What's for Dinner?

Problem Summary

Problem:

- Sometimes you have a stocked fridge or just a few random ingredients but you have no idea what to make. Manually searching for different recipes can be time consuming

Objective:

- Is to provide an easy, simple tool for at-home beginner cooks, that after some user input returns recipe options that best align with the users ingredients and preferences

Metrics for Success

- Cosine similarity
- Accuracy of ingredient overlap
- Accuracy of ingredient substitution
- Would I make it ?

Overall Approach

1. What's my problem ?
2. Scrape recipes from allrecipes.com
3. Clean & preprocess the data
4. Feature engineer and train my model
5. Recommendation pipeline
6. Build the app !

The Data

Where it came from:

- Allrecipes.com

What it consists of:

- Over 2,300 rows, 10 columns

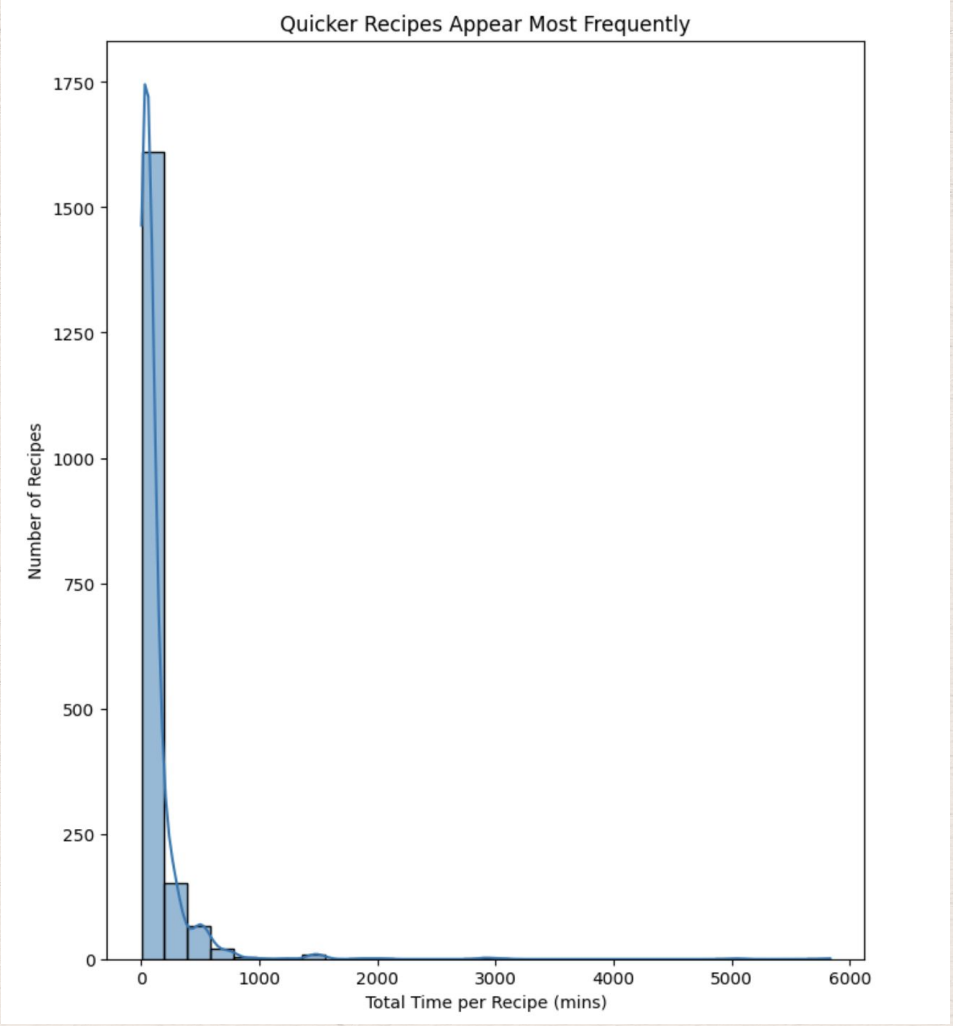
It's challenges:

