DATA 5901: Capstone 1
Final Report

MSDS 25.4 “Voice of Market (VoM)”

Faculty Advisor: Dr. Elio Zhang
Sponsor Organization: Costco Wholesale

Team Members

Mustafa Shabbir Bhavanagarwala
Hrishikesh Deepak Dhole
Lakshit Gupta
Priyal Sunil Joshi
Tejaswi Neelapu

1. Executive Summary

In today's digital landscape, customer opinions shared on social media platforms like Reddit, Instagram, and Twitter (X) significantly impact brand perception and business strategy. Costco Wholesale, one of the world's largest retailers, is committed to enhancing customer satisfaction and improving its product and service offerings. However, extracting meaningful insights from the vast volume of unstructured social media data is a challenge. Traditional methods such as surveys and focus groups are often biased, slow, and lack real-time responsiveness, making it difficult to adapt to evolving consumer needs.

This project leverages Natural Language Processing (NLP) and AI-driven sentiment analysis to transform raw customer feedback into structured insights. An automated pipeline is developed to collect, preprocess, and analyze data using advanced techniques such as VADER for short-form sentiment analysis, BERT for contextual understanding, and AI-powered summarization.[1][2][7]

Key deliverables include a sentiment analysis report, an AI-powered chatbot, and a web-based visualization tool. These solutions will provide Costco with data-driven insights to improve decision-making, anticipate customer expectations, optimize its product line, and maintain a competitive edge in the evolving retail industry.

Table of contents

2. Introduction

Costco Wholesale is the third-largest retailer in the world known for providing valuable memberships to its customers. As of 2024, it serves approximately 137 million customers worldwide.[4] High-quality and in-demand products are provided by Costco at very reasonable prices.

Traditionally, companies rely on surveys, reviews from their own websites, and focus groups to understand customer needs. While these strategies do provide some information, they are time-consuming, often biased, and not real-time responsive. Also, Costco continuously looks for innovative ideas that can improve customer satisfaction and happiness. This can be done by anticipating customer needs through their opinions.

In today's fast-evolving digital world, social media platforms and direct company websites have become key channels for customer feedback. Customers express their thoughts on Costco's products and services through reviews, product ratings, survey responses, forum discussions, and social media posts. They share their experiences whether it's praising a new product, requesting improvements, or expressing dissatisfaction in the form of comments, tweets and posts. These insights come from various sources, including Costco's website, Instagram, Reddit discussion threads, and Twitter from customers worldwide. However, extracting meaningful insights from this vast amount of unstructured text data presents a significant challenge.

To address this challenge, this project leverages sentiment analysis and AI-powered summarization to systematically analyze large volumes of social media data. Sentiment analysis plays a crucial role in identifying whether customer feedback expresses positive, negative, or neutral sentiments, enabling businesses to track sentiment trends, assess customer satisfaction, and respond proactively to emerging concerns.[17]

A critical part of the data preprocessing pipeline involves natural language processing (NLP) techniques, such as tokenization, stopword removal, and lemmatization, to clean and structure the text before sentiment analysis.[1] Tokenization is the process of breaking down text into individual words or phrases to facilitate further analysis. Stopword removal eliminates common words that do not contribute significantly to meaning, reducing noise in the data. Lemmatization reduces words to their base form, ensuring consistency in word variations. These preprocessing techniques enhance the quality of text data, making sentiment classification and topic extraction more accurate and efficient.

For short-form text commonly found on platforms like Reddit and Twitter, we will use VADER (Valence Aware Dictionary and Sentiment Reasoner), a rule-based sentiment analysis tool specifically designed for informal and social media language.[2] VADER is highly effective in processing text that includes slang, emojis, punctuation-based intensity, and capitalization, making it well-suited for analyzing customer discussions on Costco's products and services. By incorporating VADER into the sentiment analysis pipeline, we can extract quick and reliable sentiment scores from customer feedback, allowing Costco to gain actionable insights into consumer perceptions.

This project also includes an AI-powered chatbot that allows interactive querying of sentiment trends, topic analysis, and insights gathered from the data to aid in stakeholder decision making. The chatbot makes data exploration more effective and actionable by enabling stakeholders to rapidly retrieve key takeaways, sentiment summaries, and emerging discussion themes in place of manually analyzing reports. This will ultimately increase member satisfaction and drive business growth.

3. Methodology

3.1 Data Collection

To gain a comprehensive understanding of customer sentiment toward Costco's products and services, we gathered unstructured text data from a variety of online sources where customers frequently share their experiences. These included platforms such as Reddit, Instagram, X (formerly Twitter), Costco.com, and Sitejabber. Given the informal and large-scale nature of user-generated content on these platforms, we used a hybrid data collection strategy involving web scraping and public APIs. Tools like BeautifulSoup and Scrapy were used to automatically extract comments and related metadata from web pages, while public APIs were used wherever available for structured data retrieval and to avoid excessive server load.[18][19]

The collected dataset consisted of over 400,000 + rows, with each record representing a unique comment. Each entry included fields such as the comment body, source URL, post and comment timestamps, platform of origin, and a manually or algorithmically assigned category label indicating whether the post related to a product or a service. Additional fields included numerical tokens for text processing, sentiment scores derived from models, and categorical sentiment labels (Positive, Neutral, or Negative). This dataset was deemed sufficient for identifying patterns in consumer sentiment and behavior across various Costco offerings.

To ensure ethical and lawful data collection, we followed all platform-specific terms of service and respected user privacy by excluding personally identifiable information. We adhered strictly to international data protection laws, and incorporated appropriate safeguards during scraping and storage. As an added assurance of quality and relevance, we plan to update this dataset on a monthly basis to reflect emerging trends and customer feedback dynamics.

Each row in the dataset which is a unique comment on a post contains the following fields:

Field Name	**Description**
Id	Unique identifier for each post
post_url	The URL of the post
comment_body	The text of the comment
post_title	The title of the post
source	Source from which the post originates
post_year	The year the post was posted
post_month	The month the post was posted
comment_year	The year the comment was posted
comment_month	The month the comment was posted
category	Related product or service
token	Numerical token used for text processing
sentiment_score	Sentiment score related to the following comment
sentiment	Positive, negative or Neutral

Table 1. Dataset Columns

3.2 Data Preparation

Once the raw data was collected, a multi-stage preprocessing pipeline was implemented to prepare it for downstream analysis. The first step involved deduplication and noise reduction to remove repeated entries, promotional content, or irrelevant discussions. Spam filtering was handled using a layered approach that combined rule-based filters, machine learning models, and contextual classifiers. For instance, posts with excessive punctuation, spammy links, or common spam patterns were flagged using rules, while a logistic regression model trained on TF-IDF features helped identify statistical anomalies.[3] Also, BERT-based classifiers were used to catch contextually deceptive spam content that rule-based methods might miss.[7]

Text normalization was then applied to clean the remaining content. This process included converting text to lowercase, stripping away punctuation and emojis, and removing special characters and URLs. Tokenization and lemmatization were performed using SpaCy to break down the text into linguistic units and reduce inflected forms to their root words. Stopword removal was applied to filter out commonly occurring but semantically irrelevant words, improving the effectiveness of the modeling stage.[1]

After cleaning, the dataset was segmented based on whether the comment related to products (e.g., electronics, groceries) or services (e.g., checkout, membership). This categorization was achieved through a combination of keyword heuristics and unsupervised topic modeling using Latent Dirichlet Allocation (LDA), which allowed us to discover dominant themes and group similar content.[6] This structured pipeline ensured that only high-quality, relevant, and well-organized text data was used for analysis.

