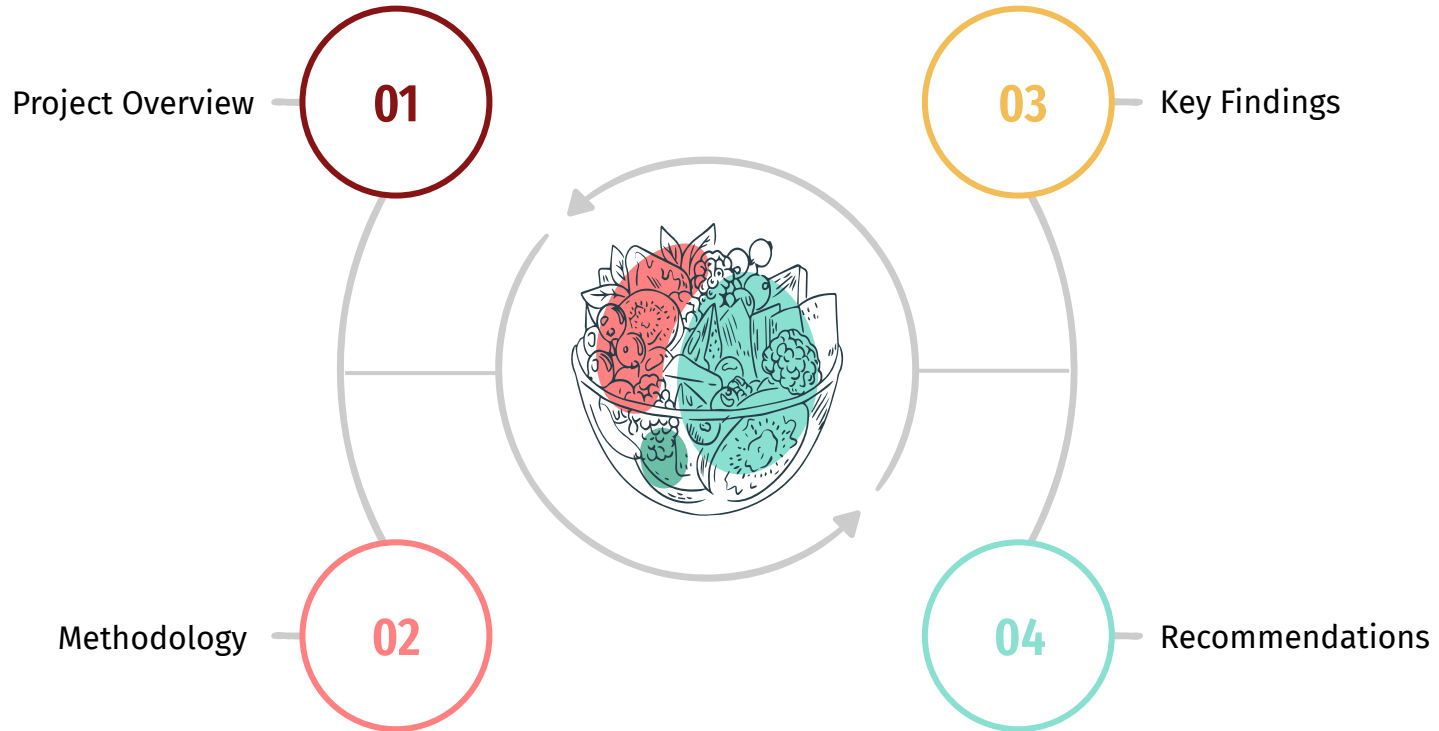


Recipe Site Traffic

Yusuf Okunlola

DataCamp Certified Data Scientist Professional

Table of Contents



Project Overview

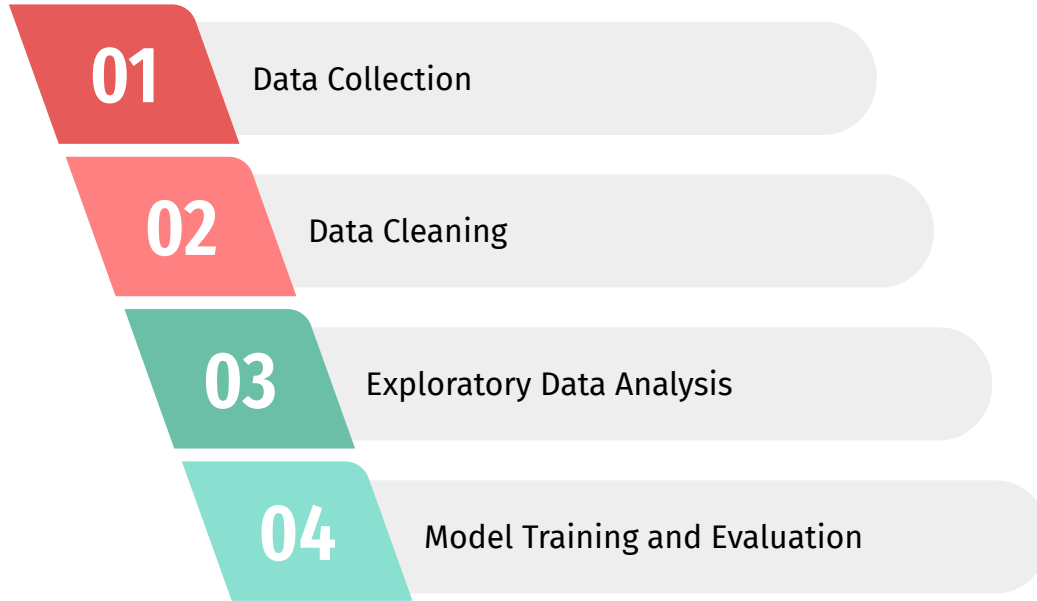
Launched in 2020, Tasty Bytes began as a recipe search tool, helping users make the most of available ingredients they had at home. Now an established company, they provide comprehensive meal plans via subscriptions, ensuring nutritious meals regardless of budget. Premium subscribers enjoy doorstep ingredient delivery.

They've observed a 40% increase in overall website traffic when a popular recipe is featured on the homepage. However, determining recipe popularity remains a challenge, crucial for boosting subscriptions and company success.

Goal

Our goal is to predict recipes driving high website traffic and attain 80% accuracy in predicting high-traffic recipes.

Methodology



Initial Steps

Data Collection

