BrandMatch1A Cohere Commerce

AI Studio Final Presentation

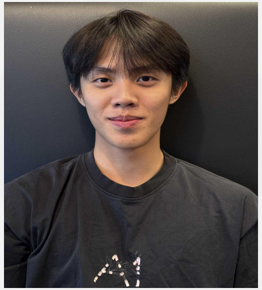
Break Through Tech @ UCLA
December 4, 2024

cohere commerce

Introductions

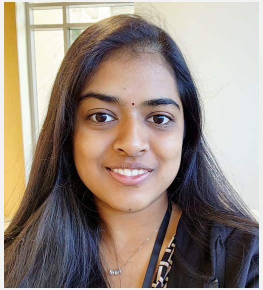
Meet Our Team!

Our Team



Nhan Tri Danh (Jack)
Pasadena City College

Carina Curiel Alaniz
USC



Abhi Sathvika Goriparthy
UCSD

Rhea Jethvani
UCI

Ashita Singh
UCLA

Our Supporters

Yu-Hsin Weng
AI Studio TA

Mike Chen
Challenge Advisor

Sonrisa Chen
Challenge Advisor

Agenda

1. Overview
2. Business Impact
3. Solution Overview
4. Data Understanding
5. Modeling & Evaluation
6. Insights & Key Findings
7. Final Thoughts & Next Steps

01

Overview

Project Overview

Bailly
Walmart
DEVIANT ART
Velasca MILANO
asos
Etsy
wish

DisplayRabbit
meow meow tweet
AliExpress
NATURISIMO Home of ethical beauty
iHerb
UPPERCASE
KOTN

amazon.com
ebay
Pink Lily
IKEA
Then I Met You
BEST BUY
POSHMARK

SUGAR
BLK & BOLD SPECIALTY BEVERAGES
Suta
ARTISAIRE
Target
Ketnipz
tentree

"This is OVERWHELMING!"

3,202,077

Retail Businesses in US

https://www.ibisworld.com/industry-statistics/number-of-businesses/retail-trade-united-states/

71% of Consumers

expect tailored experiences, but retailers lack tools to meet these expectations!

The Evolution Of Retail In 2024: A Glimpse Into The Future

24% Less Likely

to purchase online from a business that doesn't also have a storefront

The Future Of Retail: What The Stats Say About Retailers In 2023

Build a recommendation engine which recommends new brands to stores by analyzing retailer reviews of products.

Our Solution

Simplify Decision-Making

Leverages sentiment analysis, ratings, and reviews to identify the best brands for each retailer.

Drive Personalization

Matches brands with retailers based on customer interests and market fit.

Empower Retailers

Provides actionable insights that improve buyer confidence, streamline vendor selection, and boost sales

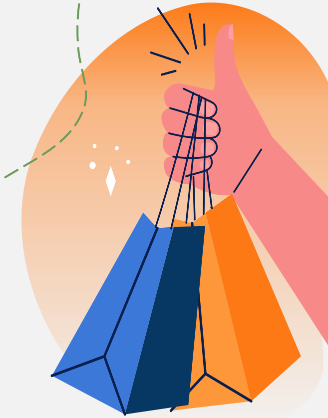
Our Goals

Recommendation System

Develop a robust recommendation model to match brands with retailers based on customer preferences.

Service Integration

Build an API to connect the ML Pipeline of Recommendation models to Cohere Commerce Website, allowing for brand discovery



02

Impact

Business Impact

Business Impact

Convenient Search and Find

Simplifies finding suitable brands based on user preferences, saving time and effort.

Improved User Search Experience

Personalized brand recommendations lead to higher customer satisfaction and engagement.

Increased Business Growth

Cohere Commerce offers a dynamic platform for businesses because now retailers and brands can easily expand their reach by connecting with the right products

03

Solution Overview

Summary and High-Level Approach

Approach

Pre-Processing

Handling Missing Values, Outliers.
Tokenizing Reviews Data.

Recommendation Pipeline

Combining all the components into a single pipeline that passes through cleaning, connects models, to a final MCDM model.

A → B → C → D

Building Models

Leveraging BERT model for Sentiment Analysis. Utilizing Collaborative Filtering and Matrix Factorization to find similar brands based on reviews.

