Food project 2 - data wrangling

1. scrape data from Reddit
   how: using GoogleScript: https://www.labnol.org/internet/web-scraping-reddit/28369/
   https://pushshift.io/api-parameters/
2. raw csv
3. extract ingredients(keywords) from the column "title" by Python library to do NLP
   ("food NLP/food_NER.ipynb")
4. add a new column "keyword" in the dataset
   ("food NLP/Covidcookery_keywords.csv?)
5. divide the column "keyword" into separated columns by excel function
   ("food Network/covidcookery.csv")
6. filter the data from Aug to Nov and manually clean the keywords
   ("food Network/curated food data")
7. transform the data into the format that network visualization requires by Python
   (food network.ipynb)
8. convert the dataset into Json by using "convert.py"