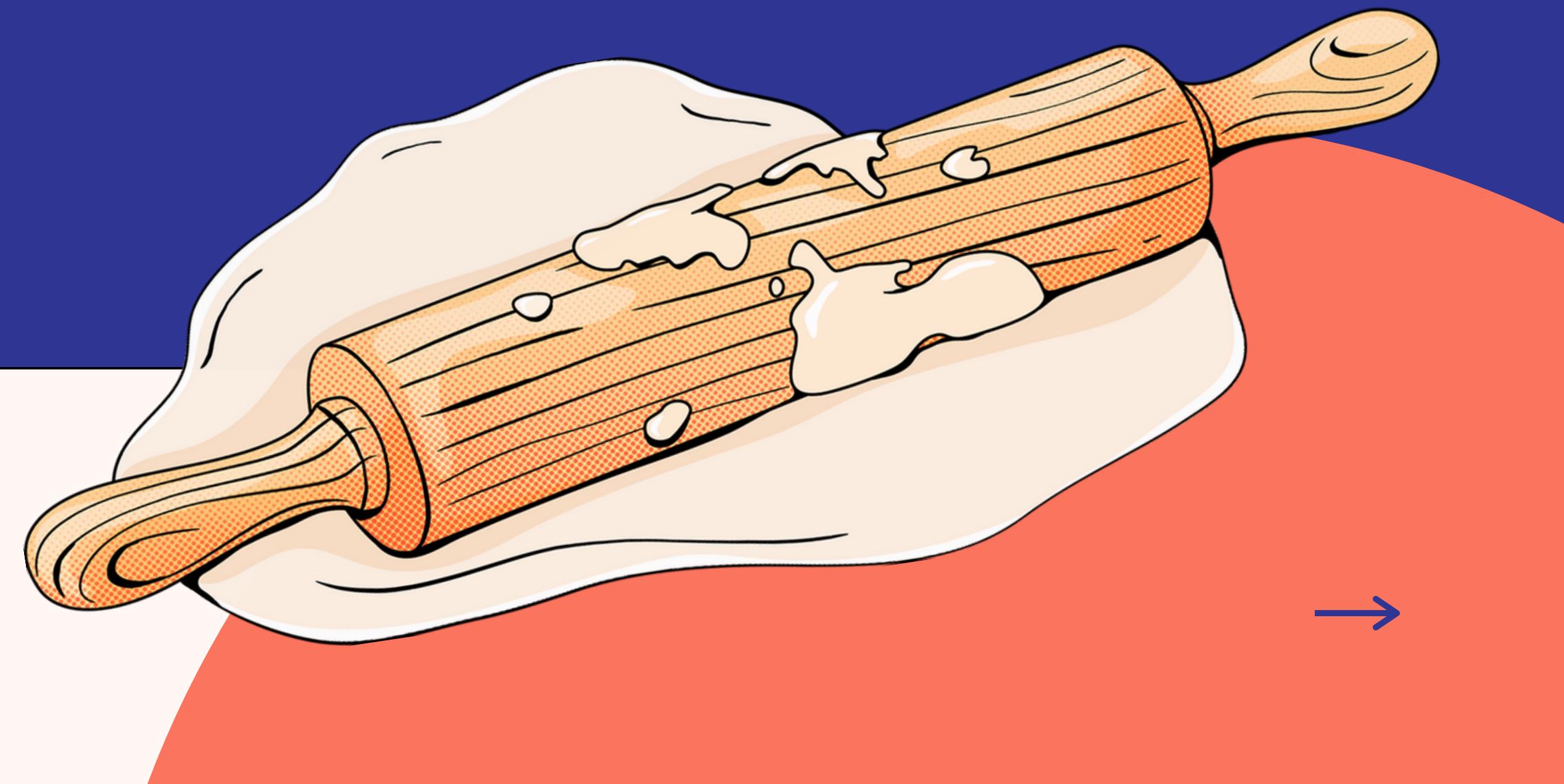


# HOMECHEF

## DIGITIZING THE FOOD CHAIN

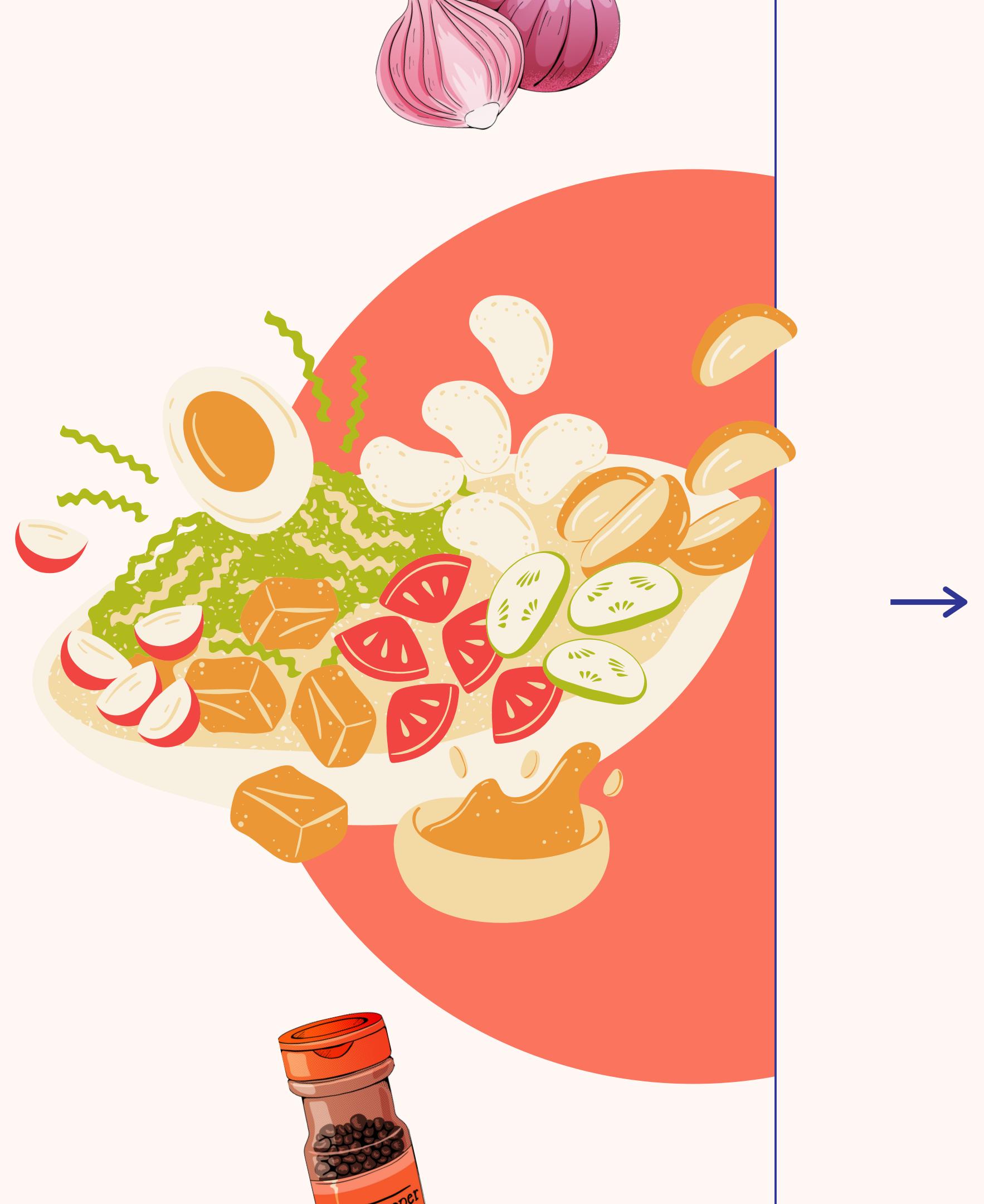


DSI CAPSTONE PROJECT  
MUBINA KAPASI



# MEAL KIT SERVICE: AN OVERVIEW

- An intersection between meal and grocery delivery services
- Enjoy the benefits of a home-cooked meal without the hassle of planning meals, buying groceries, and worrying about food wastage



# BUSINESS MODEL

## BASIC PLAN

- Novice home cooks can select recipes every week and ingredients will be provided to their door

## FLEX PLAN

- Home cooks looking for a challenge can purchase their own ingredients a la carte from our HomeChef's marketplace
- HomeChef will offer suggestions for cuisines, recipes, and other ingredients users may want to purchase based on their current selection



# PROBLEM STATEMENT



1. Gain insights on the best cuisines, ingredients, and recipes to provide consumers
2. Use multiclass classification algorithms to predict the cuisine of a dish based on available ingredients
3. Provide recommendations for recipes based on available ingredients

# DATA COLLECTION



1. Yummly: 15K+ recipes from 18 different cuisines
2. Remove cuisines and recipes with minimal data
3. Ingredient parser to extract list of ingredients from recipes

