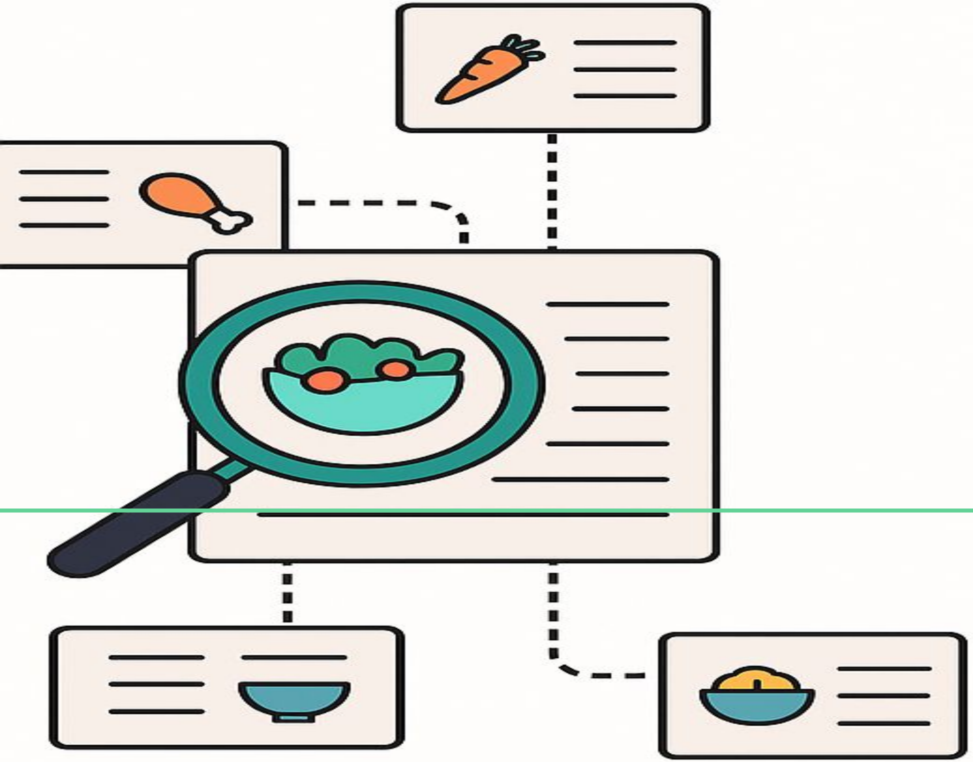


Recipe Clusterin Project

Chitra Gopalaiah

Project Overview

- Goal: Identify latent recipe clusters based on nutritional profiles
- Dataset: Food.com Recipes & Reviews (~500K recipes, 1.4M reviews)
- Methodology: K-Means clustering, PCA visualization, ingredient frequency analysis
- Outcome: Meaningful flavor/nutrition-based clusters and a demo recommendation system

Dataset Overview

Source: Food.com Recipes and Reviews Dataset (Kaggle)

Recipes Dataset:

- 522,517 recipes
- 28 columns, including recipe metadata and nutritional info
- Examples: RecipeId, Name, Author info, Calories, FatContent, ProteinContent, SugarContent

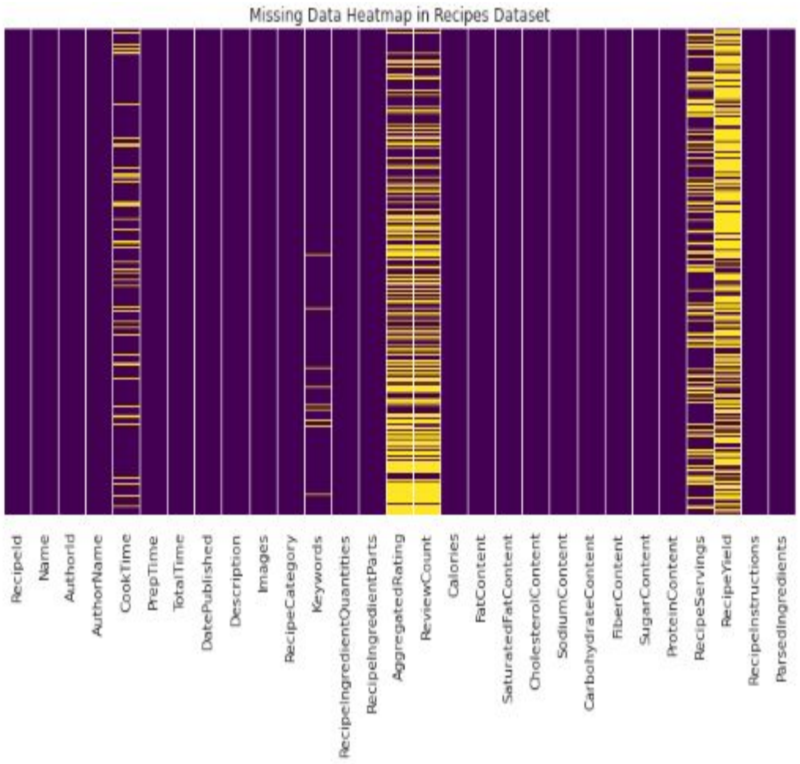
Reviews Dataset:

- 1,401,982 user reviews
- Ratings (1 to 5) and review texts linked to recipes

: Data Cleaning

Data Cleaning:

- Handled missing values in nutrition and ingredient fields
- Standardized nutritional features to same scale for clustering
- Removed or imputed incomplete entries to improve data quality



Data Preparation

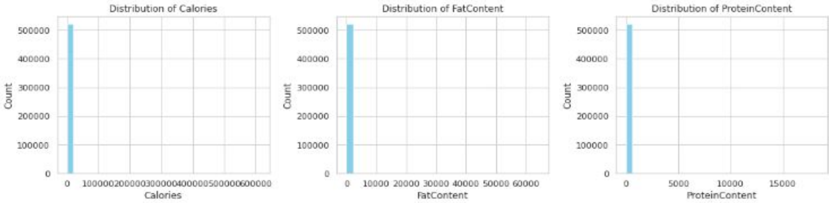
Data Preparation:

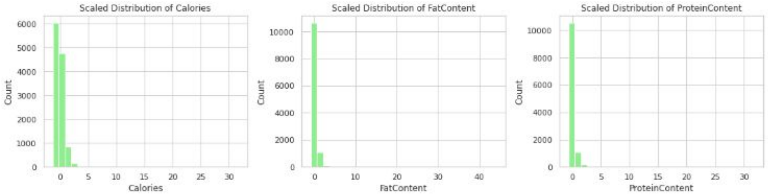
- Standardized nutritional features to the same scale
- Extracted key nutritional features (Calories, FatContent, ProteinContent, etc.)
- Normalized data using standard scaling for unbiased clustering

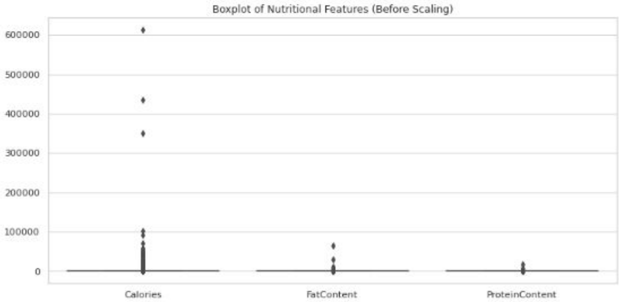
Result:

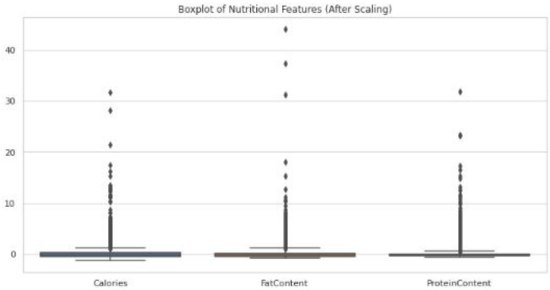
- Ready-to-use dataset for unsupervised clustering and analysis

