Recipe Website Project

Bryan Tjandra

Table of Contents

PROJECT BACKGROUND
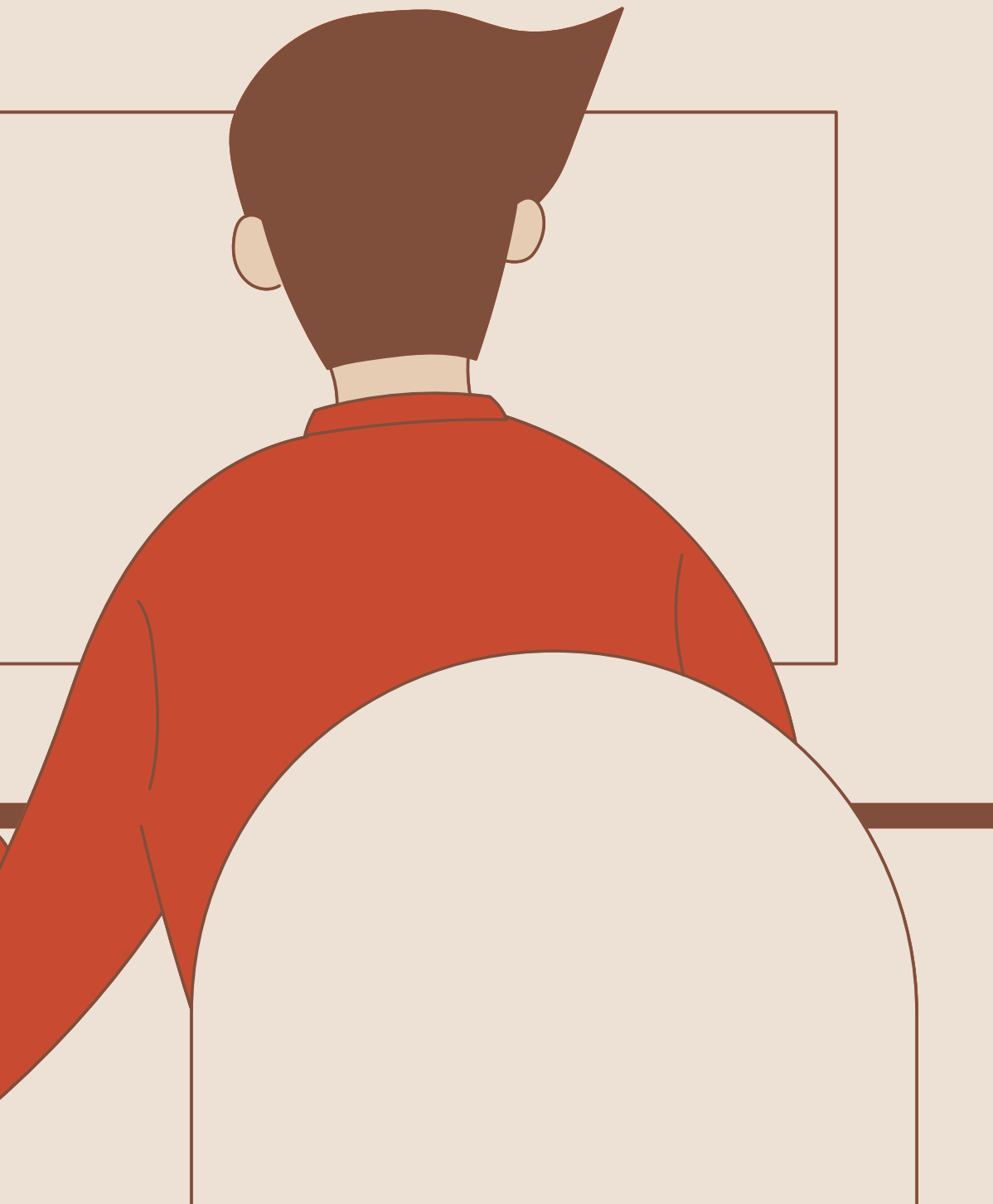
- Tasty Bytes, an online recipe startup features new recipes on their homepage website every day.
- On days that they feature a popular recipe, traffic increases by as much as 40%.
- However, it is difficult to predict in advance which recipes will be popular.
- Recipes are considered to be popular if they receive a high score.