API Development

Developing an API that can be accessed through Cohere Commerce Website to access the AI Functionality

Resources We Leveraged

Tools & Libraries	Languages & Frameworks
• Google Colab	• Python
• Sci-kit Learn	• FastAPI
• Google BERT	• JavaScript
• GitHub	• HTML

Summary of Insights & Key Findings

Key Findings:

- Sentiment analysis from review data helps refine brand preferences.
- Collaborative filtering significantly improves brand suggestions by identifying similar preferences.
- SVD matrix factorization uncovers hidden patterns in user-brand interactions.

Results:

>An API Endpoint that connects to our ML Pipeline

04

Data Understanding

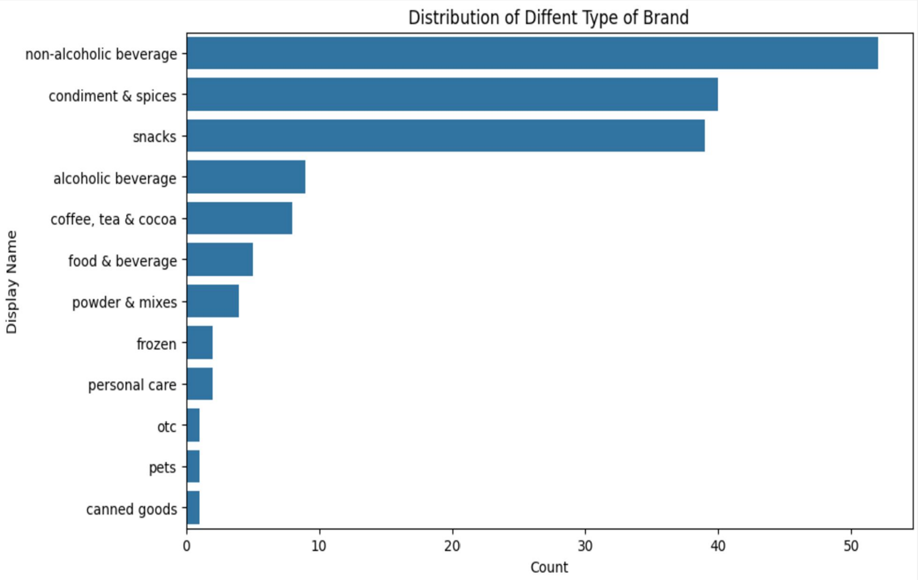
Pre-Processing and Data Cleaning

Data Understanding

164 Unique Brands

12 Product Types

6 Rating Types



display_name	count
non-alcoholic beverage	52
condiment & spices	40
snacks	39
alcoholic beverage	9
coffee, tea & cocoa	8
food & beverage	5
powder & mixes	4
frozen	2
personal care	2
otc	1
pets	1
canned goods	1


```
['rating_shipping', 'rating_service', 'rating_price', 'rating_quality', 'rating_packaging', 'rating_taste']
```

Data Preprocessing

Cleaning

Removed stopwords, special characters, and performed unified lowercasing.

Lemmatization/ Stemming

Reduced words to their base form to minimize vocabulary size.

Vectorization

Converted text into numerical features using TF-IDF, Word Embeddings (BERT, Word2Vec)

