Yelp Recommendation

Ayush Agarwal, Kevin Li, Grace Xie, Yumin Zhang

Data Source

The Yelp dataset is a rich collection of data obtained from Yelp, a platform that connects people with local businesses, including restaurants.

- Reviews: User-generated reviews and ratings for various businesses.
- Business Data: Detailed records of local businesses listed on Yelp, covering aspects such as location, categories (e.g., type of cuisine for restaurants), and attributes.
- Sampling Method: Philadelphia, Restaurants currently open, User Reviews more than 20
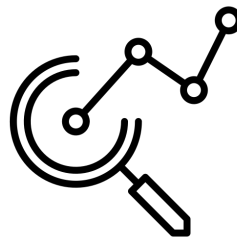
Summarize Details of Dataset
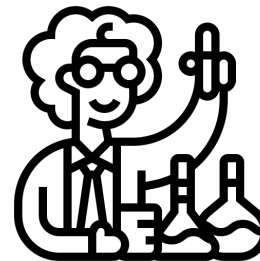
Number of samples in train and test data
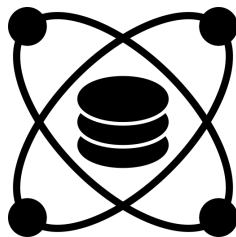
Number of features

22 features, including review_id, user_id, business_id, review_star etc.

Recommendation target
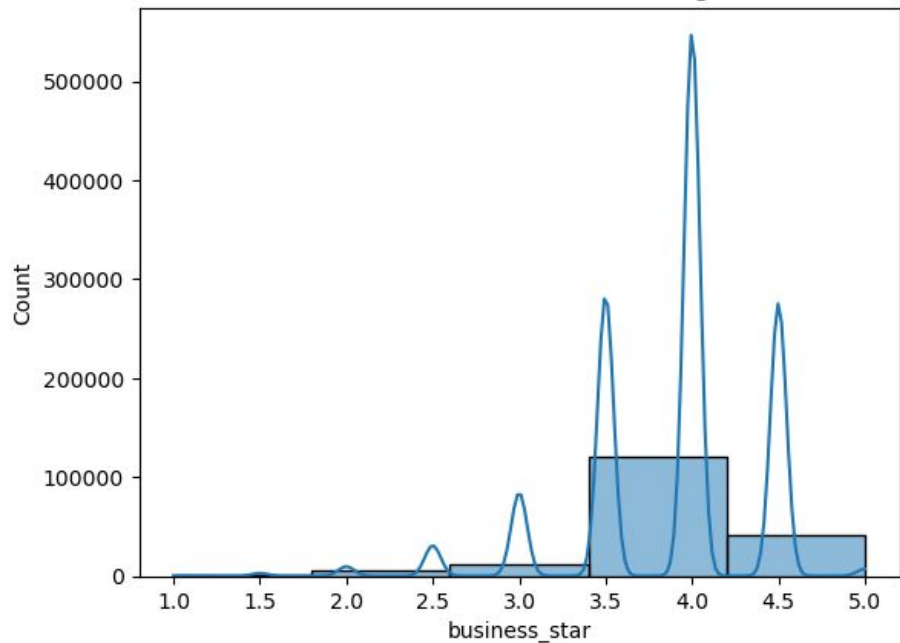
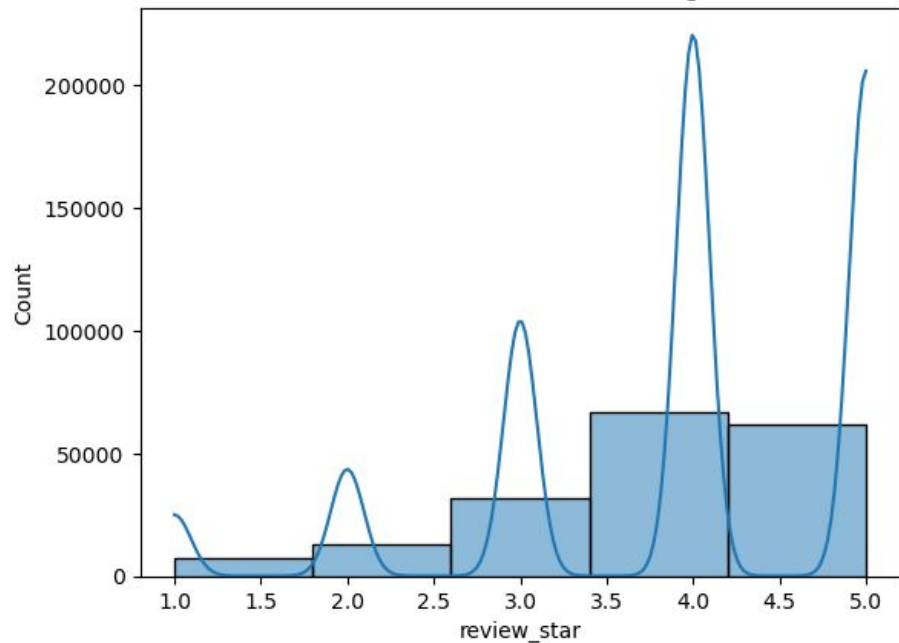Customized Restaurant Recommendations