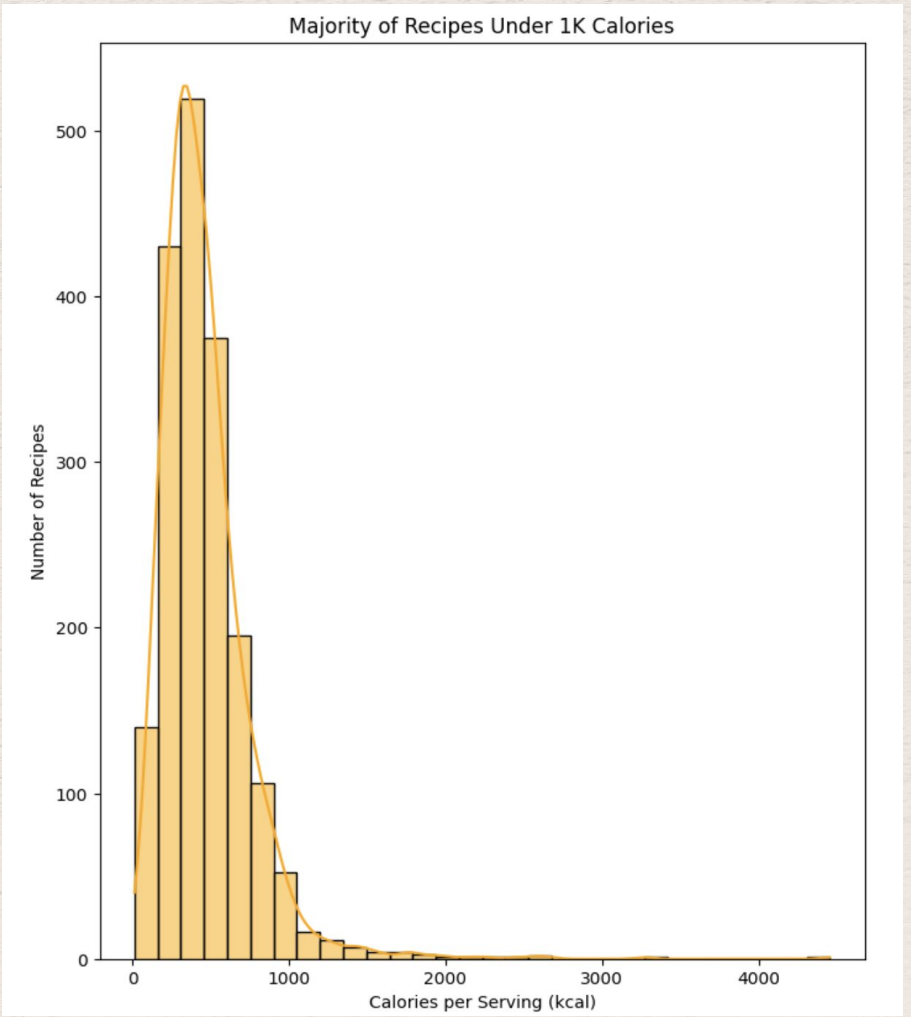
- Ingredient naming inconsistencies
- Extra fluff in ingredients

