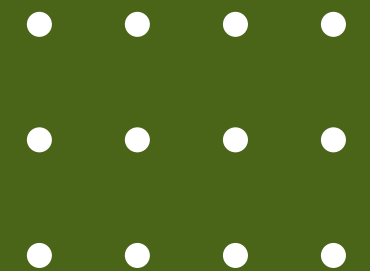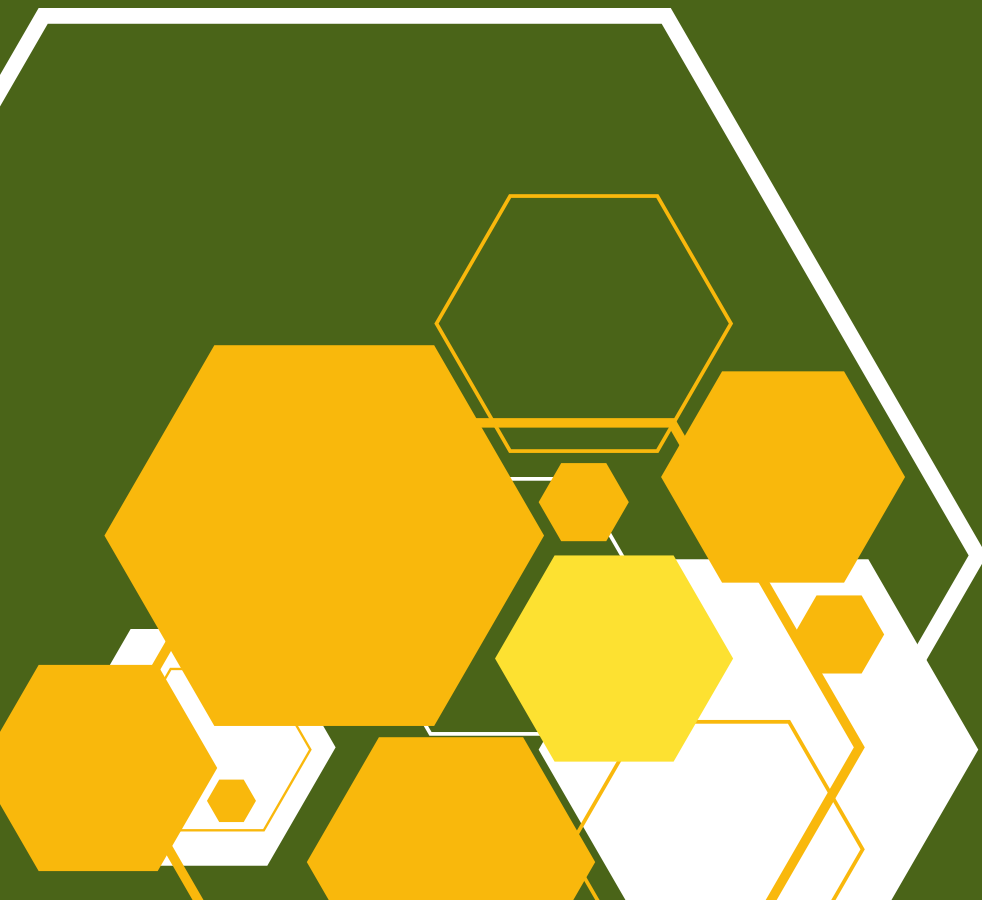