3.3 Modeling Approaches

Our modeling framework focused on three main areas: sentiment analysis, topic modeling, and text summarization. For sentiment analysis, we used a hybrid ensemble approach consisting of three Natural Language Processing (NLP) models: VADER, BERT, and CardiffNLP. VADER (Valence Aware Dictionary and sentiment Reasoner) is a lexicon-based model designed specifically for analyzing sentiments in social media and short-form texts. It is highly effective in detecting polarity through emojis, slang, and punctuation cues. For longer and more nuanced texts, we used BERT (Bidirectional Encoder Representations from Transformers), which is pre-trained to understand the context of words bidirectionally and is capable of identifying sarcasm, ambiguity, and complex expressions. CardiffNLP, a transformer model trained on social media content, was included to better handle informal text structures commonly found on platforms like Instagram.[2][7][8]

Each model classified the text into one of three sentiment categories: Positive, Neutral, or Negative. VADER was applied to shorter, informal posts, while BERT and a variant of CardiffNLP (twitter-roberta-base-sentiment) were used for longer, more complex reviews. The ensemble approach allowed us to cross-validate results and ensure a higher degree of confidence in the final sentiment labels.

In addition to sentiment analysis, we conducted topic modeling using LDA to discover latent discussion themes within the dataset. This technique helped categorize feedback into broader areas such as product satisfaction, service complaints, pricing, or store experience. It enabled us to identify trends, common concerns, and popular topics without prior labeling.

To improve accessibility and reduce information overload, we implemented both extractive and abstractive text summarization. Extractive methods such as TF-IDF and TextRank were used to identify and highlight the most relevant sentences within a comment, while abstractive methods like T5 and PEGASUS generated human-like summaries that rephrased content in a concise and readable manner.[9][10][11] This two-pronged summarization approach allows stakeholders to quickly interpret large volumes of feedback.

3.4 Model Implementation Workflow

Due to the lack of labeled ground truth data, we did not conduct traditional train-test data splits or cross-validation. Instead, we focused on fine-tuning pre-trained models for the Costco domain using domain-specific vocabulary adaptation and applied manual validation techniques to verify output quality.

For performance evaluation, we selected metrics that were suited to each modeling task. F1-score was used to assess the precision and recall balance of our sentiment classifiers on a manually annotated validation set. For topic modeling, coherence scores (UMass and UCI) were calculated to measure how semantically meaningful each topic was. In the case of summarization, we evaluated output quality using ROUGE metrics (Recall-Oriented Understudy for Gisting Evaluation), comparing generated summaries to human-labeled references.[12][13]

Manual evaluation played a critical role in verifying sentiment predictions and summaries. Feedback from stakeholders was incorporated through review cycles to refine model performance and improve alignment with business use cases. This iterative, feedback-driven process allowed us to adjust model thresholds and improve overall interpretability.

3.5 Data Storage and Cloud Computing

Initial development, data processing, and model experimentation were conducted on local machines using Jupyter notebooks and standard Python-based libraries. All preprocessing, sentiment analysis, topic modeling, and summarization pipelines were executed in a local development environment, which provided sufficient computational resources for the project scope.

While we initially planned to scale our infrastructure using Google Cloud Platform (GCP), for this phase of the project, cloud usage was limited to chatbot deployment. The AI-powered chatbot was containerized using Docker and deployed on Microsoft Azure, allowing for secure, on-demand access to chatbot services by stakeholders. Azure was selected for its simplicity, compatibility with Docker containers, and availability through academic resources.

No cloud-based data pipelines, BigQuery storage, or Vertex AI components were used in the current implementation. However, these services remain viable options for future scalability, especially if real-time data ingestion, sentiment model retraining, or production-scale deployments are required. This hybrid approach for local development with cloud-based deployment of key components ensures flexibility, portability, and future readiness.[21]

3.6 External Tools and Visualization

To translate analytical results into actionable insights, we used Power BI to create an interactive and user-friendly dashboard. Power BI was selected due to its seamless integration with various data formats (CSV, SQL, JSON), ease of use, and existing adoption within Costco's internal systems.[20]

The dashboard included key visualizations such as:

- Sentiment trends over time
- Heatmaps comparing sentiment across products and services
- Interactive filters for platform, time period, sentiment polarity, and topic
- Knowledge graph representing relationships between customer sentiment, products, and service attributes

These visualizations were tailored for both technical and non-technical stakeholders, enabling intuitive exploration of complex sentiment data. The combination of powerful modeling and clear visual representation made our findings highly accessible and immediately useful to business decision-makers.

4. Results

4.1 Overview of Customer Engagement

Figure 1. Dashboard Overview – Total Posts, Comments, and Likes

Across Reddit, Instagram, X (Twitter), Costco.com and various other social media platform reviews, we analyzed 61,000 + posts and 400,000 + comments over an 11-year period (2014–2025). These interactions gathered over 4 billion post likes and 3 million comment likes, reflecting substantial public engagement with Costco's brand presence.

4.2 Annual Sentiment Trends

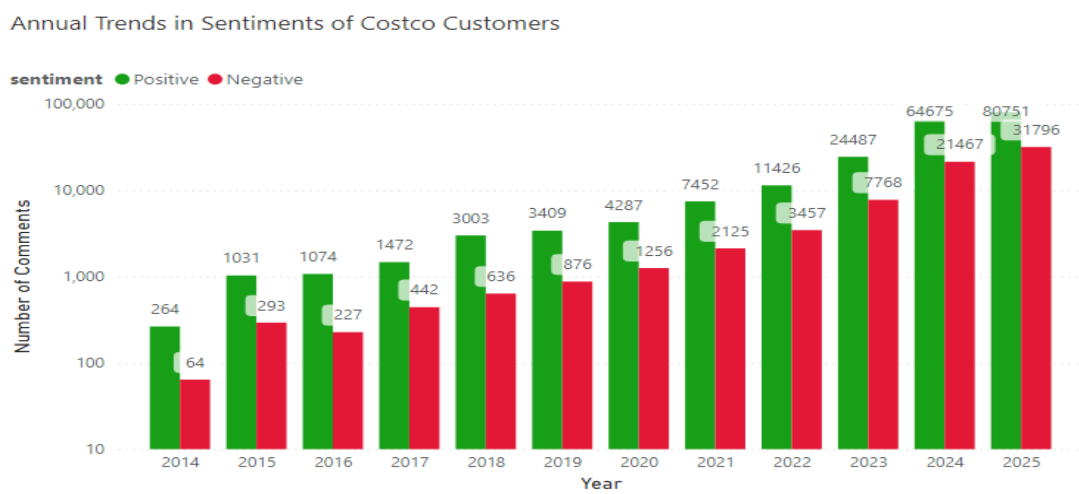

Figure 2. Annual Trends in Sentiments of Costco Customers

An increasing trajectory in both comment volume and sentiment polarity was observed. Notably, positive sentiment steadily grew, reaching over 80,000 positive comments by 2025, more than doubling the negative sentiment count for the same year. This suggests a rising level of satisfaction among Costco's customer base.