The project's dataset was obtained and comprised 947 rows and 8 columns. The attributes encompass recipe, calories, carbohydrate, sugar, protein, category, servings, and high_traffic.

Python libraries were utilized for both data analysis and machine learning purposes.

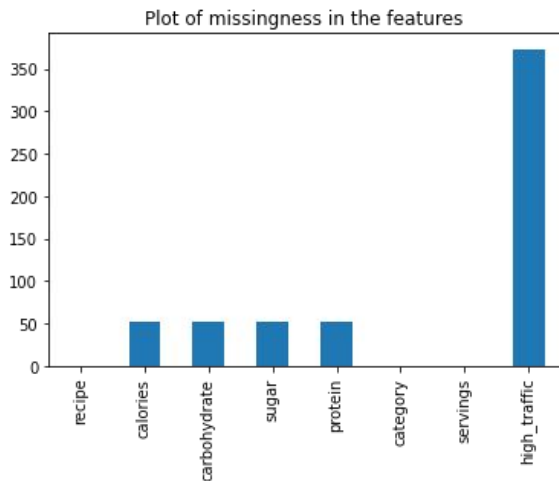


Figure 1

Data Cleaning

This step involved preserving and improving the dataset quality. The dataset was duly assessed visually and programmatically to find anomalies that required data cleaning. The subsequent actions encompassed:

- The 'servings' and 'category' features were formatted to meet the data dictionary criteria.
- Missing values in the 'calories', 'carbohydrate', 'sugar' and 'protein' fields were replaced with mean value of respective features.
- The 'recipe' column was dropped.
- Rows that contain null values in 'high_traffic' field were replaced with 'Low'.
- The 'high_traffic' column was renamed to 'traffic'.
- Categorical fields such as 'category' and 'traffic' were remapped.