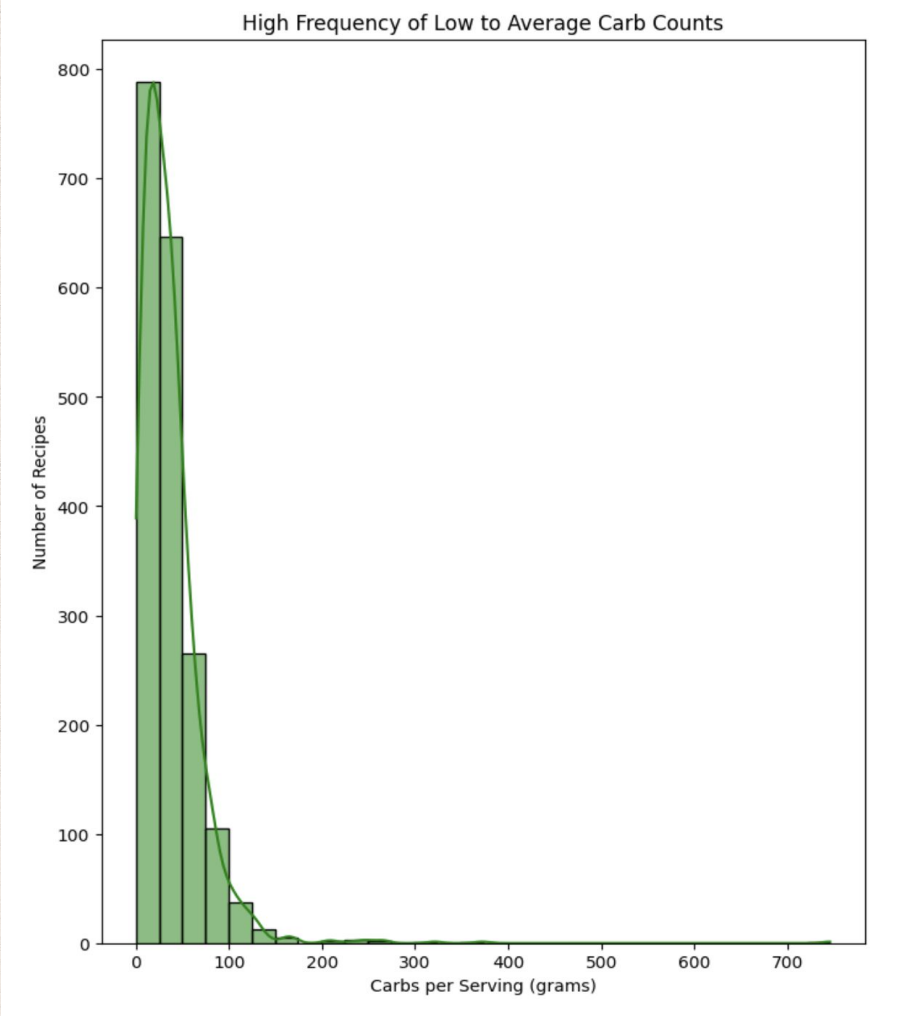
Data Munging

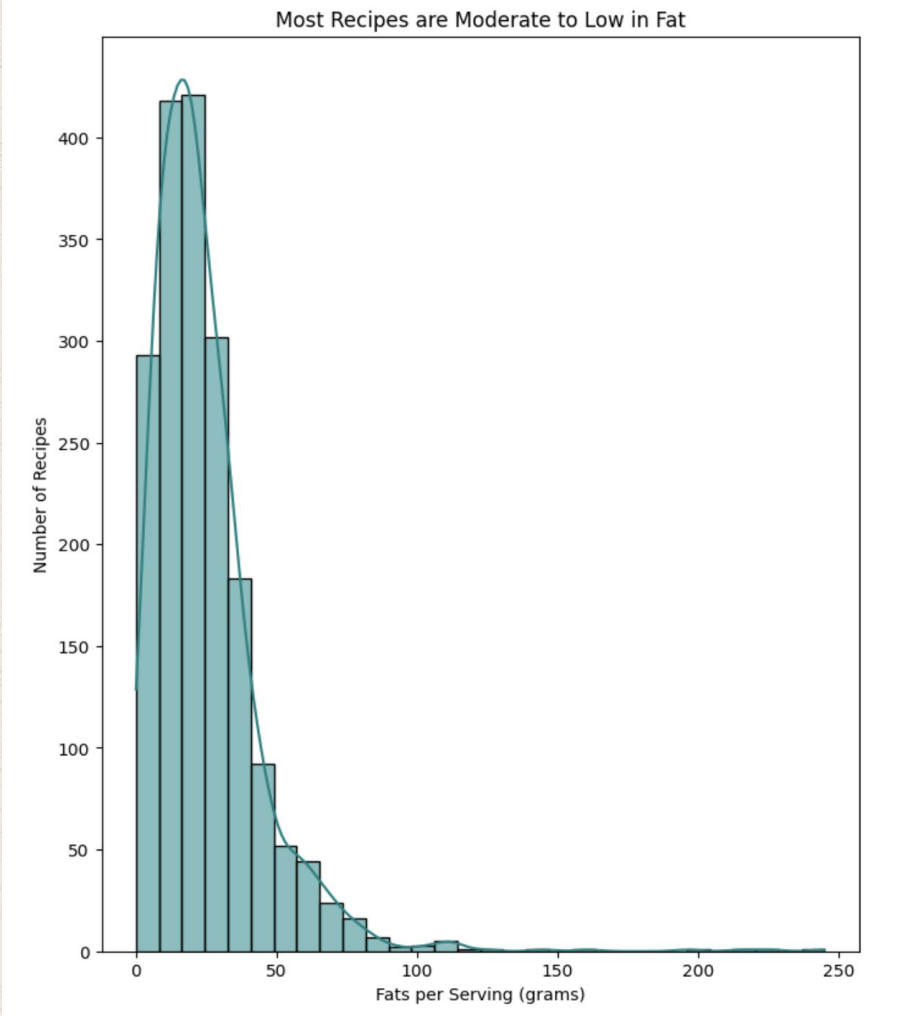
Data preprocessing:

