A Tourist's Taste

Analyzing Korean Food Discussions on Reddit's r/KoreanFood
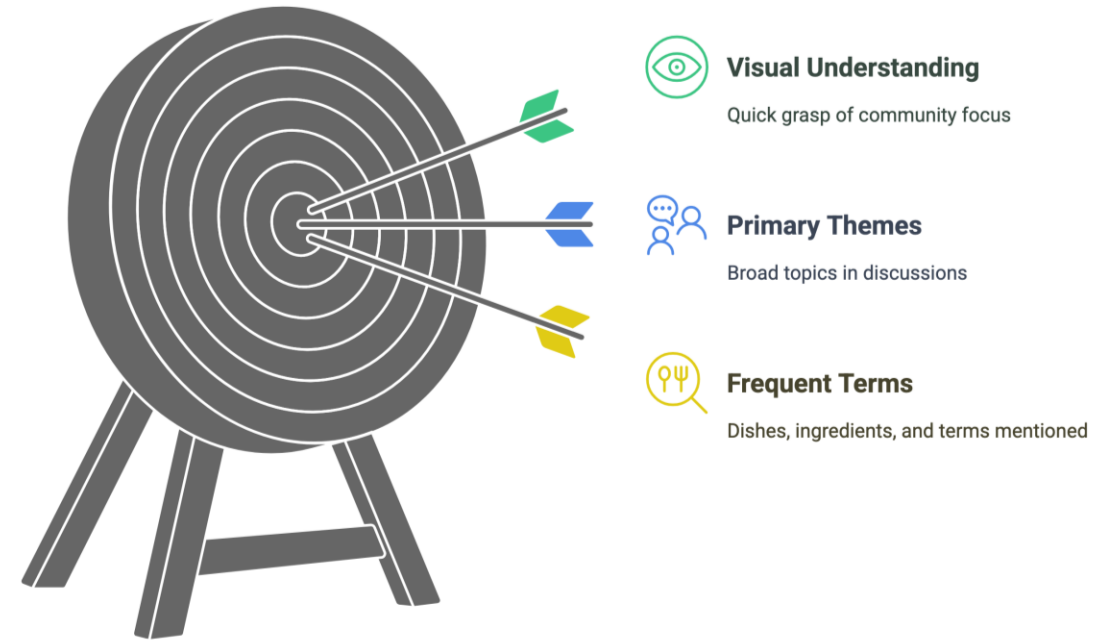
Introduction & Project Goal

- **Background:** Online communities like Reddit are rich sources of user-generated content, offering insights into public opinion and experiences.

- **Focus:** This project analyzes discussions on the *r/KoreanFood* subreddit to understand what users (likely including many tourists or those interested in Korean cuisine) **say about Korean food.**

- **Project Goal:** To identify the most discussed food items, ingredients, experiences, and common themes related to Korean cuisine as presented on *r/KoreanFood.*

Research Questions?

- This project aimed to answer:

1. What are the most frequently mentioned Korean dishes, ingredients, and food-related terms on **r/KoreanFood?**

2. What are the primary themes or **topics** that emerge from discussions about Korean food on this subreddit?

3. Can we visualize the **most prominent terms** to get a quick **understanding of the community's focus?**

👁 **Visual Understanding**
Quick grasp of community focus

💬👥 **Primary Themes**
Broad topics in discussions

🔍 **Frequent Terms**
Dishes, ingredients, and terms mentioned

Methodology - Data Collection & Source

Data Source: Reddit, specifically the r/KoreanFood subreddit.

Reasoning: Chosen for its high concentration of discussions specifically about Korean cuisine, ingredients, and recipes, likely attracting both enthusiasts and those seeking information (like tourists). **Collection Tool:** PRAW (Python Reddit API Wrapper).

Initial Dataset: [500] posts, resulting in [10] rows of text data before cleaning.

Data Scope:

- Fetched the top [500] posts from the last year.

- Included post titles, self-text, and the top 5 comments for each post to capture a broader range of discussion.

Korean Food for everyone
Foods of the entire Korean diaspora to the Korean peninsula.