Model candidates

EDA

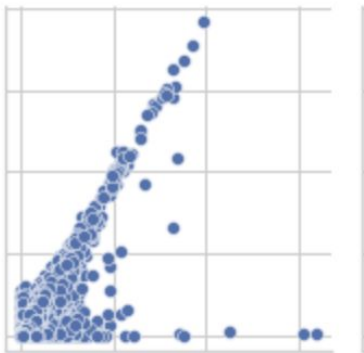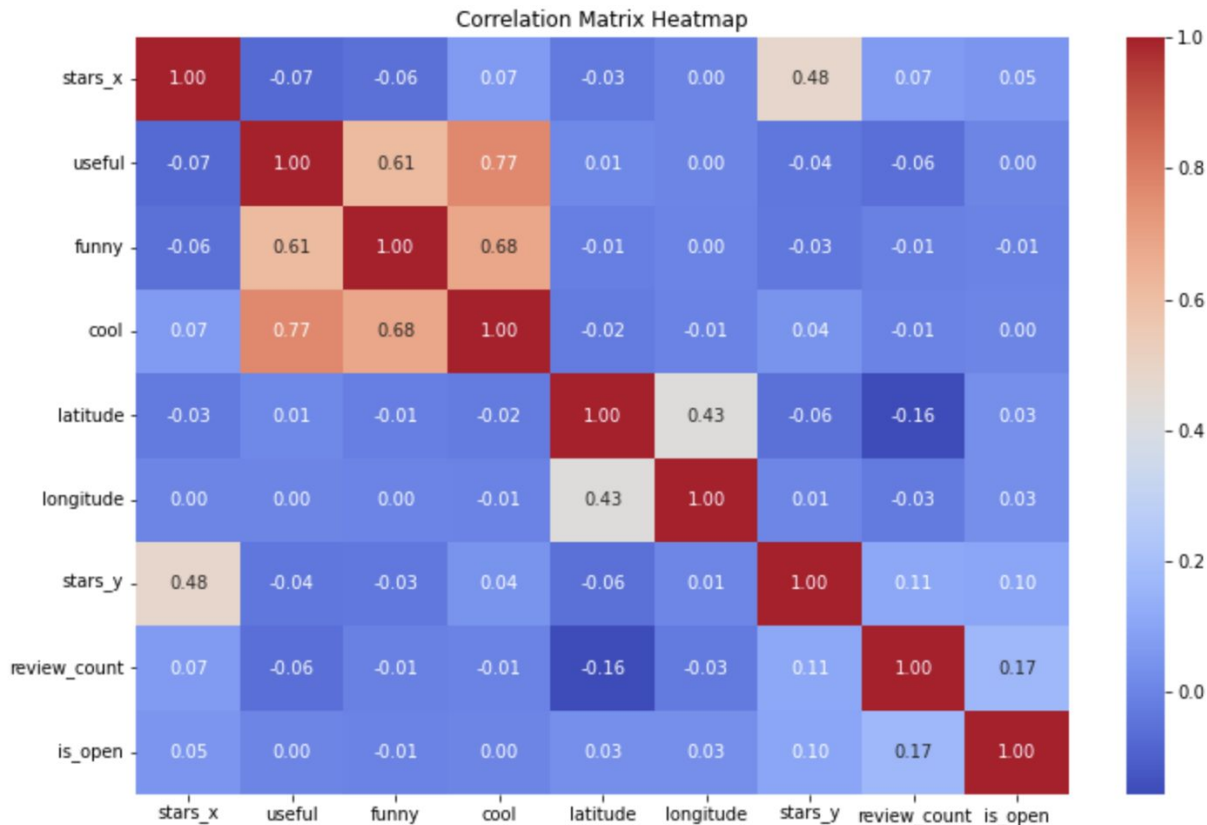- Distribution of ratings

Correlation Matrix Heatmap

Useful v.s. Cool

Funny v.s. Cool
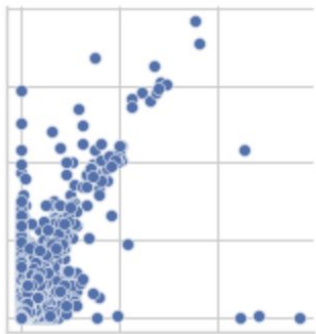
EDA cont.