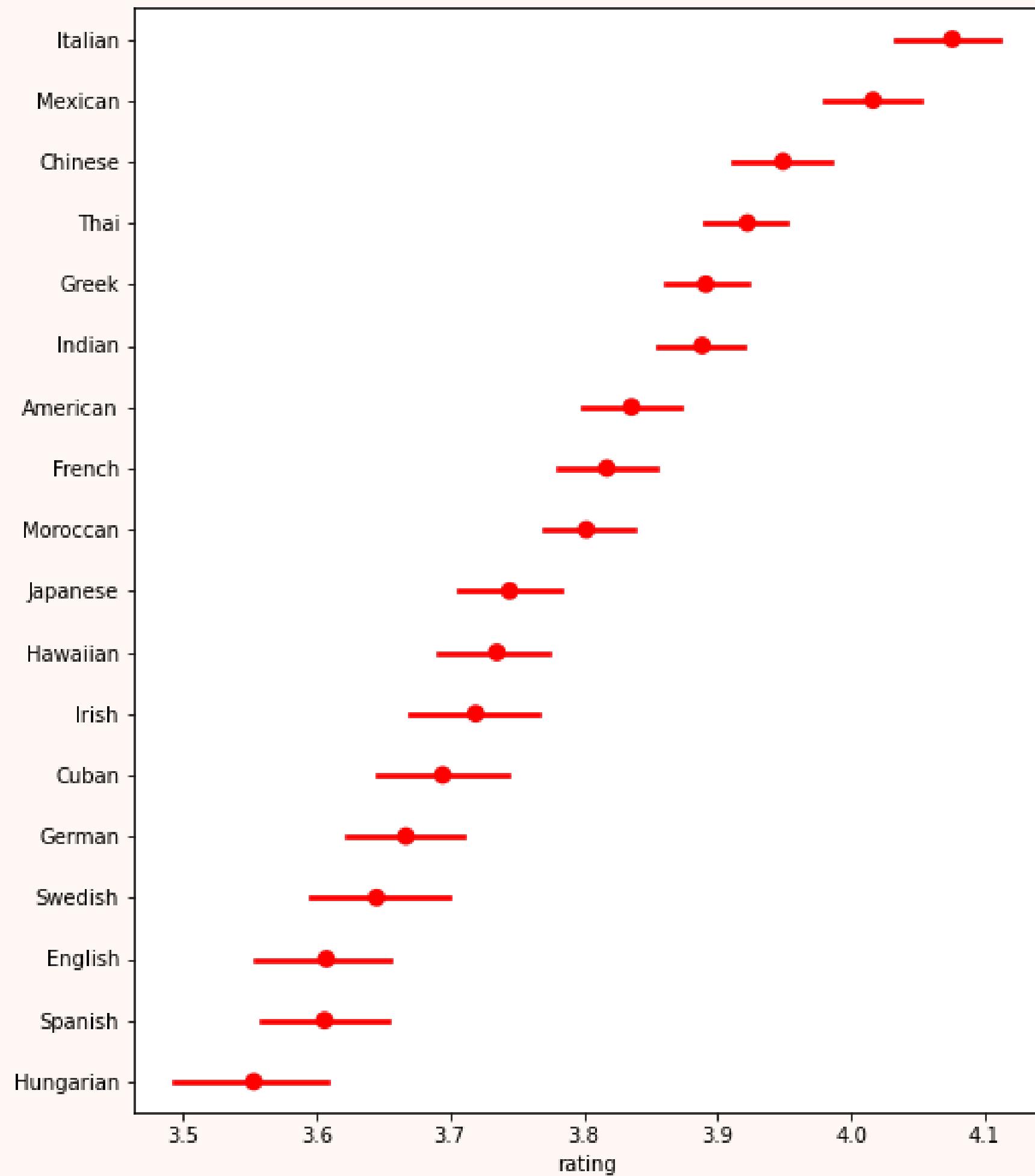


# COMMON INGREDIENTS



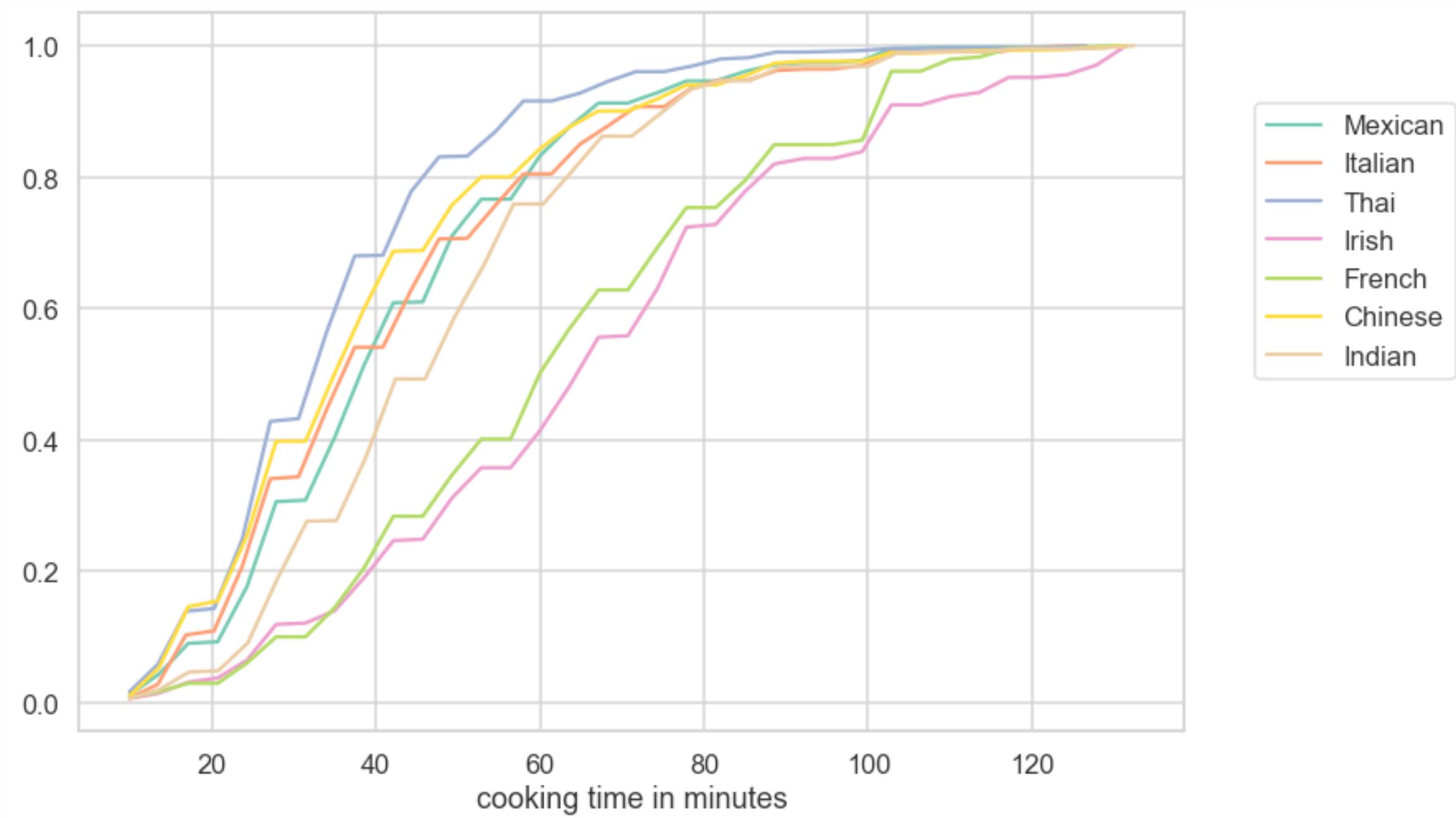
# BEST RATED CUISINES

Italian and Mexican recipes have the highest ratings, while Hungarian recipes have the lowest rating