4.3 Product Category Analysis

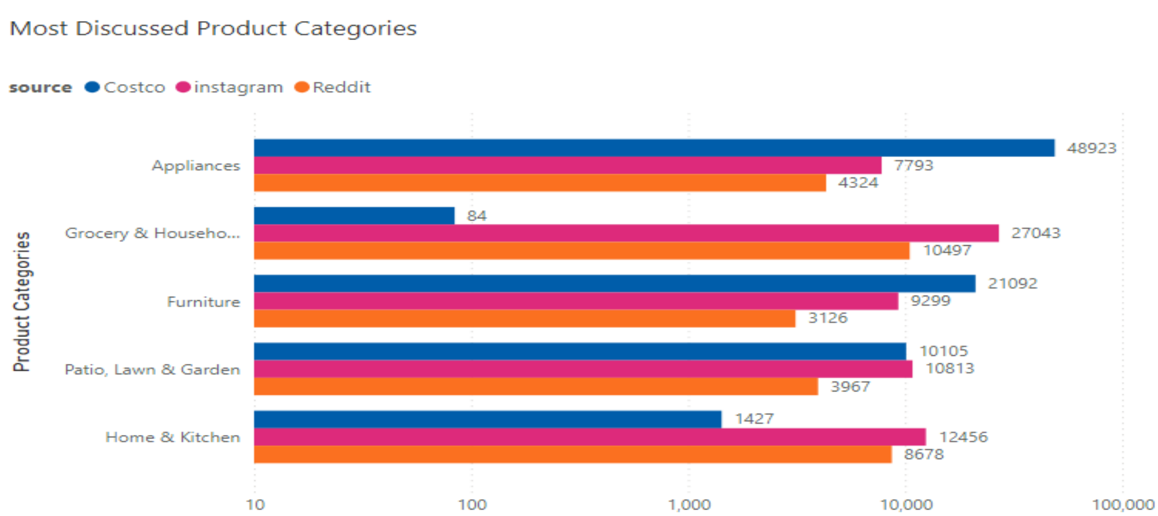

Figure 3. Horizontal bar chart comparing platform-wise category mentions

The most frequently mentioned product categories were:

- Appliances (48,000 + mentions)
- Grocery & Household Essentials (27,000 +)
- Furniture (21,000 +)

Platform-specific patterns emerged, Instagram discussions focused more on Grocery and Household items, while Costco.com highlighted appliances. This variation demonstrates how different platforms serve different audience segments.

4.4 Sentiment Analysis by Platform

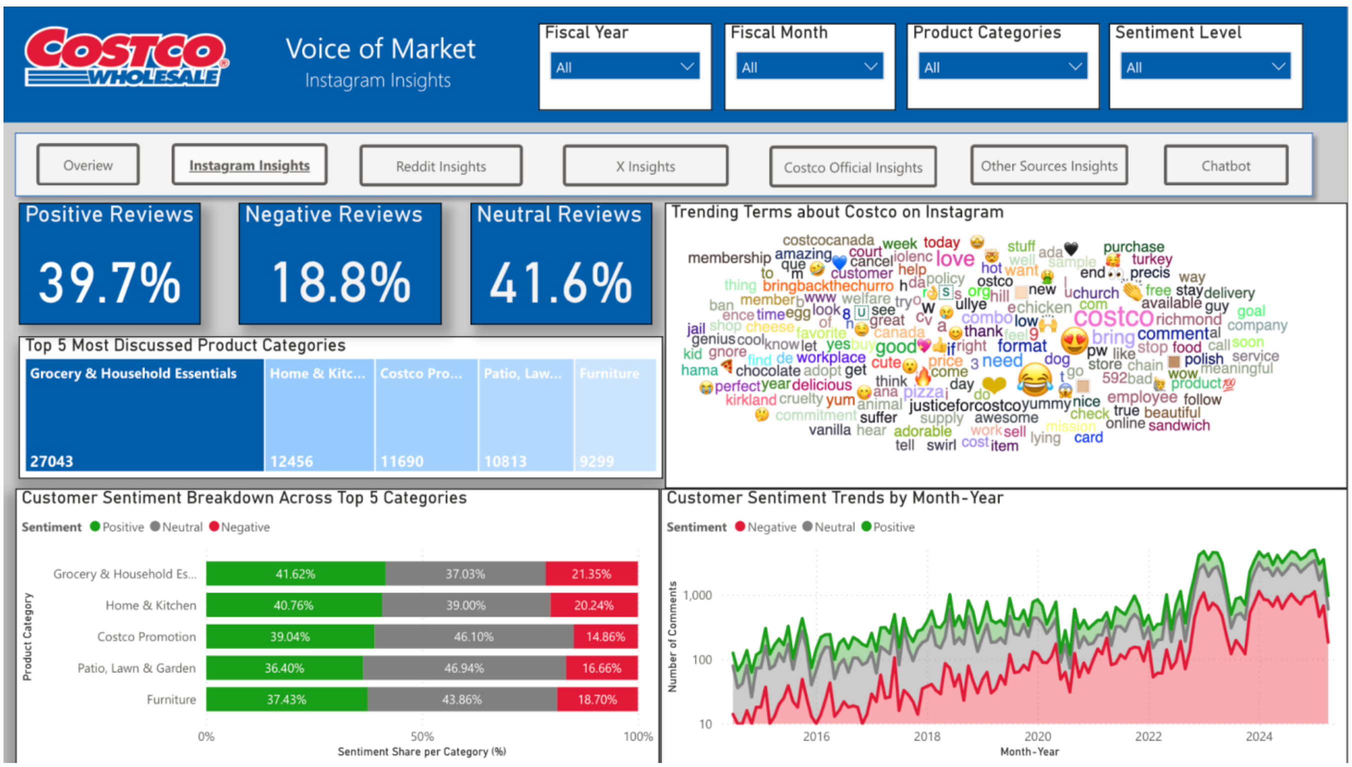

Figure 4. Combined Sentiment Analysis for Instagram

Instagram Insights

- Overall Positive sentiment: ~39.7%; Neutral: ~41.6%; Negative: ~18.8%
- Top categories: Grocery & Household Essentials (27,000 +), Home & Kitchen (12,000 +), Costco Promotions (11,000 +)
- Common topics: “membership”, “love”, “price”, “bring”, ”stop”

- Sentiment by category: Grocery & Essentials and Home & Kitchen >40% positivity; Promotions have larger neutral share
- Time trend: Increasing activity, peak in 2024–2025, slight rise in negative over positive sentiment