PROJECT GOALS

Predict whether a recipe will receive a high score based on data that has been collected from previously published recipes. The success criteria is to correctly categorize 75% of unpopular recipes.

PROJECT SUMMARY

I. Data Cleaning

The recipes contain many unusual content on it, need to be cleaned. For example:

- Zero ingredients on recipe receives a high score
- There are a recipe that has more thatn 25k calories on 1 servings
- Clean duplicate recipe

II. Analyze Insight

Extract some insights that could be achieved from the data. For example:

- Which recipe category that people like most
- On what scale people do like sugar in their recipe
- Test if ingredients affect the recipe populariity

PROJECT SUMMARY

III. Transforming Data

- Encode the recipe category
- Filter and Transform the recipe that contains abnormal ingredient so it won't create bias

IV. Recipe Popularity Prediction

- Select the appropriate model to categorize unpopular recipe correctly
- Evaluate the model

KEY FINDINGS & RESULTS

Key Findings

- 65% of total recipes are popular
- 25 categories always being popular
- Several recipes contain abnormal amount of ingredient
- No significance content difference in popular or unpopular recipe
- Duplicate recipes with different amount of ingredient
- Missing category on some recipes

Results

- 4 out of 5 recipes are correctly predicted unpopular

KEY FINDINGS

Ingredient and Popular Recipe Relationship
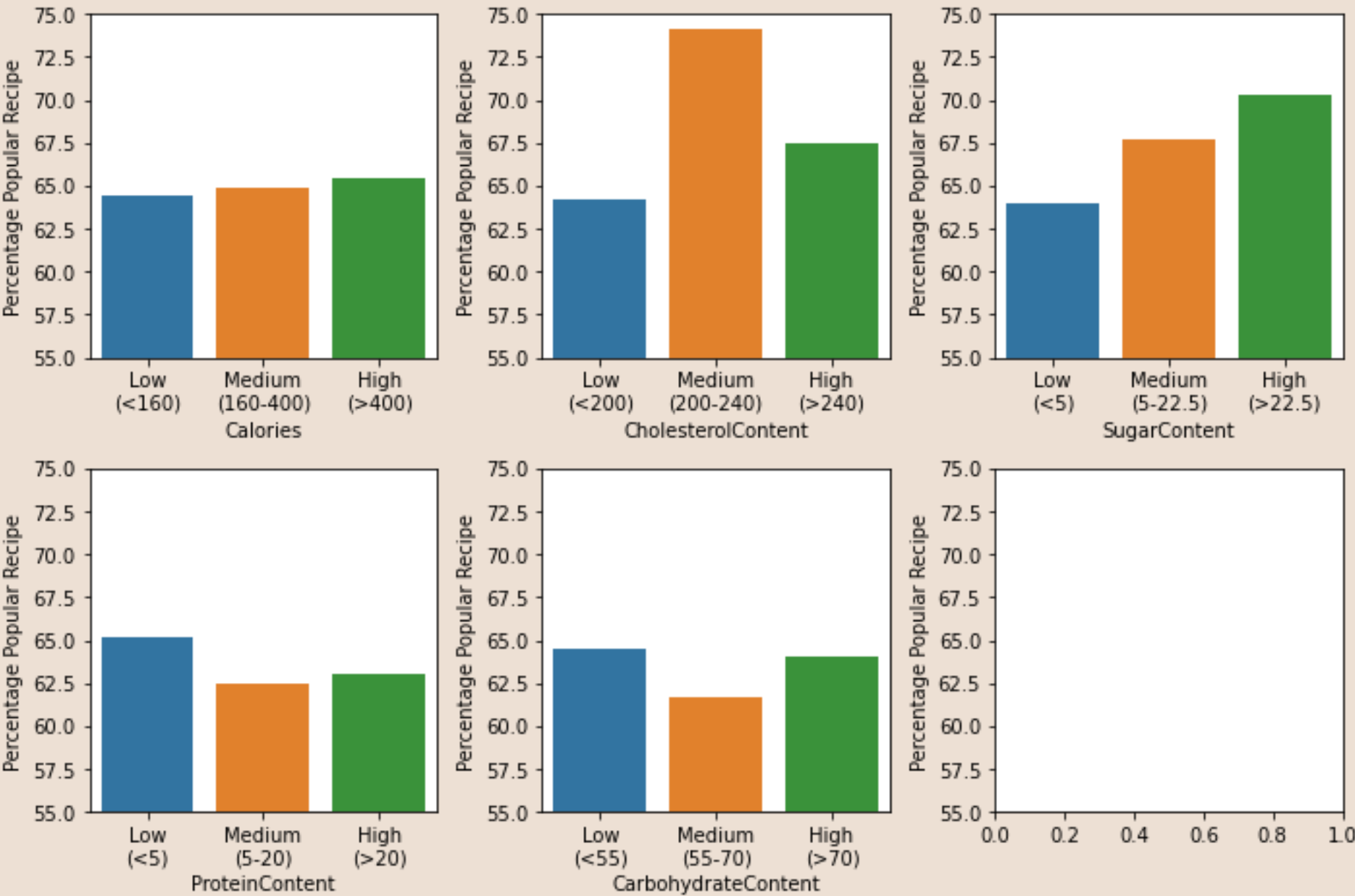
The recipes are more popular in containing:

- High calories
- Medium cholesterol
- High sweet
- Low Protein
- Low or High Carbohydrate

Total popularity proportion is 65%

- Zero sugar recipes receives 3% more popularity
- Zero carbo recipes receives 5.8% more popularity
- Zero protein recipes receive 5.7% more popularity

Zero Sugar Popularity

Zero Carbohydrate Popularity

Zero Protein Popularity
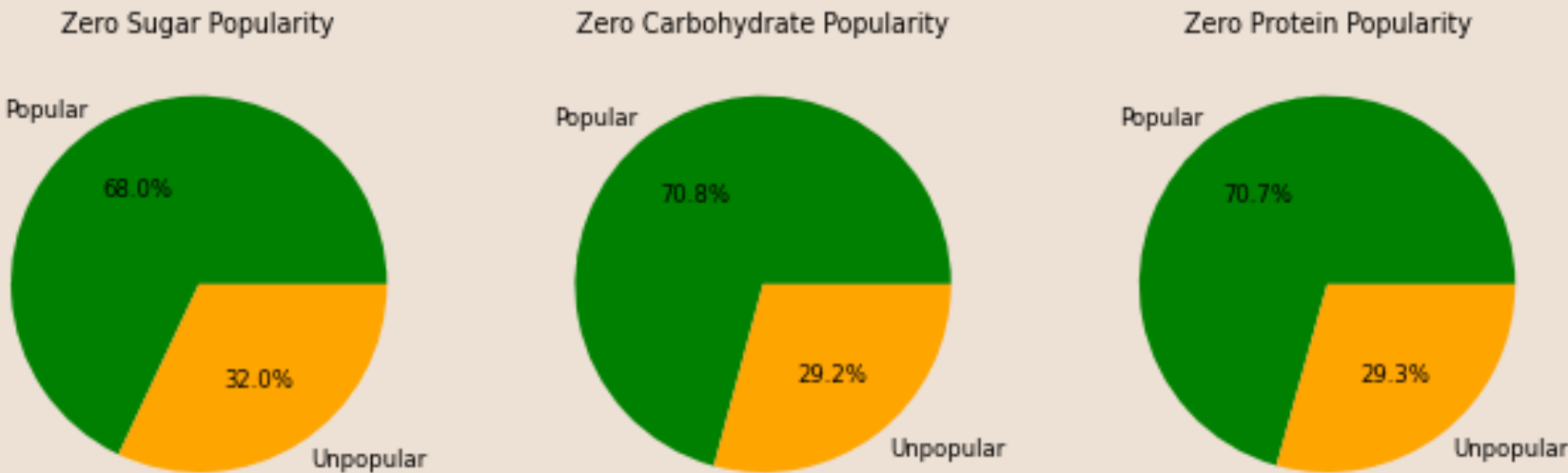
CONCLUSION & RECOMMENDATION

Conclusion

The model correctly predicts 80% of unpopular recipe. Although it's moderately high, the model still can't distinguish the target well , since the characteristic of each content and the distribution with the popular and unpopular recipes are similar

Recommendation

- Adding more recipe information to the data is helpful to make the model differentiate the recipe popularity better.
- Every recipe that comes new needs filtering because some recipes are too huge or around zero, making it is harder to tell what happened in the recipe and could create a bias.