# CUMULATIVE COOKING TIME

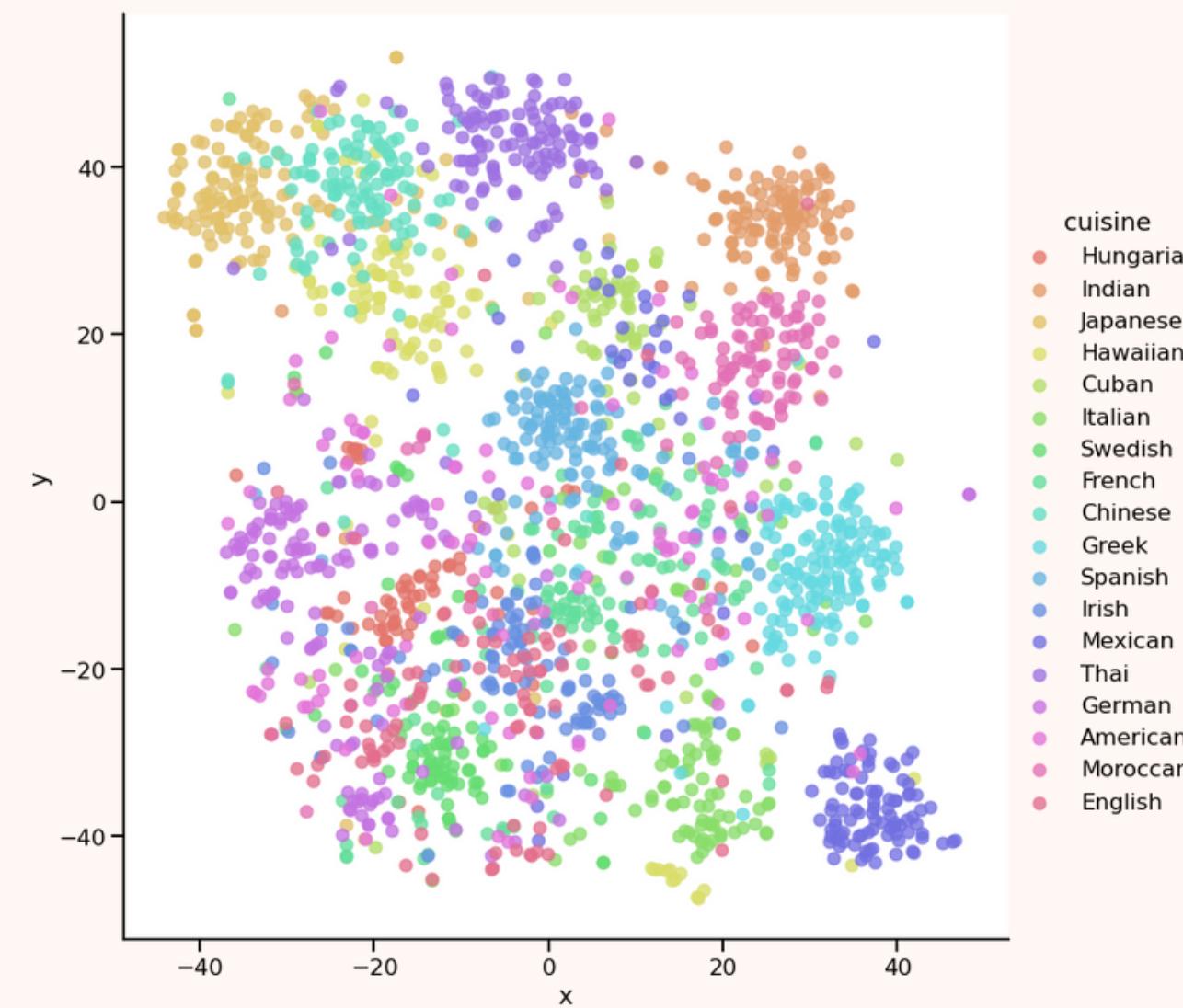
90% of Thai recipes take less than 60 minutes to make, while only 40% of Irish recipes can be completed within the same duration