- Distribution of review counts

Distribution of Review Counts
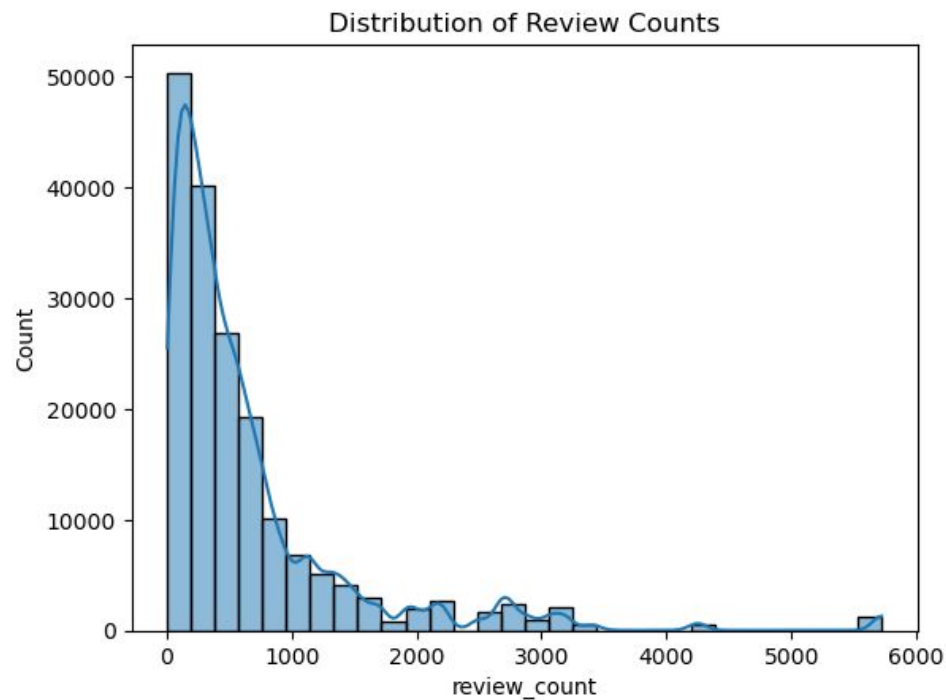
EDA cont.

- Plot number of reviews over time

Number of Reviews Over Time
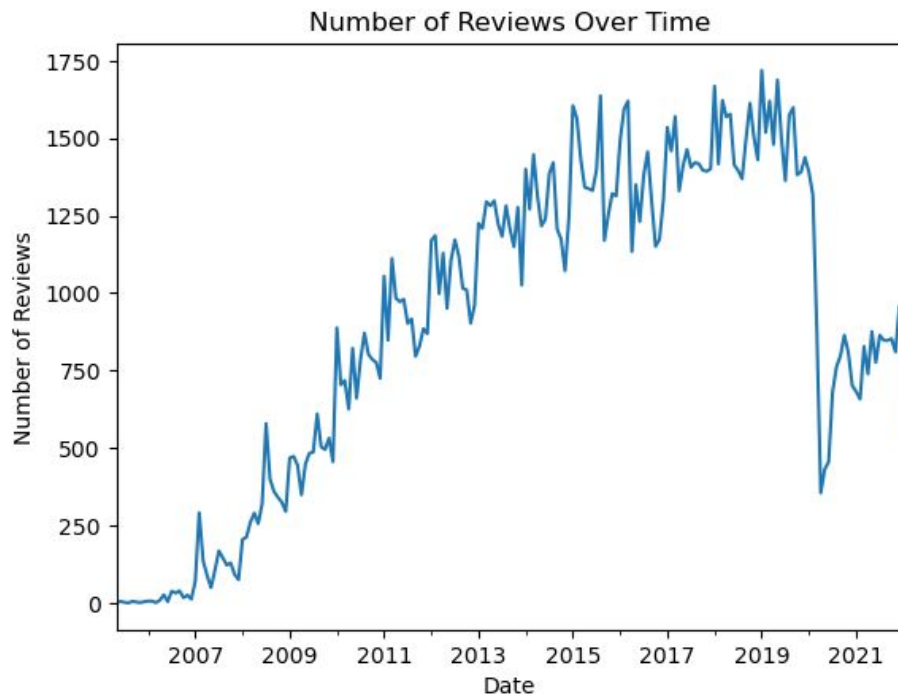
EDA cont.

