Grocery Store Review Analysis

Introduction : Why analyze grocery store reviews?

- **Customers rely on online reviews to choose grocery stores:** Digital reviews serve as trusted sources of information, influencing customer decisions.

- **Understand satisfaction & dissatisfaction drivers:** Identifying what customers love (e.g., quality, service) and areas of frustration (e.g., wait times, stock issues) helps improve the experience.

- **Review insights inform pricing, layouts & staff management:** Reviews reveal customer perceptions of value, product placement preferences, and service quality, guiding strategic decisions in pricing, store design, and staffing.

Project Objective

Goal: Identify factors influencing customer satisfaction & business performance in grocery stores

- **Objective:** Understand the key drivers of positive and negative customer experiences, with the aim of improving both customer satisfaction and overall business success.
- **Actionable insights for grocery store managers & marketing teams:**
  - Provides data-driven guidance to help managers make informed decisions about store operations, product offerings, and customer service.
  - Helps marketing teams refine strategies to enhance customer engagement, attract new shoppers, and increase brand loyalty.
- **Helps businesses address negative feedback & enhance customer retention:**
  - Identifying pain points from customer feedback allows businesses to proactively address issues (e.g., long checkout lines, poor service) and prevent losing customers.
  - Understanding satisfaction drivers enables stores to continuously improve their offerings, fostering customer loyalty and repeat business.

Research Questions

1. What are the most common factors affecting customer satisfaction?
2. How do ratings correlate with store location, pricing, and attributes?
3. What insights can be derived from sentiment analysis?
4. What marketing strategies can stores use based on customer feedback?

Industry Statistics & Data Overview

- **Dataset Source: Yelp reviews for grocery stores in Arizona**
  - **Focus:** Analysis of customer feedback from one of the most popular review platforms for local businesses.
  - **Relevance:** Yelp is frequently used by consumers to share their shopping experiences and make informed decisions.

- **Total Reviews Analyzed: 8000**

- **Most Reviewed Stores:**
  - **Trader Joe's:** Known for its unique products and friendly atmosphere.
  - **Target:** Popular for a variety of goods, including groceries and household products.
  - **Safeway:** Recognized for its wide range of grocery items and convenient locations.
  - **Fry's:** A go-to option for competitive pricing and diverse product offerings.

Dataset Summary

Dataset Overview

- 8,000 customer reviews

- Contains review texts, ratings, business names, locations

Data Cleaning

- **Missing Values**: 10 missing in attributes, 64 in business hours

- **Outliers**: Found in review counts & ratings (addressed using boxplots

Insights & Visualization

- **Sentiment Score Analysis**: Mean, median, standard deviation

- **Word Cloud**: Highlights frequently mentioned words in reviews

Most Common Words in Reviews

Suitability & Bias

Challenges & Tackling

Why This Dataset is Useful

• Identifies trends in **customer behavior & business performance**

• Helps businesses understand customer satisfaction & areas for improvement

Potential Biases

• **Review Bias**: Extreme opinions (either really positive or really negative) dominate, neutral opinions are often missing

• **Geographical Bias**: Data limited to specific locations

• **Business Selection Bias**: Large businesses overrepresented

• **Fake Reviews**: Some businesses may manipulate ratings (boost their ratings by posting or encouraging artificial reviews)

Challenges in Working with Text Data

• Messy, full of slang, hard to interpret

• Requires cleaning (removing noise, slang, inconsistencies)

• Handling missing values & extreme outliers

Processing Techniques Used

• **Sentiment Classification**: Categorizing reviews as positive, neutral, or negative

• **Topic Modeling**: Identifying key themes in reviews

• **Predictive Analytics**: Helping businesses track trends & make data-driven decisions

Named Entity 👀 👀 Recognition

Noun	Frequency		Adjective	Frequency
0	store	7496	0	good	2946
1	time	3018	1	great	2469
2	customer	2868	2	friendly	1981
3	grocery	2497	3	other	1844
4	item	2422	4	nice	1456

Verb	Frequency		Named Entity	Frequency
0	have	8560	Target	2041
1	go	4589	Safeway	2018
2	get	3674	one	1104
3	find	2385	Fry's	1098
4	shop	2312	Tucson	705

Extracted key entities like store names, locations, and competitor mentions.

Insights

○ Target, Safeway, Fry's were the most mentioned stores, indicating customer engagement.

