Make this!

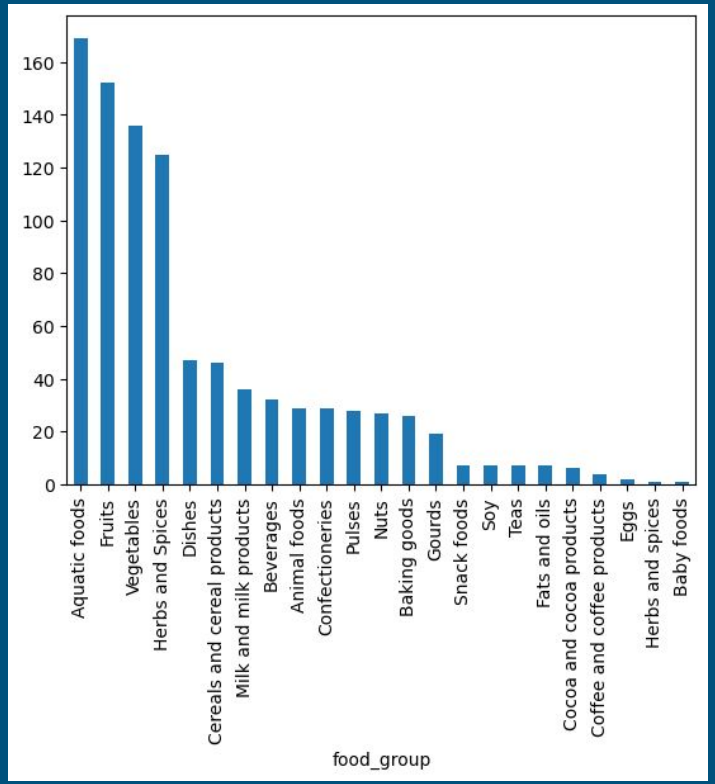
data meets cooking

Problem statement

Data

“Food” table
“Content” table

FooDB Data


160
140
120
100
80
60
40
20
0
Aquatic foods
Fruits
Vegetables
Herbs and Spices
Dishes
Cereals and cereal products
Milk and milk products
Beverages
Animal foods
Confectioneries
Pulses
Nuts
Baking goods
Gourds
Snack foods
Soy
Teas
Fats and oils
Cocoa and cocoa products
Coffee and coffee products
Eggs
Herbs and spices
Baby foods
food_group

FooDB Data

- LabelEncoder for “food_group”
- TFIDFVectorizer for “all_text”
- Classifier models worked great!