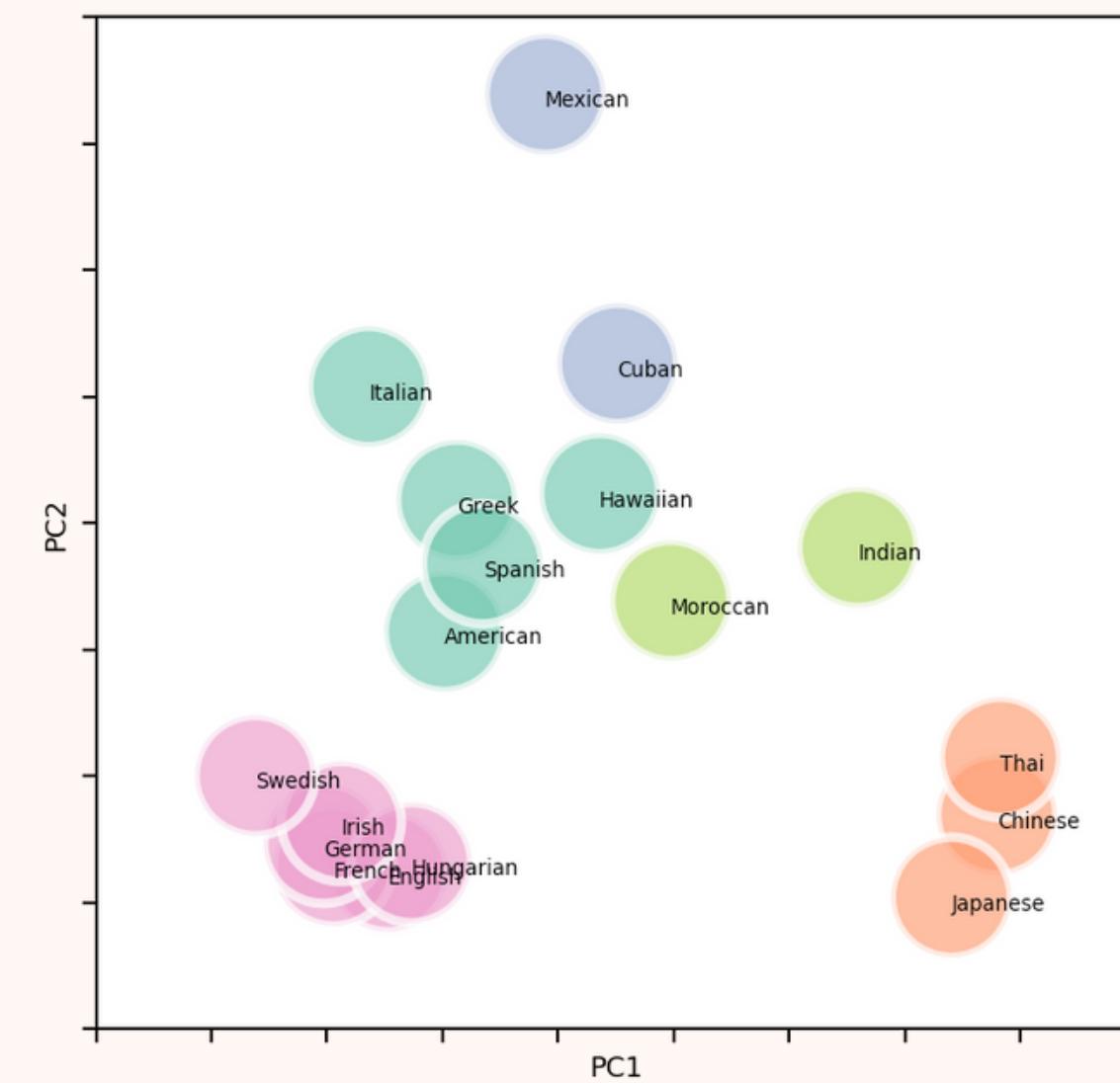
# CLUSTERING ANALYSIS

## T-SNE AND KMEANS

t-stochastic neighborhood embedding for all recipes



K-Means clustering by cuiisine, k = 5



Some cuisines are more easy to distinguish than others: Mexican, Indian, and Thai recipes do not overlap that much with other cuisines, while English and American recipes are scattered

With k=5, clusters seem to correspond to different geographic regions





explore cuisines  
interactively!



