Abdeljalil Kabbaj

Building Haku: A Cooking Recipe Recommender

Background Information: Making Connections

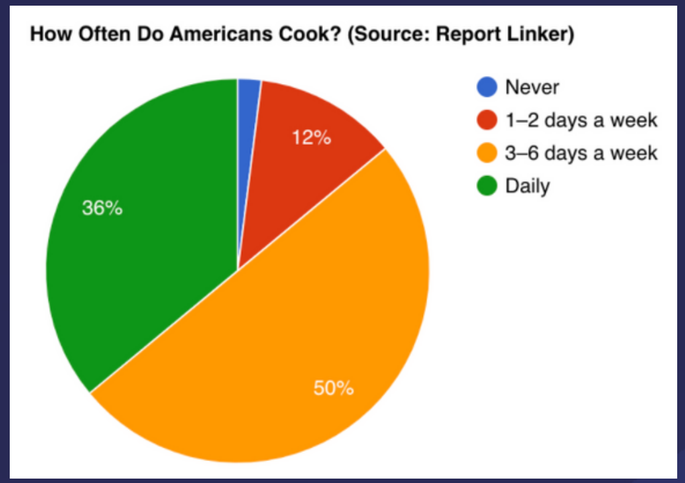
- Lizeth Aranda from Harvard Business Publishing says : “The cooking process for me is the clearest example of how we develop widely applicable skills while doing something we love. I think cooking not only develops our learning agility, but it also builds analytical skills and mathematical thinking. Among all the skills we acquire in the process of cooking, I think creativity and innovation are the most powerful.”

- Problem solving is the strong ability to find relationships (connections) between the resources we have and what we want to solve. I believe that grasping the idea of creating new experiences through our taste buds strengthens our ability to make connections when we try to solve a problem.

- I enjoy experiencing how taste can be changed by using the same ingredients. I believe that food can be transformed in countless ways, it is just a matter of making connections. For my project, I wanted to build a model that was able to help individuals make these connections.

Addressing a Problem

- Only about 36% Americans cook on a daily basis.
- Only 13.7% prepare their meals from a passion for cooking.
- Retirees are most likely to cook at home while individuals that work daily lack the time to do so.



Problem Statement

The objective is to build a recommender system able to bring forth various recipes an individual can make with the ingredients they have at home.

Because this is an unsupervised learning model, the metric of evaluation for success will be entirely reflected on the satisfaction of the user.

Data Cleaning

- Data Set original size: 2,231,142 rows and 7 columns
- Objective: Ingredient column must contain only the name of the ingredient (no brand, verb/adverb, measurements)
- Here is how I cleaned it:
  - Reduced recipes by half
  - Used spaCy to locate verbs, adverbs , and unnecessary adjectives. Added those words together and removed them from the ingredients column
  - Removed unwanted punctuation
  - Reduced rows by eliminating recipes containing more than 20 ingredients
- Data set final size: 1,095,610 rows and 4 columns