○ Mentions of locations (e.g., Tucson:705) highlighted regional sentiment differences.
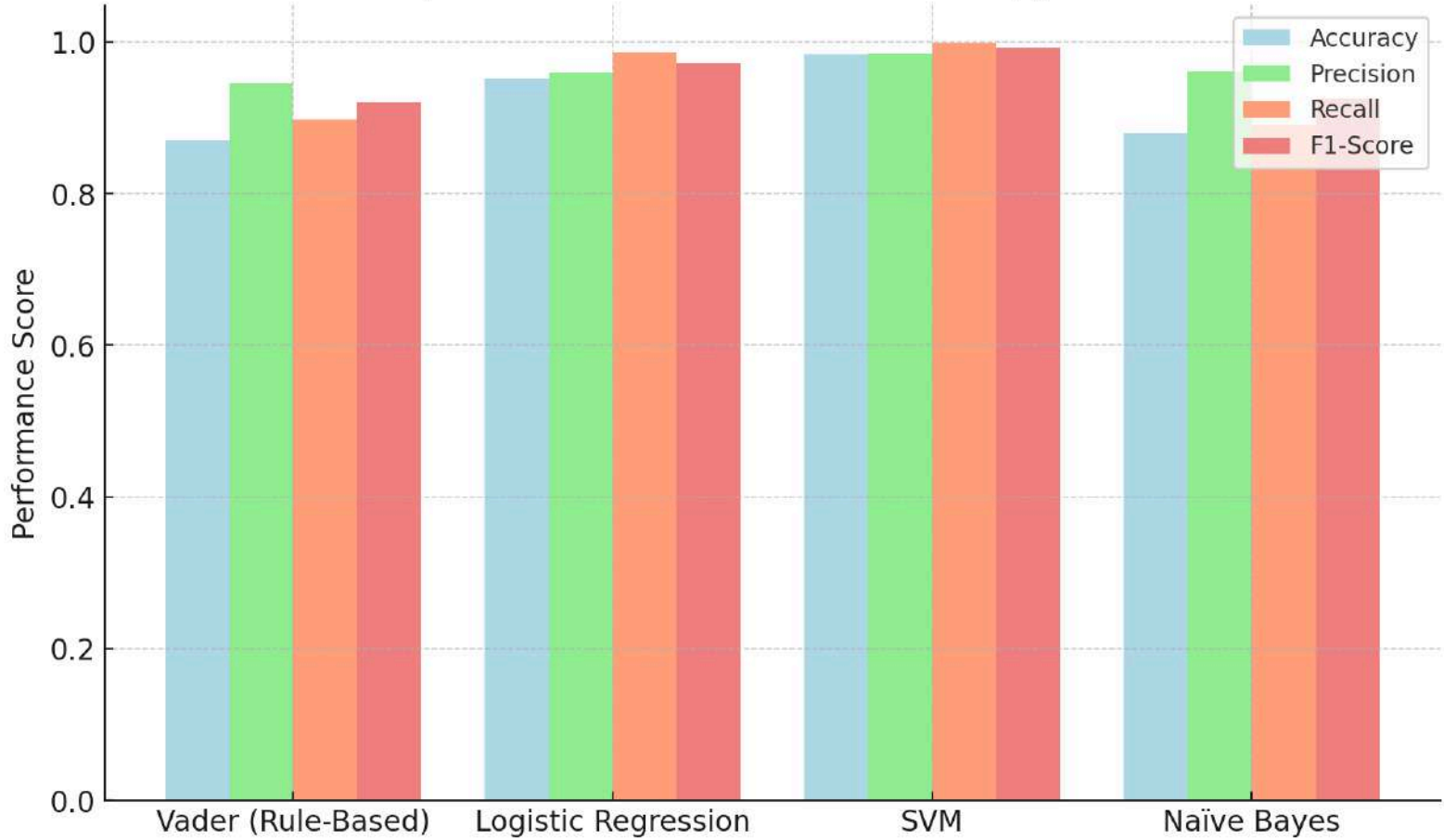
Top Words in Reviews

- **Nouns:** "store," "time," "customer" → Focus on shopping experience and service
- **Adjectives:** "good," "great," "friendly" → Positive customer sentiment
- **Verbs:** "have," "go," "get" → Action words reflecting purchase behavior
- **Named Entities:** "Target," "Safeway," "Fry's" → Most mentioned stores

Insights

✔ Customers talk about store experience & time spent shopping.

✔ Positive descriptors highlight customer satisfaction drivers.

✔ Stores should optimize service based on frequent mentions.

✔ Helps grocery stores understand brand perception and competitive positioning.

Comparison of ML vs. Rule-Based Approaches

Metric	Vader	Logistic Regression	SVM	Naive Bayes
0 Accuracy	0.870625	0.952500	0.983750	0.879375
1 Precision (Positive)	0.944939	0.958863	0.984825	0.961028
2 Recall (Positive)	0.897692	0.986154	0.998462	0.891538
3 F1-Score (Positive)	0.920710	0.972317	0.991597	0.924980

Models used in Sentiment Analysis

Sentiment analysis helps businesses *understand customer emotions* in reviews.

- Traditional approaches (Rule-Based Methods) struggle with context
- Machine Learning (ML) models for better accuracy and adaptability

Models Used in This Project:

- **Vader** (predefined lexicon) – Fast and efficient, but lacks contextual understanding
- **Naïve Bayes** – Works well for simple text but struggles with complex sentence structures
- **Logistic Regression (High Precision: 98.4%)**
    - Learns relationships between words & sentiment.
    - Performs well with structured data but can struggle with non-linear sentiment expressions
- **Support Vector Machine (SVM) (High Accuracy: 98.4%)**
    - Finds optimal decision boundaries between sentiment classes.
    - Best performer is with the highest accuracy and robustness.