# MODELING: CUISINE CLASSIFICATION

## 1. TF-IDF Vectorizer

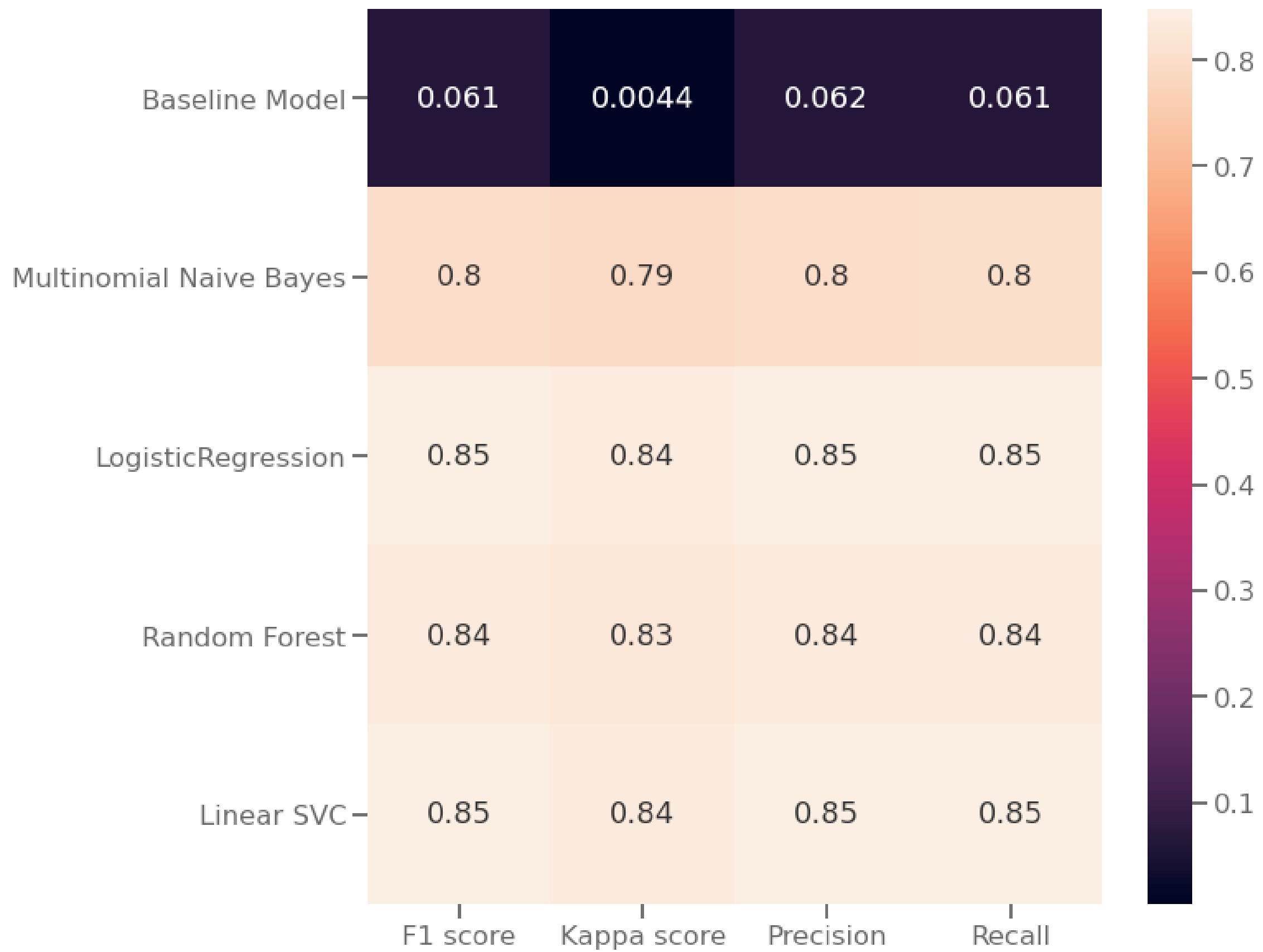
- maximum frequency of an ingredient occurring in a recipe is 1
- maximum document frequency is 95%

## 2. Multiclass Classification Model

- Baseline Model
- Multinomial Naive Bayes
- Logistic Regression
- Random Forest Classifier
- Support Vector Classifier