EDA - Word Count Distribution

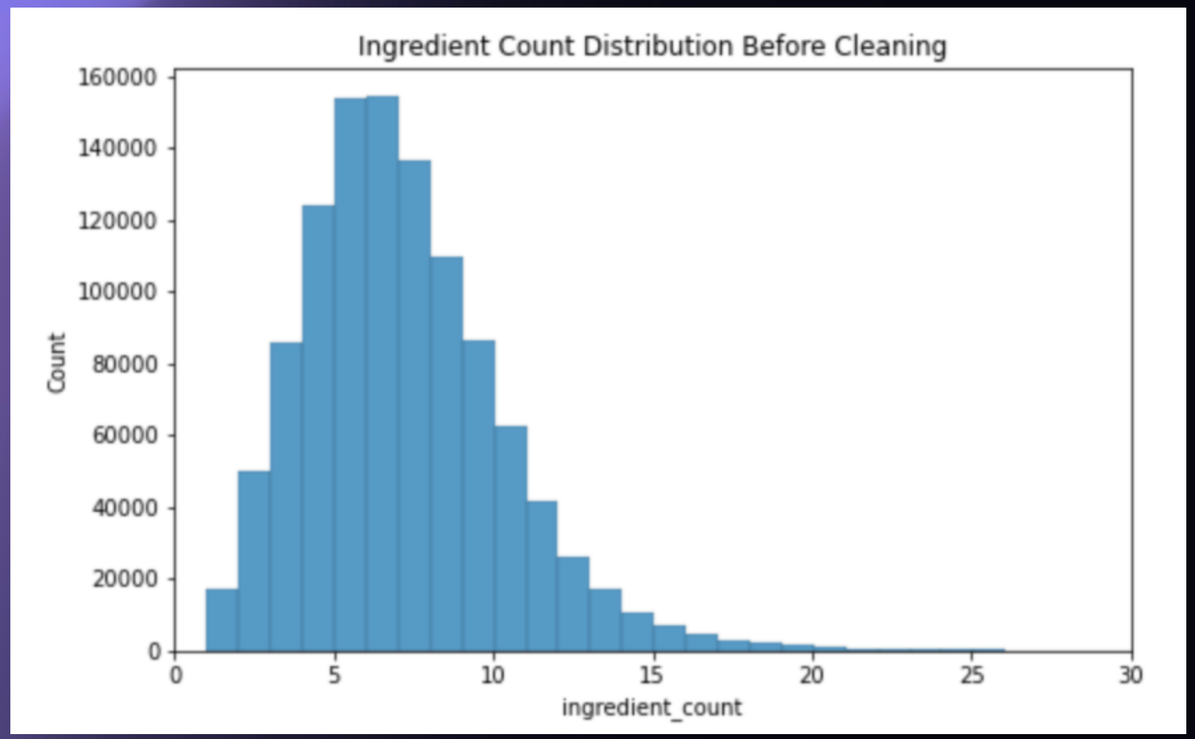


Figure 1.

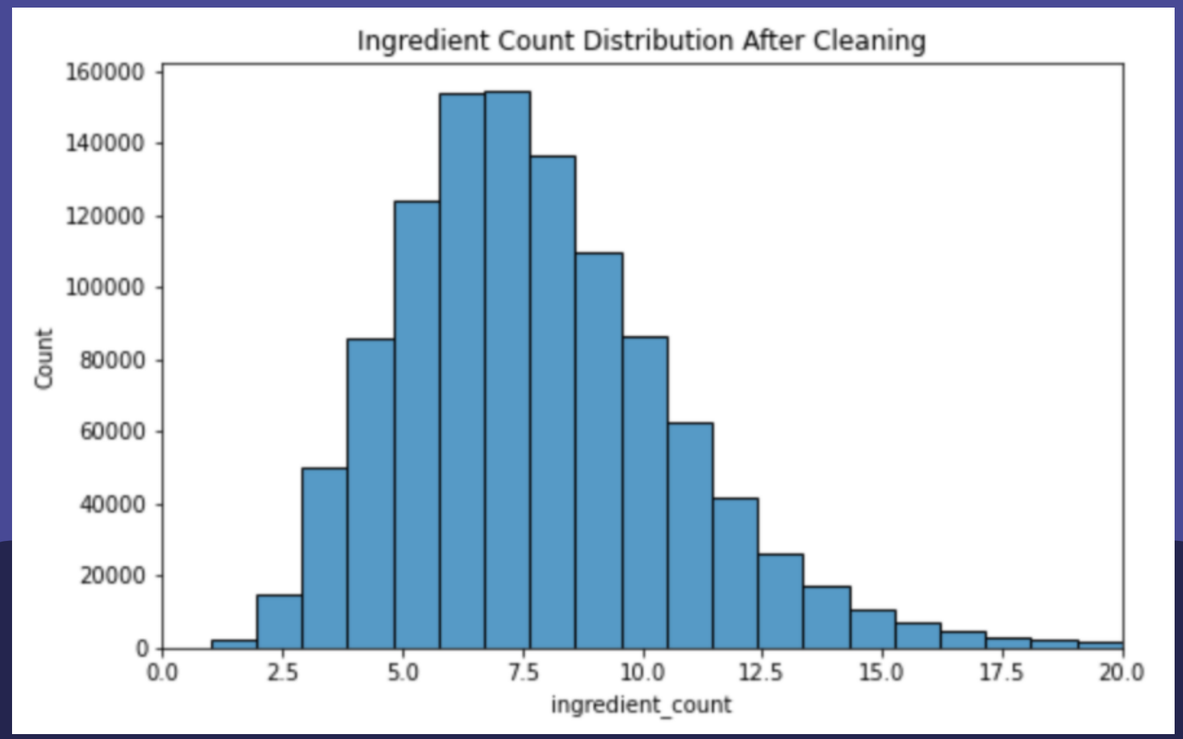


Figure2.