🗄 Created Aug 20, 2010
🌐 Public

1.2M
korean food eaters

8
🟢 food connoisseurs online

Top 1%
Rank by size ⬈

USER FLAIR

Double_Analyst_7896

Methodology - Preprocessing

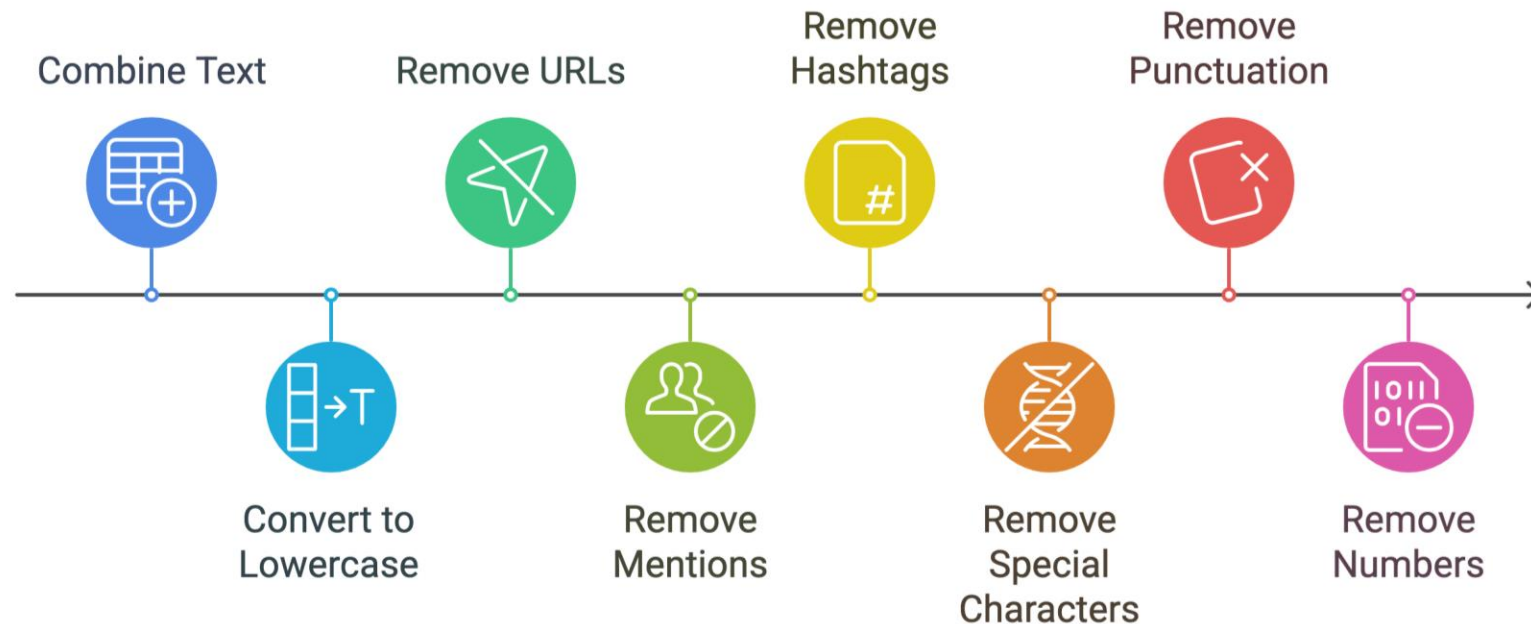- To prepare the text data for analysis, the following steps were taken (**using Python and NLTK):**

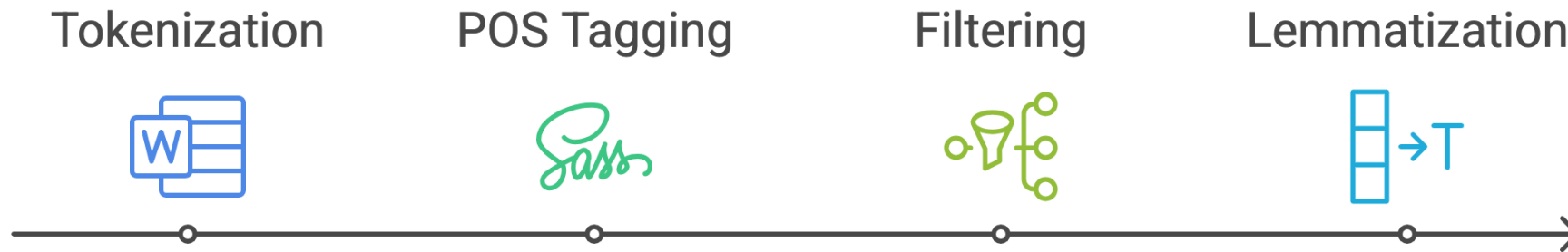1. **Combined Text:** Post titles, self-text, and comments were merged into a single 'full_text' field per post.