- Business with most

  five star review stars

Top Businesses by Count of Five-Star Reviews

Count of Five-Star Reviews (y-axis)

Business Name (x-axis): Reading Terminal Market, Zahav, Barbuzzo, Han Dynasty, Parc, Talula's Garden, Amada, Cafe La Maude, Sampan, Beiler's Bakery
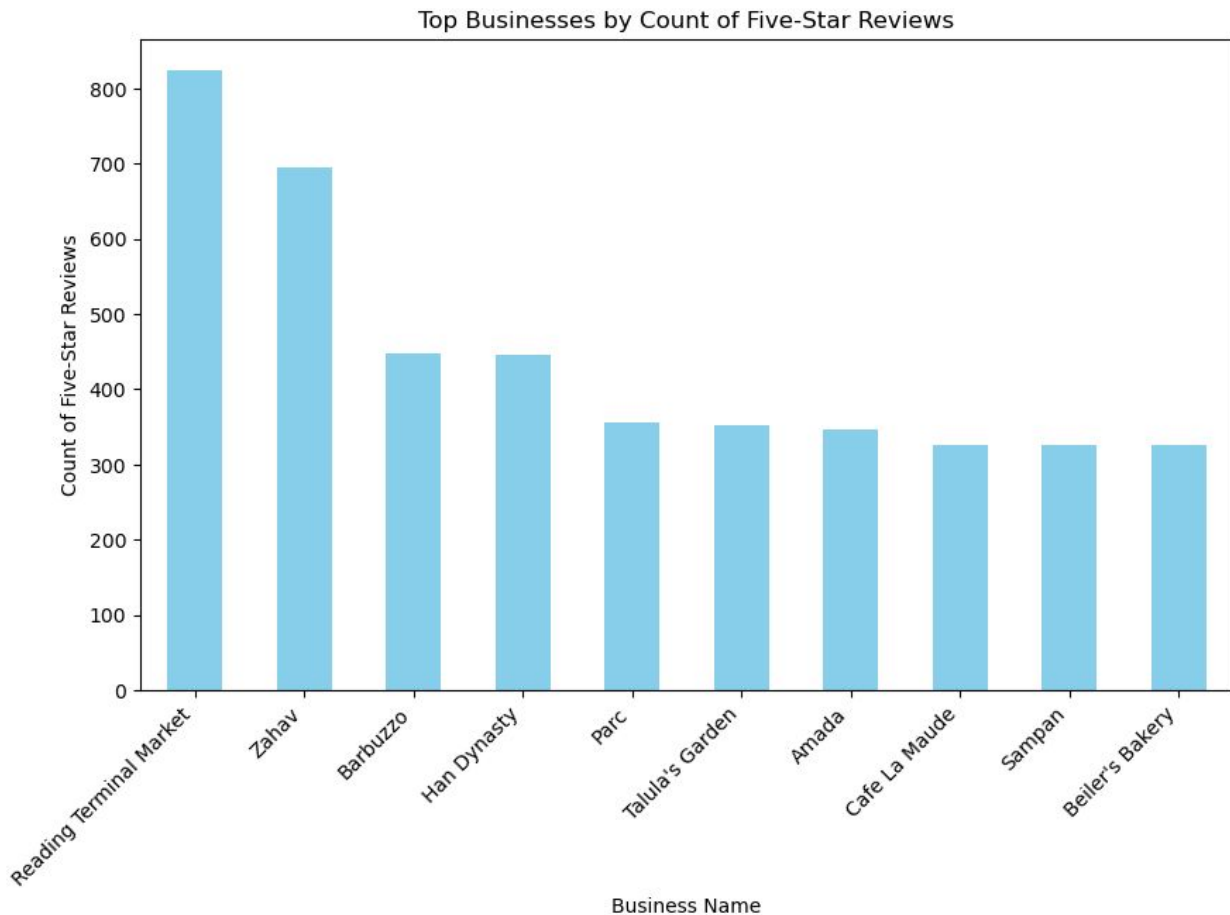
EDA cont.