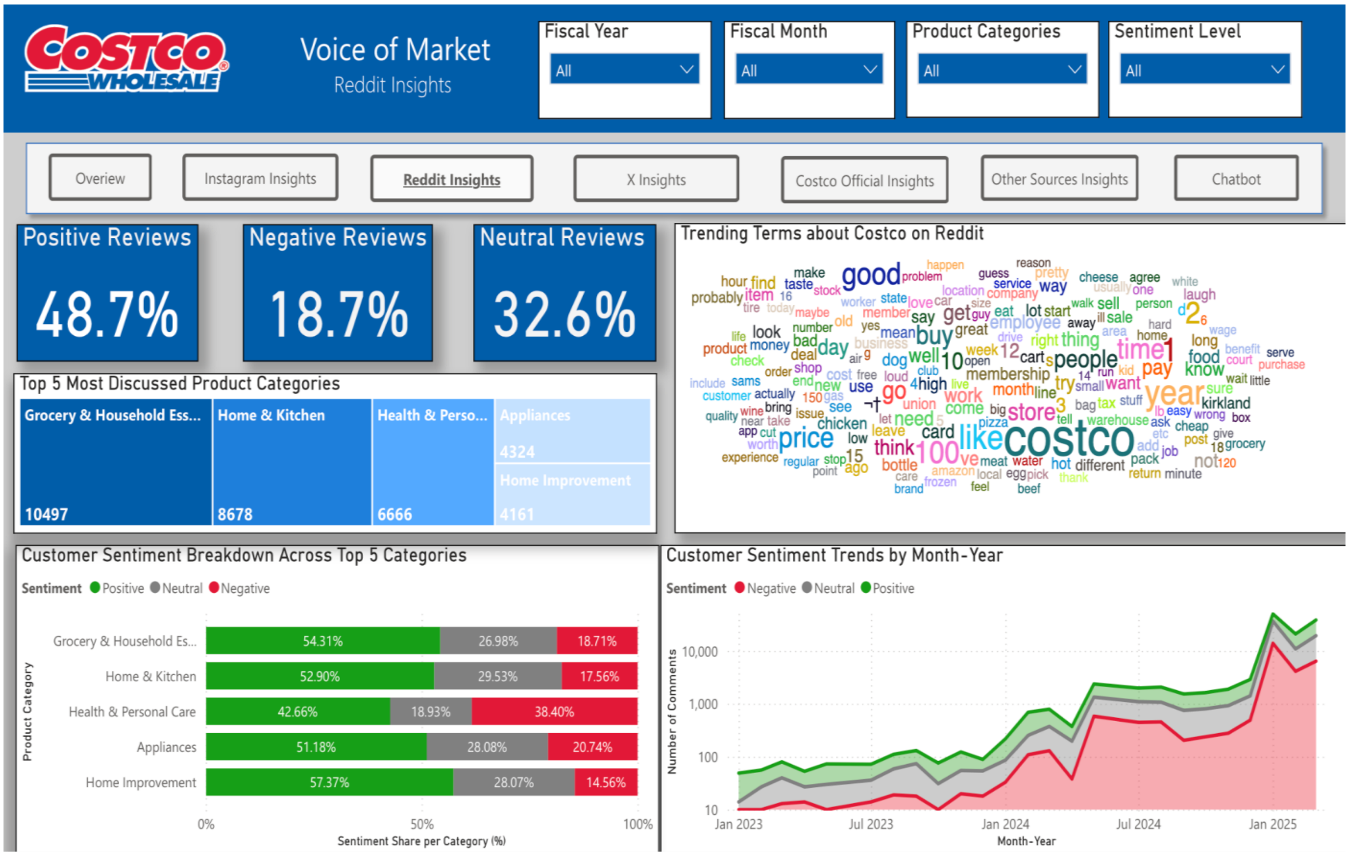

Figure 5. Combined Sentiment Analysis for Reddit

Reddit Insights

- Overall Positive sentiment: ~48.7%; Neutral: ~32.6%; Negative: ~18.7%
- Top categories: Grocery & Household Essentials (10,000 +), Home & Kitchen (8,000 +), Health & Personal Care (6,000 +)
- Common topics: “good”, “buy”, “time”, “membership”, “store”
- Sentiment by category: Home Improvement (57%) and Grocery (54%) strongest positivity; Health & Personal Care shows 38% negative
- Trend: Growing engagement, increasingly polarized discussions

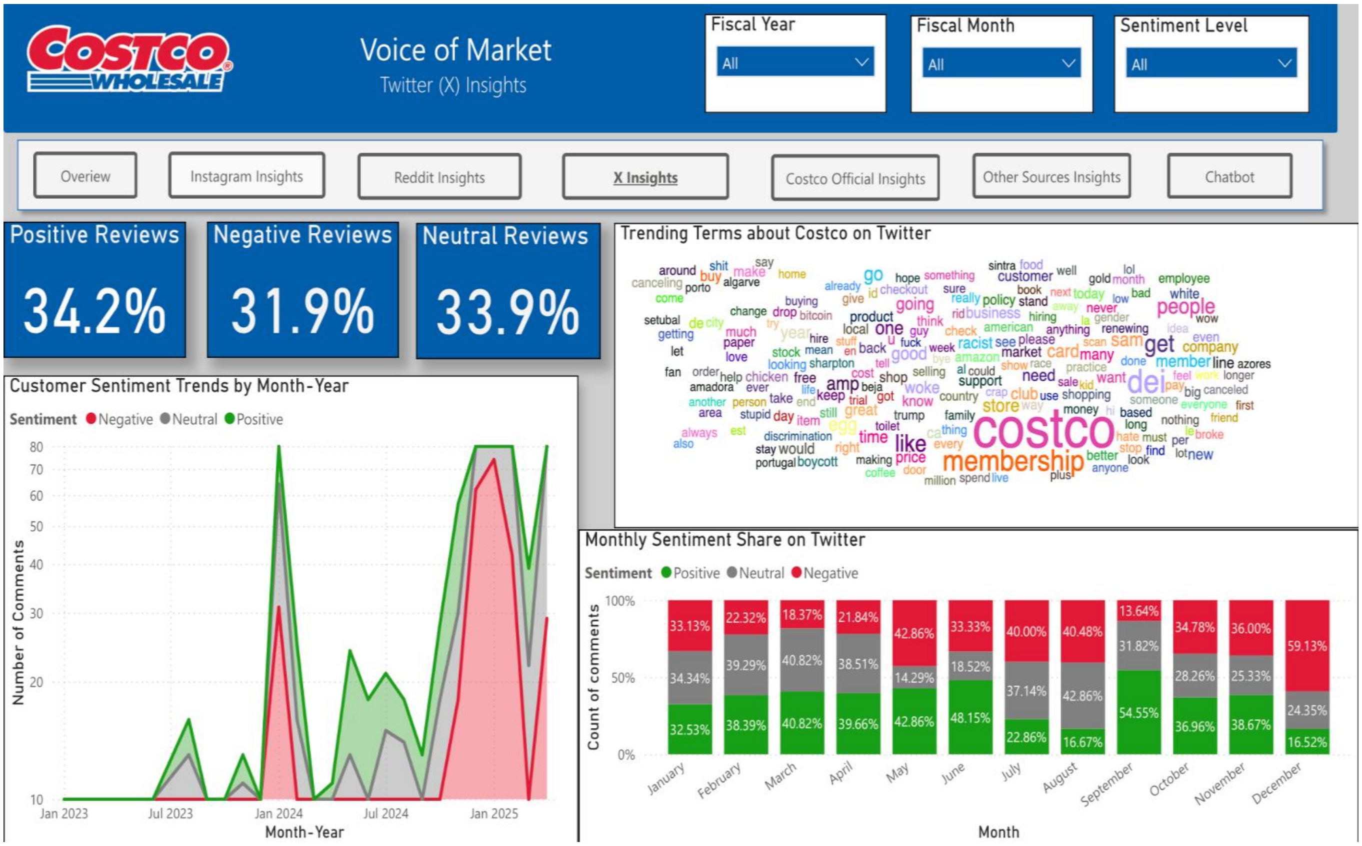

Figure 6. Combined Sentiment Analysis for X (Twitter)

X (Twitter) Insights

- Overall Positive sentiment: ~34.2%; Neutral: ~33.9%; Negative: ~31.9%
- Monthly spikes: December (59.13% negative), May (42.86% negative)
- Common topics: “membership”, “costco”, “time”, “price”
- Trend over time: Late 2024–2025 negative sentiment overtakes positive, indicating reputational risks
- Platform character: Fast-moving, volatile public opinion

