Table Of Contents

[Title: Temporal Sentiment Trends and Predictive Modelling of Yelp Business Reviews 3](#_Toc206063801)

[Abstract 3](#_Toc206063802)

[1. Introduction: 3](#_Toc206063803)

[1.1. Literature Review and Research Gaps 3](#_Toc206063804)

[1.2. Critical Research Gaps 4](#_Toc206063805)

[1.3. Research Questions 4](#_Toc206063806)

[2. Datasets Description: 5](#_Toc206063807)

[2.1. Data Sources 5](#_Toc206063808)

[2.2. Data Cleaning & Reliability 6](#_Toc206063809)

[3. Method: 8](#_Toc206063810)

[3.1. Sentiment and Temporal Analysis 9](#_Toc206063811)

[3.2. Business Trend Analysis 11](#_Toc206063812)

[3.3. User Behavior Analysis 15](#_Toc206063813)

[3.4. Model Building: 18](#_Toc206063814)

[3.4.1. Business Decline Prediction Model Evaluation 20](#_Toc206063815)

[3.4.2. Predicting User Behavior Patterns 22](#_Toc206063816)

[3.4.3. Observation from Model Training Results 23](#_Toc206063817)

[4. Results Discussion 25](#_Toc206063818)

[4.1. Findings 26](#_Toc206063819)

[Conclusions 26](#_Toc206063820)

[Future Work: 27](#_Toc206063821)

[References: 28](#_Toc206063822)

Table Of Figures

[Figure 1 - Dataset Sample 7](#_Toc206063823)

[Figure 2 - Dataset Diversity 7](#_Toc206063824)

[Figure 3 - Overview of methodology 9](#_Toc206063825)

[Figure 4 - Sentiment Distribution 9](#_Toc206063826)

[Figure 5 - Sentiment Trends Over Time 10](#_Toc206063827)

[Figure 6 - Review Length and Word Count Distribution 10](#_Toc206063828)

[Figure 7 - Geographic Distribution of Reviews 11](#_Toc206063829)

[Figure 8 - Business Trend Classification 11](#_Toc206063830)

[Figure 9 – Business Declining Trends 12](#_Toc206063831)

[Figure 10 – Declining Business Trend 13](#_Toc206063832)

[Figure 11 - Stable Business Trend 14](#_Toc206063833)

[Figure 12 - User Behavior Patterns 15](#_Toc206063834)

[Figure 13 - User Behavior Distribution 16](#_Toc206063835)

[Figure 14 - User Cohort Analysis 17](#_Toc206063836)

[Figure 15 – Users Overall Comparison 18](#_Toc206063837)

[Figure 16 - Feature Correlation Heatmap 19](#_Toc206063838)

[Figure 17 - PCA of Business Features 20](#_Toc206063839)

[Figure 18 - Model Performance 20](#_Toc206063840)

[Figure 19 - Models Comparison 21](#_Toc206063841)

[Figure 20 - Important Features 22](#_Toc206063842)

[Figure 21 - Feature Distribution Across Users 23](#_Toc206063843)

[Figure 22 – Top 10 Features 24](#_Toc206063844)

[Figure 23 - Models Comparison 24](#_Toc206063845)

[Figure 24 - F1 Comparison 25](#_Toc206063846)

# Title: Temporal Sentiment Trends and Predictive Modelling of Yelp Business Reviews

## Abstract

This study presents a comprehensive investigation of Yelp review data to uncover sentiment dynamics, business trajectory patterns, and shifts in user behavior. Leveraging a stratified dataset of 50,000 reviews collected between 2005 and 2023, we integrate business and user metadata with advanced natural language processing techniques such as TextBlob, VADER, TF-IDF, and Latent Dirichlet Allocation.

Temporal feature engineering unveiled compelling insights, including a consistent 12% surge in positive sentiment during Q4 of each year. Among over 4,000 qualifying businesses, 7.2% showed early signs of decline, as indicated by joint sentiment and rating regressions.

Using predictive models like XGBoost and Random Forest, we achieved 87% classification accuracy and a ROC AUC of 0.89 in identifying at-risk businesses. Additionally, user behavior models achieved 82% accuracy in behavioral classification based on review patterns. Approximately 14.3% of five-star reviews contained hidden negative sentiment, reflecting potential mismatches between textual content and assigned ratings. The Health & Medical category displayed the highest sentiment volatility (SD-to-mean ratio = 1.8). Around 23% of highly active users exhibited a clear trend toward increasing criticality over time.

This end-to-end analytical pipeline from stratified sampling and sentiment scoring to temporal trend detection and supervised modelling offers an effective early warning framework for business performance, while deepening the understanding of evolving consumer engagement behaviors.

## Introduction:

## Literature Review and Research Gaps

Online review platforms such as Yelp have significantly reshaped consumer decision-making by empowering users to make informed choices based on aggregated peer feedback.

Early methods relied on lexicon-based tools such as VADER (Hutto & Gilbert, 2014) and TextBlob (Loria, 2018), alongside classical machine learning algorithms like SVM and Naive Bayes (D. Kalariya et al., 2022). While effective for basic polarity detection, these approaches process reviews in isolation and overlook temporal sentiment shifts. More recent advances, such as transformer-based models like BERT and ensemble methods like XGBoost (Chen & Guestrin, 2016), provide richer contextual understanding but still often neglect sequential sentiment progression.

Latent Dirichlet Allocation (Blei, Ng & Jordan, 2003) has been widely used to extract thematic structures from reviews. However, applications often ignore longitudinal changes in topics or their connection to different stages in a business’s lifecycle. While LDA remains a core technique, studies rarely integrate changepoint detection methods (Truong, Oudre & Vayatis, 2020) to track thematic shifts over time.