2. **Cleaning:**
   - Converted text to lowercase.
   - Removed URLs, @mentions, hashtags, and special characters (like | used as a comment separator).
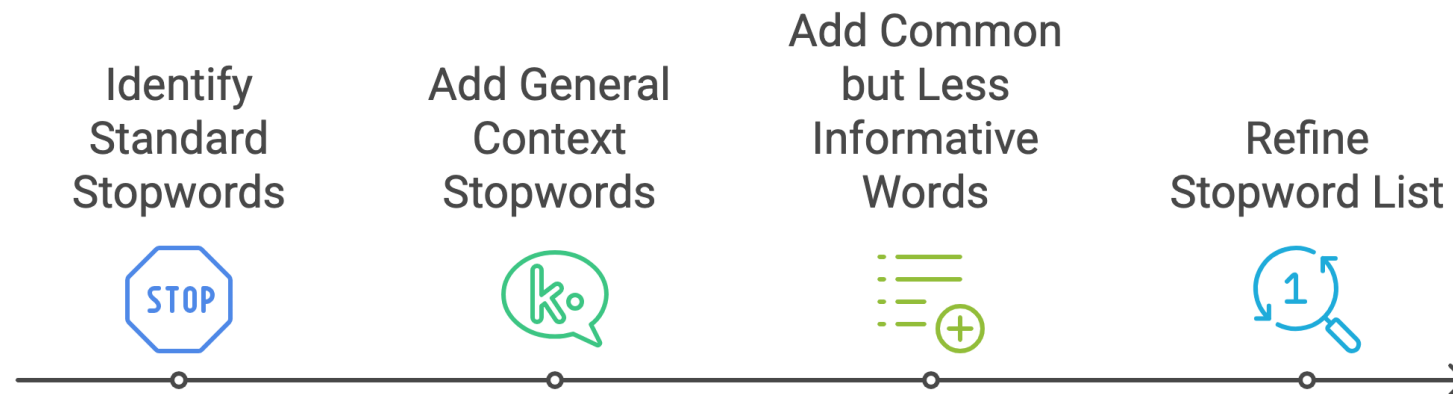   - Removed punctuation and numbers.

Methodology - Preprocessing

- **Tokenization:** Text was split into individual words (tokens).

- **Part-of-Speech (POS) Tagging & Filtering:** Words were tagged with their grammatical role. Only Nouns (NN) and Adjectives (JJ) were retained, as they often carry the most descriptive meaning.

- **Lemmatization:** Words were reduced to their base or dictionary form (e.g., "eating" -> "eat").