LOGISTIC  
REGRESSION IS THE  
BEST PERFORMING  
MODEL



# HOW DO CUISINES INFLUENCE EACH OTHER?

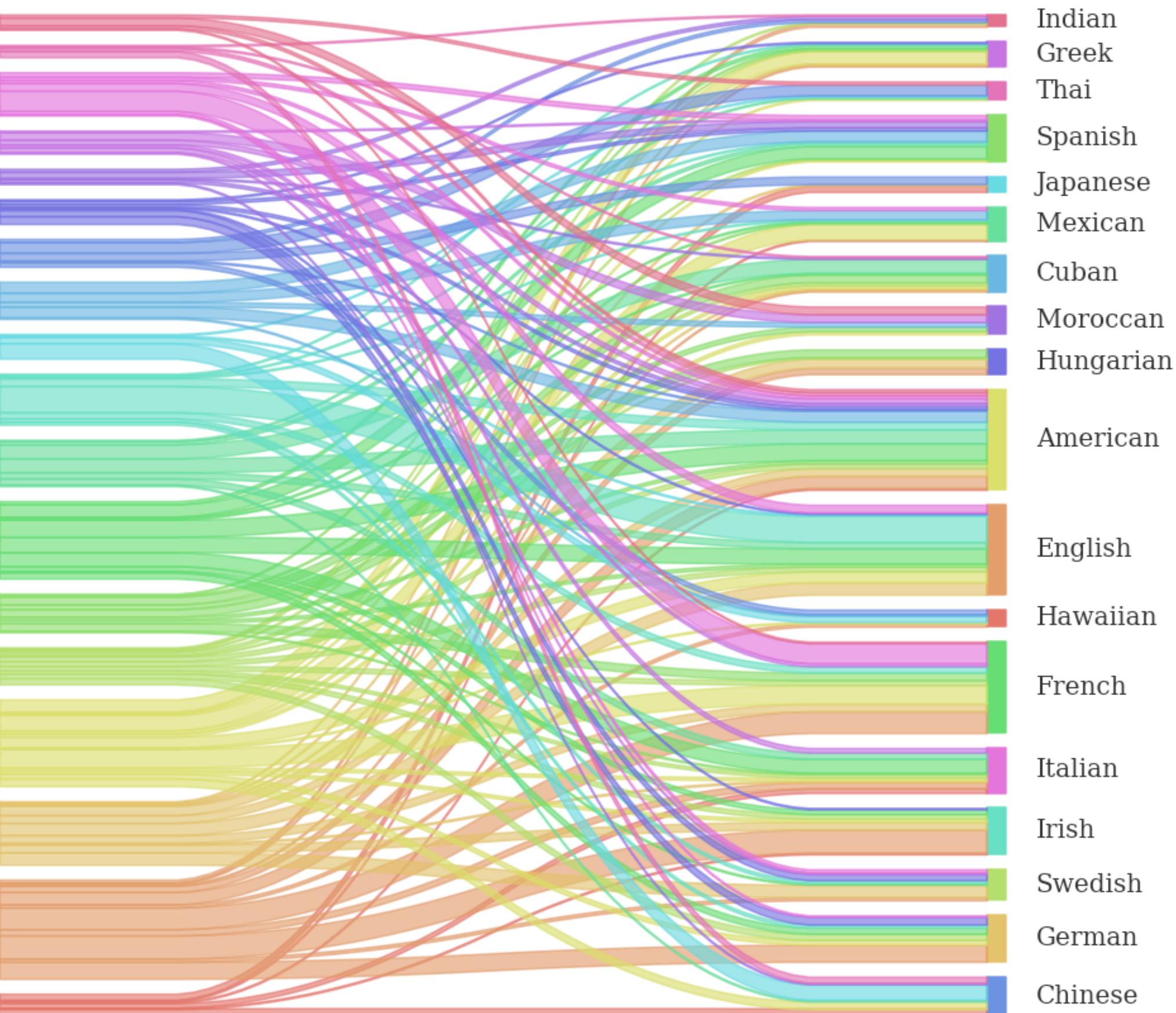
Most of the cuisines that are misclassified  
are geographically close



true labels

Indian  
Thai  
Italian  
Greek  
Moroccan  
Hungarian  
Chinese  
Cuban  
Japanese  
Irish  
Mexican  
French  
Spanish  
Swedish  
American  
German  
English  
Hawaiian