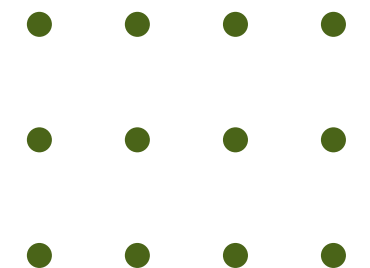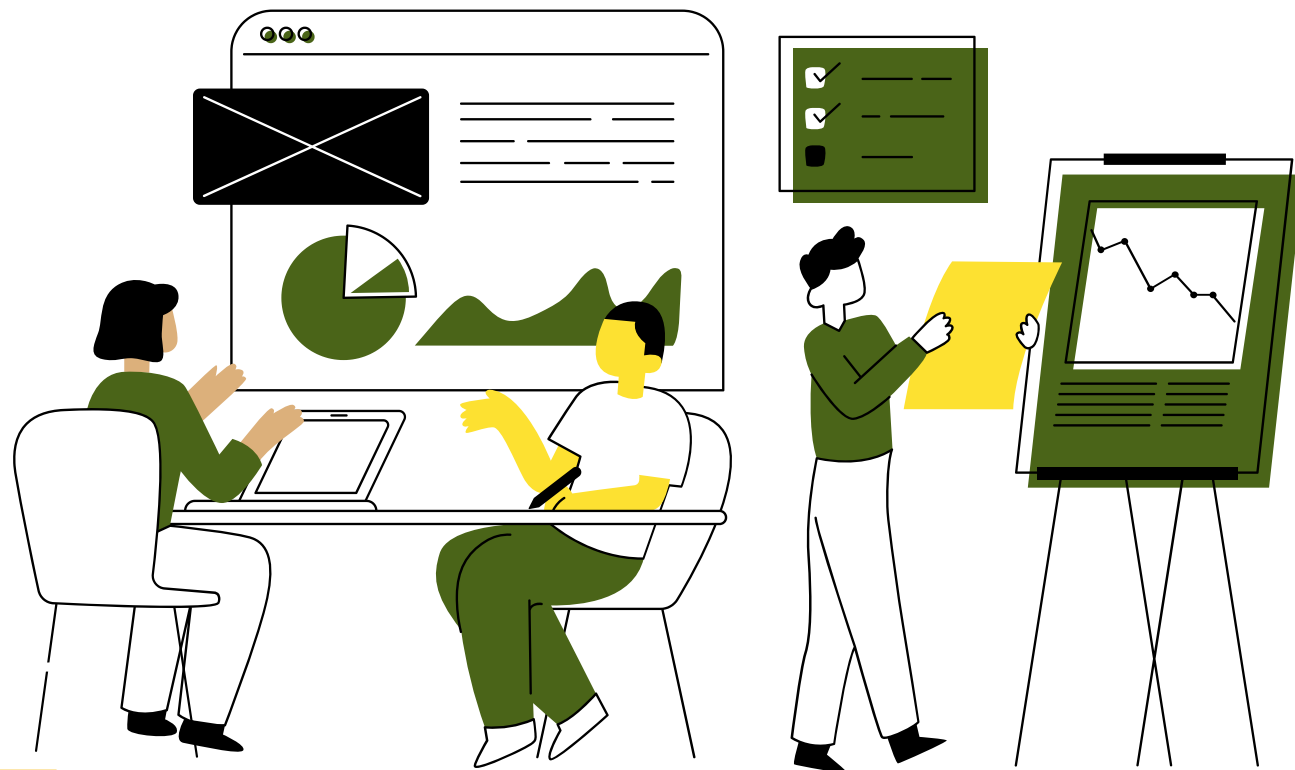# Taylor Baker