- Top 10 most popular categories

Top 10 Most Popular Categories
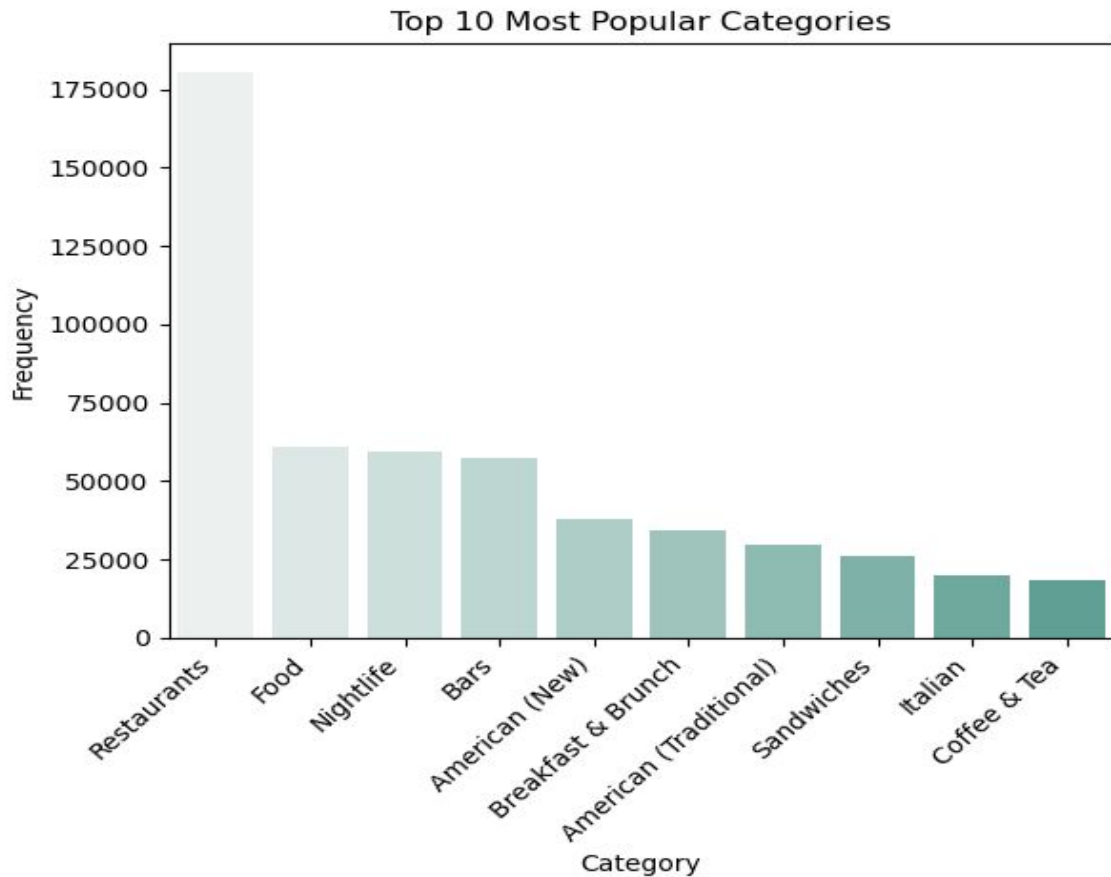
EDA cont.

- Key Terms in Customer Reviews of Dining Experiences

Problem Definition & Importance

- Personalized restaurant recommendations in Philadelphia from Yelp review (Hybrid dual-cascade Model)

- A

We aim to revolutionize the way people discover and interact with local dining options, and how restaurants reach out to their potential customers.

Business Problem 1

Model Architecture
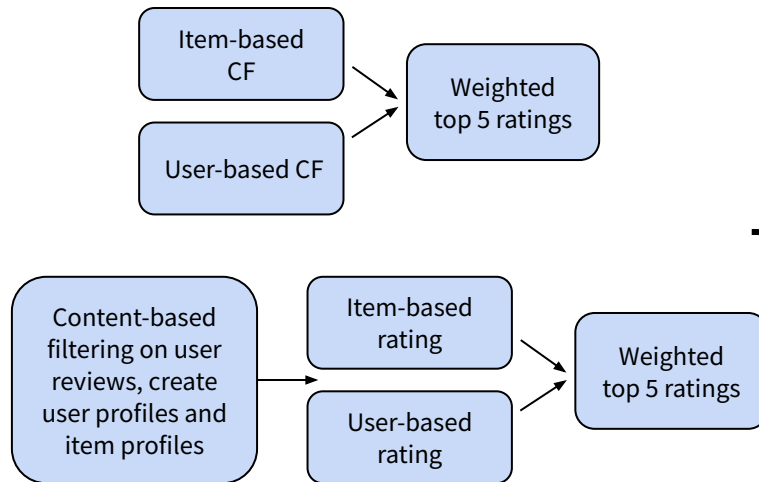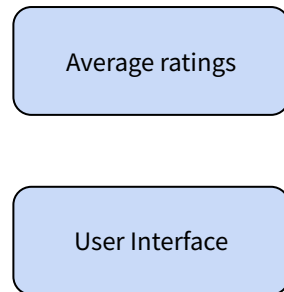
Collaborative Filtering