EDA - Word Count Box Plot

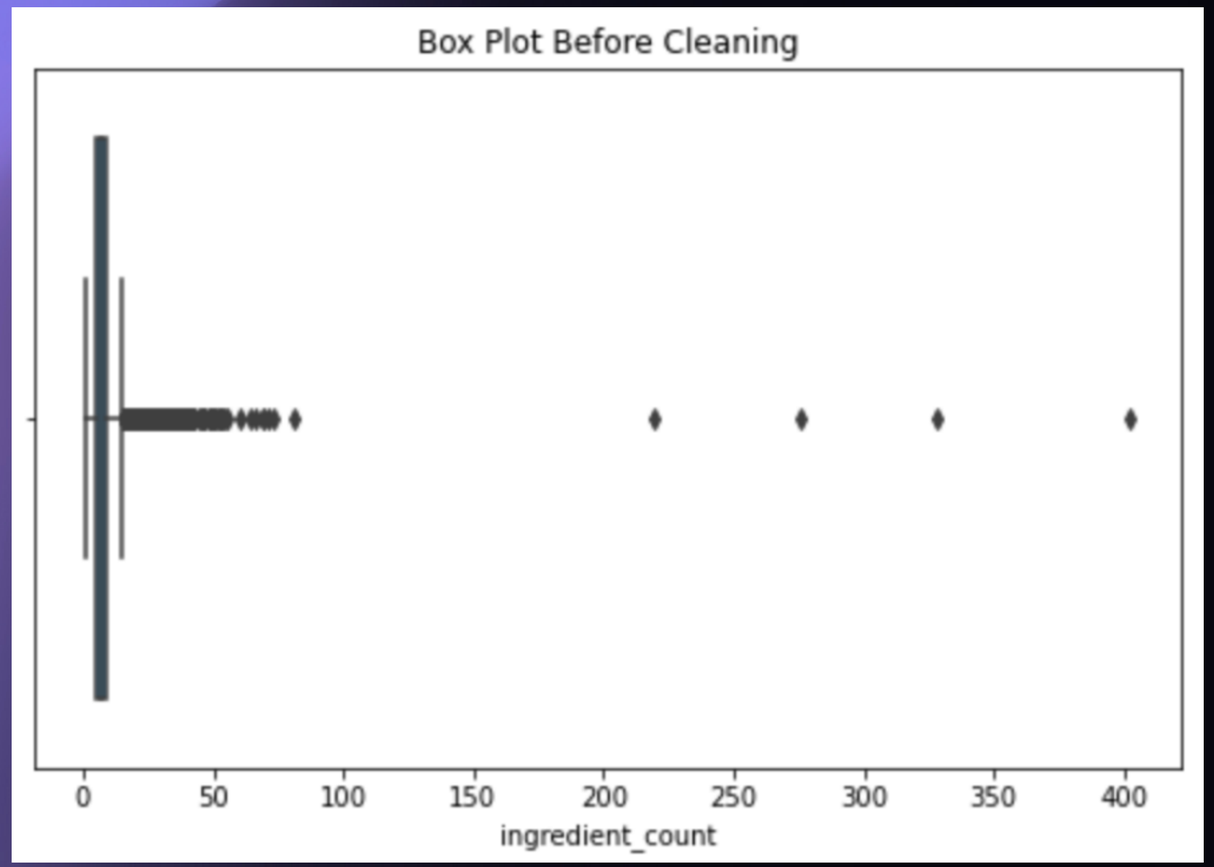


Figure 1.