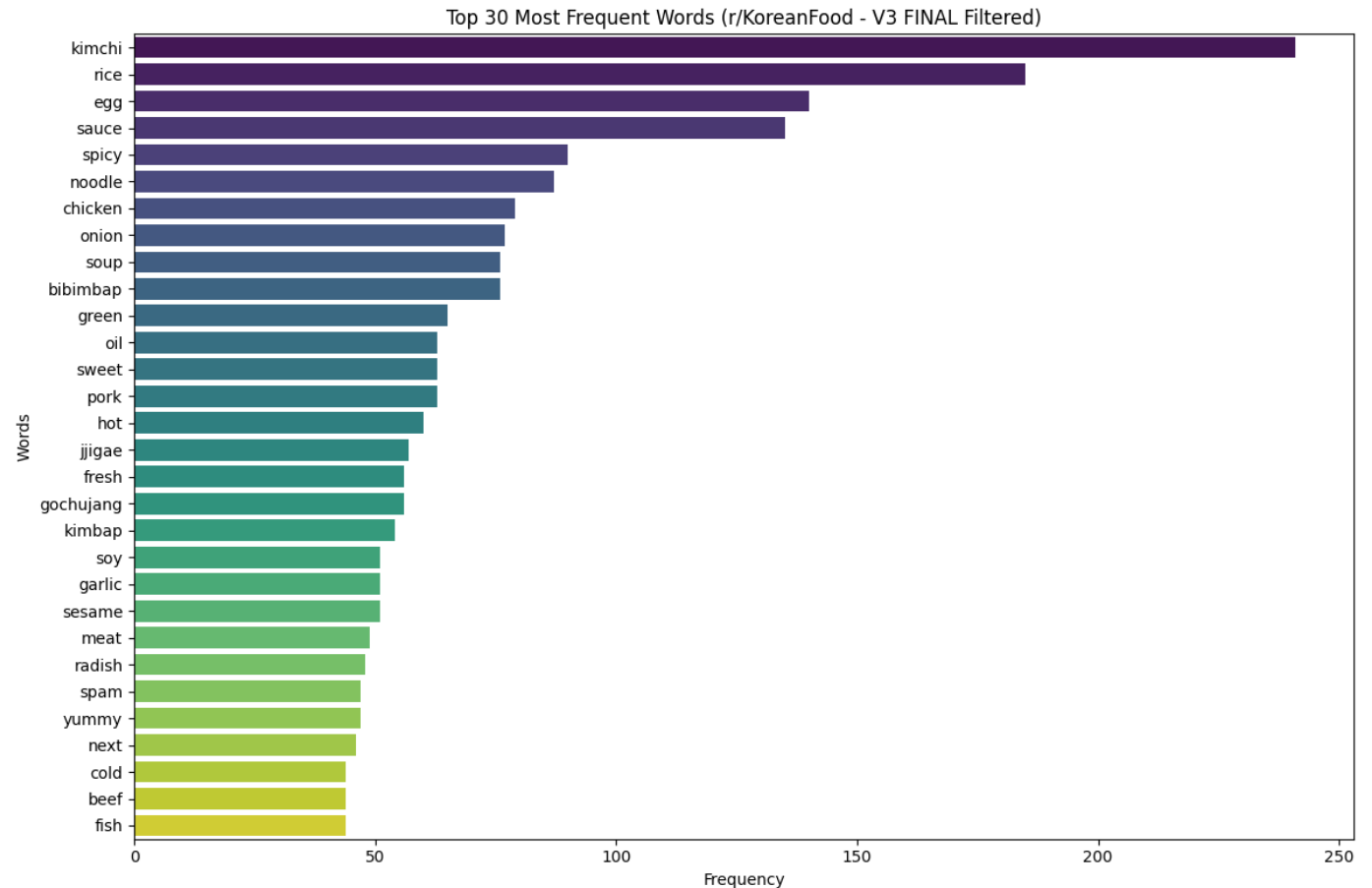| Tokenization | POS Tagging | Filtering | Lemmatization |

Methodology - Preprocessing

- **Stopword Removal:** Standard English stopwords were removed (e.g., "the", "is", "a").

- **Custom Stopwords (Crucial Step):** Several layers of custom stopwords were iteratively added to remove:

  - General Reddit/travel/Korea context words (e.g., "korea", "food", "restaurant", "trip", "question", "recommend").

  - Common but less informative words identified from initial frequency lists (e.g., "station", "hotel", "dish", "favorite", "lol").