1. **User-Based vs. Item-Based Collaborative Filtering**
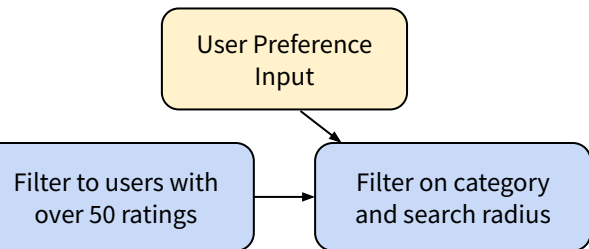   a. User-based: Recommends items by finding similar users based on their item interactions.
   b. Item-based: Suggests items similar to those a user has previously liked, focusing on item similarities.

2. **Hyperparameter Tuning in Collaborative Filtering**
   a. KNNBasic() Method: A fundamental algorithm for both user-based and item-based approaches.
   b. Key Hyperparameters:
   - k: The number of nearest neighbors to consider.
   - min_k: The minimum number of neighbors to take into account for prediction.

Content Based Recommendation
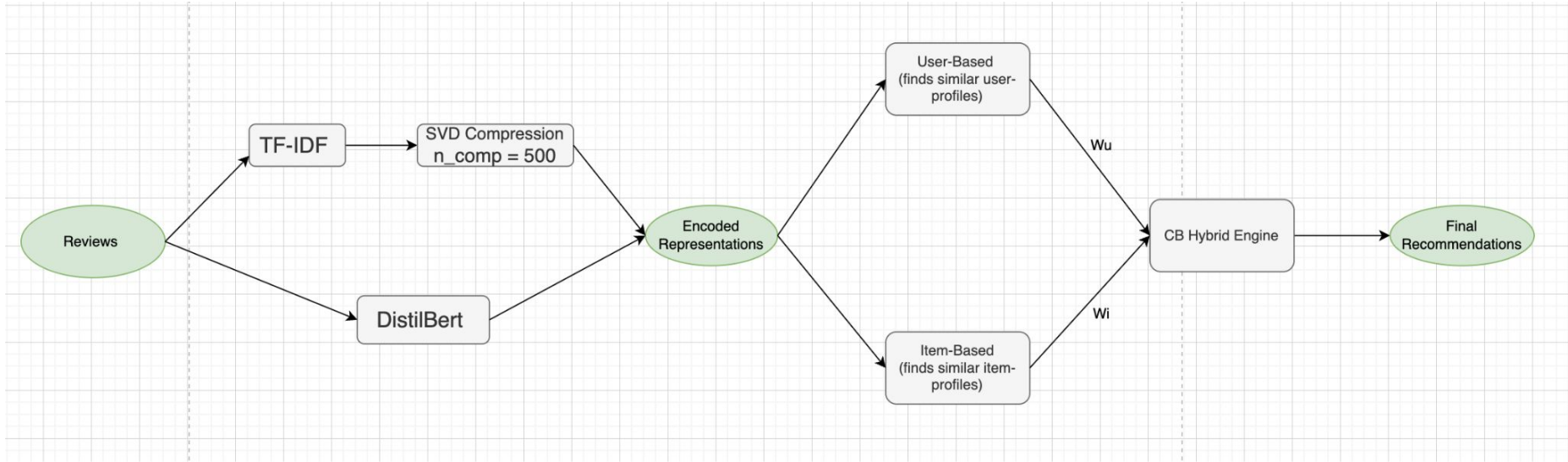
MSE of Different Recommendation Approaches
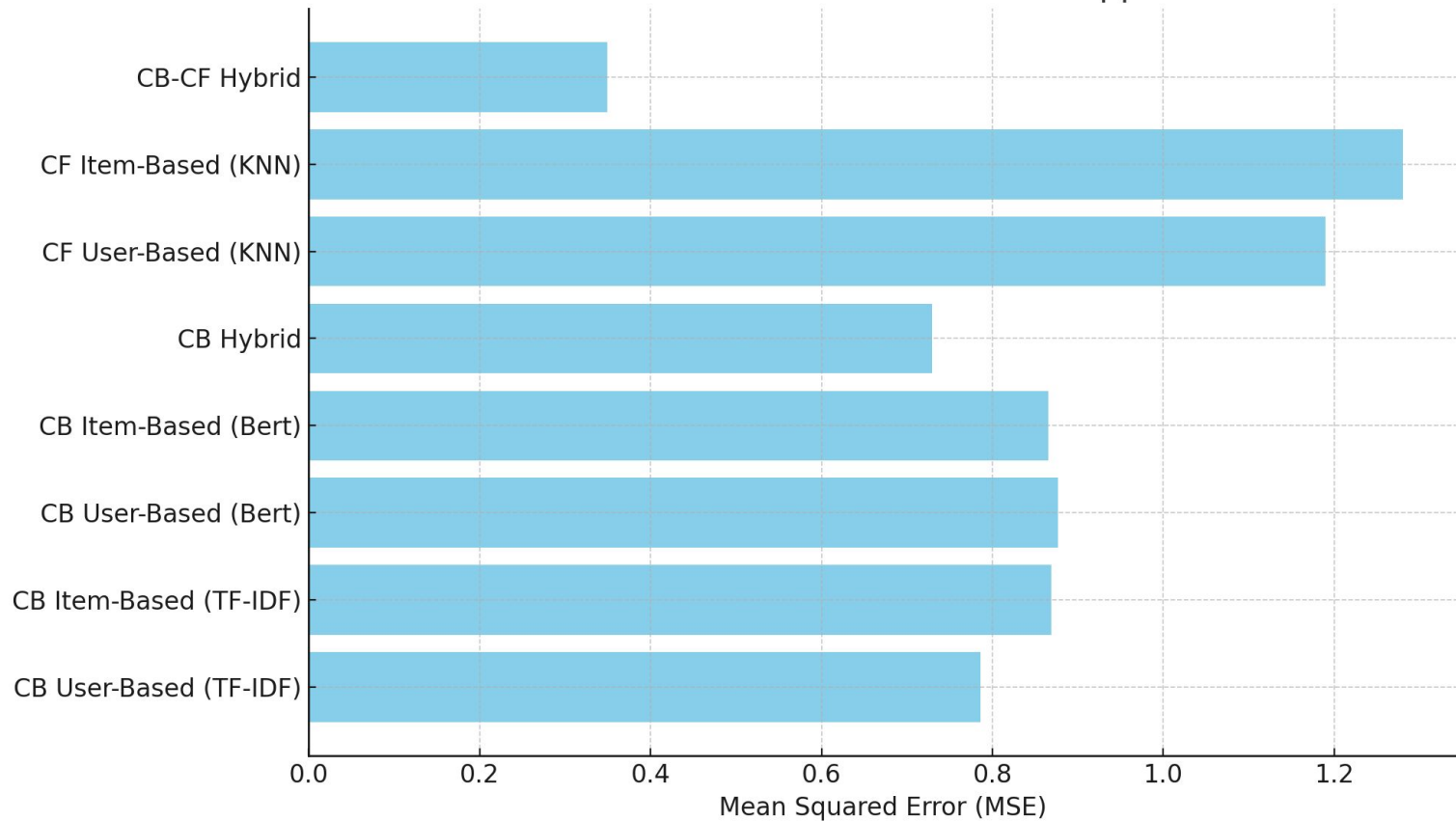
Performance and Comparison