Sentiment Comparison
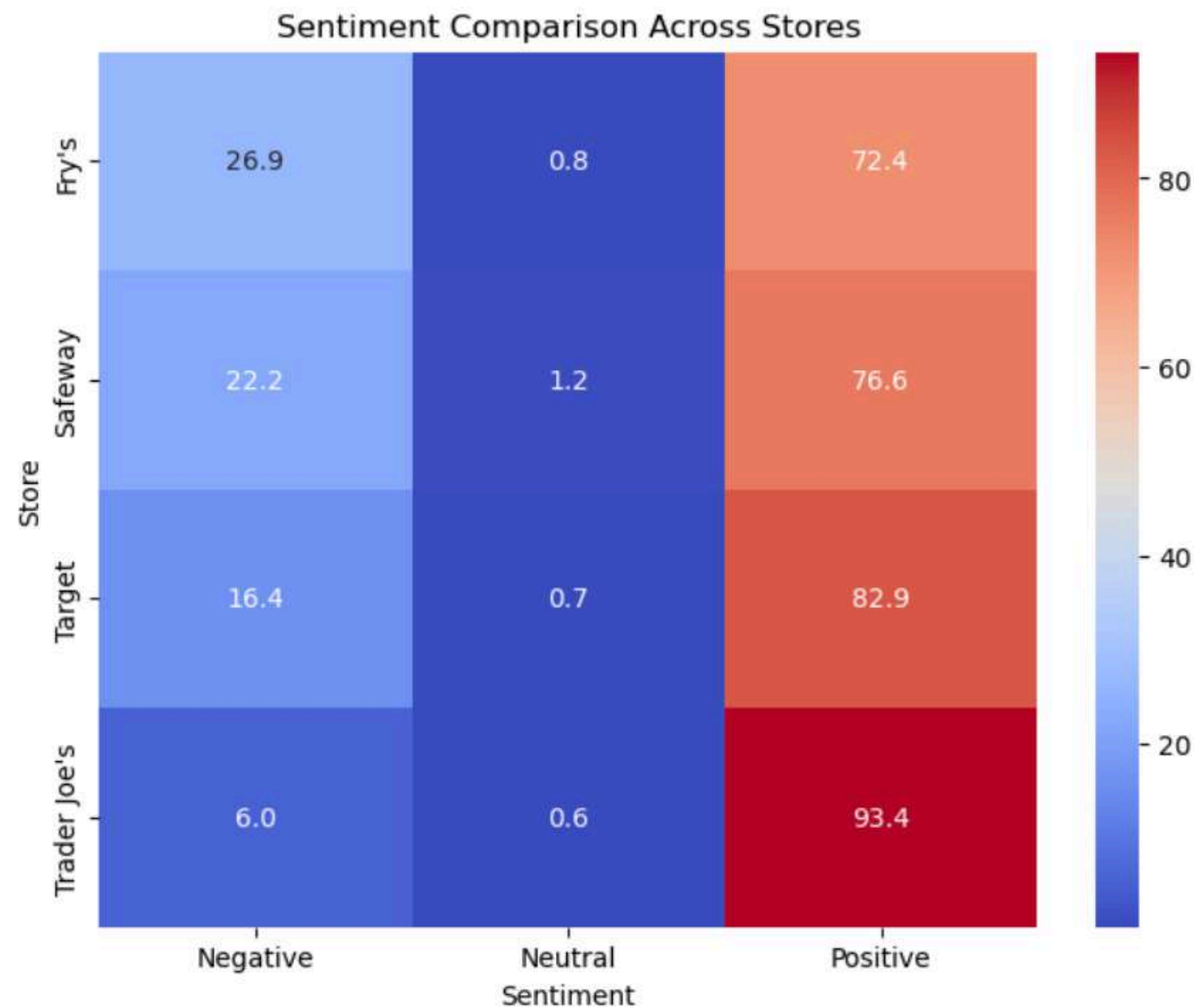

Sentiment Comparison Across Stores

Takeaways from Sentiment Analysis

- **Trader Joe's**: *Highest positive sentiment* (**93.4%**), strong brand loyalty.
- **Target**: *High customer satisfaction* (**82.9%**), minimal negative sentiment.
- **Safeway**: Moderate satisfaction (76.6%), but some pricing & service concerns.
- **Fry's**: Highest negative sentiment (26.9%), indicating service issues.

Business Insights

- Fry's should focus on improving customer service and product availability.
- Safeway needs to address complaints about pricing and quality.
- Target should work on eliminating minor negative feedback to enhance reputation.
- Trader Joe's should maintain its strong customer engagement and unique shopping experience.

Deep Learning Models

	Model	Accuracy	Precision	Recall	F1-Score
0	ANN + TF-IDF	99.25	99	99	99
1	RNN + Word Embeddings	89.40	90	89	88
2	LSTM + Word Embeddings	2.80	97	11	20
3	Improved LSTM + GloVe	8.35	60	45	51
4	One-Hot Encoding + RNN	0.85	0	0	0
5	BiLSTM + GloVe	99.15	99	99	99

best *for* **Topic Modeling**

bad

Best Performing Models : High accuracy and contextual understanding

- **BiLSTM + GloVe** - GloVe embeddings helped with semantic representation, crucial for distinguishing topics
- **ANN + TF-IDF** - Fast & efficient especially when combined with LDA (Latent Dirichlet Allocation) or NMF (Non-negative Matrix Factorization)