Photo by Tungsten Rising on Unsplash

Recipe data

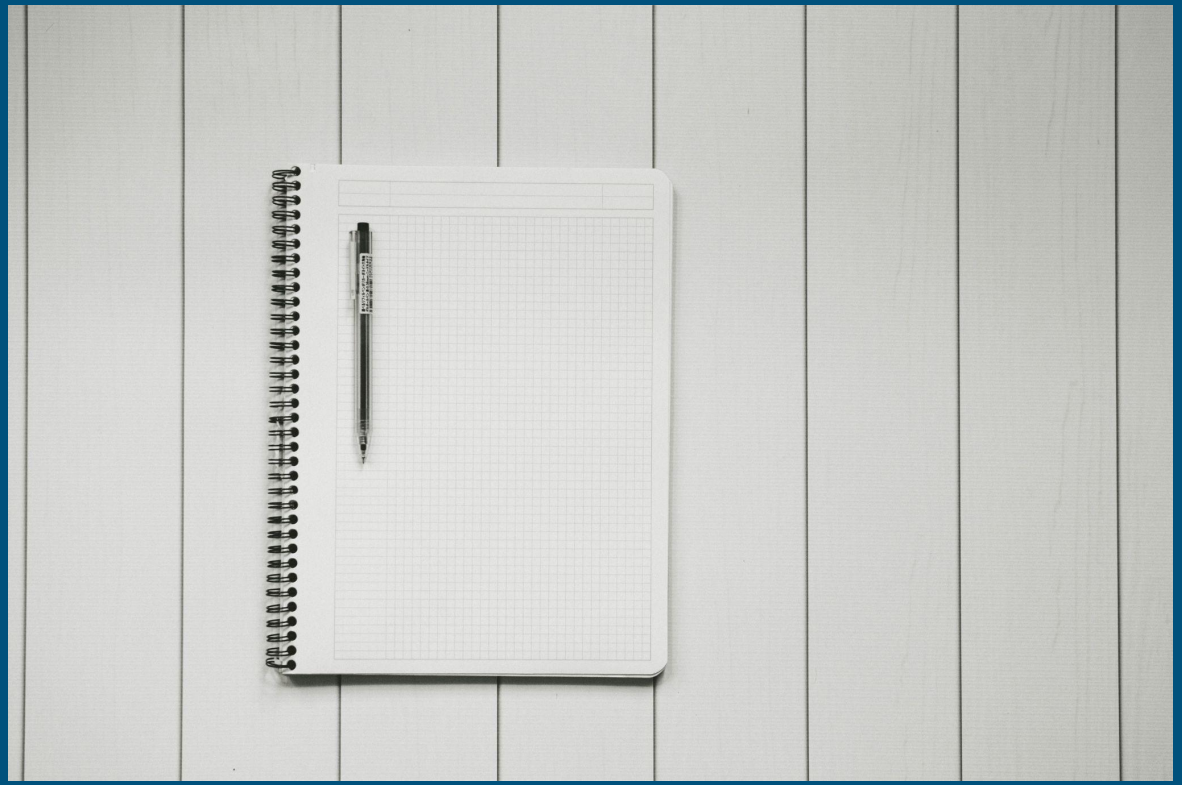
