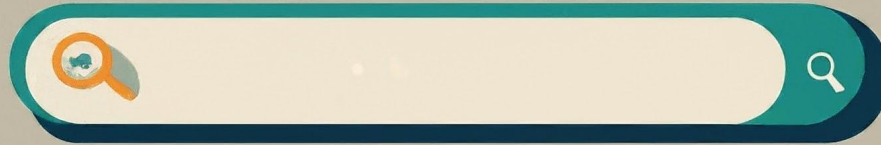


What's On The Menu?



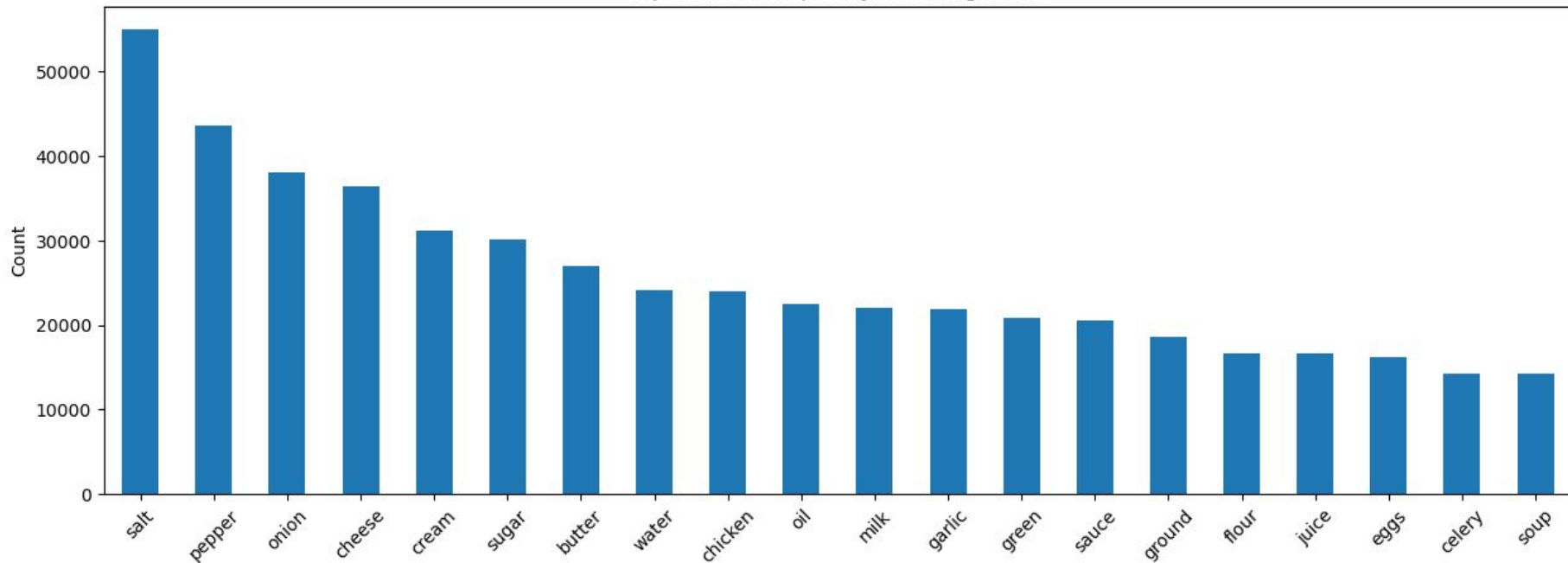
Crystal Gil Herrera

Problem, Solution and Impact

- **Problem:** Households often face the challenge of managing food resources efficiently. A significant amount of food is wasted due to over purchasing, improper storage, and the inability to use ingredients before they spoil. Also, individuals often struggle to plan meals that make the best use of what they already have, leading to unnecessary expenditure on groceries.
- **Solution:** With data science it is possible to develop a recipe recommendation tool that provides relevant recipe options based on ingredients on hand which helps reduce expenses and food waste.
- **Impact:** Cost savings as the user can input ingredients that they already have on hand which helps reduce expenses and food waste.