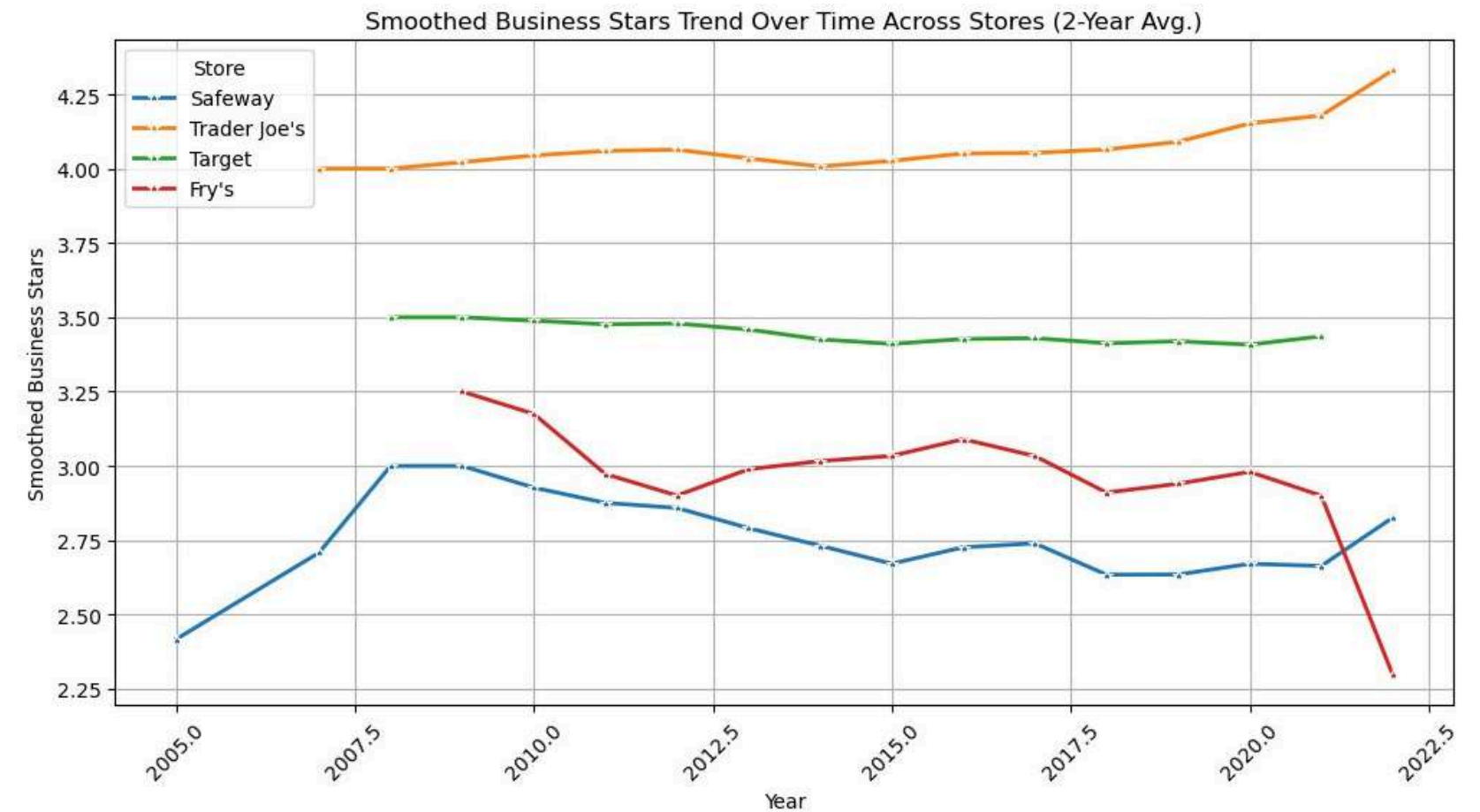Poor Performing Models : **LSTM + Word Embeddings** & **One-Hot Encoding + RNN**

- Training issues, Low Recall & Accuracy, does not retain word meanings, making it ineffective

Sentiment & Business~OverTime

Smoothed Sentiment Trend Over Time Across Stores (2-Year Avg.)

