Dinner decisions based on time and ingredients<hr>

Introduction<hr>

With fast pace living style, people may not want to spend too much time on cooking dinner and shopping grocery, especially during weekdays. Many recipes websites collected a great number of recipes, and nicely categorized them so users could find the dishes they want to cook by main ingredient type (beef, chicken, vegetarian, etc.). However, few of them, if any, allow users to choose recipe based on the time needed to get the dish ready, and/or the ingredients available at hand. I scrapped the simplyrecipes.com website for their dinner recipes and focused on the following:

1. performed exploratory analysis on distribution and correlation of time (cooking and preparing) and ingredients with various dish type (beef, chicken, pork, seafood, lamb, and vegetarian)
2. write a script to recommend the user with recipes based on one or more criteria:
   1. dish type desired
   2. time intended for preparing & cooking
   3. ingredients at hand

<hr>Analysis<hr>

1. prep time and cook time are not correlated: 0.0384
2. at least one type of dish has significantly different cooking and preparing time than other types (pair-wised t-test in progress)
3. onion, pepper, and tomato are among the most popular auxiliary ingredients, get those and you could cook most of the meat/vegetarian dishes; for different types of meat, the auxiliary ingredients greatly overlap

<hr>Usage<hr>

Another python script: recipe\_recommendation.jpynb could be used to interact with user. The user could enter the type of dishes (allow multiple selection, allow any), the maximum that he/she wants to prepare and cook (all any), and the ingredients he/she has in hand (allow any, assume have necessary seasonings and spice), and the script would return recipes that meets the criteria for the user to choose from.

<hr>Future improvements<hr>

1. Currently the wrangling of ingredients and its NLP is pretty preliminary and have a lot of human manipulation involved (adding to the customized stop words list), which is subject to human error and time consuming. In the future, could fine-tone this process to be more rigorous
2. statistical tests could be more thorough: e.g., pair-wise t-test could be formed for cook/prep time by dish type
3. Some other recipe websites also have recipe rating data (this also has rating data, but all recipes are 5 stars, which render this piece of information unhelpful) and calorie data, which could add on to perspectives of analysis and recommendation based on user’s preference
4. make the output csv file into a SQL database for future reference

<hr>

PS: In order to get dish type, I need to scrap the website a second time because the categorization info is not on the webpage of each individual recipe. I currently performed the second scrap by starting a new project “category”, hence the 2 github links. In the future, I would attempt to merge the scraps into 1 project with multiple spiders.

<hr>

Github links:

<https://github.com/microwolf/Scrapy-simplyrecipes.git>

<https://github.com/microwolf/Scrapy-category.git>