- This iterative stopword refinement was key to focusing on specific food terms.
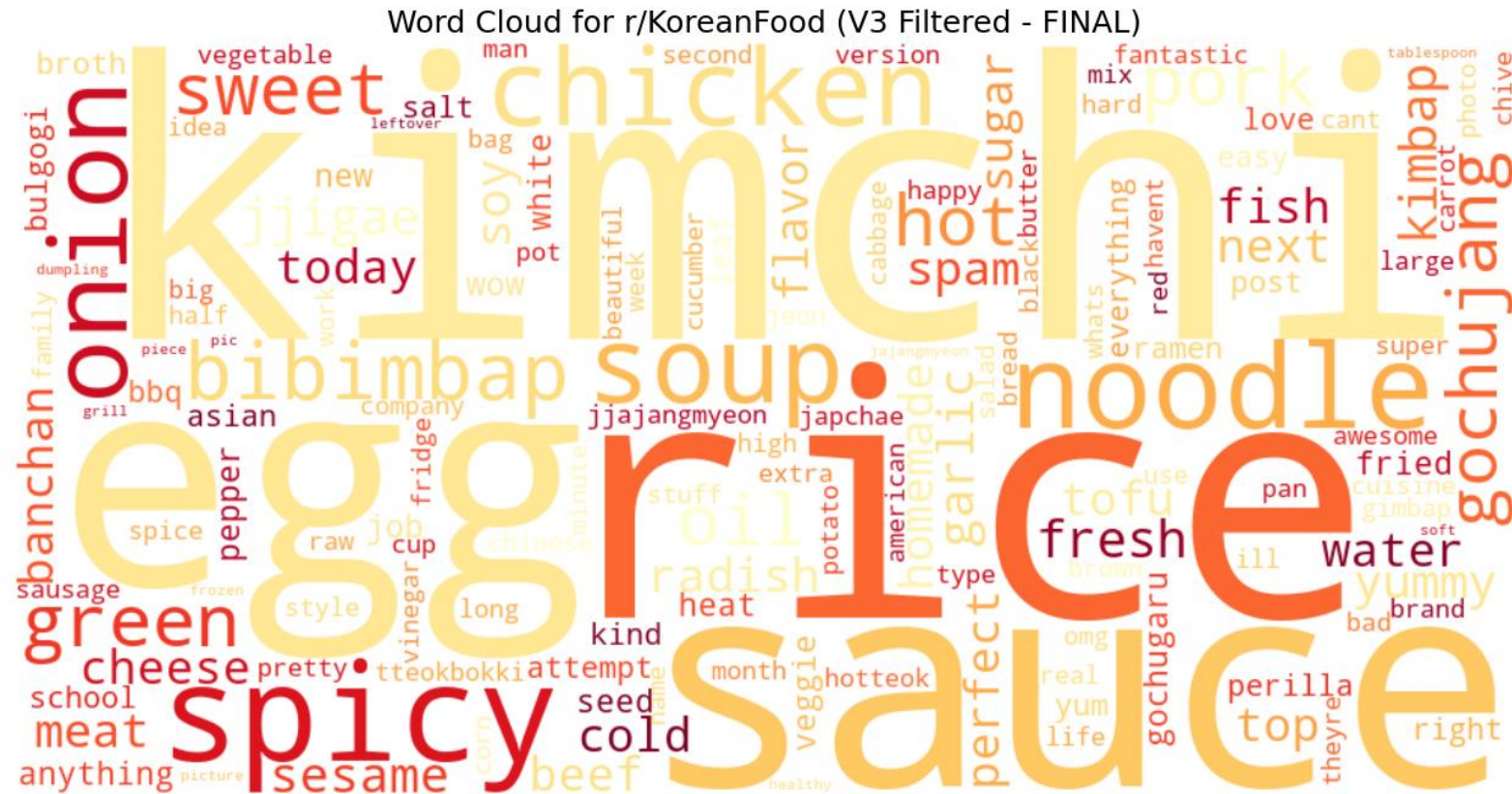
Results - Top Frequent Words

- After preprocessing and multiple layers of **stopword filtering**, the most frequent terms provided a clear focus on specific food items and ingredients.

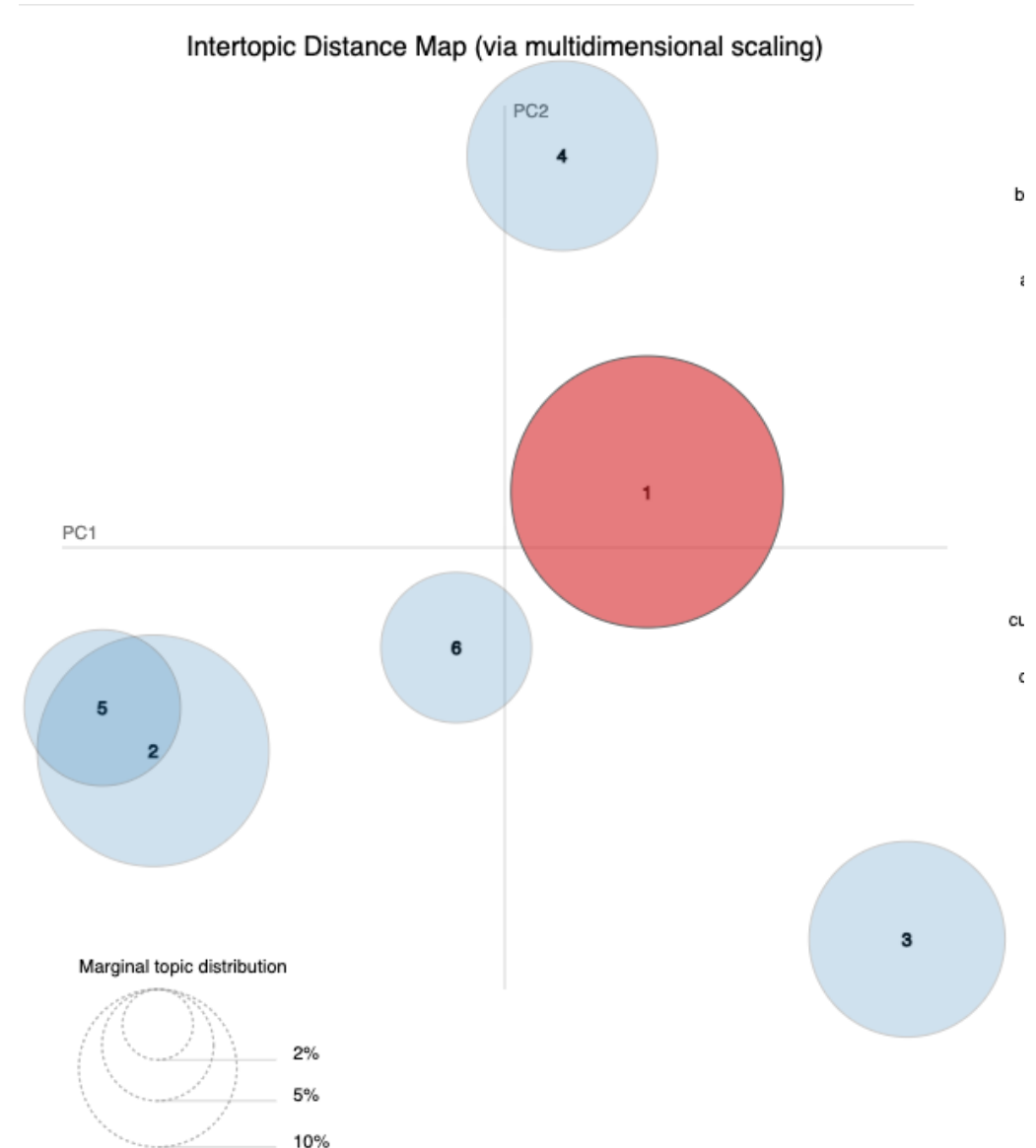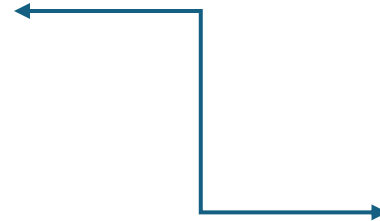Top 30 Most Frequent Words (r/KoreanFood - V3 FINAL Filtered)