After Cleaning

brand_id	review_id	avg_rating	is_product_recomme	review_content	is_anonymous	display_name	rating_packaging	rating_price	rating_quality	rating_service	rating_shipping	rating_taste	sentiment_score
00411460f7c92d212	2598	5	TRUE	basically stick peanu	FALSE	Snacks	5	5	5	5	5	5	1
00411460f7c92d212	3718	5	TRUE	carry currently stock	TRUE	Snacks	5	5	5	5	5	5	5
00411460f7c92d212	3719	4	TRUE	carry whim chocolat	TRUE	Snacks	4	4	4	4	4	4	3
00411460f7c92d212	2004	5	TRUE	good flavor selling fa	FALSE	Snacks	5	5	5	5	5	5	4
00411460f7c92d212	3717	4	TRUE	still carry whim sellin	TRUE	Snacks	4	4	4	4	4	4	5
00411460f7c92d212	3716	4	TRUE	peanut butter cup market continue really		Snacks	4	4	4	4	4	4	5
00411460f7c92d212	3131	5	TRUE	love sharing whim cu	FALSE	Snacks	5	5	5	5	5	5	5
00411460f7c92d212	2604	5	TRUE	little island middle t	FALSE	Snacks	5	5	5	5	5	5	3
00411460f7c92d212	2603	5	TRUE	well selling well ever	FALSE	Snacks	5	5	5	5	5	5	4
00411460f7c92d212	2601	5	TRUE	well little one really	FALSE	Snacks	5	5	5	5	5	5	4
00411460f7c92d212	2600	5	TRUE	look like still pretty g	FALSE	Snacks	5	5	5	5	5	5	4
00411460f7c92d212	2599	5	TRUE	sell pretty well would	FALSE	Snacks	5	5	5	5	5	5	3
00ec53c4682d36f5c	2901	4	TRUE	okay kind slow right l	FALSE	Condiment & Spices	4	4	4	4	4	4	3
00ec53c4682d36f5c	2969	4.75	TRUE	sauz sold really well	FALSE	Condiment & Spices	4.75	4	5	5	5	4.75	4
00ec53c4682d36f5c	2904	4	TRUE	okay started end disp	FALSE	Condiment & Spices	4	4	4	4	4	4	3
00ec53c4682d36f5c	2903	3	TRUE	half gone selling hav	TRUE	Condiment & Spices	3	3	3	3	3	3	1
00ec53c4682d36f5c	2902	3	TRUE	ive shelf week ive sol	TRUE	Condiment & Spices	3	3	3	3	3	3	3
00ec53c4682d36f5c	2900	4	TRUE	almost sold shelf one	FALSE	Condiment & Spices	4	4	4	4	4	4	3
00ec53c4682d36f5c	2899	4	TRUE	okay doesnt run shelf	FALSE	Condiment & Spices	4	4	4	4	4	4	3
00ec53c4682d36f5c	2898	4	TRUE	okay started end disp	FALSE	Condiment & Spices	4	4	4	4	4	4	3

Data Summary

Dataset Overview

2032 reviews, 164 brands, 12 product categories with significant imbalance (e.g., 52 beverage brands vs. 1 pet product brand)

Data Preprocessing

Cleaned and aggregated review content, applied tokenization, stopword removal, and lemmatization.
Extracted features such as aggregated review content, average ratings, and multi-criteria ratings (e.g., packaging, price).

Sentiment Analysis

Used BERT to read the review content and grade each review on scale 1-5

Next: Recommendation Models

05

Modeling

Building ML Model

Modeling

Sentence Transformer

Converted reviews data into word embeddings (vectors)

Weighted Features

Rating types assigned different weights based on user preferences

Trained SVD Model

Performs Matrix Factorization on review embeddings to reveal underlying key patterns and latent factors (essential for recommendation systems)

Matching Similar Brands

Using SVD model's top predicted output and associated rating weights, top brands are selected by score

Recommendation

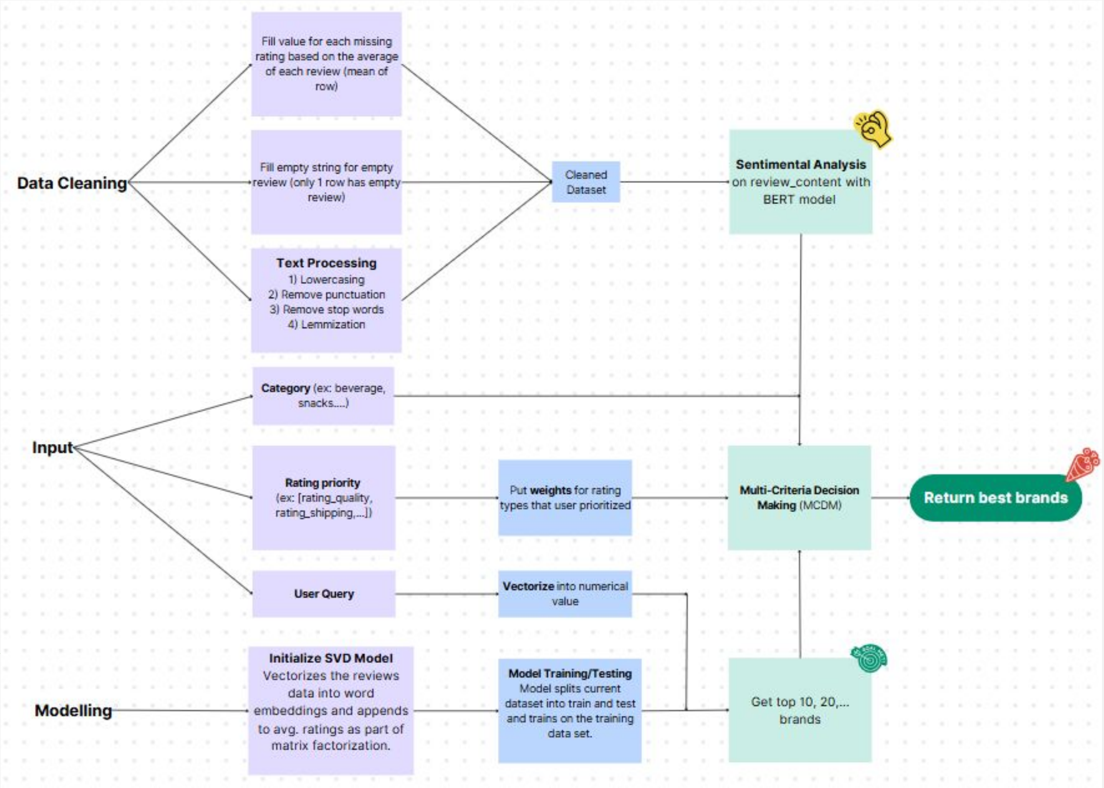
The Top 5 brands according to user preferences are shown!