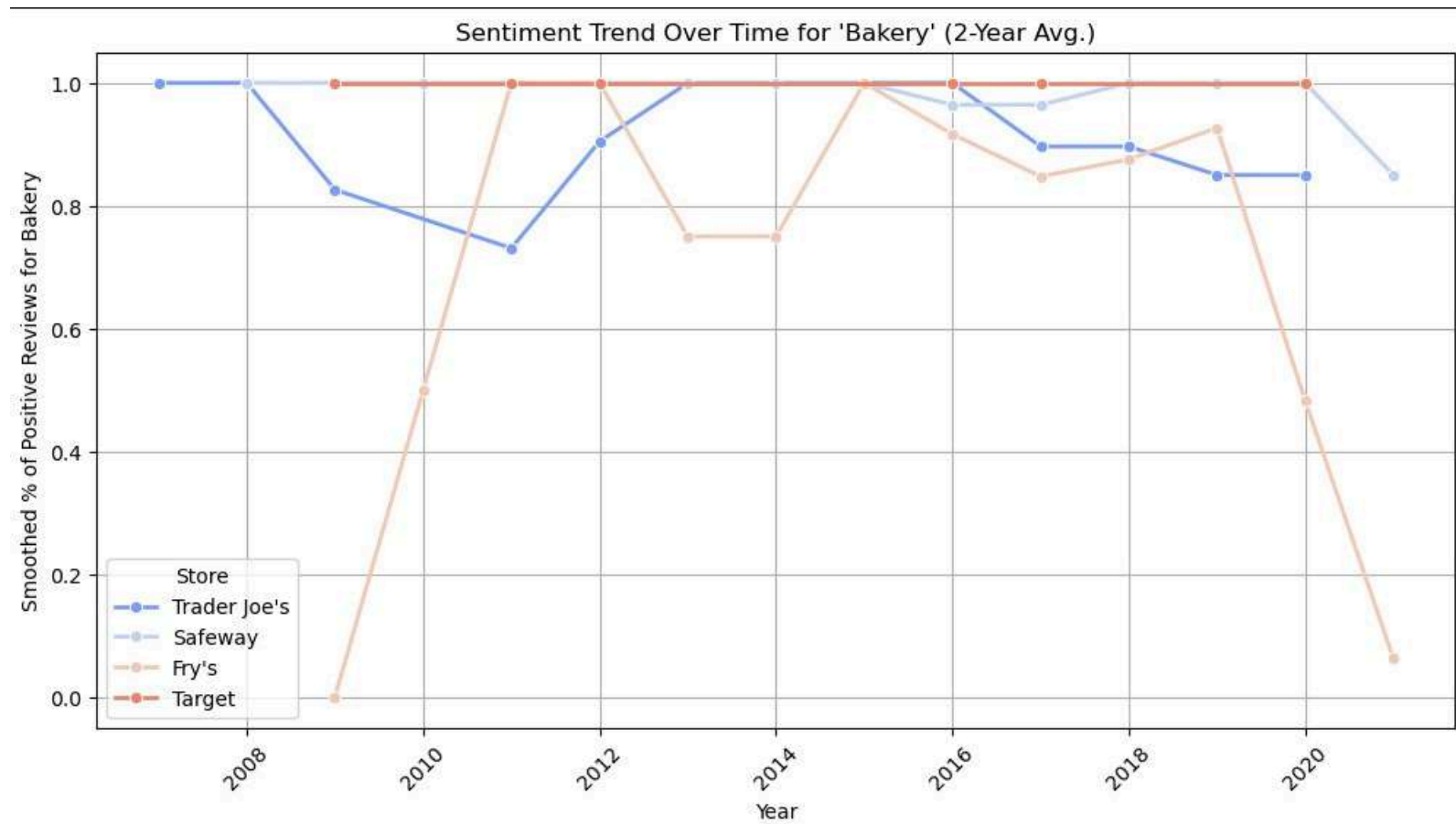Smoothed Business Stars Trend Over Time Across Stores (2-Year Avg.)

Key Takeaways

- **Trader Joe's** maintains high sentiment & ratings, showing strong customer satisfaction.