Results - Word Cloud

- A word cloud was generated from the final filtered tokens to visually represent the most prominent terms.

Word Cloud for r/KoreanFood (V3 Filtered - FINAL)

LDA Topic Modeling

- Latent Dirichlet Allocation (LDA) was used to identify underlying themes within the r/KoreanFood discussions.

- **Number of Topics Chosen:** 6

Attached As HTML file

Intertopic Distance Map (via multidimensional scaling)

Discussion & Insights

- The analysis of **r/KoreanFood reveals** a vibrant community focused on specific dishes, ingredients, and cooking methods.

- Unlike broader travel forums, the discussion here is deeply centered on the food itself, with less emphasis on general travel logistics after targeted stopword removal.

- Key dishes like **Kimchi, Bibimbap, Korean BBQ (chicken, pork), Jjigae, and Kimbap** are frequently discussed.

- Essential ingredients like **gochujang, soy sauce, sesame oil, garlic, and onion** are central to many conversations.

- Descriptors like **"spicy," "sweet," and "hot"** are commonly used, highlighting taste profiles.

- The topics identified by LDA provide a structured view of the main conversational clusters, ranging from specific dish types to core ingredients and cooking discussions.

Challenges & Limitations

- **Initial Data Focus:** Early attempts using broader subreddits **(r/koreatravel)** or general news sources yielded less specific food-related terms, requiring a pivot to **r/KoreanFood.**

- **Iterative Stopword Refinement:** Achieving a clear focus on specific food items required multiple iterations of adding custom stopwords based on observed frequency lists.

- **Reddit Slang/Informality:** The informal nature of Reddit text (e.g., **"lol", "youre"**) can sometimes appear in initial analyses if not aggressively filtered.

- **Sentiment Analysis (Not Performed in Detail):** While basic sentiment could be inferred, a dedicated sentiment analysis was not the primary focus of this iteration but could be future work.

- **Limited Scope:** Analysis is based on a snapshot of **'top' posts** from one year and may not capture all nuances or long-term trends.

Conclusion

- This project successfully identified the key food items, ingredients, and discussion themes within the **r/KoreanFood** subreddit.

- By focusing the data source and iteratively refining preprocessing (especially stopwords), it was possible to extract clear and specific insights directly related to **Korean cuisine.**

- The results demonstrate that **r/KoreanFood** is a rich resource for understanding popular Korean dishes and the community's engagement with them.

- The dominant terms and topics revolve around iconic dishes like **Kimchi, BBQ, and Bibimbap**, essential ingredients **like gochujang, and the experience of preparing and enjoying these foods.**