06

Key Findings

Insights and Key Results

Model Diagram



Demo

Sample Demonstration of how our the model is returning specific brand ids to match with user preferences through an API call from our sample WebPage.

Insights & Key Findings

Key Findings:

- Data
  - Preprocessing: Cleaned and aggregated review content, applied tokenization, stopword removal, and lemmatization
  - Extracted features such as aggregated review content, average ratings, and multi-criteria ratings (e.g., packaging, price).
- Analysis
  - Used BERT to classify reviews into 5 sentiment categories
- Recommendation Models: SVD, KNN, Matrix Factorization

Summary: Identified top brands in categories (e.g., snacks, beverages) based on multi-criteria scores and sentiment, aligning with user-defined priorities.

We Found...

1. Comprehensive recommendations balancing qualitative (sentiment) and quantitative (ratings) data.
2. Actionable insights for retailers to optimize brand selection and improve customer satisfaction

07

Final Thoughts

Next Steps

What We Learned

Gained Proficiency

Learning New Tech

Challenges

Learned about new ML models (SVD, MCDM), building an API to access model, creating UI/Frontend

Managing Data Sparsity while training models, and balancing model performance with computational costs

Gained expertise in integrating collaborative filtering, content-based analysis, and decision-making frameworks.

Next Steps

Real-Time Adaptation

Enable the system to adapt to new reviews and user feedback dynamically with continuous feedback from users in real time. Adapt to new rating types and preferences in the future.

Strengthen Data Sources

Incorporate more data (e.g., sales performance, seasonal trends) for more contextual recommendations for users to make better decisions

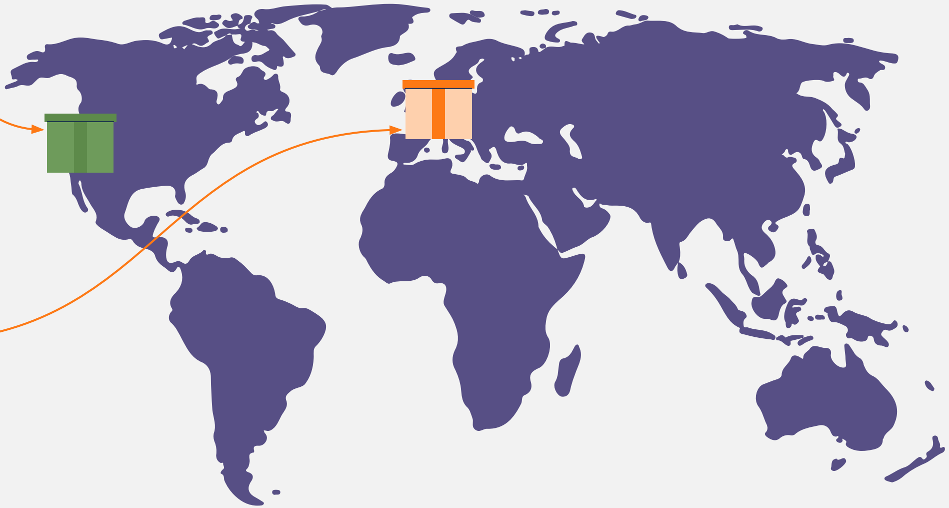
Next Jumps

U.S.

Enable the system to be widely used by ~3 million retail businesses

Europe

Able to adapt recommendation system to other markets.



Thank You!

Any Questions?

BREAK THROUGH TECH

cohere