- **Fry's** has a declining trend, indicating growing dissatisfaction.
- **Target & Safeway** show moderate fluctuations but remain relatively stable.
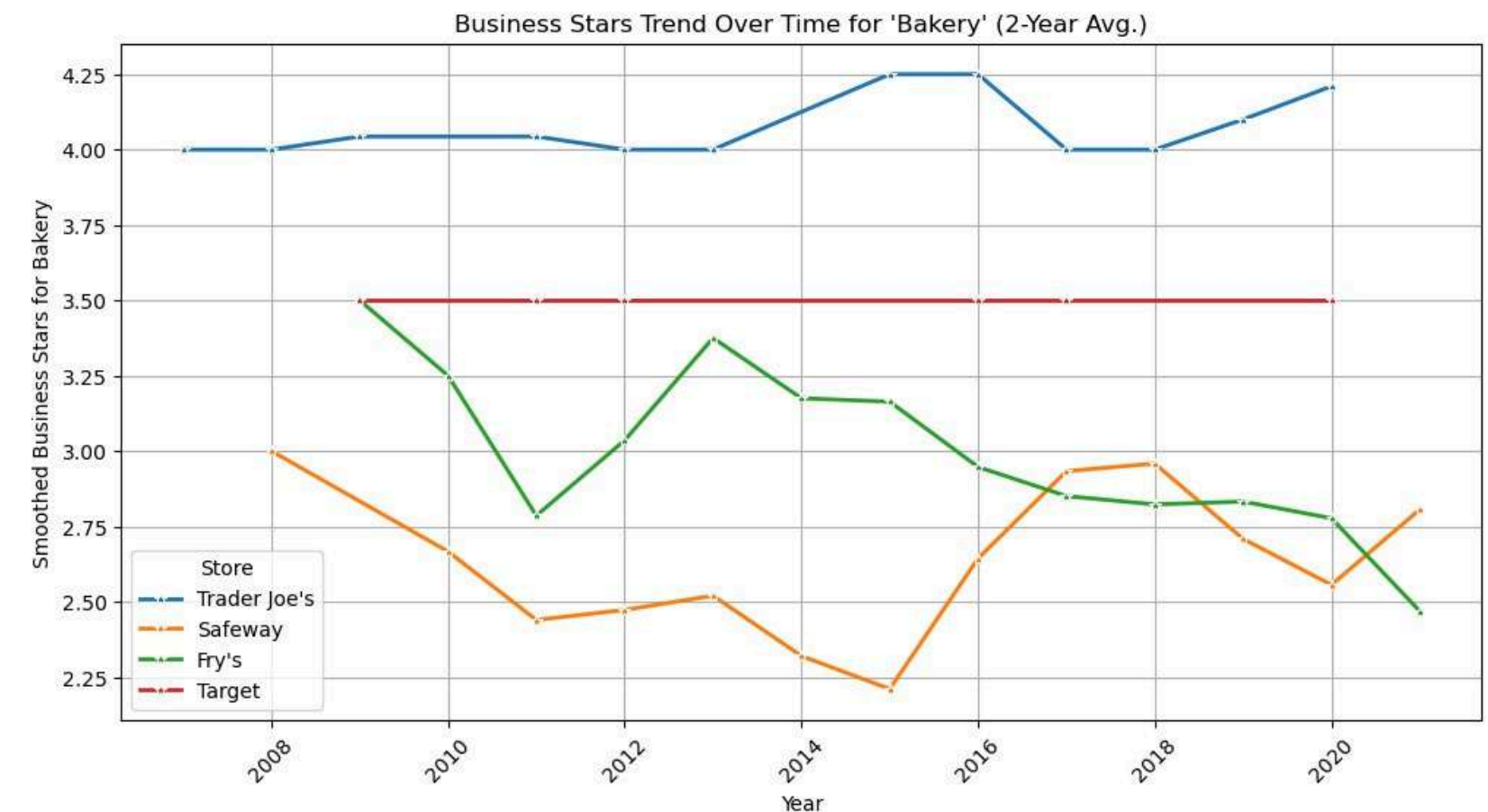
Business Insights

