**Please limit yourself to 4 hours time!**

**Place your submissions to the Submissions sub-folder with the naming convention: lastname\_firstname**

## **The Goal**

You've just joined the data team at an online publishing company. One of your verticals is a food publication. A product manager on your team wants to build a feature for this vertical that enables users to query by cuisine, not just by ingredients. All of [your recipes](https://drive.google.com/file/d/1N71j6lZmwNIDrgpj4NTX253ocbIFSc8H/view?usp=sharing) are unlabeled, and it's infeasible to label them by hand.

Design and execute a method to group the recipes by cuisine. How could you design this to be robust enough to understand similarities / substitutions between ingredients? For each each major cuisine, what are the driving ingredients that characterize it?

Your product manager indicates a likelihood that you will only need to write a guideline for an outsourced team to hand label the remaining corpus. How would you go about writing this guide for a few major cuisines?