Exploratory Data Analysis

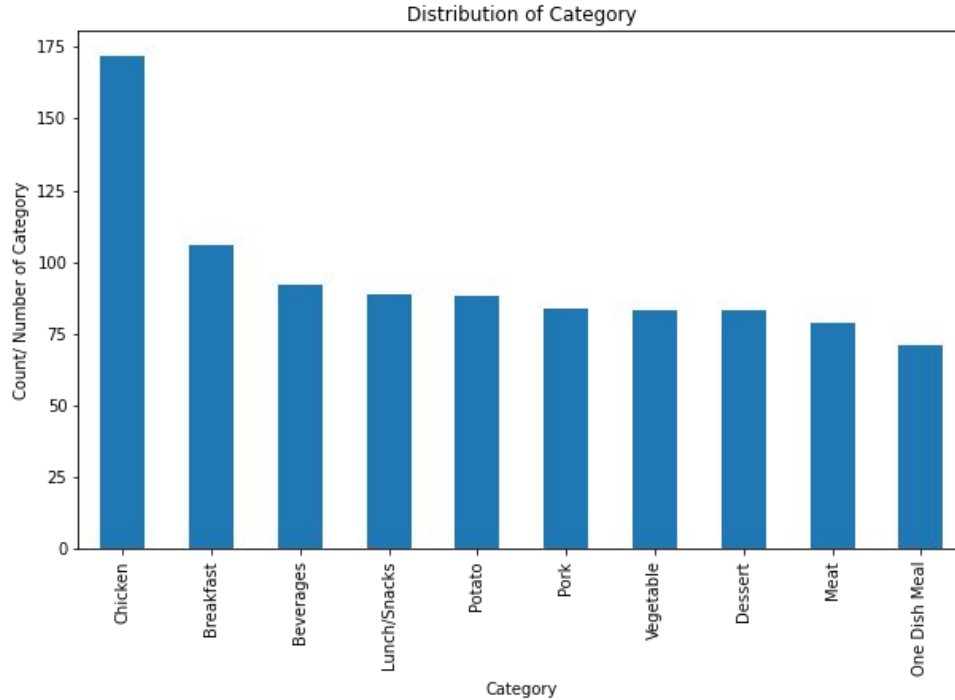


Figure 2

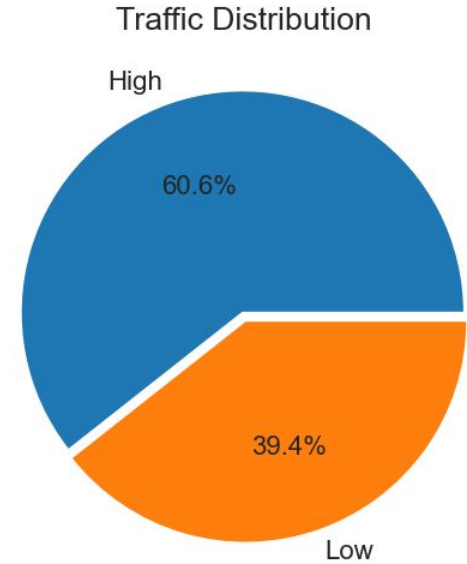


Figure 3

Exploratory Data Analysis

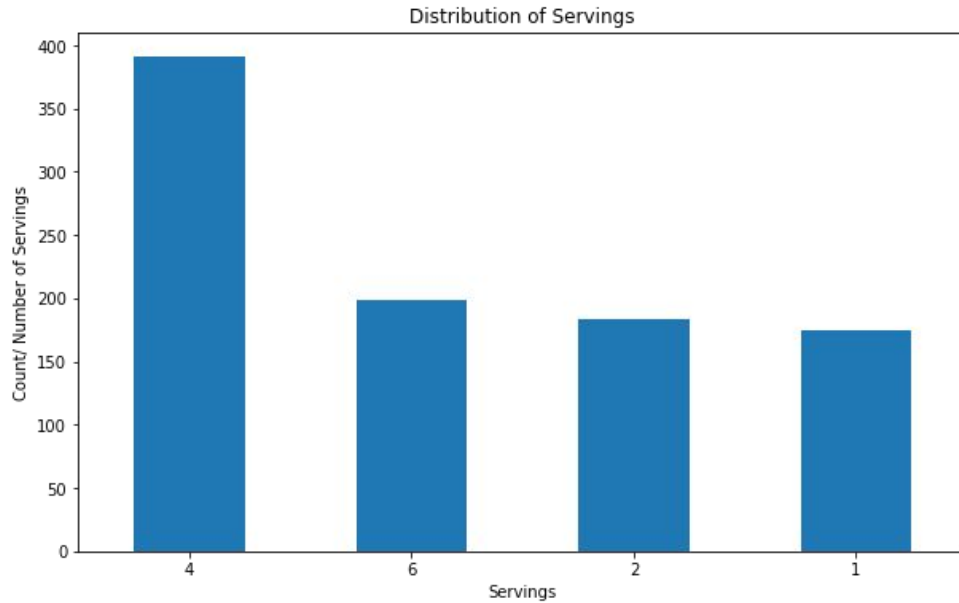


Figure 4

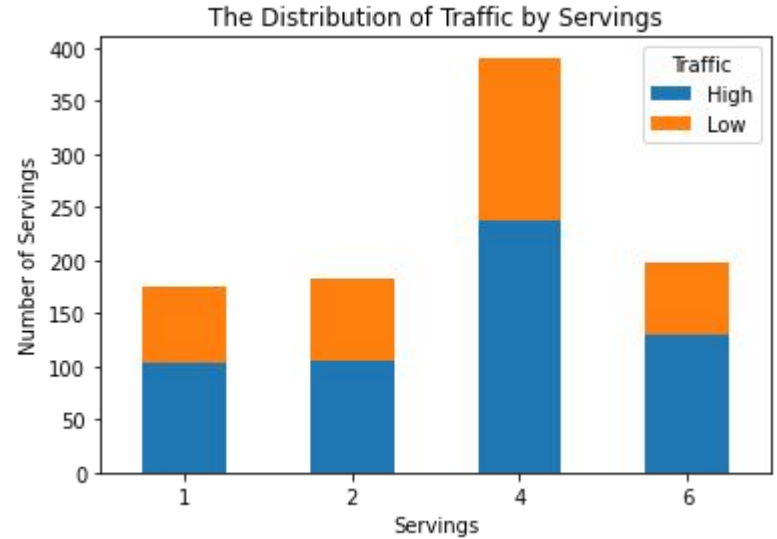


Figure 5

Exploratory Data Analysis

The Distribution of Traffic by Category

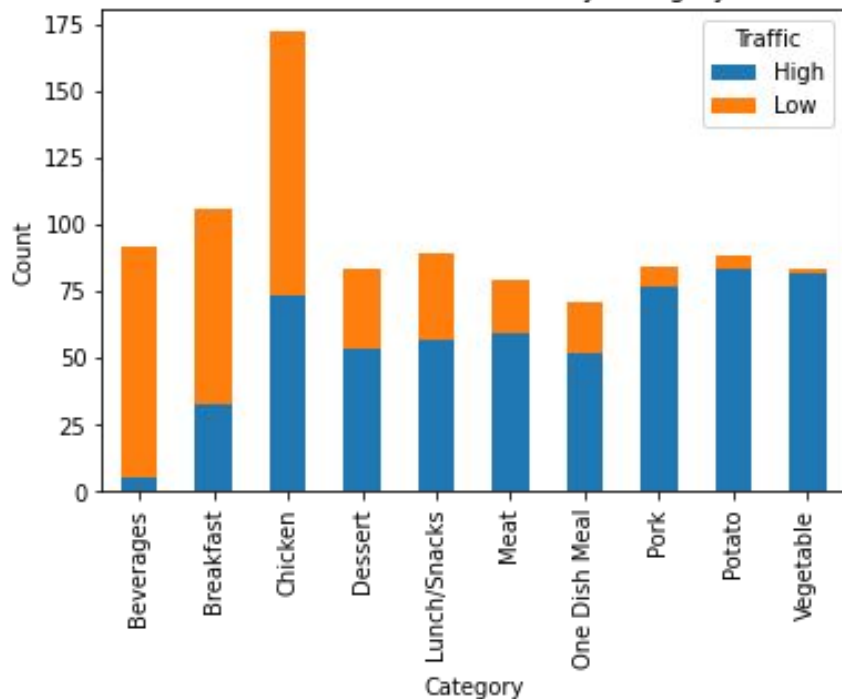


Figure 6