Data Prediction

## ON HIGH WEBSITE TRAFFIC

Project Overview

- Develop a predictive model that can identify popular recipes with an accuracy of at least 80%.

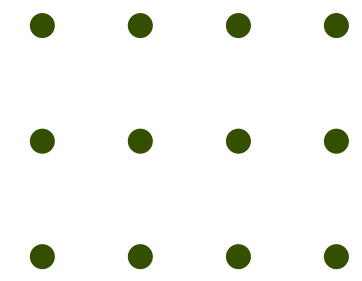- Minimize the chance of displaying unpopular recipes to users, enhancing the overall user experience.

Business Goal

Improve user engagement and satisfaction on the website creating recipes that are likely to be popular, thereby increasing user subscriptions.
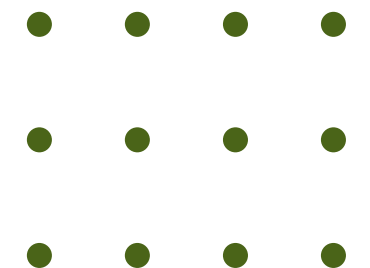
Data Exploration And Pre - Processing

01
Data Cleaning
Fill missing values, address any outliers

02
Correlation
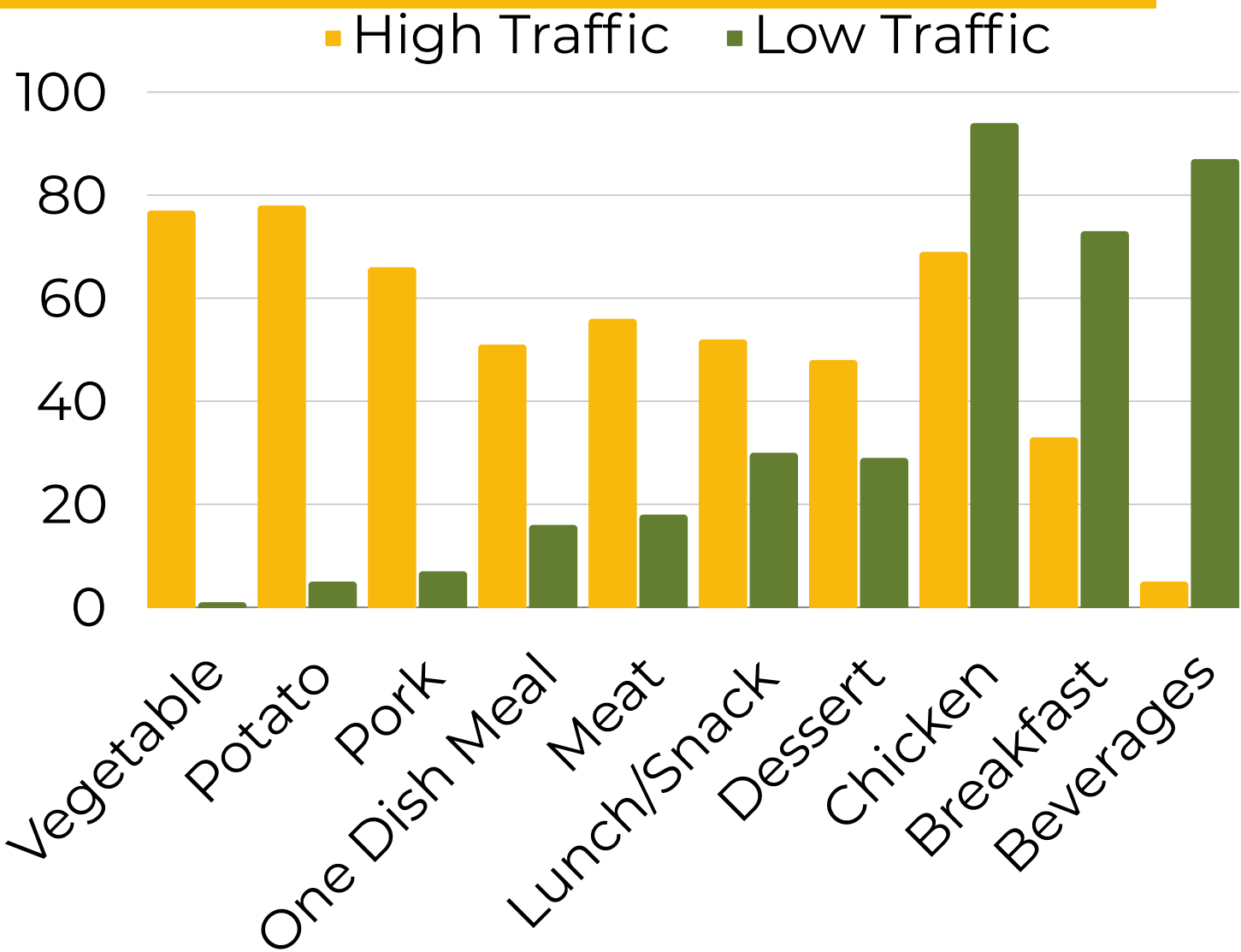Explore correlations between features and the target variable identify any patterns or trends
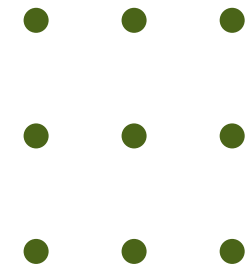
03
Model Prediction
Explore and test classification models

Meal Category Vs Traffic

■ High Traffic   ■ Low Traffic

100
80
60
40
20
0

Vegetable, Potato, Pork, One Dish Meal, Meat, Lunch/Snack, Dessert, Chicken, Breakfast, Beverages

One of the biggest factors when choosing a recipe is what the type of recipe is.

The largest high traffic to low traffic ratio is displayed left to right.

Chicken, Breakfast, Beverages are all poor traffic performers.

Vegetable has a 99% high traffic ratio.