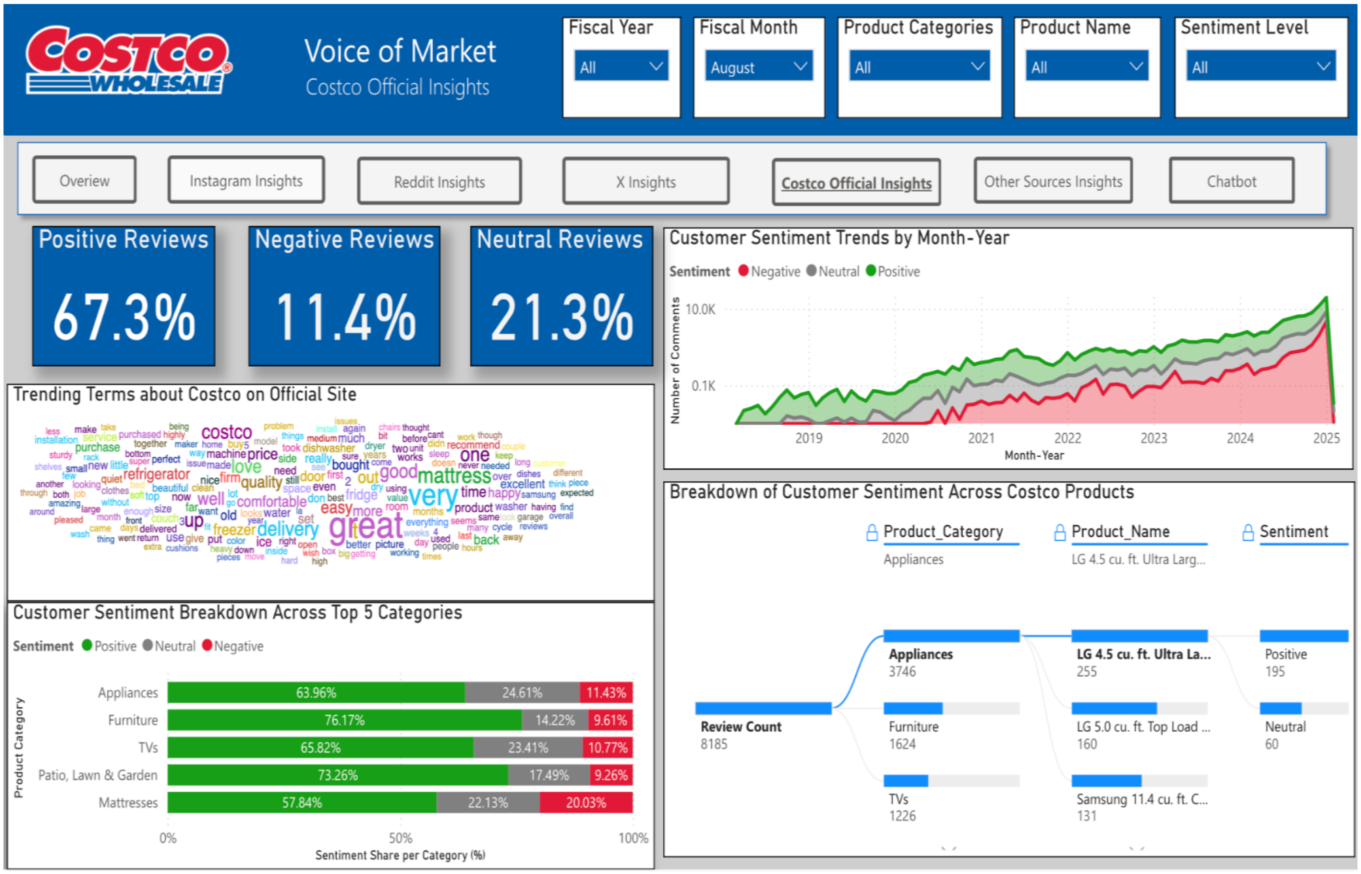

Figure 7. Combined Sentiment Analysis for Costco.com

Costco.com Insights

- Overall Positive sentiment: ~65.6%; Neutral: ~20.4%; Negative: ~14.0%
- Top categories: Appliances (48,000 +), Furniture (21,000 +), TVs (18,000 +), Patio & Garden, Mattresses
- Common topics: “delivery”, “mattress”, “love”, “amazing”, “installation”, “freezer”
- Sentiment by category: Furniture and Patio & Garden show >70% positivity; Appliances, TVs, and Mattresses ~60–65% positive
- Time trend: Steady growth in review volume over time, peaking in recent years, with positive sentiment consistently leading over negative and neutral

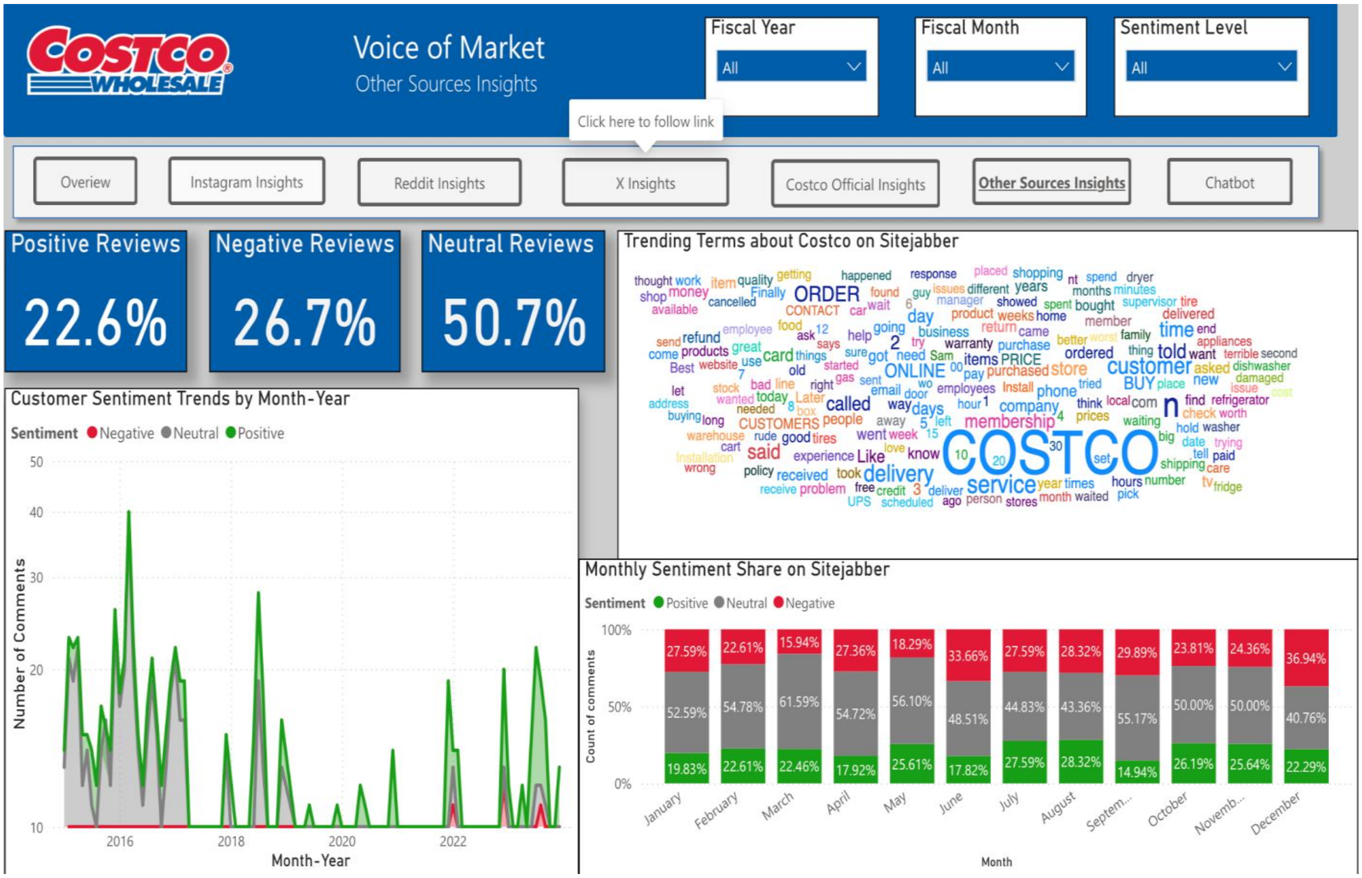

Figure 8. Combined Sentiment Analysis for Sitejabber

Sitejabber Insights

- Overall Positive sentiment: ~22.6%; Neutral: ~50.7%; Negative: ~26.7%
- Top topics: “order”, “delivery”, “service”, “membership”, “price”, “customer”, “called”, “online”
- Sentiment by month: Positive sentiment consistently low (~15–28%) across months, with neutral sentiment dominating (>50%) and negative sentiment peaking especially in September–December (~28–36%)
- Time trend: Low and fluctuating comment volume over the years, with occasional spikes in negative sentiment and sparse positive/neutral trends after 2020
- Notable: Sitejabber feedback centers heavily on service issues, order handling, and customer support frustrations, reflected in the high neutral + negative share and common complaint terms like “delivery,” “called,” and “wait”

