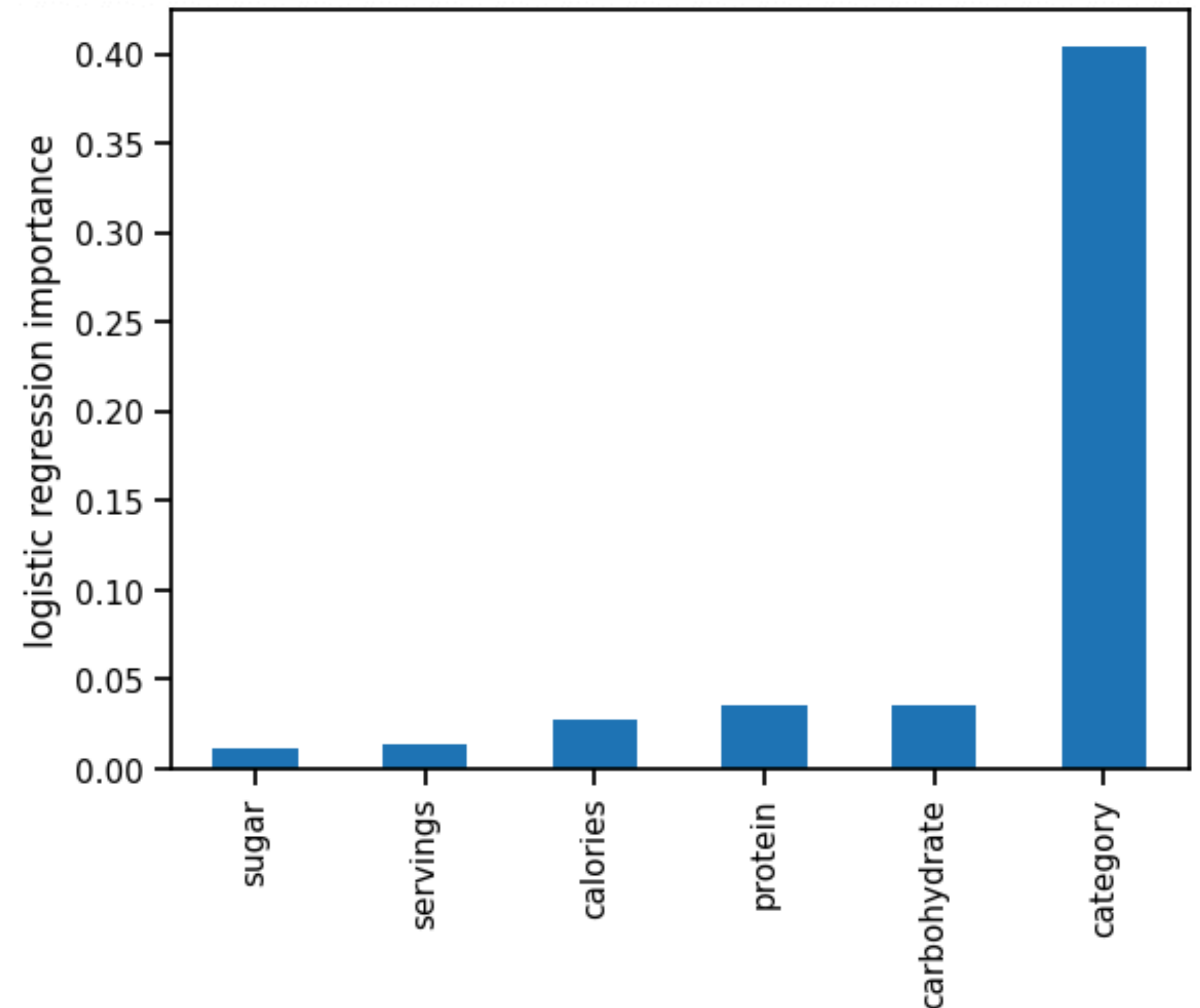
Recommendations

Data analysis and machine learning models suggest that the **category of a recipe is a defining feature**.