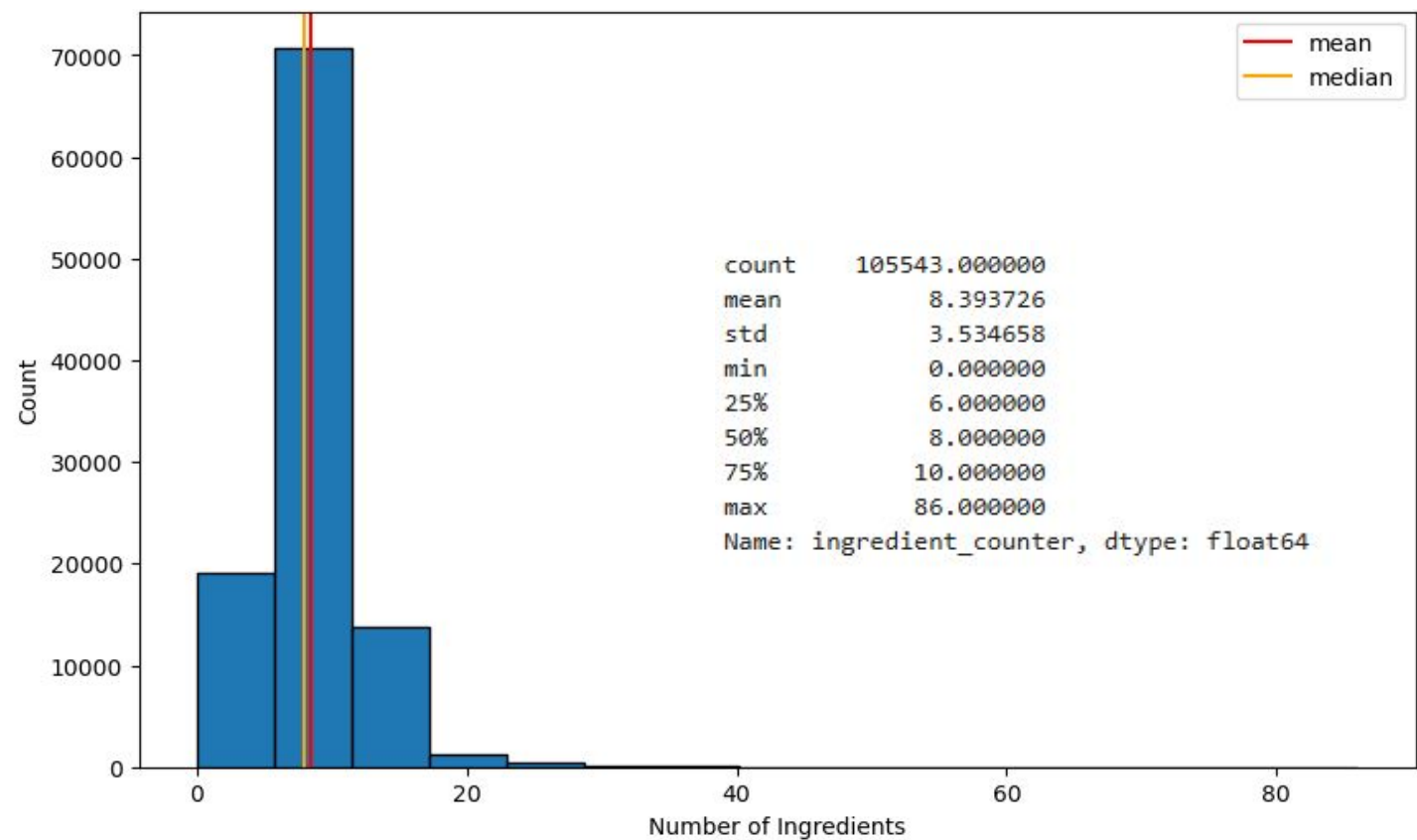
Dataset Description

Top 20 most frequently occurring words



- Mostly text data, 105,543 unique recipes with 5,529 unique tokens, average number of ingredients per recipe is ~ 8 .

Distribution of Number of Ingredients in Recipes



Preprocessing: Recipes Labelling

1. First round of labelling using the name of recipes in the dataset:

Used .loc to label recipes based on keywords in their names (i.e. recipes with the word 'breakfast' in their title were labeled as breakfast).

Able to label approximately half of the dataset.

2. Second round of labelling involved training a Random Forest Classifier model with supervised learning:

Trained on the ingredients list of the recipes that were initially labeled.

	precision	recall	f1-score	support
breakfast	0.83	0.65	0.73	1562
dessert	0.83	0.91	0.87	2088
dinner	0.76	0.85	0.80	2718
drink	0.89	0.84	0.87	444
lunch	0.83	0.72	0.77	1831
salad	0.82	0.85	0.84	1598

Sparse Matrix to One Hot Encoded Dataframe

```
<Compressed Sparse Row sparse matrix of dtype 'int64'  
with 1261886 stored elements and shape (105543, 5529)>
```



	__	__	__	__	abalone	absolut	abt	aburaage	aburage	acacia	...	zesty	zinfandel	zinger	ziplock	ziti	zone	zucchini	zucchini	zucchini	zwieback
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...
105538	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
105539	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
105540	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
105541	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
105542	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

105543 rows × 5529 columns

Content-Based Recommendation System

Cosine_similarity.

Dissimilar

Test → Recipe Title: “Pumpkin Dream Pie”.

Similar

	recipe	similarity
60806	BBQ Chicken Quesadillas	0.0
42	Poached Chicken Breast In Wine	0.0
40	Rhubarb Cake	0.0
39	Reuben Casserole	0.0
37	Ambrosia Mold	0.0
36	Black-Eyed Mullet	0.0
60800	Bogart's Scalloped Carrots Recipe	0.0
60798	Alaska Snow Crab Potato Salad Recipe	0.0
60797	Baked Enchilada Chicken Recipe	0.0
60796	Beef Roast in Gravy	0.0

	recipe	similarity
35	Pumpkin Dream Pie	1.000000
4540	Quick Pumpkin Pudding	0.790569
24717	Fat-Free Pumpkin Dessert	0.744208
13758	Pumpkin Mousse	0.744208
4069	Gingersnap Pumpkin Pie	0.700000
34547	Pumpkin Pudding	0.657376
17125	Banana Pudding	0.645497
8662	Banana Pudding	0.645497
34074	No Crust Pumpkin Pie	0.645497
519	Dream Pie	0.632456

Concerns and Next Steps

- Due to memory issues, baseline recommendation model was calculated cosine similarity of a reduced dataset of only recipes which includes ≤ 8 ingredients. In case I cannot expand the dataset to include the remaining recipes with big data tools, I will continue with this dataset.
- Integrate the labels into the recommendation system. Baseline model only considered the one-hot-encoded ingredients list.
- Perform further model evaluation.
- Investigate how to include ranking based on the match of ingredients initially inputted, output the recipes that more closely match this and add an additional output to indicate missing ingredients (like a grocery list of the remaining ingredients that would be needed to prepare that recipe).
- Productionalization: Streamlit.