Ratings from 1-5

Approach	MSE	MAE
CB User-Based (TF-IDF)	0.7864	0.6884
CB Item-Based (TF-IDF)	0.869	0.730
CB User-Based (Bert)	0.8764	0.6498
CB Item-Based (Bert)	0.8651	0.7281
CB Hybrid	0.7294	N/A
CF User-Based (KNN)	1.1902	0.8316
CF Item-Based (KNN)	1.2801	0.8620
CB-CF Hybrid	**0.3491**	**0.4312**

Business Problem 2

Prompt Engineering Advertising System for Restaurants Based on Existing User Reviews and Yelp Attributions

Model Architecture

Apply to restaurants with sufficient user reviews (>100) → Apply text preprocessing and construct TF-IDF matrix → Use Association Rules Mining to find frequent keywords → Combine with restaurant attributes, hours, and location → Feed into prompt engineering system to obtain final advertisement
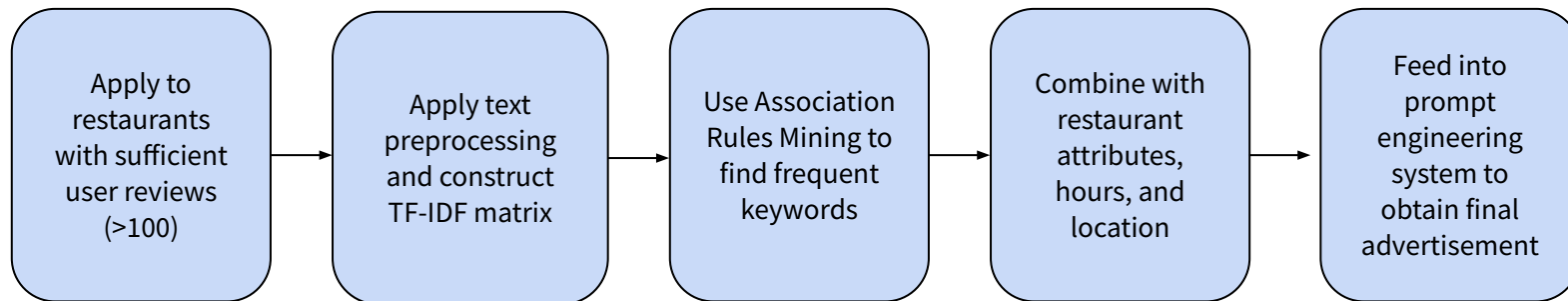
Association Rule Mining

1. Frequent Itemset Mining with Apriori
   a. Applies the Apriori algorithm to a binary matrix derived from a TF-IDF matrix.
   b. Identifies sets of terms that frequently occur together in at least 10% of documents, using a minimum support threshold of 0.1.

2. Association Rules Generation
   a. Generates association rules from the frequent itemsets to discover relationships between terms.
   b. Uses "lift" as the metric to measure the strength of these associations, focusing on rules with a lift value indicating associations stronger than random chance (min_threshold = 1).
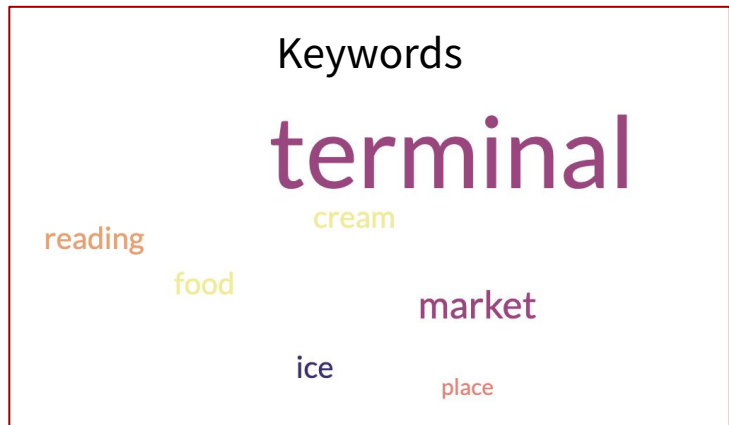
3. Identifying Top Keywords
   a. Extracts top keywords from the rules with the highest confidence.
   b. Aims to highlight the most predictive and significant terms, providing insights into key themes or patterns in the data.
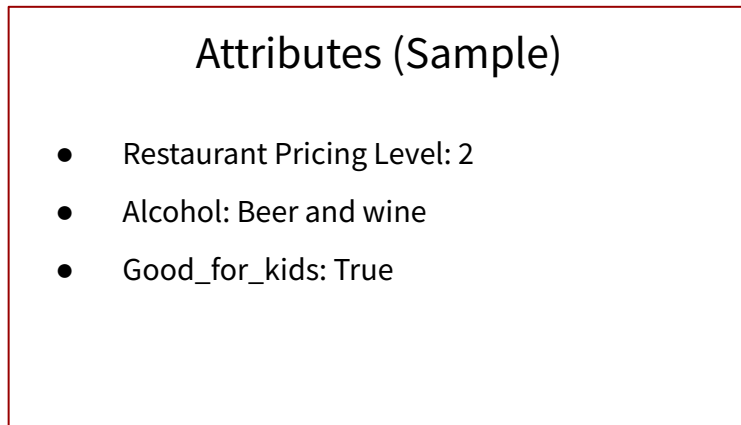
Use Case Demonstration

Reading Terminal Market in Philadelphia, with approx. 750 user reviews

Use Case Demonstration

Keywords

terminal

cream

reading

food

market

ice

place

+

Attributes (Sample)

- Restaurant Pricing Level: 2
- Alcohol: Beer and wine
- Good_for_kids: True

"Come dine at our restaurant, where we serve up delicious cream and ice treats in a casual and lively atmosphere! Our menu is filled with fresh and love-filled dishes that will satisfy your cravings. With a market-style setup, you can choose from a variety of options at our terminal. Whether you're looking for a good place to enjoy a meal with friends or family, or just want to grab some food to go, our terminal is the perfect spot. Plus, we offer beer and wine options to complement your meal. Visit us today for a taste of Philly in Reading!"

Thank You