Removing chicken, breakfast, and beverages increases our overall high traffic ratio to 75% and low traffic 25%.

Vegetable, potato, and pork give the highest chances of having a recipe get high traffic. (94% high traffic)
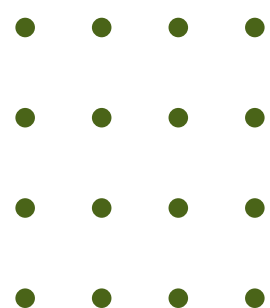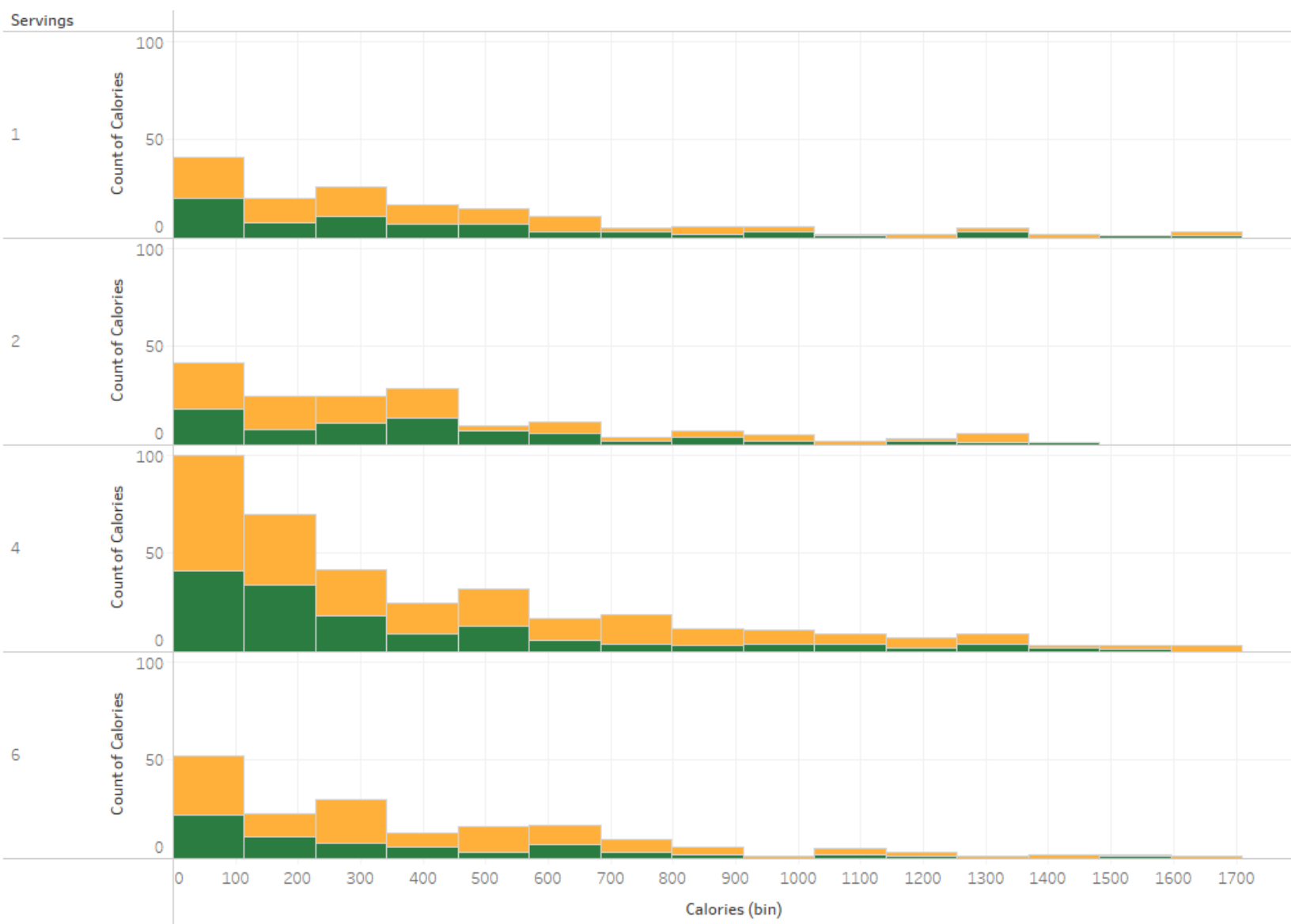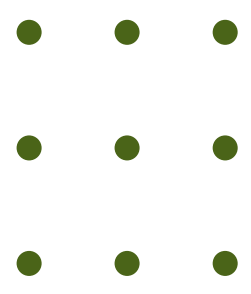
Serving Size and Calories

The other large factor is how many people are you cooking for.