4.5 Chatbot-Driven Customer Insights

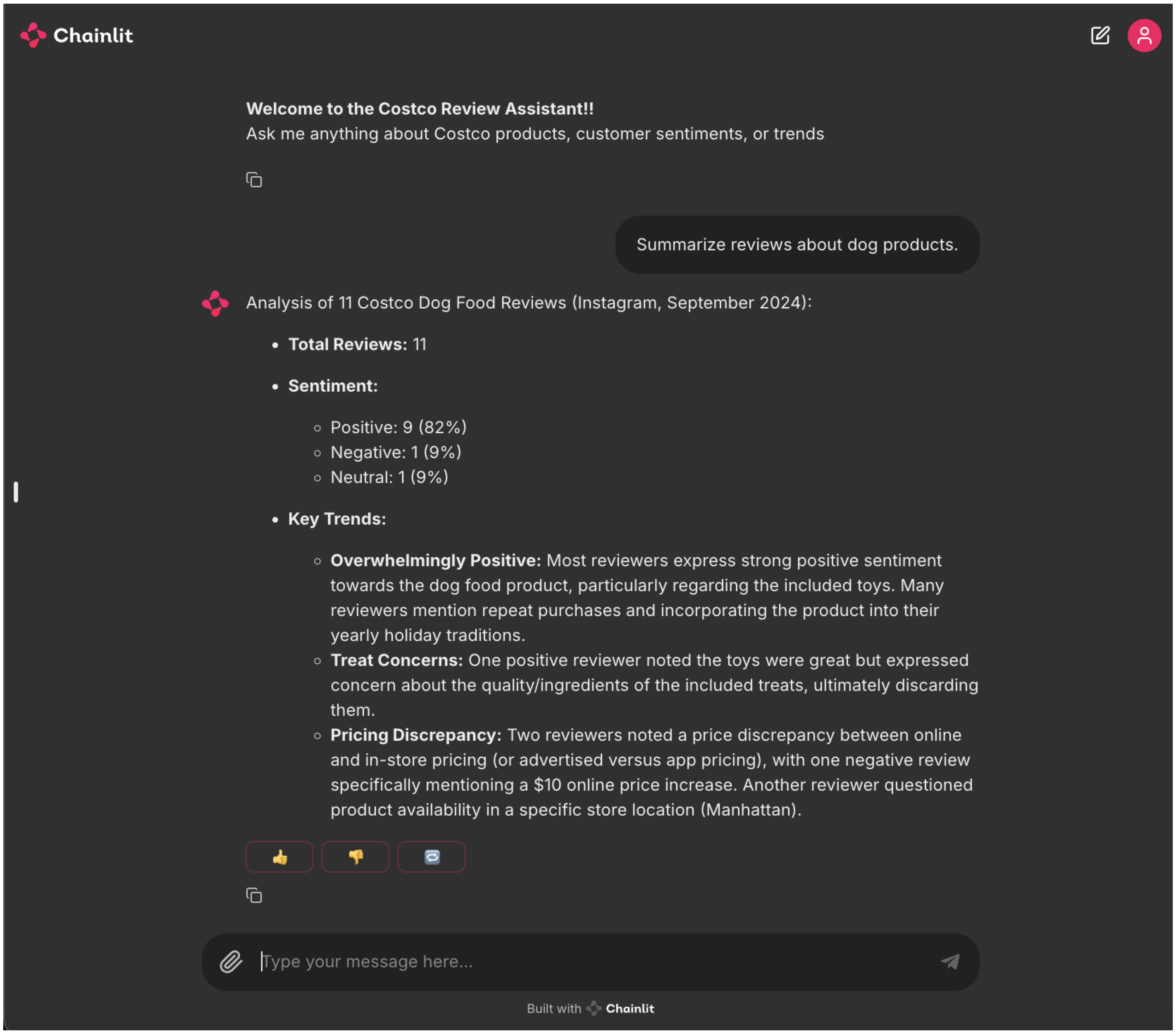

Figure 9. AI Chatbot Interface – Chainlit Implementation

A custom AI chatbot is developed using Chainlit and Gemini Pro 1.5, capable of extracting contextual insights from product reviews. The chatbot dynamically summarizes sentiment, highlights trends, and answers product-specific queries in natural language.

Key features:

- Context-aware sentiment classification

- Real-time topic detection
- Interactive response mechanism with user feedback

Summary:

- Figure 8. shows the Costco Review Assistant chatbot built using Chainlit.
- Query: Summarize Instagram dog food reviews.
- Total reviews analyzed: 11
- Sentiment breakdown:
    - Positive: 82%
    - Negative: 9%
    - Neutral: 9%
- Key insights identified by the bot:
    - Strong positive sentiment, especially regarding included toys.
    - Minor concerns about treat ingredient quality.
    - Notable complaints about a $10 price discrepancy (online vs. in-store).
    - Product availability issues raised in specific store locations (e.g., Manhattan).

This demonstrates the chatbot's ability to deliver actionable sentiment summaries and trends from social media data.

5. Discussion

5.1 Overview of Evaluation Goals

The primary objective of this project is to transform unstructured customer feedback from platforms such as Reddit, Instagram, and X (formerly Twitter) into actionable insights that can be used by Costco Wholesale to inform business strategy. Since traditional feedback mechanisms like surveys and focus groups are often time-consuming, biased, and lack real-time responsiveness, this project utilized modern natural language processing and artificial intelligence tools to bridge that gap.

Technically, we evaluated our sentiment models VADER and BERT to determine their effectiveness in interpreting the informal, diverse, and often emotionally charged language used by Costco customers online. To validate model performance, we manually labeled the comments by assigning sentiment tags positive, neutral and negative based on the contextual meaning of each comment. This labeled dataset served as a benchmark to compare model predictions against human judgment and helped us assess how accurately the models reflect real-world sentiments. Through this process, we also identified recurring patterns of misclassification, especially in comments containing sarcasm, slang, or emojis, elements that are frequently found on social media platforms. These insights helped us implement improvements to our overall sentiment analysis workflow and contributed to the development of a more reliable hybrid modeling approach.

We also assessed how well key components of our system, such as the dashboard and the AI-powered chatbot, helped non-technical stakeholders explore insights independently. These tools allow users to drill down by platform, product category, and time period, as well as access concise summaries and sentiment breakdowns without requiring technical expertise. By evaluating both the accuracy of our sentiment models and the usability of these components, we ensured the overall solution was not only technically sound but also actionable and accessible for Costco's business teams.

5.2 Model Performance Comparison

We compared predictions from two NLP models, VADER and BERT (CardiffNLP), with a manually labeled sentiment column (bart_summary) on Reddit data to evaluate the accuracy of sentiment classification.[2][7][8] Based on the contextual understanding of each comment, this benchmark represents how each comment is expected to be classified (positive, neutral, or negative).

We selected models based on linguistic patterns observed across platforms. In datasets from Reddit, Costco.com, Sitejabber, and X, we observed a high prevalence of slang, abbreviations, and informal language, but limited use of emojis. Due to these characteristics, we used VADER, a rule-based sentiment analysis model tailored for short-form, informal language. VADER handled basic polarity and punctuation cues well and was effective on comments like "love the price but its service sucks." However, it often failed to capture nuances in sarcasm, negation, and complex sentiment structures, such as "Costco's delivery is too fast... didn't even get a call."