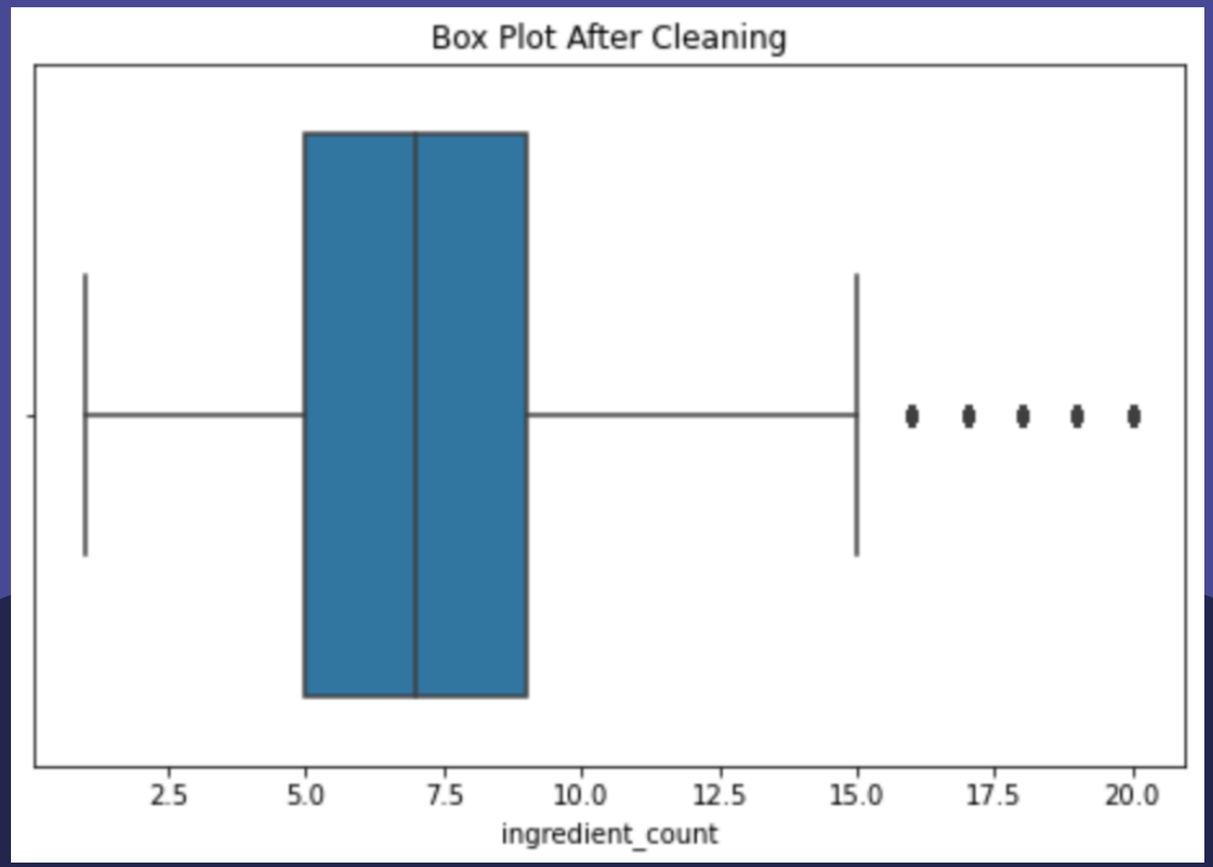


Figure2.