For example, top-3 categories with highest traffic:

- Pork
- Potato
- Vegetables

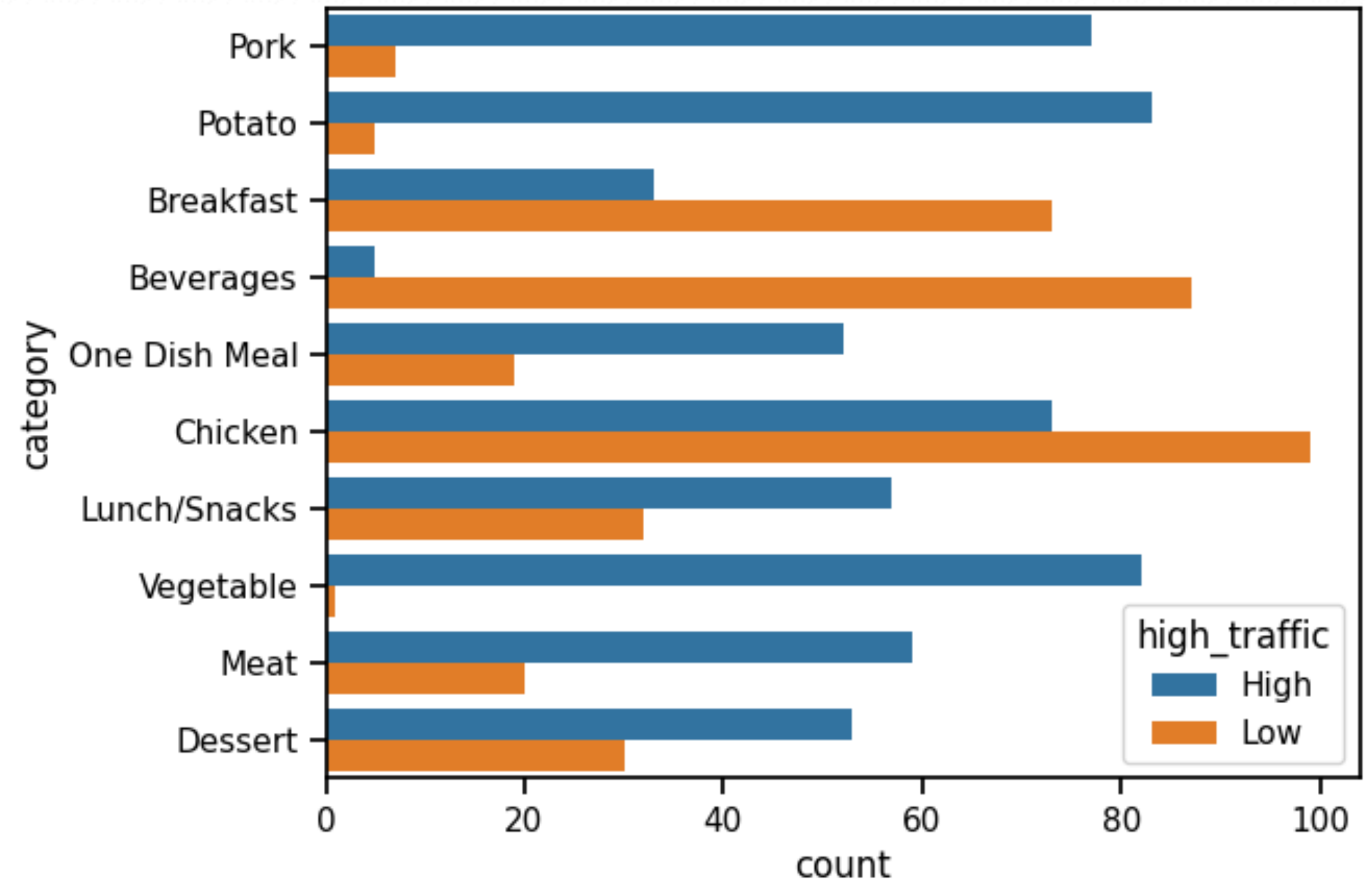
Beverages is a category with the lowest rate of high traffic.



Recommendations

Further suggestions

- more example of top categories
- use online learning approach (when we constantly enhancing the model as new data arrives)
- add relevant features such as cost of the recipe, time to prepare and cook the meal, history of user's visited recipes.



Thank you