The Distribution of Servings by Category

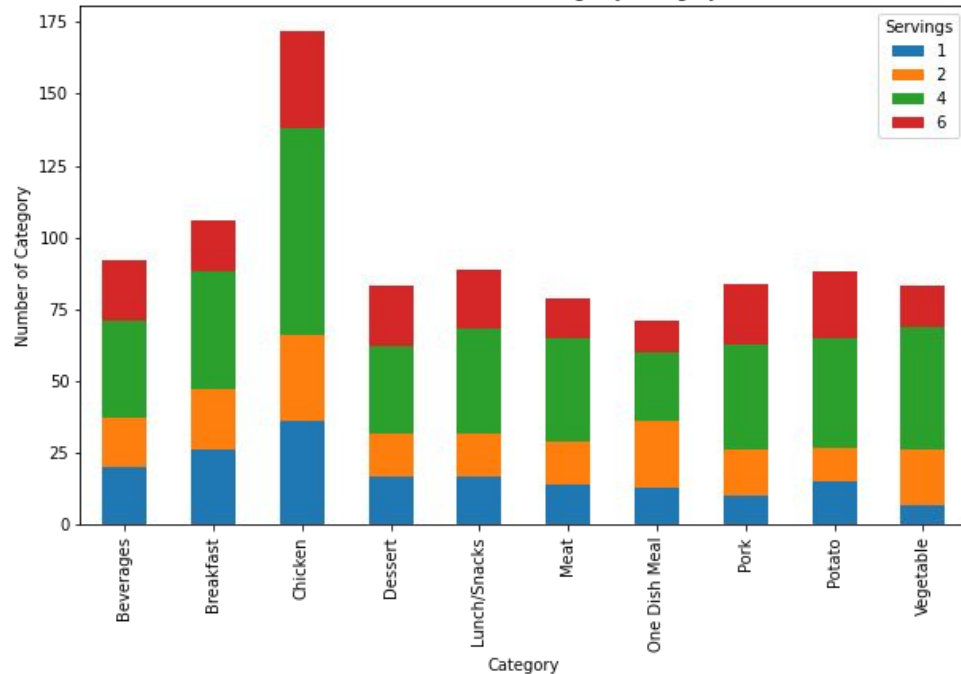


Figure 7

Model Training and Evaluation

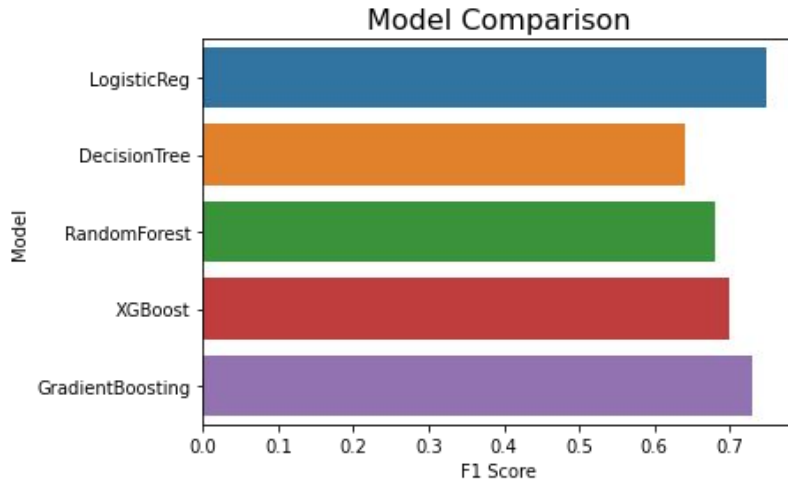


Figure 8

- The explanatory variable (X) and response variable (y) were specified from the dataset and thereafter normalized using a MinMax scaler.
- The dataset was split into a training set and a testing set of 80:20.
- Five classification models were trained on the dataset using default parameters and evaluated with the Accuracy and F1 scores.
- The F1 score metric was used as it gave an optimal blend of recall and precision metrics for classification task.
- Our preferred model, Logistic Regression model had a F1 score of 0.75 (75%), as shown in Figure 6.
- The model was tuned using GridSearchCV and a F1 score of 0.80 (80%) was obtained.