EDA - Count/TFIDF Vectorizer

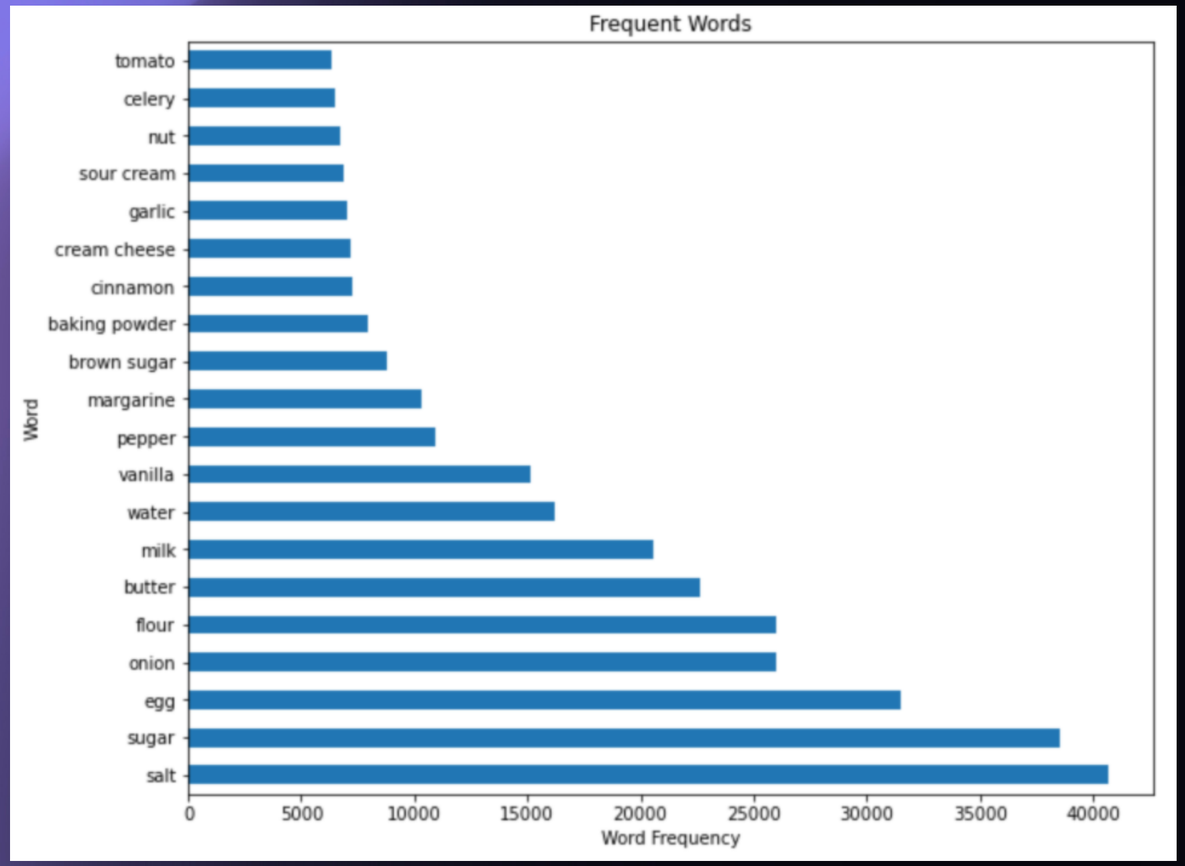


Figure 1.

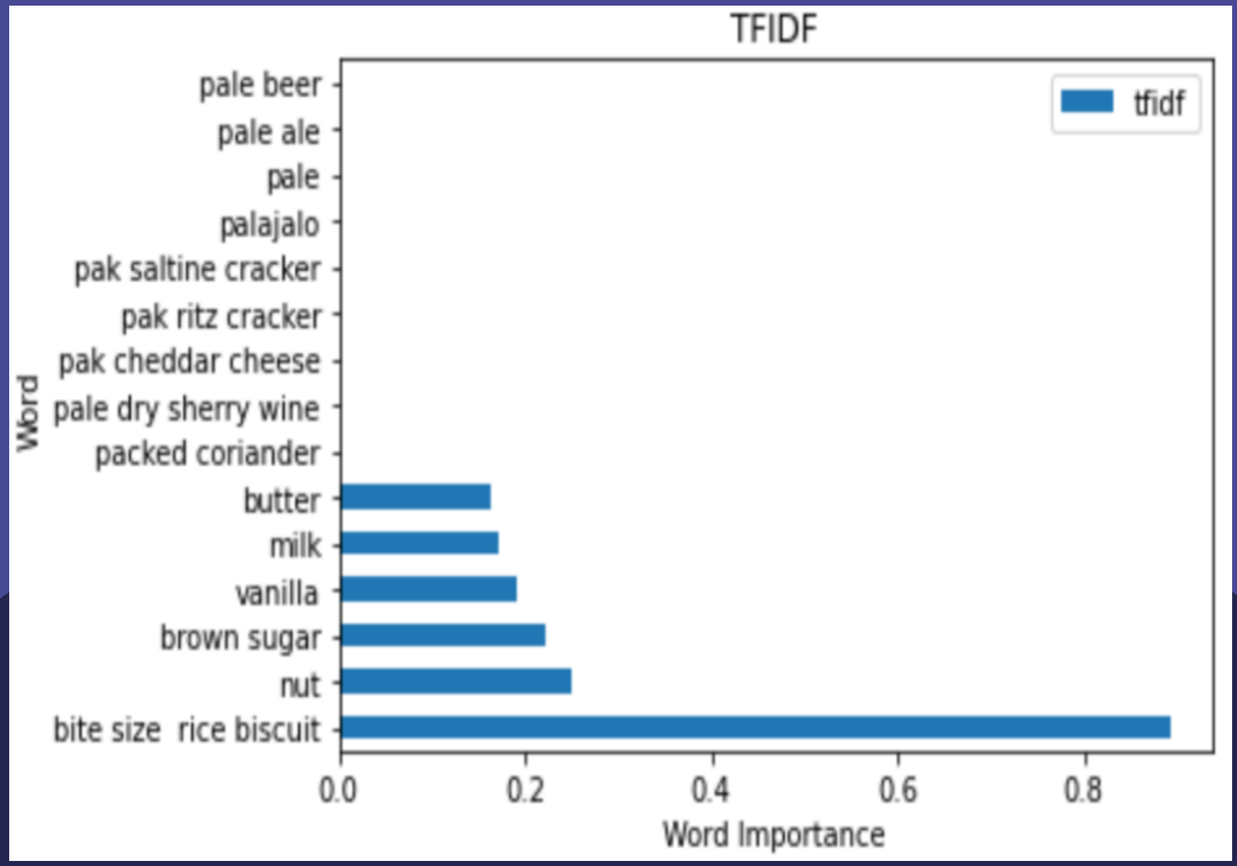
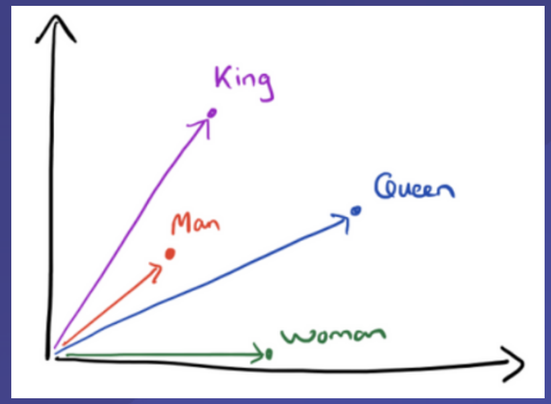