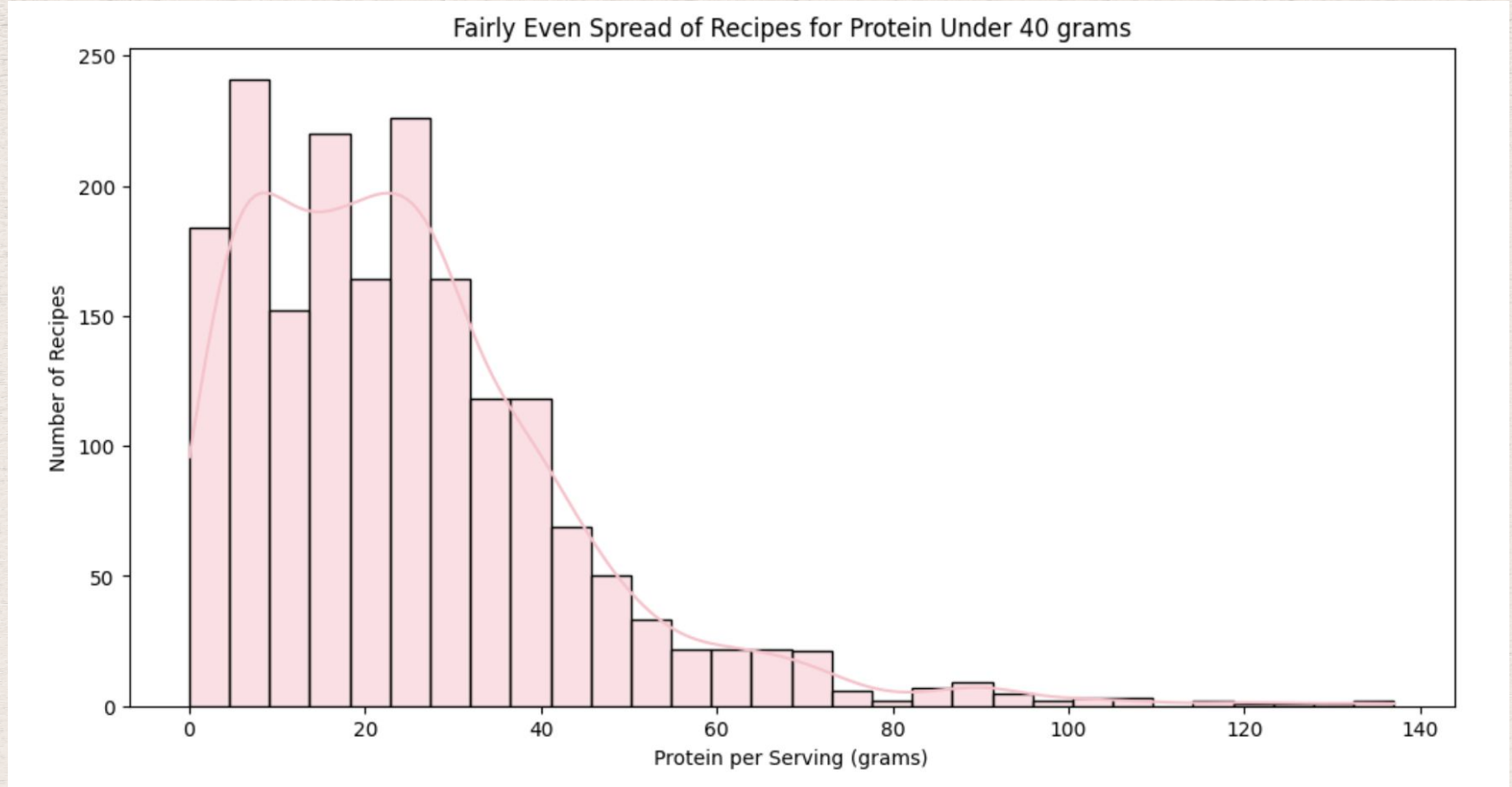
- Clean nulls, duplicates, & data types
- Time conversion
- Normalize ingredients
- Lemmatize ingredients
- Feature engineering clustering columns