- **Trader Joe's** → Strong & consistent customer approval
- **Fry's** → Needs improvement in service & quality
- **Target** → Gradual improvement in sentiment & ratings
- **Safeway** → Stable ratings but fluctuating sentiment

Sentiment & Business ~ Bakery

Sentiment Trend Over Time for 'Bakery' (2-Year Avg.)

Business Stars Trend Over Time for 'Bakery' (2-Year Avg.)

Key Takeaways

- **Trader Joe's** → Consistently high positive sentiment & ratings, indicating strong customer loyalty.
- **Fry's & Safeway** → Declining sentiment and ratings, reflecting growing dissatisfaction with bakery products.
- **Target** → Stable star ratings but fluctuating sentiment, suggesting inconsistent customer experience

Business Insights

- **Trader Joe's** maintains quality & customer satisfaction in bakery items.
- **Fry's & Safeway** need improvements in bakery quality & service to regain customer trust.
- **Target's** bakery sentiment fluctuates, requiring further analysis of customer expectations.

Business Changes! Future Sentiment?

Why Predict Future Sentiment ?

- Enables grocery stores to anticipate customer reactions before making business decisions.

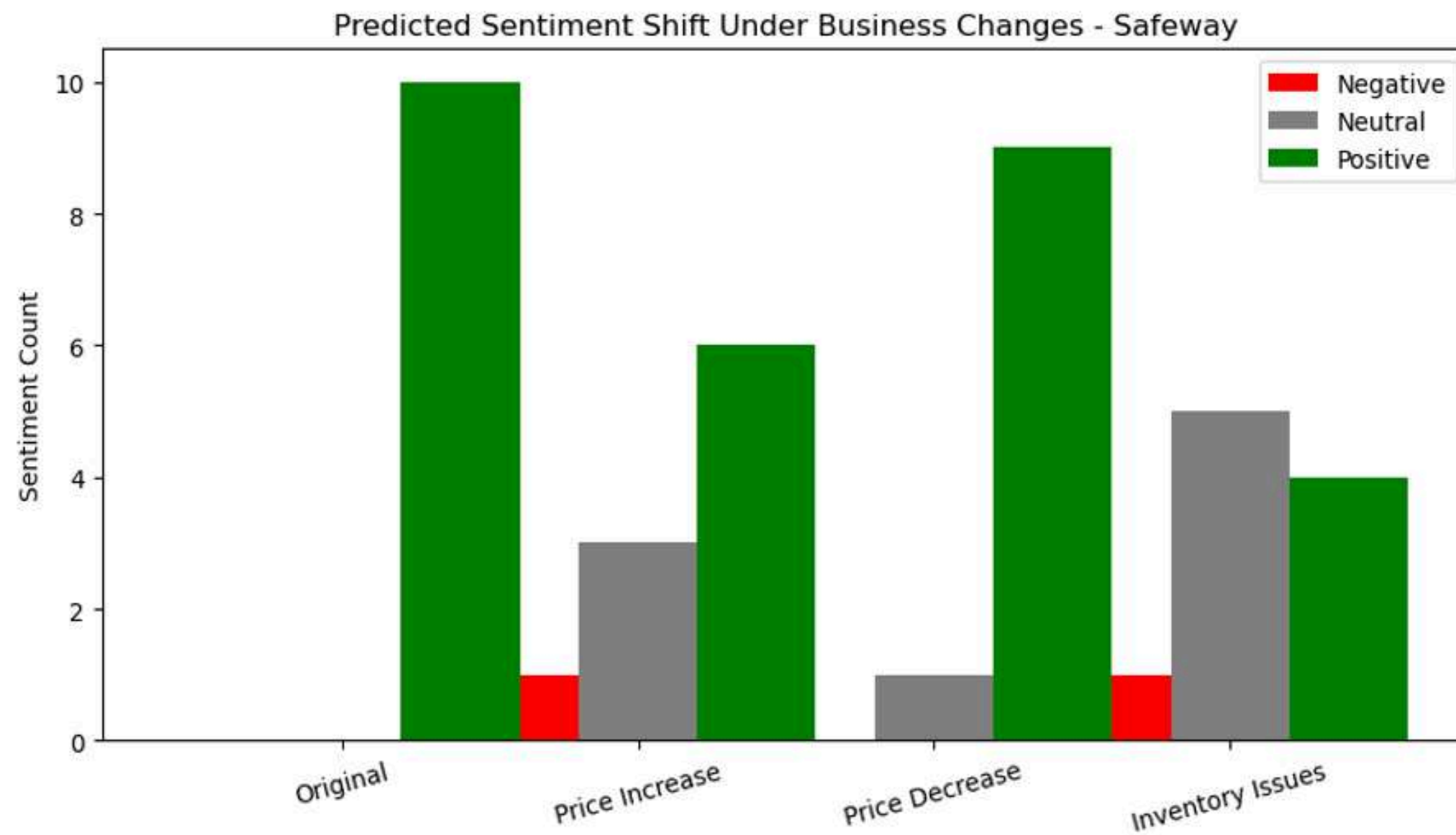- Provides actionable insights to improve pricing and inventory strategies.

Used **BiLSTM model** to predict sentiment shifts under these scenarios

Business Scenario	Findings from Sentiment
Price Increase	Lead to more negative reviews
Price Decrease	Improved sentiment but not drastically
Inventory Issues	Resulted in a noticeable rise in negative reviews

Changes! Future Sentiment?

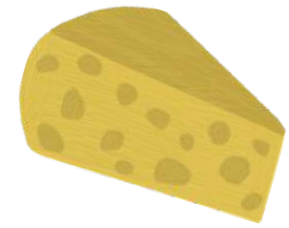Predicted Sentiment Shift Under Business Changes - Safeway

Key Observations

- **Price Increase:** Decrease in positive sentiment, rise in neutral & negative reviews.
- **Price Decrease:** Slight improvement in positive sentiment, minimal effect on neutral/negative.
- **Inventory Issues:** Significant drop in positive sentiment, sharp rise in negative reviews.

Takeaways

- Customers react negatively to price hikes & stock issues.
- Optimizing pricing & inventory can help maintain customer satisfaction.

Model Errors & Challenges

Review	Actual	Predicted
I love waiting 30 minutes in a checkout line.	Negative (Sarcasm)	Positive
The staff was friendly, but the prices were outrageous!	Negative	Neutral
Decent selection, but the aisles are always messy.	Neutral	Negative
Good variety of products, but customer service could be better.	Neutral	Positive

Vader Misclassified

Naïve Bayes Misclassified

LSTM Misclassified

RNN Misclassified

Misclassification

Some models struggled to detect neutrality, often classifying it as positive or negative

Understanding

Models like Vader & Naïve Bayes misinterpreted sarcastic comments

Long Reviews

RNN and LSTM-based models faced training difficulties with longer reviews

Business Insights

Recommendations

1. Brand Engagement & Sentiment Trends

- Trader Joe's and Target have strong customer loyalty, while Bashas' and Safeway struggle with engagement.
- Sentiment analysis shows fluctuations over time, with dips in 2017 and 2020, possibly due to economic or store policy changes.

2. Key Drivers of Positive & Negative Reviews

- Positive factors: Staff friendliness, store cleanliness, and product variety.
- Negative factors: Pricing concerns, long checkout times, and stock shortages.