In contrast, BERT (Cardiff NLP variant) was only used for Instagram data due to its distinct linguistic features. Unlike other platforms, Instagram comments frequently include emojis, emotional expressions, and multi-sentence posts, especially when promoting products or sharing user experiences. As these comments often convey sentiment through visual symbols or an implied tone, VADER has difficulty interpreting them properly. The Cardiff NLP model, which is trained on large-scale Twitter data, is well-suited to decoding social media language and can capture both contextual relationships and sentiment conveyed through emojis. It excels at understanding nuanced feedback, such as "this combo is 🔥 🔥 but shipping was 💀," where emotional intent is embedded in symbols and phrasing. Therefore, while VADER proves more suitable for platforms characterized by short, slang-heavy language, BERT demonstrated broader effectiveness across all datasets due to its ability to capture subtle tone shifts, implied sentiment, and domain-specific vocabulary, making it particularly valuable for platforms like Instagram with higher linguistic complexity and emotional nuance.

To validate model performance, we selected Reddit as the evaluation platform. We manually labeled a subset of its comments with sentiment tags positive, neutral, or negative based on contextual interpretation. This annotated dataset served as a benchmark to assess how accurately each model aligned with human understanding.

As expected, VADER showed reasonable performance in short, informal comments where sentiment was explicitly stated. However, it often misclassified comments with sarcasm, subtle tone shifts, or layered meaning, which are common in Reddit discussions. Its rule-based design, while effective for surface-level polarity, lacks the depth to interpret more nuanced expressions.

In contrast, BERT demonstrated strong alignment with human-labeled sentiment, especially in context-heavy or semantically ambiguous comments. Its ability to model relationships between words and detect implied sentiment gave it a significant advantage in capturing more complex emotional tone. However, BERT was also more computationally intensive and slower to process large volumes of data.

To leverage the complementary strengths of both models, we implemented an ensemble approach that combined their outputs using a weighted average, prioritizing BERT's predictions. This strategy consistently achieved the highest match rate with manually labeled sentiment and effectively reduced misclassifications caused by any single model.

By tailoring model usage to platform-specific language characteristics and validating outputs through manual annotation, we ensured that our sentiment analysis pipeline was both linguistically appropriate and operationally robust, providing accurate, scalable insights for Costco's customer feedback analysis.

5.3 Insights from Sentiment Trends

To understand how customer sentiment varied across Costco's products and services, we examined customer feedback trends over time and by category. A number of categories, such as appliances, furniture, and televisions, consistently received high levels of positive sentiment, indicating high levels of satisfaction with large, high-value purchases. Often, these categories included praise for product reliability, perceived value, and delivery experience. In contrast, Gift Cards & Tickets and Health & Personal Care exhibited more balanced or mixed sentiments. Although customers appreciated the convenience of gift cards and the variety of wellness products, negative feedback frequently cited issues related to redemption, availability, and packaging, especially with respect to supplements.

Additionally, sentiment trends reflect seasonal shifts and operational challenges. A notable spike in positive sentiment was observed in December, largely as a result of holiday promotions and popular seasonal products. However, this was accompanied by increased negative feedback related to shipping delays, out-of-stock items, and in-store congestion. A high level of engagement was observed in January, with both praise and frustration surrounding the return of items and delays in customer service. There has been a slight decline in engagement between February and March, however feedback volumes have remained steady, particularly with regards to everyday essentials and home improvement items.

A further illustration of these trends is provided by real user comments. Positive sentiments included remarks like: “Love my Kittleman Kirkland espresso machine—great value for the price,” and “Costco’s snack section is undefeated.” On the other hand, concerns were raised in comments such as: “Gift card didn’t work at checkout. Embarrassing,” and “The vitamin bottle is huge for just 30 pills. This example illustrates a spectrum of sentiment ranging from product satisfaction to dissatisfaction with service.

Additionally, platform-specific behavior plays an important role in sentiment expression. Generally, community-based platforms focused on long-format, critical feedback, particularly around services, whereas platforms specializing in visuals and lifestyles tended to promote more positive product reviews, often with emojis and emoticons.

By identifying where and when sentiment shifts, Costco can maintain strong-performing categories and deal proactively with customer service, product availability, and operations pain points. As a result, real-time business decisions and strategic improvements across the organization are informed by structured categorization and model-driven sentiment tagging.

5.4 Topic-Specific Interpretation

In order to gain an understanding of customer sentiments, we began by identifying the underlying topics in user-generated content. Our goal was to identify specific areas of praise or concern within Costco's product and service offerings by connecting these topics with associated sentiments.

Initially, Latent Dirichlet Allocation (LDA) was applied to extract topics from post and comment text across platforms like Reddit, Instagram, X, and Sitejabber. LDA, a probabilistic algorithm, groups frequently co-occurring words into latent topics, which can then be interpreted and labeled manually. However, LDA poses significant challenges in our context. Social media content is typically short, informal, and noisy, lacking word co-occurrence density. This led to incoherent or generic clusters where unrelated themes such as snacks, mattresses, and electronics were grouped together. As a result, many topics could not be clearly mapped to business-relevant categories, reducing interpretability and limiting actionability for stakeholders.

To overcome these limitations, we adopted a semantic similarity-based classification method using sentence transformer embeddings.[14][16] Instead of discovering topics probabilistically, we directly matched each post to a predefined Costco subcategory using cosine similarity between sentence embeddings. The subcategory list was curated from Costco's official online product taxonomy, spanning over 400+ subcategories under 25+ main categories. Each post was embedded with a neutral prompt to standardize the semantic context. If the highest similarity score exceeded a set threshold of 65%, the post was assigned to that subcategory; otherwise, it was labeled "Brand Promotion".

This approach has proven to be more effective than LDA for several reasons. First, it mapped posts directly to known Costco product categories, aligning with the company's business taxonomy. Unlike LDA's abstract clusters, this ensures interpretability and practical relevance.

Second, it handled short, informal, or ambiguous text more effectively. Posts with minimal content or slang that confused LDA were accurately classified using semantic embeddings.

Third, it allowed consistent categorization even when posts used emojis or shorthand. For instance, “these Costco snacks are difficult" was correctly mapped to Grocery & Household Essentials, and “need that Samsung washer ASAP?" to Washers & Dryers.

By pairing these mappings with sentiment labels, we uncover actionable insights. Grocery & Household Essentials had strong positive sentiment around snacks, coffee, and Kirkland products highlighting opportunities for targeted promotions. Gift Cards & Tickets received mixed feedback, with concerns about redemption clarity and availability.

As a result of these insights, Costco can highlight top-rated items, address frequent complaints, expand its product line, or change suppliers based on these insights.

Ultimately, by connecting sentiments with structured product categories, this approach enables data-driven customer feedback prioritization. With these insights embedded into our reporting and conversational interfaces, Costco can continuously refine product offerings and enhance customer experience based on real-time feedback.

5.5 Final Deliverables

We developed two key deliverables to help translate large volumes of unstructured customer feedback into clear, actionable insights: a Power BI dashboard and a chatbot powered by artificial intelligence. Through these tools, stakeholders can independently explore customer sentiment and emerging discussion themes especially those without technical backgrounds.

The interactive dashboard provides an overview of sentiment trends, insights about product categories, and platform-specific engagement over millions of customer interactions. Using the filters, users can explore data by product category, sentiment level, time period, and platform dynamically. There are several key visual components of the report, including sentiment breakdowns by category, trending terms from customer comments, and time-series charts that highlight periods of positive or negative sentiment. These features allow business users to quickly identify which products are receiving praise, which services are under scrutiny, and how customer sentiment evolves over time.

Furthermore, the dashboard's cross-platform design allows users to compare customer opinions across all social media platforms and other sources. For example, users can investigate why furniture might receive stronger praise on the official site while facing more criticism on third-party review platforms. This multi-source visibility ensures Costco can make balanced and informed decisions based on customer voices.

To complement the dashboard and provide a more conversational interface, we developed an AI-powered chatbot using Chainlit and Gemini Pro 1.5.[15] This chatbot allows users to interact with the dataset in natural language, making it possible to retrieve insights through simple queries like, “What are customers saying about patio furniture in 2024?” or “Show sentiment trends for electronics last summer.” The chatbot responds with synthesized summaries, sentiment breakdowns, and relevant product or service mentions, all grounded in the underlying data.