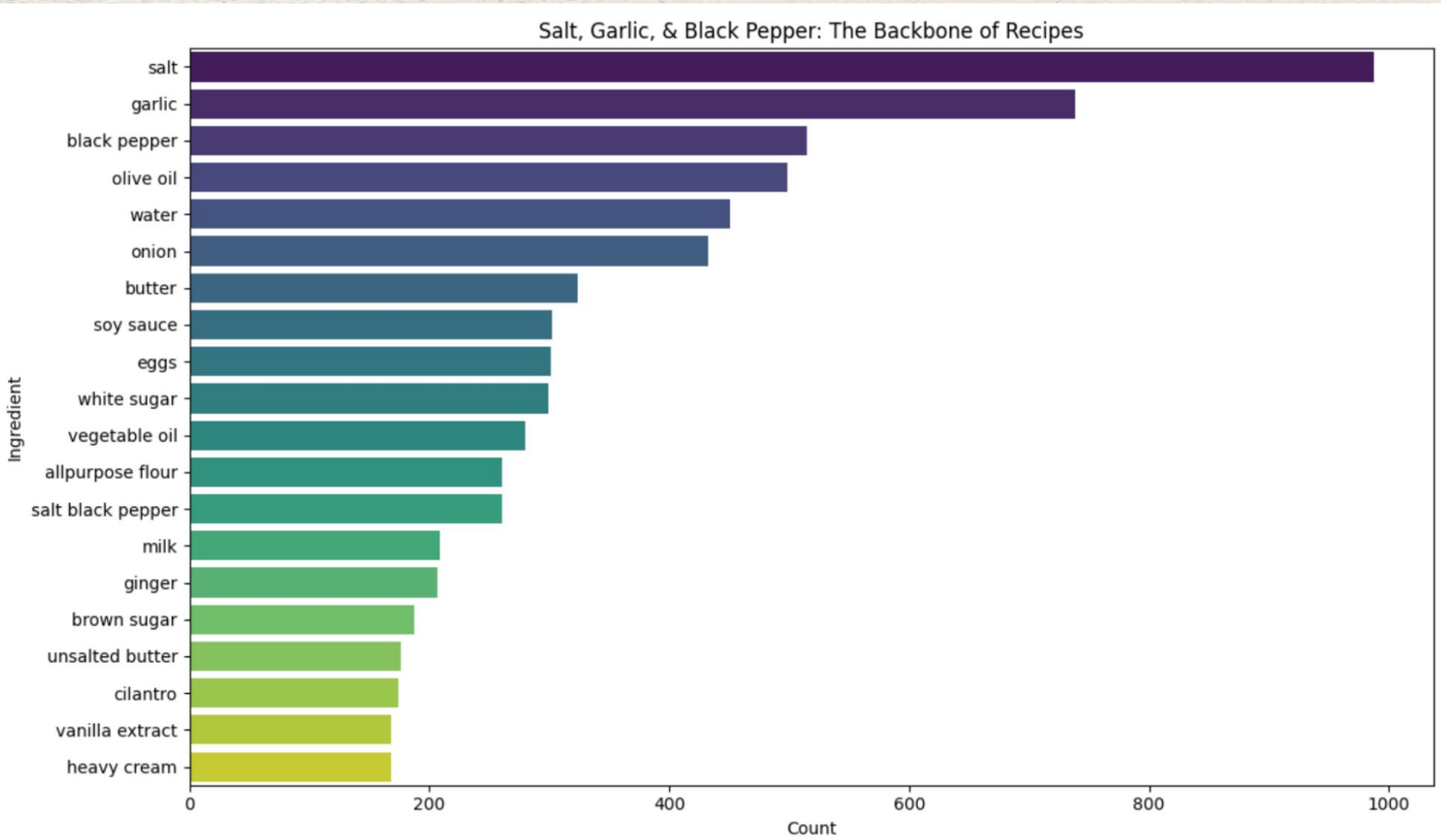

Quicker Recipes Appear Most Frequently
Number of Recipes
Total Time per Recipe (mins)
0
250
500
750
1000
1250
1500
1750
0
1000
2000
3000
4000
5000
6000


Majority of Recipes Under 1K Calories
Number of Recipes
Calories per Serving (kcal)
0
100
200
300
400
500
0
1000
2000
3000
4000


High Frequency of Low to Average Carb Counts
Number of Recipes
Carbs per Serving (grams)
0
100
200
300
400
500
600
700
800


Most Recipes are Moderate to Low in Fat
Number of Recipes
Fats per Serving (grams)
0
50
100
150
200
250
300
350
400


Fairly Even Spread of Recipes for Protein Under 40 grams
Number of Recipes
Protein per Serving (grams)
0
50
100
150
200
250
0
20
40
60
80
100
120
140


Salt, Garlic, & Black Pepper: The Backbone of Recipes
salt
garlic
black pepper
olive oil
water
onion
butter
soy sauce
eggs
white sugar
vegetable oil
allpurpose flour
salt black pepper
milk
ginger
brown sugar
unsalted butter
cilantro
vanilla extract
heavy cream
Ingredient
0
200
400
600
800
1000
Count

Model Evaluation

Chosen Model:

