Brandon Woo
004265218
CS 170A

Project Proposal: Most Popular Ingredients for Cooking

(1) a description of a dataset that you constructed;
- Takes the ingredients for the most popular recipes and finds the most common and popular ingredients

(2) a log of the analyses you attempted (including models and methods covered in CS170A);
- Find the most popular ingredients to make the most recipes with
- Find the most popular ethnic groups of food. ie. ginger and soy sauce used in common together for a lot of Asian recipes
- Find the best price per weight for common recipes. ie. Salmon at $9 a pound and chicken at $3 a pound but say 100 recipes for salmon vs 10 recipes for chicken. Thus, dollar value per recipe is 9 cents per salmon recipe vs 30 cents per chicken recipe.
- Average amount of each ingredient used in each recipe. ie. 1lb of chicken per chicken recipe. This can be balanced with the above analysis of price per weight.
- Can use least squares and PCA to find trends among ingredients that are the most common and correlate in frequency together. Can also weigh how much each factor plays into the analysis.

(3) a list of results you obtained,
- Results would be the above observations

(4) a summary of effort put in, results gotten out, and lessons learned
- The main takeaway would be to figure out which ingredients to shop for at grocery shops to make the most recipes with different factors in play such as which ingredients give the most recipes, are the most budget conscious, or lead to the best ethnic types of recipes.

1. Data familiarity/interest (Why is it that are you knowledgeable about/interested in this data?)
- I meal prep every week and love trying out new recipes. Sometimes I buy an ingredient I think I'll use multiple times except then the surplus just says in my fridge until it expires. From this analysis I will be able to figure out which ingredients are the most popular while staying in my budget.

2. Data acquisition effort (How much effort is required to construct the dataset?)
- Webscraping from allrecipes.com, foodnetwork.com, and food.com into a normalized database

3. Data novelty (In what ways is looking at this kind of data new or innovative?)
- From some popular ingredient lists for cooking, there are no factors taken such as how to eat on a budget for the most recipes. Adding these factors in will help make more informed decisions when buying ingredients.

4. Analysis effort (How much work is needed to analyze the data?)
- A lot of the work is in the analysis section since I will be analyzing the data from multiple different approaches.

5. Analysis methods (Which methods are most unusual, novel or worthwhile?)

- Correlating the data to see which ingredients appear most commonly together is novel. From this you can figure out which ingredients are best to buy to get the most value out of your dollar.

6. Project difficulty (In what ways is the project most challenging?)
- Standardization of the data received from the web scraping. ie. "garlic" vs "garlic cloves" vs "garlic powder"

7. Project relevance to the course (How does the project relate to matrix methods and modeling?)
- We can manipulate the ingredient data to find how ingredients correlate based on price, recipe frequency, groupings.
- Can use least squares to determine regressions where ingredients support each other.
- PCA can be used to figure out which factors play into bringing ingredients together

8. Project novelty (Which aspects of the project are most unusual or creative?)
- Adding in the factors by which to filter through the data to find the best correlation

9. Report (Which aspects of the report require the most effort?)
- The analysis will take the most effort. The visualizations and conclusions drawn will be the most helpful.

10. Overall effort (Which aspects of the project require the most effort?)
- Analysis will take the bulk of the effort. I have done a lot of web scraping in the past.

As many college students embark on cooking for themselves for the first time, it would be practical to compile a list of the most popular ingredients for crafting the best recipes. The more often an ingredient appears from the top rated recipes, the more useful the ingredient is in someone's kitchen. For students on a budget, the most common ingredients can also be balanced out with the average price per oz or pound to figure out the most cost effective ingredients to cook with.

Data of the ingredients can be gathered from popular cooking websites such as allrecipes.com, foodnetwork.com, and food.com. Each website aggregates popular recipes and can be sorted from highest rated recipes to lowest rated recipes with other factors such as the most reviewed or viewed recipes. From here, a list of recipes can be obtained and through more webscraping, the ingredients can be gathered and put into either a csv or a database. The analysis can then begin after sorting the data based on different filters. Some difficulty could appear in normalizing the data to have consistent naming across the board (ie. "garlic" vs "garlic cloves" vs "garlic powder"). Some interesting factors that can be examined could be the average amount of ingredient used in each recipe (ie. 1 teaspoon of cayenne pepper vs 1 stick of butter) compared with the price cost for each.

After a quick search, some blogs have listed the most popular ingredients for cooking. However, none of them seem to provide data to back up their ingredient choices nor do they allow for choice based on factors like price per oz of the popular ingredient. This project would provide a guide for those on a tighter budget to make the most variety out of the ingredients they have.  As many students begin cooking for themselves for the rest of their lives, it is important to understand which ingredients are important to create the most variety in their lives.