Model Training and Evaluation

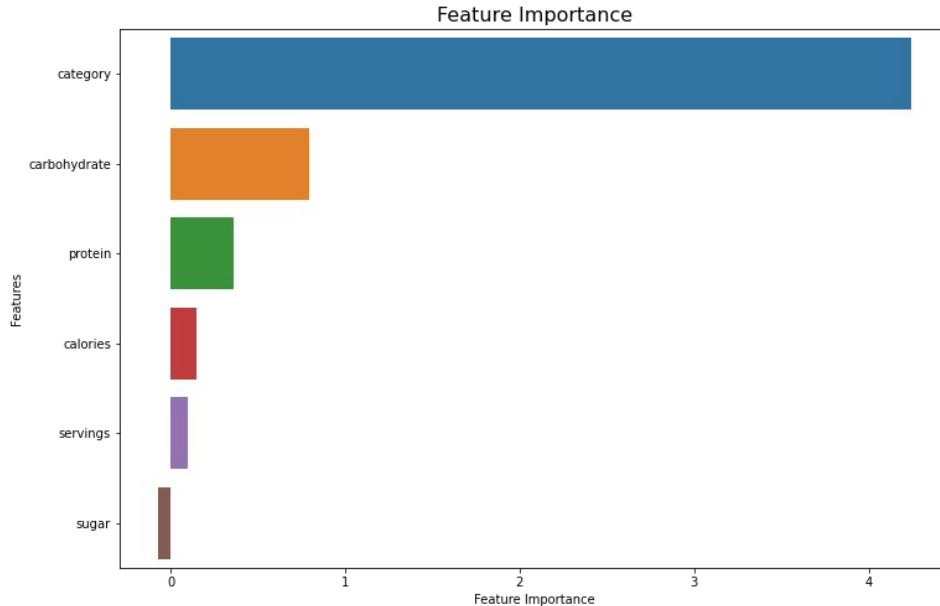


Figure 9

- Utilizing the Logistic Regression model, our aim was to uncover the main catalysts for increased website traffic. The 'category' attribute surfaced as the most influential factor, trailed by carbohydrate and protein, as illustrated in Figure 9.
- Notably, recipes with elevated sugar content were linked to reduced traffic, possibly due to the maintenance of healthy lifestyle by subscribers. Consequently, it's unsurprising that such recipes ranked lower in the analysis.

Key Findings

- Examining the pie chart, it's evident that 60.6% of recipes result in high traffic, while 39.4% yield low traffic.
- Chicken is the most frequently ordered category, with Breakfast being the next popular choice, while One Dish Meal is the least preferred. This suggests that chicken-based recipes tend to provide higher calorie content. Consequently, foods with greater calories exhibit a heightened likelihood of selection, resulting in increased traffic.
- Categories like Potato, Vegetable, and Pork are associated with high traffic and this is due to their high nutritional contents.
- Regarding servings, '4' exhibits the most frequent traffic, succeeded by '6', '2', and '1'. This suggests subscribers live an above-average lifestyle, opting for moderate food consumption to maintain health. However, insights could emerge from Tasty Bytes' operational area and customer demographics data.

Recommendations

I would recommend that Tasty Bytes should do the following in order to boost revenue:

- Using key metrics to monitor whether there is a strong sign to see the elevated website traffic.
- Cost per serving should be provided on the website for visitors to make choice.
- Display food categories with substantial carbohydrate content and protein more often.
- Display food with high calorie contents and nutritional value.
- Data Collection for in-depth analysis
 - Operational area and customer demographics
 - Time to make and cost per serving (\$) as these features might influence customer's decision.

Thank You!

Do you have any questions?

Connect via: [LinkedIn](#)

CREDITS: This presentation template was created by **Slidesgo**, including infographics & images by **Freepik**.