- Word2Vec

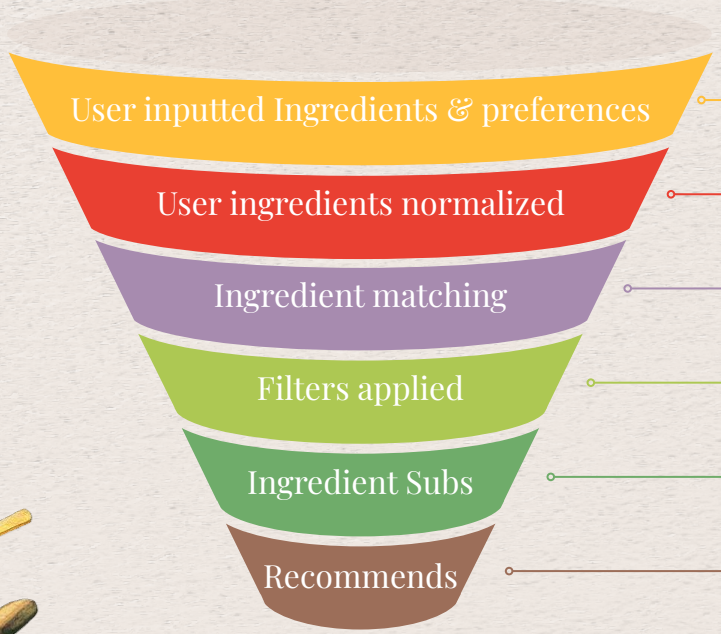
Challenges:

- Data size
- False Linkage
- Extra words in ingredients column

Performance:

- Fairly happy with performance

How the App Works



Minimum of 4 ingredients inputted (comma separated)

Lemmatize, extra white space removed, snake_case, tokenize

Tokenized ingredients fed into model (minimum similarity score of .6)

Optional user filters applied if inputted (time, calories, carbs, proteins, & fats)

Missing ingredients matched with any possible substitutions

Recommended recipes returned

APP DEMO

http://localhost:8501/
http://192.168.1.137:8502

Limitations of the Process

Data limitations:

- Small dataset
- Inconsistencies within the ingredients data

My limitations

- Time
- Knowledge

Conclusion & Future Improvements

Did the model meet my goals ?

- Yes, fairly happy with model performance

Improvements & refinements of model

- More data
- Further cleaning and substitution matching
- Dietary preference filtering
- Ingredient quantity & serving sizes


“
Questions?