COMS4111 Proposal
Group 58
Sirui Li(sl4653), Jiarui Yang(jy3036)

Our project is to create an application that helps people to figure out what they can cook with the ingredients left in their fridge. In order to make our application more functional, we will build this application based on recipe description, recipe ratings, and ingredient nutrition data sets. For our database, we would have 6 entities: Person, Recipes, Reviews, Ingredients, Tags, Subtags, and following relationships: gives_rating_to, contribute_to, contains, was_given, assigned_with, and belongs_to. The interesting part of this project is that we are going to build a relatively practical and useful application. However, since we are considering adding filter functions and pictures, it is difficult to match recipes and pictures in our datasets and categorize tags from recipes.

We found an inspirational dataset on Kaggle called Food.com Recipe and Interactions, which is collected from Food.com website. It contains 230K+ recipes and 1000K+ reviews from users. We will expand this dataset by adding nutritions dataset from USDA, recipe pictures from Food.com.

For part 3, we decided to follow the Web Front-End option. Users will get access to our database through a website and they will be able to find recipes based on searching keywords of ingredients, or any recipe's name. Users could get recipe steps and see past users reviews on it. Moreover, we hope to add more functions like filtering receipts by ratings, cuisines, cooking tools required, cooking time, level of difficulty of cooking, and nutrition of the dish. If time allows, we will build a dataset for users in order to let them store recipes they like, and make recommendations system based on user's interaction data.

Both of us will not drop this course since introduction in database is required to fulfill our graduation requirement, and we must have 4 courses to maintain full-time student visa. Hence there is no contingency plan. However, if any of us drop this course, we will include fewer entities in the database such as dropping tags or recipe pictures for further filtering function.

Update:
Since our data sets contain 231, 637 recipes and 1, 132, 367 reviews, it will be relatively time-consuming when we create tables and implement our relations over PostgreSQL for part 2 of this project. Therefore we decided to randomly sample and pick 10% recipes and reviews as our updated demo data sets. The updated dataset includes 23, 163 recipes and 113, 430 reviews. Additionally, we reversed the relationship between Subtags and Tags in the first version to make it more plausible.