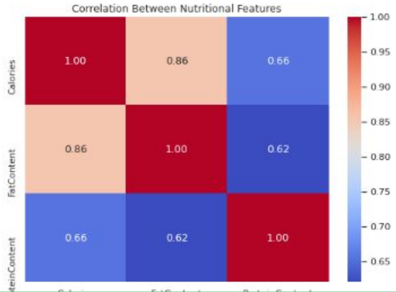
Data Preparation Visualizations

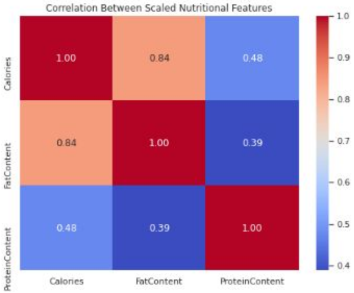






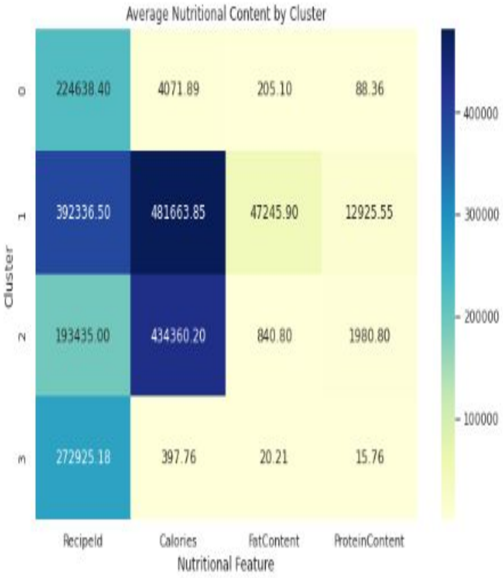


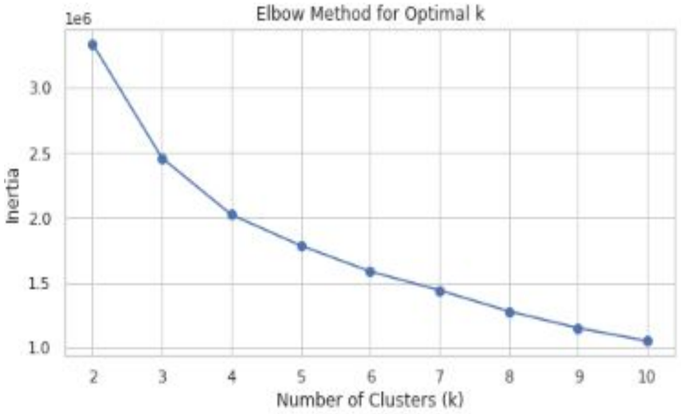


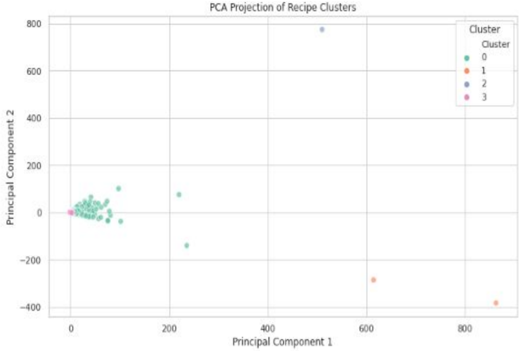


Clustering Recipes by Nutritional Profiles

- Applied K-Means clustering to group recipes based on standardized nutritional data
- Selected 4 clusters using the elbow method for optimal balance
- Visualized clusters with PCA for dimensionality reduction
- Analyzed cluster characteristics: calories, fat, protein, and user ratings