Together, these tools provide a cohesive feedback intelligence system. These technologies enable the transition between model outputs and actionable business insights, bridging the gap between data science and decision-making. By utilizing the deliverables, marketing, merchandising, and operations teams will be able to gain a better understanding of customer sentiment while also taking timely action based on evidence. Consequently, Costco is better positioned to adapt to consumer needs, monitor public perception, and maintain a competitive edge in a rapidly changing retail environment.

6. Conclusion

6.1 Empowering Costco with Data-Driven Customer Insight

The "Voice of Market (VoM)" addresses Costco Wholesale's challenge of converting huge volumes of unstructured customer feedback into structured, actionable insights. Traditional forms of feedback, such as surveys and focus groups, are often time-consuming, resource-heavy, and limited in scope. By contrast, this solution uses sentiment analysis, summarization, an AI chatbot, and a Power BI dashboard to get insights from over 400K comments on Costco's official website and social media, including Reddit, Instagram, Twitter (X), and Sitejabber.

With the scalable pipeline that includes web scraping, data cleansing, sentiment models, and topic modeling, a comprehensive solution has been made that transforms the manner in which Costco can monitor customer sentiment. The dashboard visually presents sentiment trends and trending themes over time, and the chatbot enables stakeholder engagement with precise insights. This solution enables decision-makers to align product offers and services with the true voice of the customer, increase satisfaction, and respond to emerging trends and issues more rapidly.

6.2 Strengths & Limitations

Strengths:

- End-to-end scalable pipeline: From data cleaning to chatbot delivery, the components are reusable and modular in other Costco initiatives.
- Multi-source unification: Unified data from five sites spanning 2014–2025 provides a wide range for customer opinion.
- Hybrid sentiment approach: VADER analyzes short comments, while Twitter Roberta and BERT capture rich, nuanced sentiment in longer statements.
- Visualization and reporting: The dashboard seamlessly presents sentiment breakdown, trending categories, and time-based trends.
- Summarization chatbot: The chatbot enables stakeholders to ask questions and receive direct insight without the need for technical knowledge.

Limitations:

- Data bias: Social media may overrepresent certain demographics or opinions, underrepresenting non-digital customers.
- Temporal imbalance in data: Certain months or years had considerably more activity than others, creating an unbalanced distribution in time-based analysis

6.3 Usefulness

This initiative is an excellent demonstration of the increasing relevance of AI-powered consumer insight in modern retail. Since many companies have historically used manual analysis, this solution illustrates how public, real-time opinion can be transformed into structured knowledge at scale. This approach benefits Costco merchandising, marketing, and customer service teams by revealing a deeper understanding of customer sentiment through an extensive range of unstructured channels.

The interactive dashboard helps with a high-sentiment product category identification, improving recurring pain points, and visualization of potential longitudinal sentiment change. The chatbot maximizes accessibility even further by helping stakeholders to query insights in real-time, reducing the need to manually go through big reports. These technologies together enable smarter product launches, customer loyalty-building campaigns, and more effective issue resolution.

6.4 Future Work

Taking advantage of the foundation work in this project, future expansions could significantly increase impact:

- Voice and video integration: Explore sentiment extraction from voice-based support or video reviews as future channels emerge
- Live chatbot integration: Integrate the chatbot into Costco's internal Slack to provide real-time alerts.
- Mobile app prototype: Create a light-weight interface to give Costco product teams immediate on-the-go access to key insights.

7. References

1. **Geeks for Geeks.** (2025, April 08). *Natural language processing (NLP) – Overview.* Retrieved February 13, 2025, from https://www.geeksforgeeks.org/natural-language-processing-overview/

2. **Geeks for Geeks.** (2025, May 10). *Python Sentiment Analysis using VADER.* Retrieved March 20, 2025, from https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/

3. **Arpanjeet Sandhu, & Sabah Mohammed** (April 2022). *Detecting Anomalies in Logs by Combining NLP features with Embedding or TF-IDF.* In 2023 6th International Conference on

Computing, Communications, and Security (ICCCS) (pp. 1–6). IEEE. Retrieved February 20, 2025, from https://ieeexplore.ieee.org/document/10433949

4. **Statista.** (2024). *Number of Costco card holders worldwide from 2014 to 2024*. Retrieved February 9, 2025, from, https://www.statista.com/statistics/718406/costco-number-memberships/

5. **Sandhu, A., & Mohammed, S.** (2022). *Detecting Anomalies in Logs by Combining NLP Features with Embedding or TF-IDF*. TechRxiv. Retrieved March 20, 2025, from https://www.researchgate.net/publication/359931328_Detecting_Anomalies_in_Logs_by_Combining_NLP_features_with_Embedding_or_TF-IDF

6. **IBM.** (2024, April 22). *Latent Dirichlet Allocation (LDA).* Retrieved March 20, 2025, from https://www.ibm.com/topics/latent-dirichlet-allocation

7. **Tech Target**. (n.d.). *BERT language model.* Retrieved April 15, 2025, from https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model

8. **Hugging Face.** (n.d.). *Twitter RoBERTa Base for Sentiment Analysis.* Hugging Face. Retrieved April 27, 2025, from https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment

9. **Geeks for Geeks.** (2024, May 28). *Text Summarization Techniques.* Retrieved March 24, 2025, from https://www.geeksforgeeks.org/text-summarization-techniques/

10. **Geeks for Geeks.** (2025, January 22). *Text Summarization in NLP*. Retrieved February 20, 2025, from https://www.geeksforgeeks.org/text-summarization-in-nlp/

11. **Mounish, V.** (2024, May 27). *How to Build an Effective Text Summarization Model Using Google's T5-Base?* Analytics Vidhya. Retrieved March 12, 2025, from https://www.analyticsvidhya.com/blog/2024/05/text-summarization-using-googles-t5-base/?ref=footer

12. **Muddala, S.** (2023, December 12). *NLP Model Metrics.* Medium. Retrieved May 02, 2025, from https://medium.com/@sujathamuddala1213/nlp-model-metrics-b3fa32373269

13. **Rosner, F., Hinneburg, A., Röder, M., Nettling, M., & Both, A.** (2013). *Evaluating topic coherence measures*. Retrieved April 20, 2025, from http://mimno.infosci.cornell.edu/nips2013ws/nips2013tm_submission_7.pdf

14. **Hugging Face**. (n.d.). *all-MiniLM-L6-v2 model card.* Retrieved April 26, 2025, from https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

15. **Chainlit.** (n.d.). *Overview – Getting Started.* Retrieved May 13, 2025, from https://docs.chainlit.io/get-started/overview

16**. SBERT.net.** (n.d.). *SentenceTransformers Documentation.* Retrieved February 20, 2025, from https://www.sbert.net

17**. IBM**. (n.d.). *What is Sentiment analysis?* Retrieved February 5, 2025, from, https://www.ibm.com/think/topics/sentiment-analysis

18. **Dorian, L.** (2021, March 5th). *Scraping Medium with Python & BeautifulSoup*. Retrieved February 9, 2025, from, https://dorianlazar.medium.com/scraping-medium-with-python-beautiful-soup-3314f898bbf5

19. **GitHub.** (n.d.). *Public APIs [Repository]*. Retrieved February 12, 2025, from, https://github.com/public-apis/public-apis

20. **Microsoft.** (n.d.). *Power BI overview.* Microsoft Learn. Retrieved February 22, 2025, from https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview

21. **Google Cloud. (n.d.).** *Vertex AI Platform: Innovate faster with enterprise-ready AI, enhanced by Gemini models.* Retrieved May 20, 2025, from https://cloud.google.com/vertex-ai