Figure2.

Turning words into vectors using Word2Vec:

Inputting all the ingredients from the different recipes to word2Vec, the neural network learns the association between each word and is able to output the vectorized form of these words.

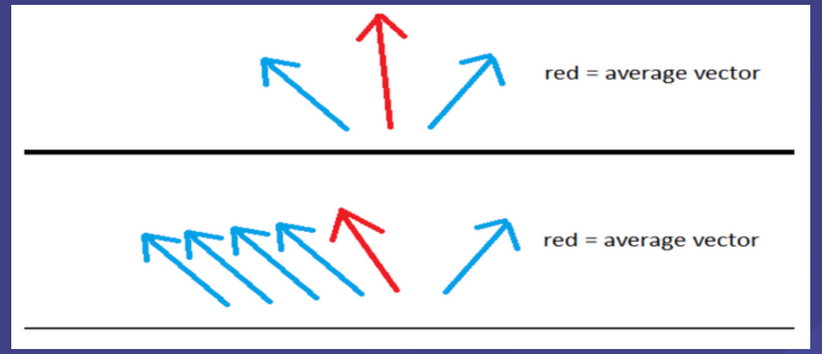
King, Queen, Man, Woman



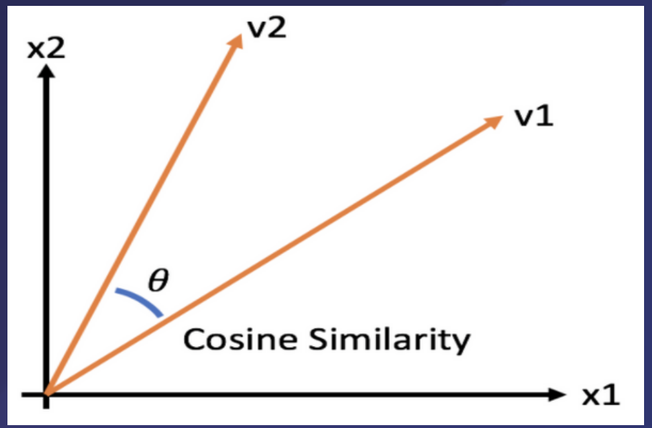
Using the cosine similarity, the angle between two vectorized words, we are able to determine whether these words are similar in meaning.

Building the Recommender System

Vectorizing lists of recipes as one vector by taking the mean of those vectors:



Calculating new cosine similarity. The greater the value for the cosine similarity the greater the match:



A look at Haku

THE METRIC OF EVALUATION FOR MY MODEL IS USER SATISFACTION WITH THAT SAID LET'S GIVE IT A TRY AND DETERMINE ITS SUCCESS.

Conclusions & Recommendations

- The Haku prototype can be used as a tool to inspire and to try new recipes even with a limited number of ingredients we can have at home.
- As a base-model, allowing the user to choose from 3 different trained models, I am satisfied with the performance of Haku.

- In order to further develop and improve the web app, I intend to work on the speed at which it outputs results.
- Work with smaller models and optimize its performance in order to deploy the app online.
- A future for Haku is a model that generates-recipes instead of matching already-existent recipes. This would be a complex and deeper investigation, but working on this project gave me an introduction to the possibilities that still exist.

Thank you.
Do you have any questions?

ABDELJALIL KABBAJ