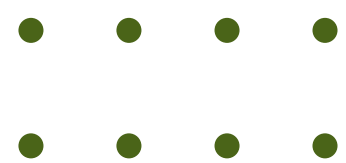**What size of meals are people that subscribe to this service looking for?**
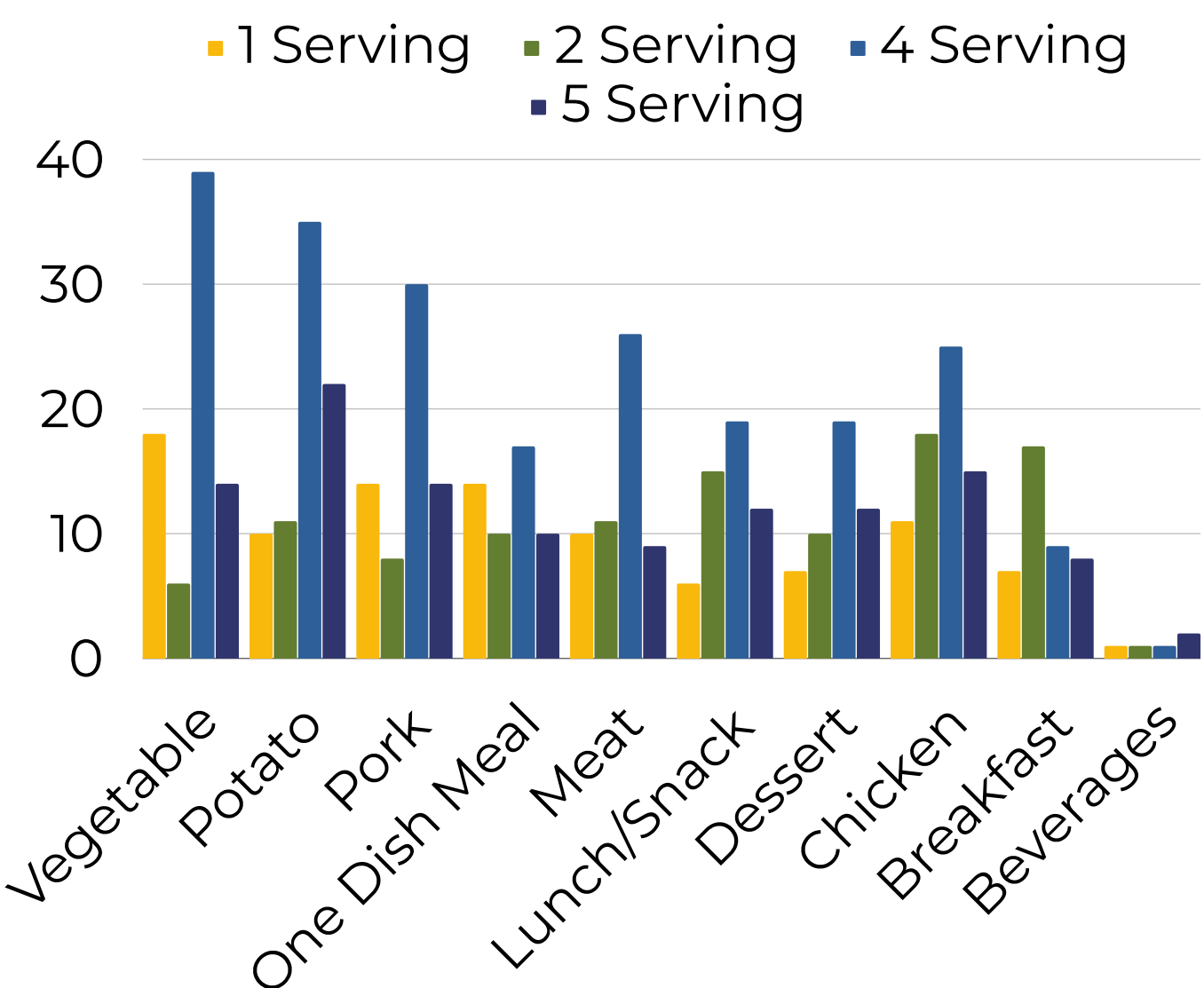

4 servings gets the largest amount of recipes

0 - 100 calories contains the most recipes for each serving size.