predicted labels





**CHINESE**

soy sauce



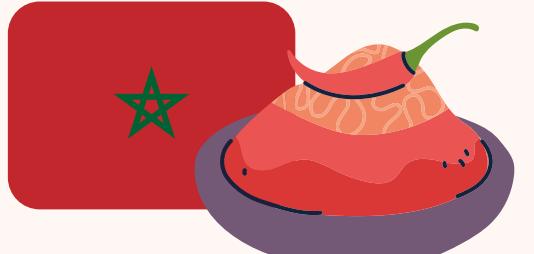
**GERMAN**

sauerkraut



**FRENCH**

thyme



**MOROCCAN**

harissa



**IRISH**

potato



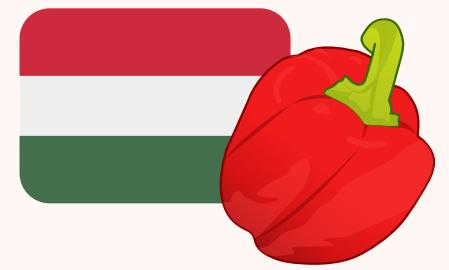
**SPANISH**

chorizo



**MEXICAN**

tortilla



**HUNGARIAN**

paprika



**ITALIAN**

pasta



**AMERICAN**

pecan



**CUBAN**

plantain



**ENGLISH**

pudding



**GREEK**

feta cheese



**HAWAIIAN**

pineapple



**INDIAN**

turmeric



**JAPANESE**

mirin



**SWEDISH**

meatball



**THAI**

fish sauce

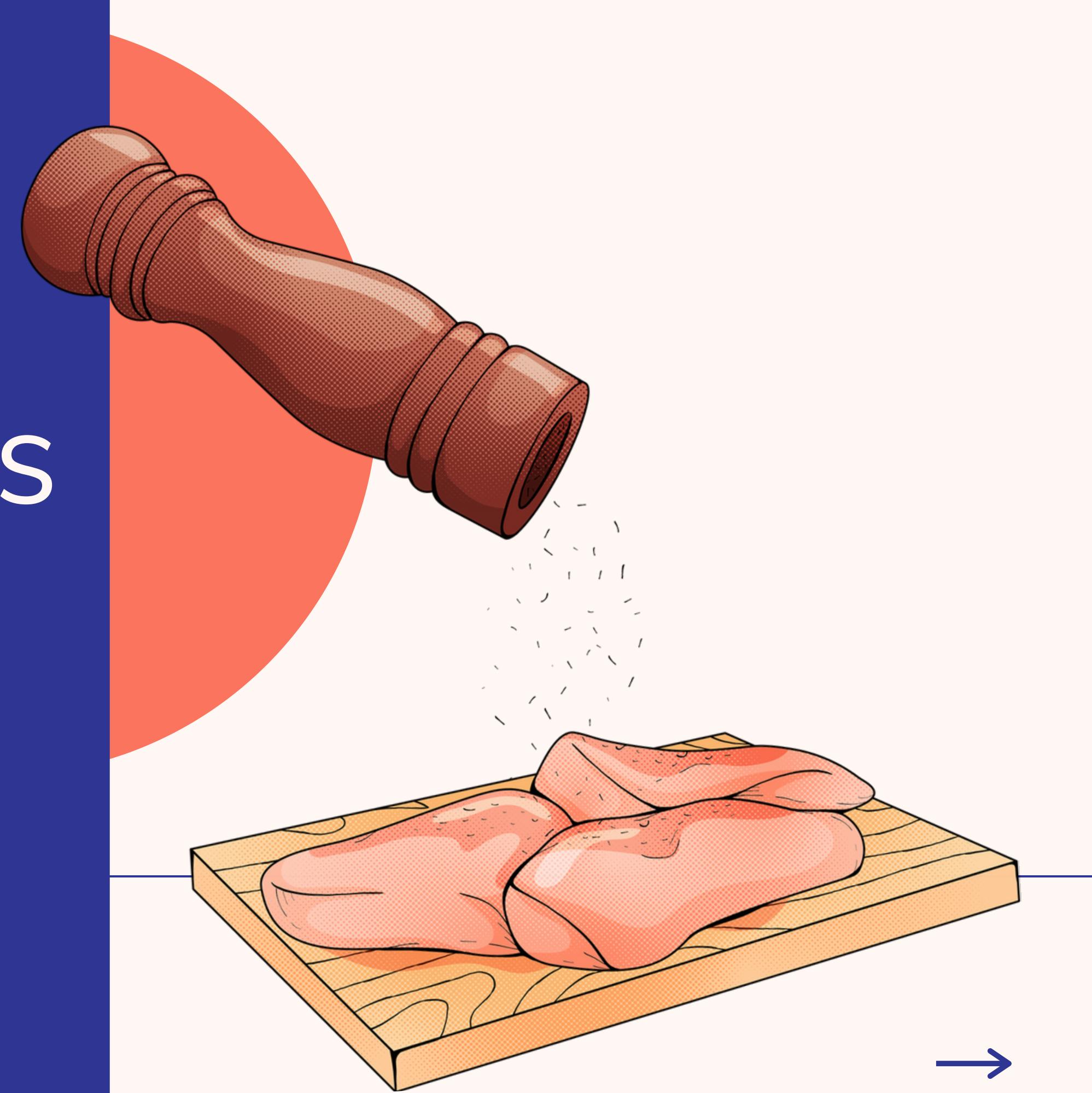
**MOST DISTINCTIVE  
INGREDIENTS**

# MODELING: RECIPE RECOMMENDATIONS

- 1.TF-IDF vectorizer
- 2.Cosine Similarity



Generate your own recipe  
recommendations on telegram!



# BUSINESS RECOMMENDATIONS

Focus on providing recipes from the following 5 cuisines:

1. Mexican
2. Italian
3. Chinese
4. Indian
5. French

In general, these cuisines have:

1. High recipe ratings
2. Manageable cooking times (except French)
3. Belong to different k-Means clusters



# FURTHER STEPS

1. Collect recipes from more sources for full scope of taste preferences and variations
2. Sentiment analysis on target audience to better understand how consumers will react to certain cuisines and recipes
3. Collect user feedback to evaluate recipe recommendation system effectively
4. Build a personalized recommendation system



# Thank you!

[click here to view my code for this project](#)

