The Home Cooks Guide to Ingredients and Spices

Final Project: Data and Analytics Bootcamp - Team 8
Charlie Willmore, Brittany Garza and Lauren Neidhardt

Selected Topic

Charlie had an idea for the average at home chef to be able to have way to know which herbs and spices go well together. We went in a more generalized direction, looking at the overall ingredients within recipes within certain cuisines.

Which ingredients occur in recipes of the same cuisine most frequently and can we use them as a predictor of cuisine?

Spoonacular

Our knowledge engineers spent years crafting our **complex food ontology**, which allows us to understand the relationships between ingredients, recipes, nutrition, allergens, and more.

The Spoonacular API is an online resource containing:

- Ingredients
- Recipes
- Product Information
- Menu Items (restaurants)

Questions we hope to answer with the data:

We hope to be able to see which ingredients most commonly occur within different cuisine types and can those ingredients predict a cuisine type when it is unknown.

Possibly put together recipes based on available ingredients?

Cook certain cuisines based on common ingredients?

Data Exploration

Initial Data Processing:

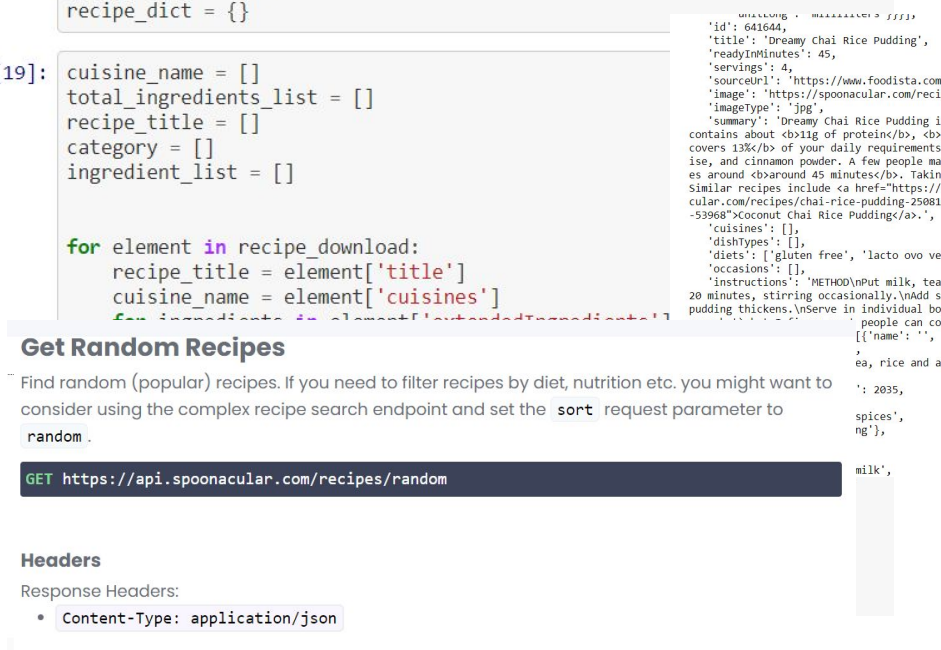
- Data downloaded via Spoonacular API
- Performing a random recipe search, 100 random recipes were downloaded in each call.
- Data cleaned in JSON to limit to recipe, cuisine, ingredient names
- For loop to pull 10k recipes, merged into single dataframe.

```
In [16]: recipe_download = recipe_download['recipes']

In [17]: ingredients_dict = {}
         category_dict = {}
         recipe_dict = {}

In [19]: cuisine_name = []
         total_ingredients_list = []
         recipe_title = []
         category = []
         ingredient_list = []


         for element in recipe_download:
             recipe_title = element['title']
             cuisine_name = element['cuisines']
```



Data Exploration

- Create single list of ingredients and their counts
- Use Regex to create a consistent, clean list of ingredients
- Drop overly common items like salt and pepper, oil

```
    elif re.search('(?:(?:^)|(?:\s))[Tt]omato(?:\ssauce|\spaste|\sjuice)', element):
        corrected_ingredient_list.append('tomato sauce')
    elif re.search('(?:(?:^)|(?:\s))[Cc]oconut(?:\smeat|\sextract|\sflake|$)', element):
        corrected_ingredient_list.append('coconut')
    elif re.search('(?:(?:^)|(?:\s))[Mm]ustard(?!\spowder|\sseed)', element):
        corrected_ingredient_list.append('prepared mustard')
    elif re.search('(?:(?:^)|(?:\s))[Mm]ushroom|mushrooms(?!\ssoup)', element):
        corrected_ingredient_list.append('mushrooms')
    elif re.search('(?:(?:^)|(?:\s))[Cc]umin(?:\sseeds|$)', element):
        corrected_ingredient_list.append('mushrooms')
    elif re.search('(?:(?:^)|(?:\s))[Ss]ugar(?:\s|$)', element):
        corrected_ingredient_list.append('sugar')
    elif re.search('(?:(?:^)|(?:\s))[Gg]arlic(?!\ssauce|\schili)', element):
        corrected_ingredient_list.append('garlic')
    elif re.findall('dried.*?chile',element):
        corrected_ingredient_list.append('dried chile')
    elif re.search('(?:(?:^)|(?:dried\s))[Cc]ilantro(?:\s|$)', element):
        corrected_ingredient_list.append('cilantro')
    elif re.search('(?:(?:^)|(?:dried\s))[Dd]ill(?:\s|weed|$)', element):
        corrected_ingredient_list.append('dill')
    elif re.search('(?:(?:^)|(?:\s))[Ff]enugreek(?:\s|$)', element):
        corrected_ingredient_list.append('fenugreek')
    elif re.search('(?:(?:^)|(?:dried\s))[Mm]int(?:\s|$)', element):
        corrected_ingredient_list.append('mint')
    elif re.search('(?:(?:^)|(?:dried\s))[Pp]arsley(?:\s|$)', element):
        corrected_ingredient_list.append('parsley')
```

e_predict
an
an
an

andom Forest

	cuisine_SP	onion	garlic	vanilla	lemon	bell pepper	tomato	chocolate	mushrooms	italian cheese
0	[]	0	0	0	0	0	0	0	0	0
1	[]	0	0	0	0	0	0	0	0	0
2	[]	0	1	0	0	0	0	0	0	0
3	[]	0	0	0	0	0	0	0	0	0
4	[]	1	1	0	0	0	1			
...	...	...	...	...	...	...	...			
9995	[]	0	0	0	0	0	0			
9996	[]	0	1	0	0	0	0			

	Predicted 0	Predicted 1	Predic 2
Actual 0	4	0	0
Actual 1	0	180	0
Actual 2	0	0	65
Actual 3	0	0	0
Actual 4	0	0	0
Actual 5	0	0	0
Actual 6	0	0	0

sis Phase I:

ɔm Forest was the machine learning model we chose to classify known ingredients by cuisine and ther
es in the data. About 7000 recipes had a null value for cuisine, 3000 had a populated value

Ingredients turned into columns, with one column for cuisine
Recipes become the index
Ingredients are given 1 or 0 if they occur in a recipe
Cuisine names normalized
Cuisines encoded
Confusion Matrix

PCA Analysis: Spices

KMeans: Spices

Market Basket Analysis: by Cuisines

Another way to understand the relationship between ingredients and cuisines.