This does not show a distinct difference between servings sizes in relation to high traffic.

Category by Serving Size

What patterns do we see when looking at only high traffic across category and servings?

High Traffic

Legend:
- 1 Serving
- 2 Serving
- 4 Serving
- 5 Serving

Categories (x-axis): Vegetable, Potato, Pork, One Dish Meal, Meat, Lunch/Snack, Dessert, Chicken, Breakfast, Beverages

4 servings has the largest amount of recipes

2 servings has the lowest amount of recipes in the categories that get the most high traffic.

5 servings has more recipes in potato than other categories.

I assumed lunch/snack would have high amounts of 1 serving recipes, but I was wrong its actually in vegetable. That might be because some vegetable recipes could be considered snacks.
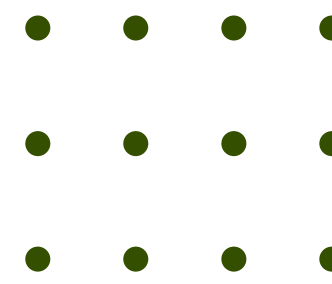
Calories Vs Traffic

Does calories have higher traffic for any specific calorie amount?

- Calories shows a higher amount of high traffic vs low traffic, but that could just be because we have more high traffic data.

- Between 0 - 100 calories we see the most recipes with high traffic.
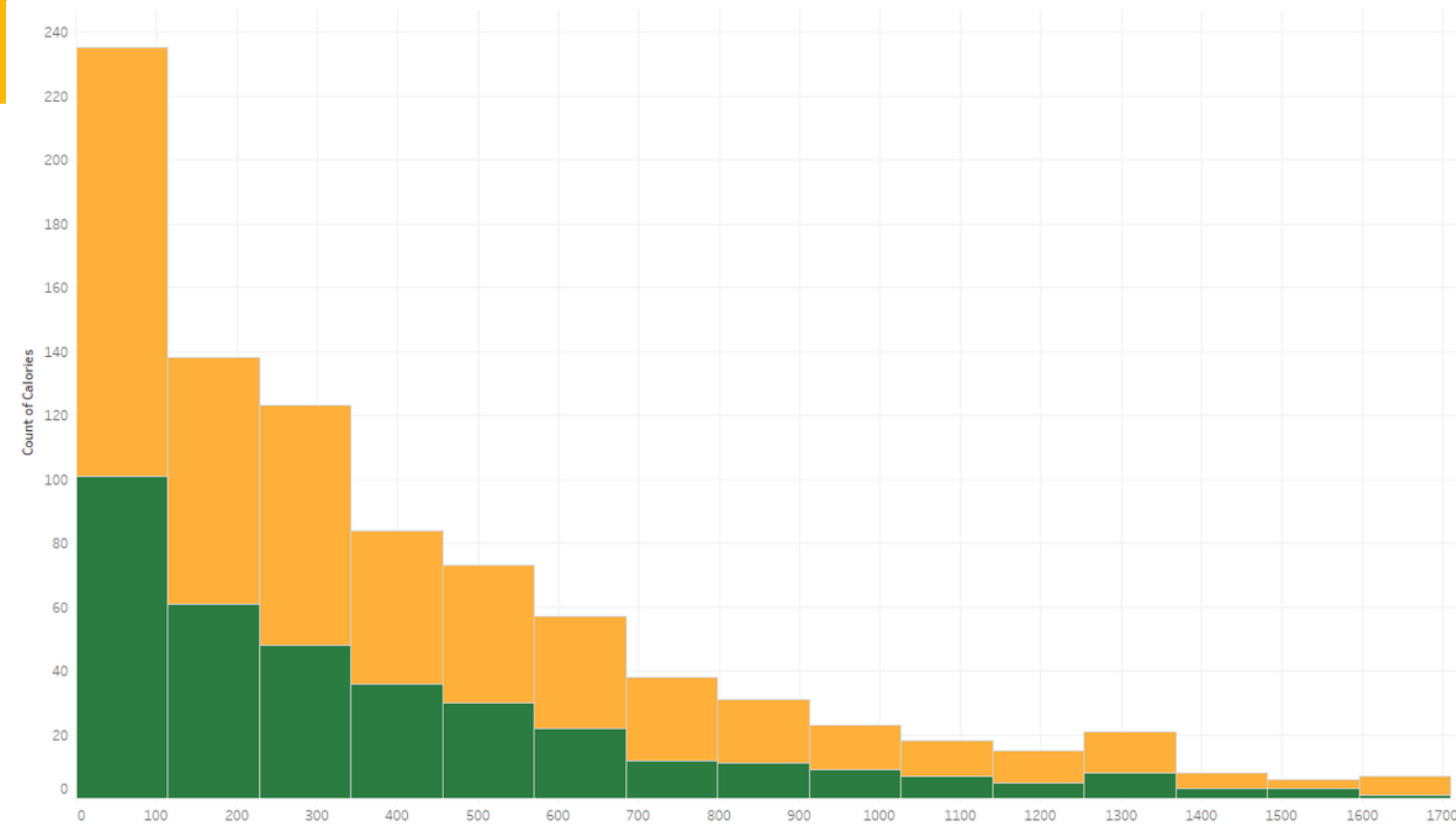
Model Predictions

