

# Examination of Food Recipes from Around the World

By: Ben Sturm

## Minimum Viable Product Summary

### Domain

For this project, I plan to explore recipes from around the world. One of the reasons I chose this topic is that I really enjoy travel and cooking, and both of these interests would be exercised for this topic.

Some of the questions I hope to answer are:

- Which cuisines are most similar and which are most different?
- Does American food have a relationship to other foods?
- What can I learn through clustering of these different recipes?

### Data

Although there are many different websites that aggregate recipes, there aren't that many that have public APIs associated with them. However, one site that does have an API is [Yummly.com](https://www.yummly.com). I was granted an academic license to this site and I have started to explore the data that is available. Below is a table which summarizes this data.

Name	Type	Description
Course	String	e.g. Lunch, Main Dish
Flavors	Dict	e.g. {sour: float, salty: float, sweet: float}
Ingredients	List of Strings	e.g. [bow-tie pasta, bacon slices]
Rating	Int	Rating from 1-5
RecipeName	String	e.g. 'Lemon Chicken Pasta'
SourceDisplayName	String	Site where recipe was sourced from

### Known unknowns

The Yummly API allows the user to search for recipes with specific search parameters, however I am only able to get 10 recipes at a time with each search request. So, one of the unknowns is how do I get around this limitation? My plan is to do a search for different cuisines and save the results of that search in separate pandas dataframes. I can then merge all of those dataframes together and run my clustering models. I plan to start with k-Means Clustering and then try other types of clustering algorithms like Agglomerative and DBSCAN. I also think it will be really interesting to do dimensionality reduction on this dataset, using a technique such as PCA, to look at how cuisines are interrelated along the different principle components.