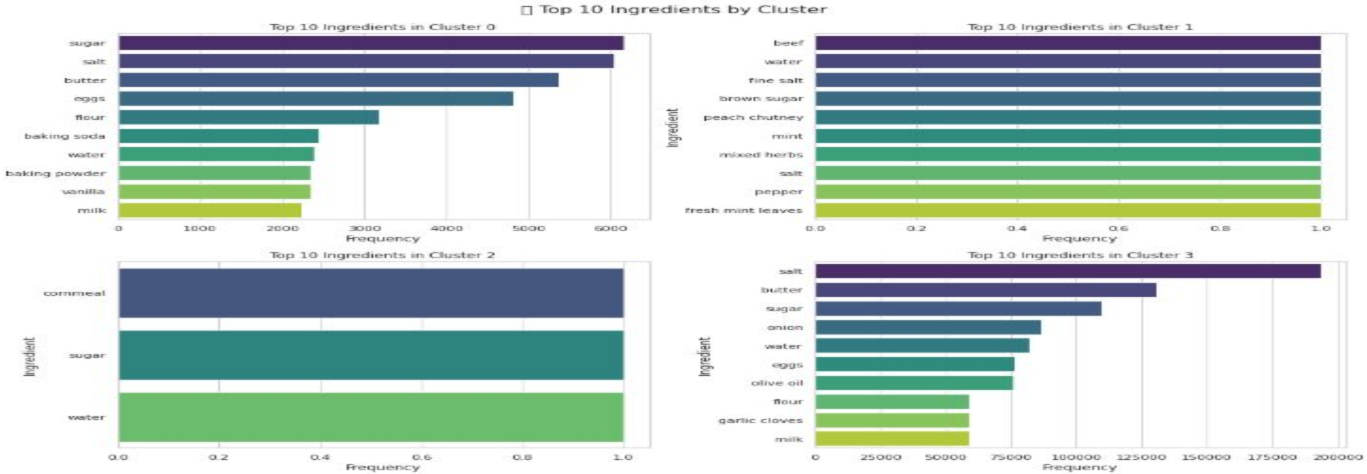






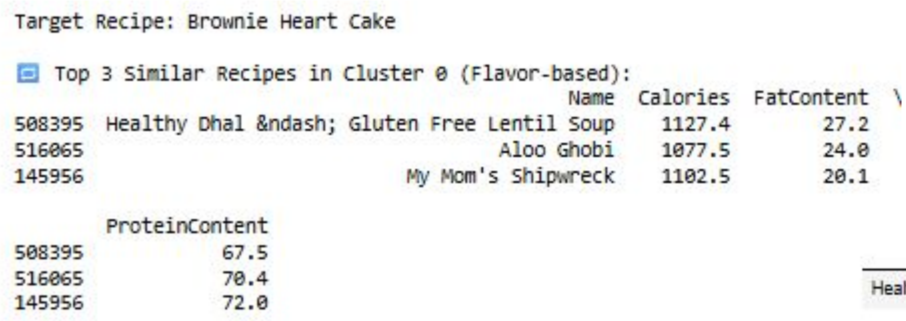
Ingredient Frequency Analysis by Cluster

- Examined the most common ingredients within each cluster to reveal culinary patterns
- Cluster 0: Baking ingredients (sugar, butter, eggs, flour) — desserts and baked goods
- Cluster 3: Savory staples (salt, onion, garlic, olive oil) — main dishes and sides
- Clusters 1 & 2: Sparse or outlier ingredient profiles — possible data anomalies or niche recipes



Flavor-Based Recommendation Demo

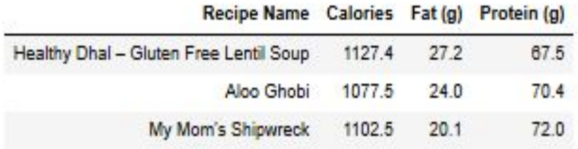
- Demonstrated how to recommend similar recipes within a cluster based on nutritional profiles
- Example: Given the recipe “Brownie Heart Cake” in Cluster 0, we retrieved 3 nutritionally similar recipes
- Recommendations consider calories, fat, and protein to match flavor and nutritional style



```
Target Recipe: Brownie Heart Cake

▣ Top 3 Similar Recipes in Cluster 0 (Flavor-based):
                                          Name  Calories  FatContent  \
508395  Healthy Dhal &ndash; Gluten Free Lentil Soup    1127.4        27.2
516065                                Aloo Ghobi    1077.5        24.0
145956                         My Mom's Shipwreck    1102.5        20.1

        ProteinContent
508395            67.5
516065            70.4
145956            72.0
```



Recipe Name	Calories	Fat (g)	Protein (g)
Healthy Dhal – Gluten Free Lentil Soup	1127.4	27.2	67.5
Aloo Ghobi	1077.5	24.0	70.4
My Mom's Shipwreck	1102.5	20.1	72.0

Conclusion & Future Work

Conclusion:

- Successfully applied unsupervised learning (K-Means clustering) to group recipes by nutrition
- Discovered meaningful clusters aligned with culinary categories (desserts, savory mains)
- Ingredient analysis and user ratings helped interpret cluster profiles
- Developed a flavor-based recommendation demo for personalized recipe discovery

Future Work:

- Experiment with other clustering methods (GMM, hierarchical) for improved cluster quality
- Incorporate semantic ingredient embeddings (TF-IDF, Word2Vec) for deeper flavor analysis
- Build interactive dashboards for user exploration of recipe clusters and filters
- Integrate user dietary preferences (vegetarian, keto, allergies) into recommendations
- Address data sparsity and clean clusters with missing or anomalous data