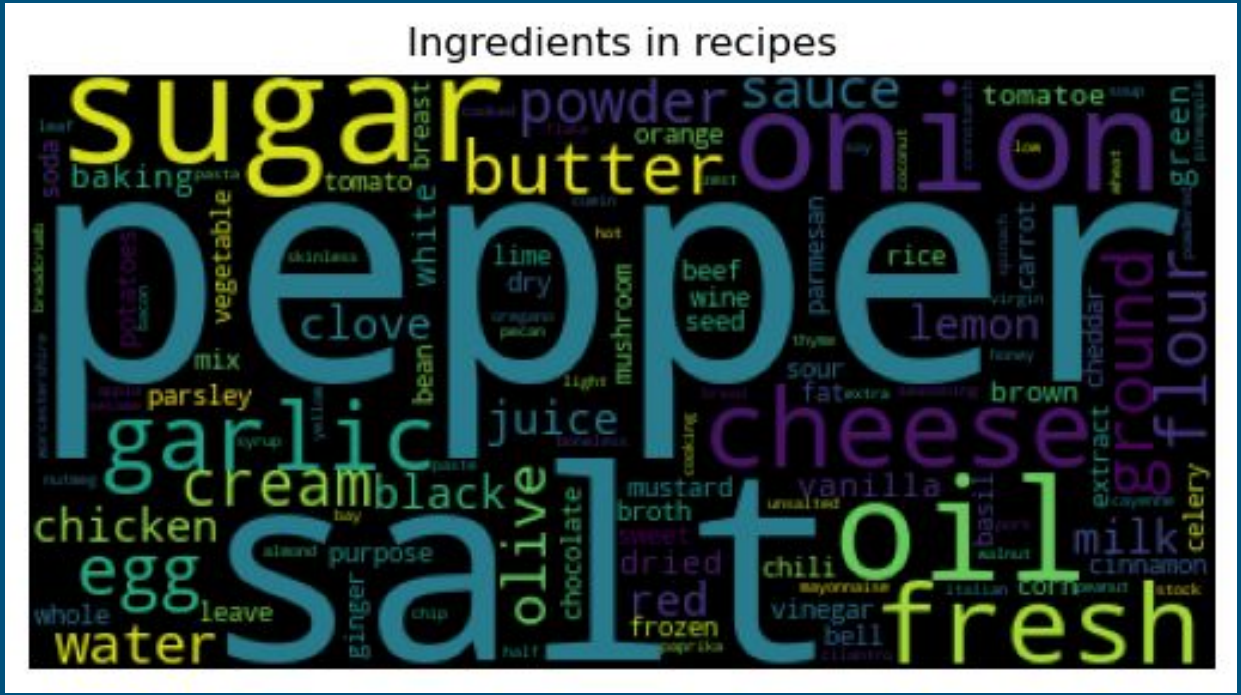
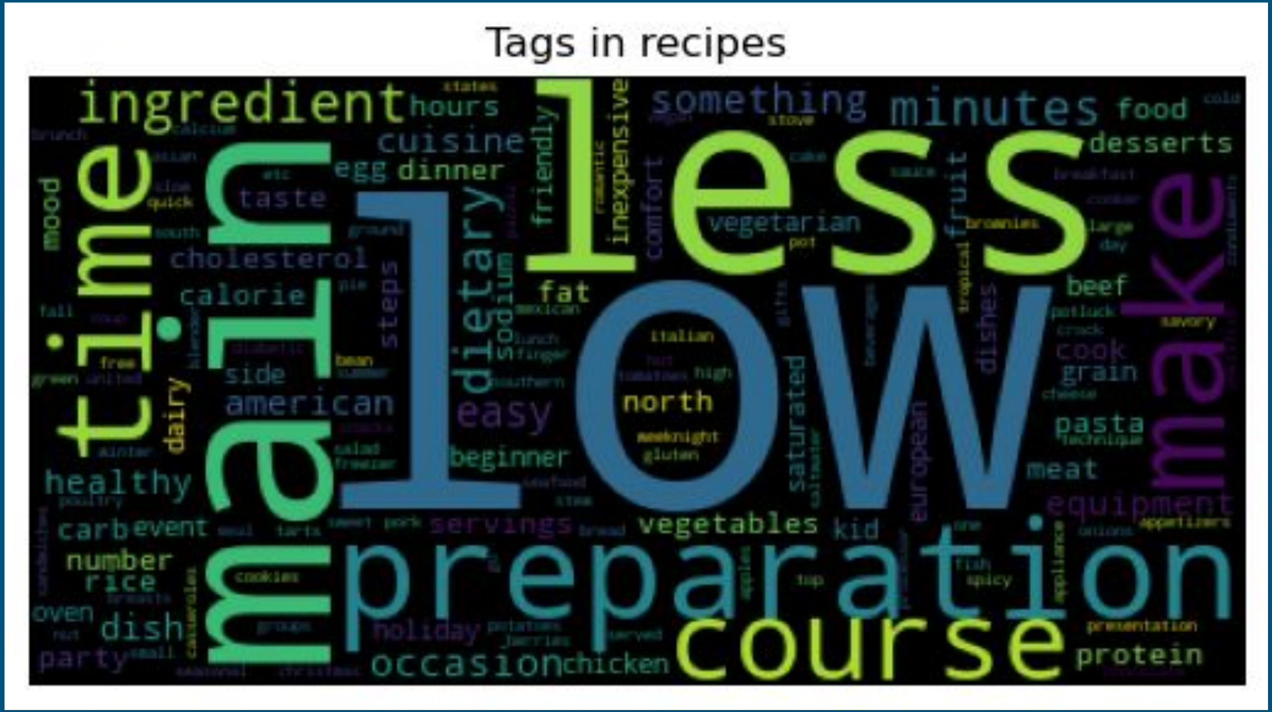