Dummy Classification Evaluation:

- Metric - Precision give the True Positive prediction %
- The model accurately predicted(Precision) the True Positive('High Traffic') 60% of the time.
- We achieved this by choosing high traffic for every prediction in the test set.
- That means we had 100 high and low traffic variables and 60 of them were high and 40 were low. We would get a 60% true positive prediction.

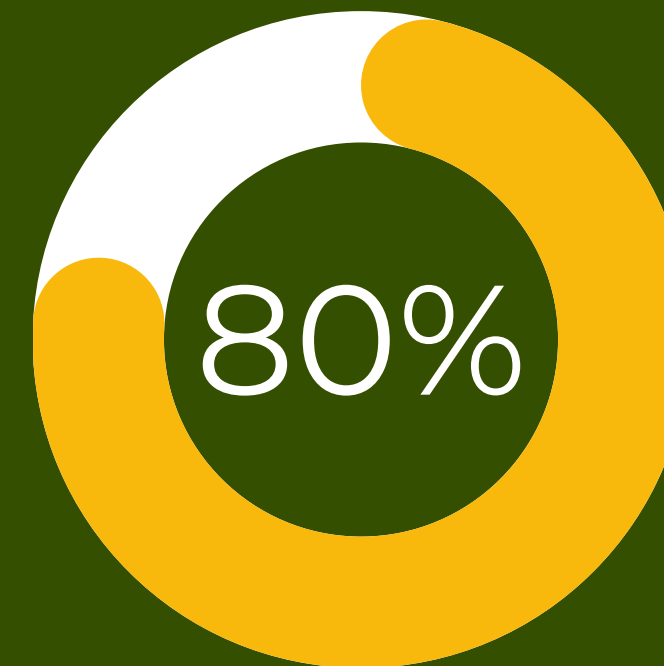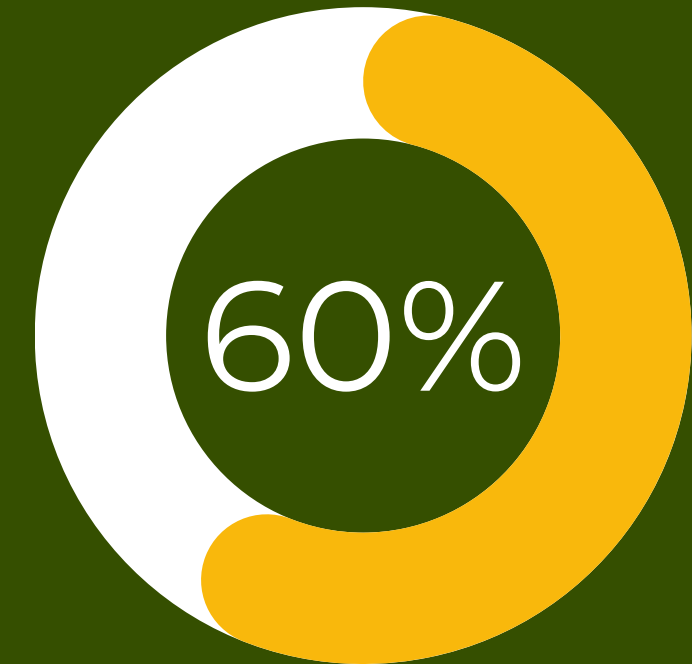**Logistic Regression Evaluation:**

