THE WHOLE WORLD COOKS

Chris Caress coming to chef Jose Andres with a new app idea for to make him money

INTRO

Recommender's are useful for home cooks of all skill levels. There are multiple ways that we can approach this problem and they can build on each other as well. I am pitching this as a way to help Chef Jose Andres to start thinking through building his own app and increase his outreach to the cooking world.

GOAL

Open the world of cooking to newcomers by giving them easy recipe recommendations based on what they have on hand and what they have enjoyed cooking in the past.

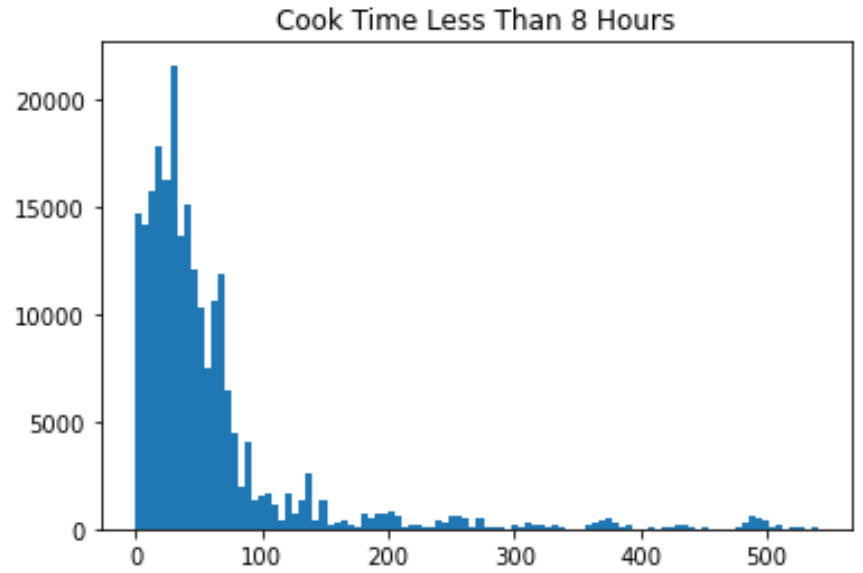
Start to build towards a collaborative recommendation model by beginning with a strictly content based approach.

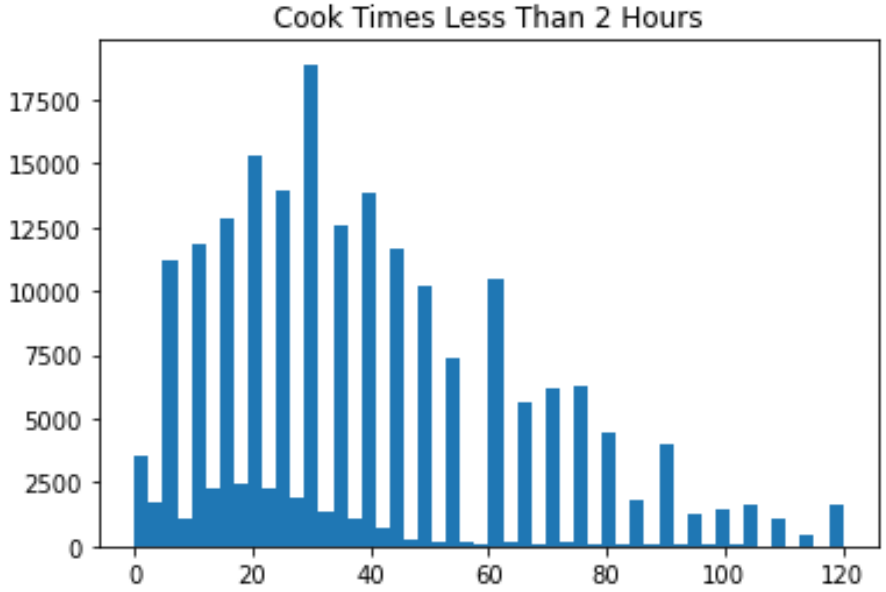
DATA

- This data comes from Kaggle and contains 230,000+ recipes
- Features: Tags, Description, Ingredients, Steps, Name, Recipe_ID
- What I focused on: Description, Ingredients, Tags
- Cleaning:
  - Separate Nutrition column into multiple columns including Calories, Total Fat, Sodium, Protein.
  - Drop NaN values from data frame.