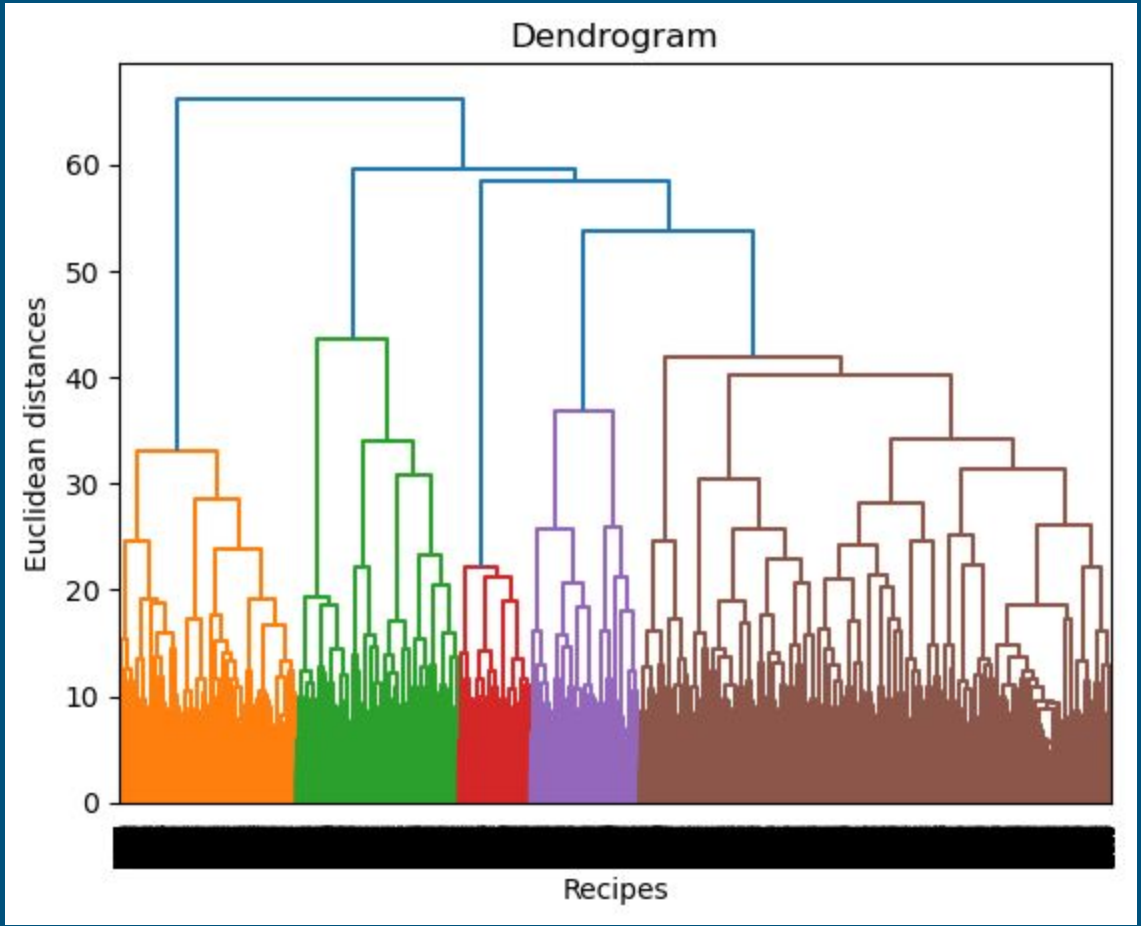
Photo by Kelly Sikkema on Unsplash

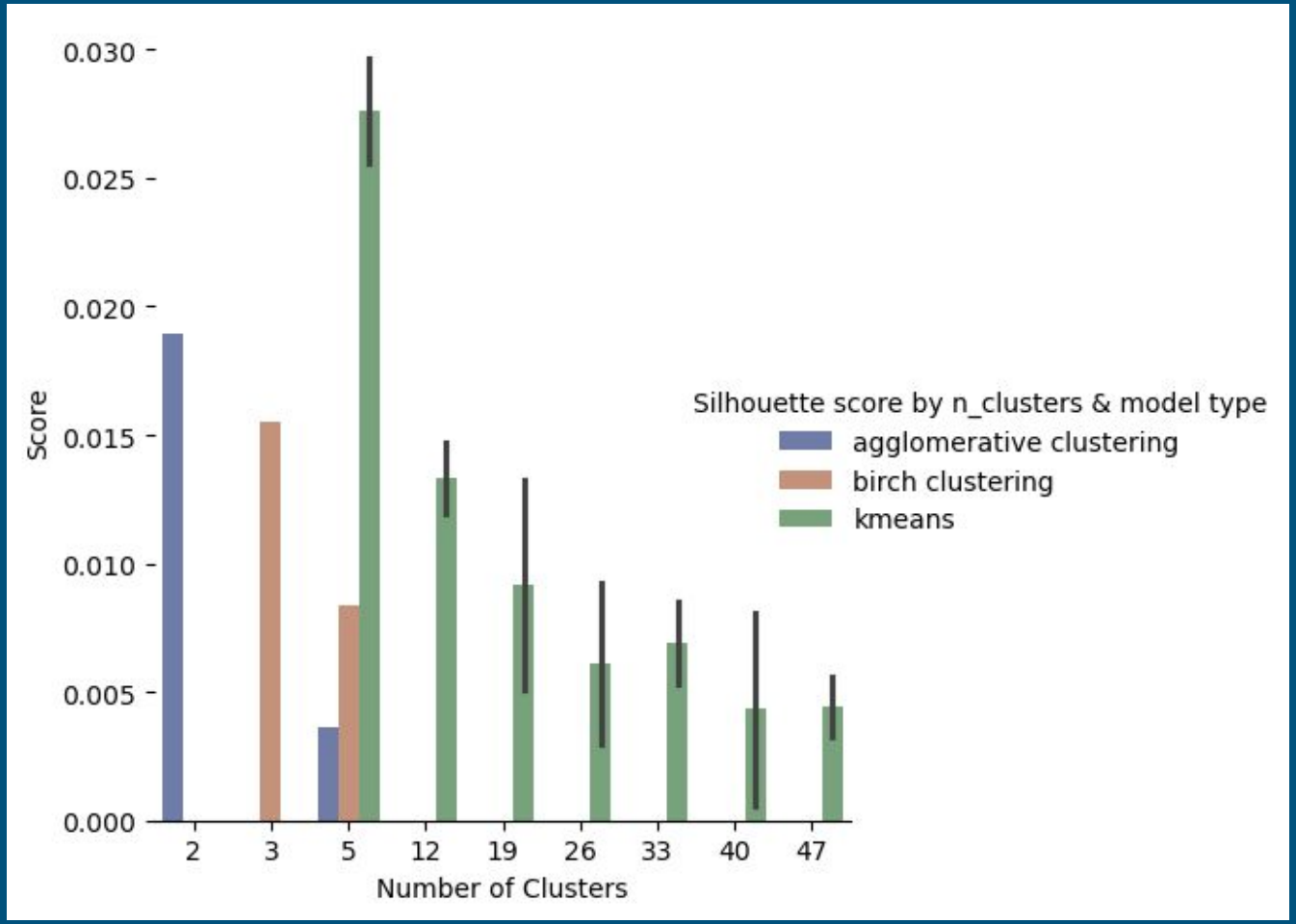
Recipe data


Ingredients in recipes
sugar
pepper
salt
onion
butter
powder
sauce
tomatoe
orange
baking
tomato
soda
vegetable
white
breast
lime
dry
beef
wine
seed
rice
parmesan
carrot
green
flour
ground
lemon
clove
mix
parsley
garlic
cream
juice
cheese
sour
fat
brown
black
chicken
egg
olive
mustard
broth
vanilla
oil
fresh
milk
celery
extract
basil
cinnamon
chili
purpose
dried
red
vinegar
frozen
bell
whole
leave
water
chocolate
ginger
bean
mushroom
cheddar

Recipe data


Tags in recipes
low
less
main
preparation
time
make
course
ingredient
something
minutes
cuisine
dietary
inexpensive
friendly
food
desserts
hours
egg
dinner
taste
cholesterol
calorie
vegetarian
comfort
fruit
mood
steps
sodium
fat
beef
north
easy
american
side
healthy
dairy
saturated
european
dishes
cook
grain
pasta
meat
equipment
beginner
carb
event
number
rice
servings
vegetables
kid
oven
dish
party
holiday
occasion
chicken
protein
italian


Dendrogram
Euclidean distances