3. Pricing & Shopping Behavior Trends

- Customers are highly price-sensitive, especially for fresh produce and specialty items.
- Growing demand for organic products, discounts, and online shopping options.
- A shift toward digital shopping and home delivery signals the need for improved e-commerce.
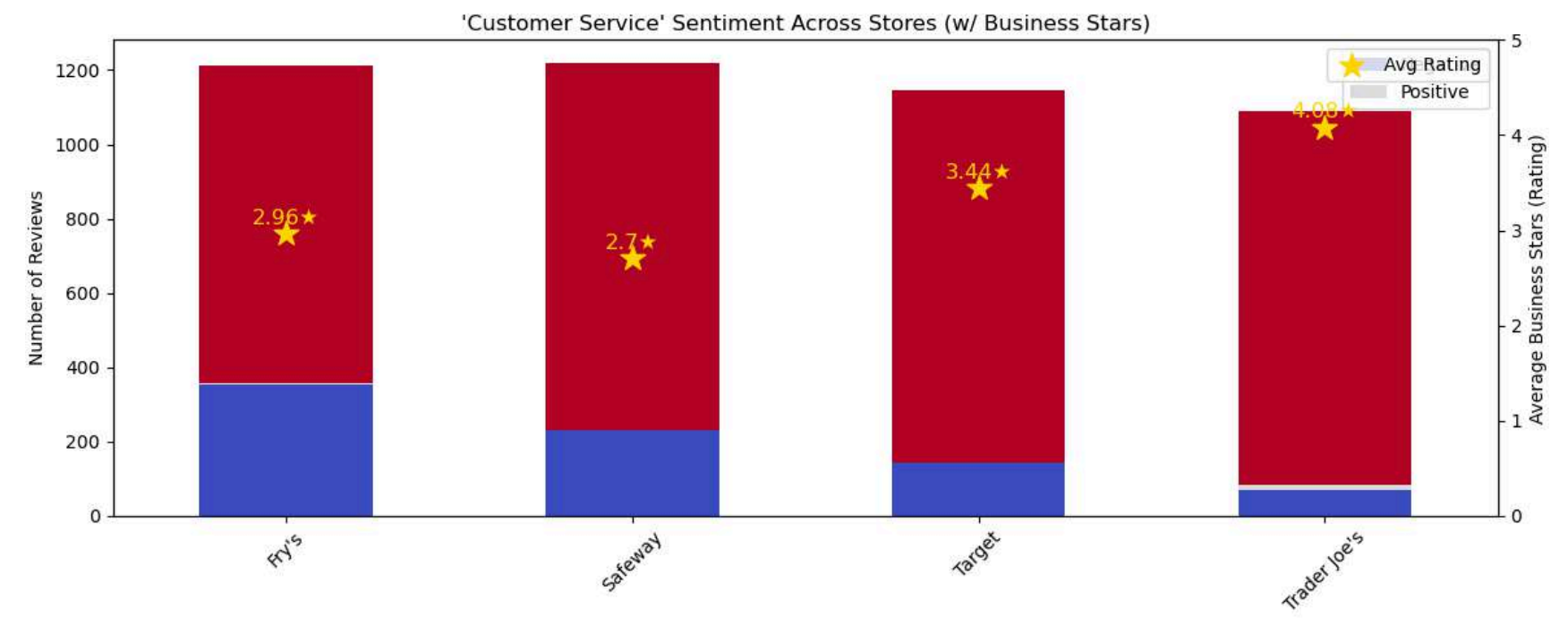
1. Improve Customer Service

- Train staff for better service & problem resolution.
- Use AI sentiment monitoring to detect issues early.

2. Optimize Pricing Strategies

- Track competitor pricing to stay competitive.
- Use dynamic pricing for frequently mentioned expensive products.
- Offer targeted discounts on high-demand items.

3. Enhance Store Efficiency

- Reduce checkout wait times with self-checkout expansion.
- Optimize store layouts for better navigation.
- Use predictive analytics to manage stock levels.

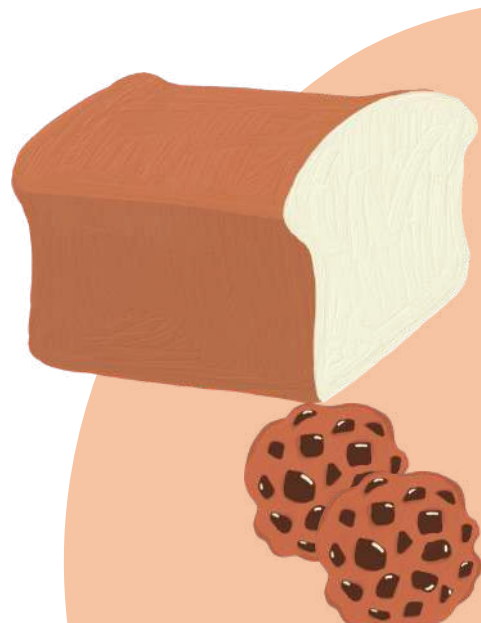'Customer Service' Sentiment Across Stores (w/ Business Stars)

Conclusion

- Our data-driven analysis of grocery store reviews highlights key areas impacting customer satisfaction and business performance.
- **Sentiment Analysis Trends:** Customer sentiment has fluctuated over time, influenced by pricing, service quality, and operational bottlenecks. Trader Joe's and Target maintain strong customer loyalty, while Bashas' and Safeway need strategic improvements.
- **Customer Experience Matters:** Friendly staff and clean stores drive positive engagement, while long checkout times and poor stock availability lead to negative feedback.
- **Pricing is a Major Concern:** Price-sensitive customers expect competitive pricing—high prices on essentials like fresh produce significantly impact sentiment.
- **Shift Towards Digital Shopping:** Demand for online orders, digital coupons, and seamless e-commerce is increasing, requiring improved technology and logistics.

Future Outlook:

- To stay competitive in an evolving retail landscape, grocery stores must adopt data-driven decision-making and AI-powered automation. By continuously analyzing customer sentiment and operational data, businesses can proactively enhance customer experience, optimize pricing, and drive long-term success.

Thank You !