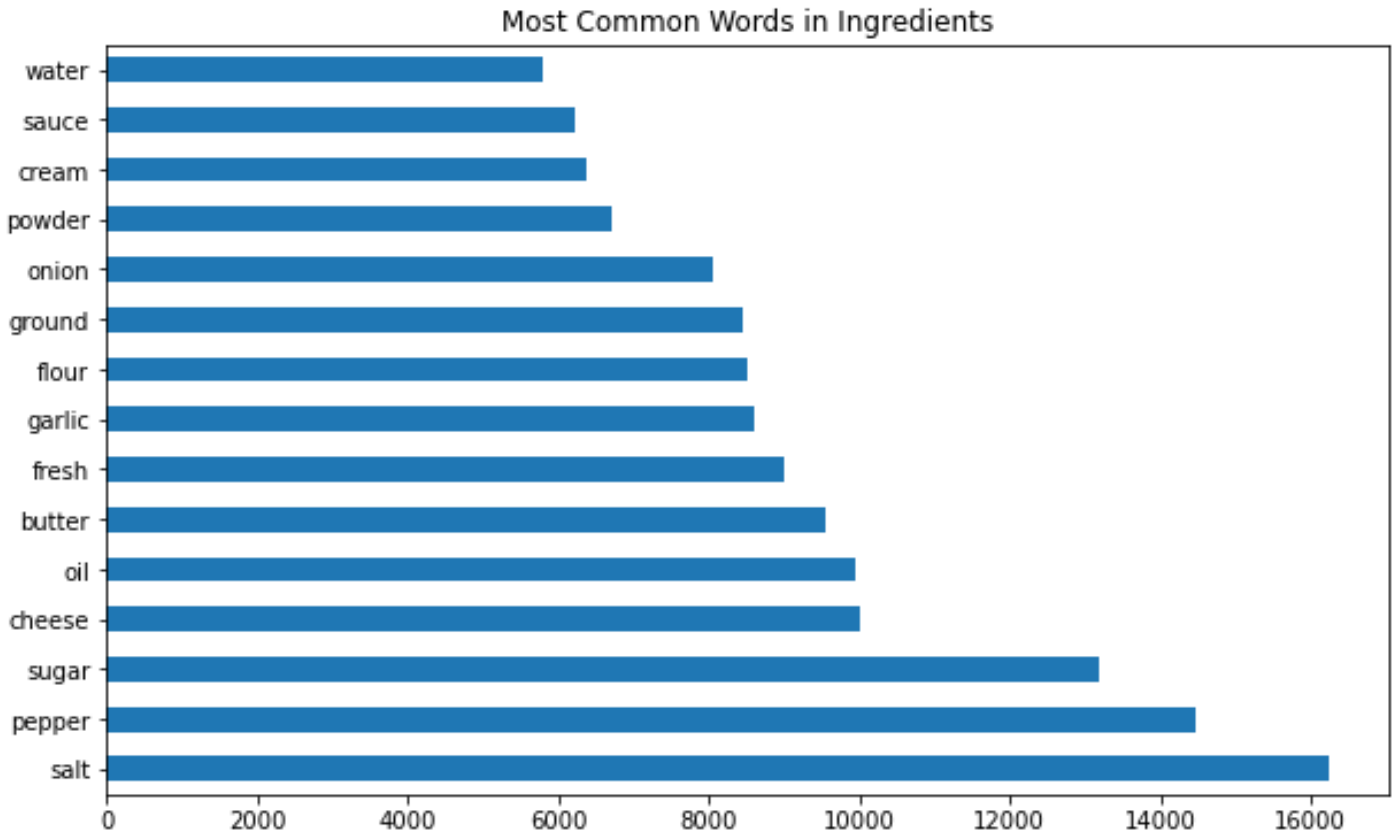
COOK TIME

- Cook times were grouped very closely for most of the recipes with 169,000 out of 230,000 coming in at under an hour.
- There were some outliers over 200,000 hours but these few were longer format or involved some sort of aging process and anything over 8 hours (still a one day recipe) was omitted from the final data being used.





COMMON INGREDIENT WORDS



MODELS: MOSTLY CLUSTERING

These models were not good, turns out there isn't a ton of separation

- DBSCAN:
  - Silhouette Score: -0.26826
  - Clusters: 20
- KMeans:
  - Silhouette Score: 0.5336
  - Clusters: 5
  - This had the best clustering of all of the models that I ran but even then the majority of recipes (nearly 11,000) were in the same grouping.

MODELLING CONCLUSION

By its nature cooking is very subjective as are recommendation systems so the lack of clear clustering is acceptable.

Unsurprisingly, the way that food is written about and the ingredients used are fairly common across the board and at the end of the day it is the method's being used to cook that really change your product.

METHODS/FUNCTIONS USED TO RECOMMEND

- What's in your pantry?
  - Based on a list of ingredients the system will recommend a recipe that includes those ingredients.
  - Using this to check against ingredient lists in the data frame we can recommend a recipe that you will be able to make.

- Cosine Similarity:
  - Input prior recipe and recommender will display new recipe.
  - CVEC used for Ingredients, TFIDF used for Descriptions.
  - Utilized cosine similarity for both description as well as ingredients.

WHAT IS IN YOUR PANTRY?

WHAT ELSE SHOULD YOU MAKE?

FURTHER WORK AND GAME PLAN MOVING FORWARD

For the time being, before we have actual user data, we can use this to get the recommendation started. There will be a cold starting issue but once we can reference users and item interactions the recommender will become stronger.

In the future we can also look to use tags to further segment our recommendations and created better clustering and better recommendations.

SOURCES

- https://www.kaggle.com/shuyangli94/food-com-recipes-and-user-interactions/tasks?taskId=164
- https://www.datacamp.com/community/tutorials/recommender-systems-python
- https://towardsdatascience.com/building-a-recommendation-system-using-neural-network-embeddings-1ef92e5c80c9