- Metric - Precition give the True Positive prediction %
- The model accurately predicted(Precision) the True Positive('High Traffic') 80% of the time.
- We achieved this by removing columns that just created noise for the model giving inaccurate results.
- The category beverages was over 90% low traffic which doesnt help us find high traffic values in this case.
- We increased our ablility to predict True Positives by 20% from our baseline which is a big jump.
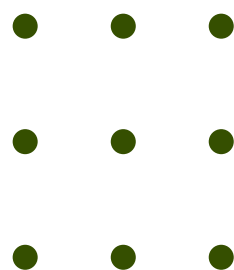
Model 1
Most Frequent
60%

80%
Logistic Regresson
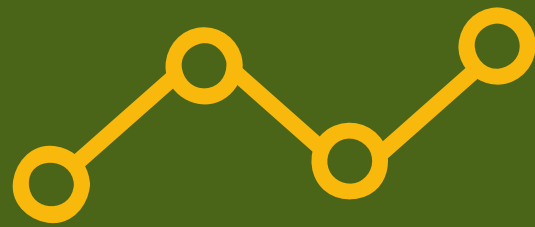
Conclusion

Key Takeaways:

1. The things that mattered the most in predicting high traffic recipes were -
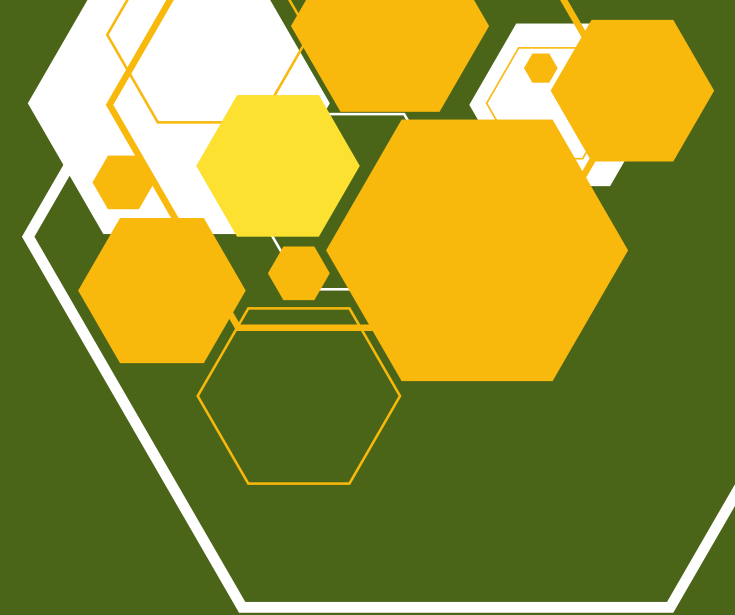   - meal category
   - servings

2. **The type of category matters when looking at high traffic.**
   - P*otato,* and *vegetable* recipes contained the highest ratio of high traffic
   - P*ork* , *meat* , and *one dish meals* come in second for high traffic
   - D*essert*, and *lunch/snack* come in third for high traffic
   - The other categories have a higher ratio of low traffic than high traffic

With the goal being high traffic then based on this data making a recipe with potato's or vegetables has a very high probability that the recipe will get high traffic. Surprisingly the nutritional information had little effect on whether a recipe got high traffic or not.

3. **Things to avoid when making recipes.**
   - Beverage recipes as they have a very low chance of getting high traffic.
   - Breakfast recipes also have a very low chance of getting high traffic.

4. **Future data collection.**
   - T*ime it takes to make the dish* - People might stray away from long prep time or cook time.
   - C*ost per serving or cost per dish* - High cost per serving might keep people from using this dish.
   - *# of steps to create the dish* - A large amount of steps keeps many people away from making meals.
   - *# of ingredients* - Simple ingredients means people probably have them and can make the dish without going to the store.
   - u*sers favorite recipes* - This would lead to being able to make recommendation systems to push recipes to specific users.
   - *user ratings* - Even if a recipe gets high traffic it might get good ratings. ratings would also help with recommendations.