0
10
20
30
40
50
60
Recipes


0.030
0.025
0.020
0.015
0.010
0.005
0.000
Score
2
3
5
12
19
26
33
40
47
Number of Clusters
Silhouette score by n_clusters & model type
agglomerative clustering
birch clustering
kmeans

Cosine similarity

	recipe_id	ingredient	in_recipe
0	17962	10	1
1	17962	11	1
2	17962	23	1
3	17962	26	1
4	17962	29	1

ingredient	10	101	1015	102	103	104	106	11
ingredient								
10	1.000000	0.266726	0.057712	0.074069	0.162367	0.104770	0.165997	0.499387
101	0.266726	1.000000	0.033912	0.038595	0.086688	0.065697	0.065789	0.007827
1015	0.057712	0.033912	1.000000	0.029814	0.014693	0.001993	0.028006	0.043037
102	0.074069	0.038595	0.029814	1.000000	0.046467	0.011231	0.045089	0.073445
103	0.162367	0.086688	0.014693	0.046467	1.000000	0.063029	0.036189	0.099807

Make this!

{ app demo }

References

API & data resources:

- Edamam API: developer page
- FooDB: downloads
- Kaggle dataset

Blogs & articles:

- Clustering 101: Understanding BIRCH Clustering Using Jupyter Notebook & Python
- Byte-Pair Encoding: Subword-based tokenization algorithm