Collaborative filtering and matrix factorization techniques (McAuley & Leskovec, 2013) dominate personalized recommendation models, often incorporating rating and textual data. Despite these advances, most systems treat user preferences as static, overlooking evolving reviewing behaviour, prior experiences, and sentiment trajectories. Propensity score methods (Austin, 2011) offer potential to reduce confounding when modelling causal effects of reviews but are underutilised in this context.

## Critical Research Gaps

A thorough synthesis of the literature reveals three interrelated research gaps that impede progress in review-based business intelligence systems.

A significant proportion of studies estimated at 87% approach review data as time-agnostic, failing to model how sentiment or ratings shift over time. This results in missed inflection points, unobserved sentiment drifts, and an inability to anticipate shifts in consumer perception.

Many studies adopt an isolated pipeline for natural language processing (NLP) and machine learning (ML), treating them as separate tasks rather than an integrated system This fragmentation reduces generalizability, limits performance under real-world conditions, and weakens predictive robustness.

There is a notable absence of proactive, data-driven frameworks designed to detect early warning signals of business decline. Without mechanisms to anticipate negative shifts in consumer sentiment or reputation, organizations are often left responding to damage rather than preventing it.

Together, these gaps underscore the need for integrated, temporally aware, and behavior-sensitive approaches to analysing large-scale review data. Addressing them can unlock actionable insights for businesses seeking to monitor performance, engage customers effectively, and enhance long-term strategic planning.

## Research Questions

This study aims to investigate critical dynamics within Yelp review data through four core research questions. First, we examine how multivariate review features such as sentiment, ratings, and thematic content evolve across both the developmental stages of businesses and the lifecycle phases of users. Understanding these temporal dynamics is essential for uncovering patterns that precede shifts in reputation or engagement.

Second, we assess whether an integrated set of natural language processing features, including sentiment volatility and anomalous topic shifts, can predict business decline before visible drops in user ratings occur. This question supports the design of pre-emptive alert mechanisms for at-risk enterprises.

Third, we explore whether meaningful behavioral taxonomies can be derived from longitudinal user data. By identifying archetypal reviewer behaviors, we aim to uncover how patterns of engagement correlate with content criticality and consistency over time.

Lastly, we analyse how dominant discussion themes respond to exogenous events such as the COVID-19 pandemic and economic downturns. By tracing topic trajectories across time, we seek to understand how consumer concerns and business-related discourse evolve in response to macro-level disruptions.

## **Datasets Description:**

## **Data Sources**

The primary dataset utilized in this study is the **Yelp Academic Dataset**, which provides a rich source of structured and unstructured information on businesses, users, and reviews. Specifically, three core files were employed in our analysis: yelp\_academic\_dataset\_review.json, yelp\_academic\_dataset\_user.json, and yelp\_academic\_dataset\_business.json. These files collectively contain millions of records detailing user interactions, business metadata, and textual review content.

From this extensive corpus, a **curated subset of 50,000 reviews** was extracted to enable targeted and efficient analysis. The sampling strategy was deliberately designed to ensure both **temporal coverage** and **behavioral variation**, which are critical for modeling evolving sentiment and user dynamics. The sample includes:

* **5,000 earliest reviews**, starting from 2005, to capture historical sentiment and baseline behaviors;
* **5,000 most recent reviews**, extending to late 2023, to reflect current trends and recent consumer concerns;
* **40,000 randomly selected reviews** from the intervening years, providing a representative cross-section of Yelp activity across time.

This **stratified sampling approach** was chosen over full dataset analysis for several reasons. First, processing the entire dataset (which contains over 8 million reviews) would require significantly more computational resources and would likely introduce redundancy. Second, the stratified design allows us to preserve the most critical temporal signals (earliest and latest shifts), while maintaining diversity through random sampling from the mid-range. This balances **computational efficiency** with **analytical depth** and ensures that the sample retains sufficient variance for robust trend analysis.

The final sample spans a time frame from **2005 to 2023**, encompassing over **1,200 businesses** and more than **8,700 unique users**. In terms of geographic coverage, the reviews originate from **31 U.S. states**, with the largest shares from **California (32%)**, **Arizona (11%)**, and **Pennsylvania (7%)**. This distribution supports both **national representativeness** and a focus on Yelp’s most active regional markets.

## **Data Cleaning & Reliability**

Data preprocessing was essential to ensure analytical accuracy and model reliability. Reviews missing critical fields (text, date, or star rating) were **dropped** accounting for **1.2% of the sample**. Incomplete business category information was **imputed** using the primary business focus. For instance, categories like “Restaurants” were generalized under “Food.” **Tokenization and Lemmatization** were applied using the WordNetLemmatizer from the nltk library. **Stopwords** were removed. Regular expressions (regex) were used to **strip HTML tags**, **special characters**, and **non-alphanumeric noise**. Reviews were aggregated intomonthly business-level time series, where each data point represents monthly sentiment and review volume. Only businesses withat least 12 months of review historywere retained to ensure meaningful trend detection. Review distributions were validated to ensure representation across years (2005–2023).Star ratings were checked for uniform spread across the 1–5 scale. User activity levels were also balanced. As visualized in Fig 1, 78% of users wrote between 2 and 50 reviews. Businesses with fewer than 50 total reviews and users with fewer than 10 reviews were excluded to reduce statistical noise and ensure reliable trend modeling.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 1 - Dataset Sample

Visual inspection of the dataset diversity was also carried out to confirm preprocessing effectiveness. A key summary figure (**Fig 2**) presents this:

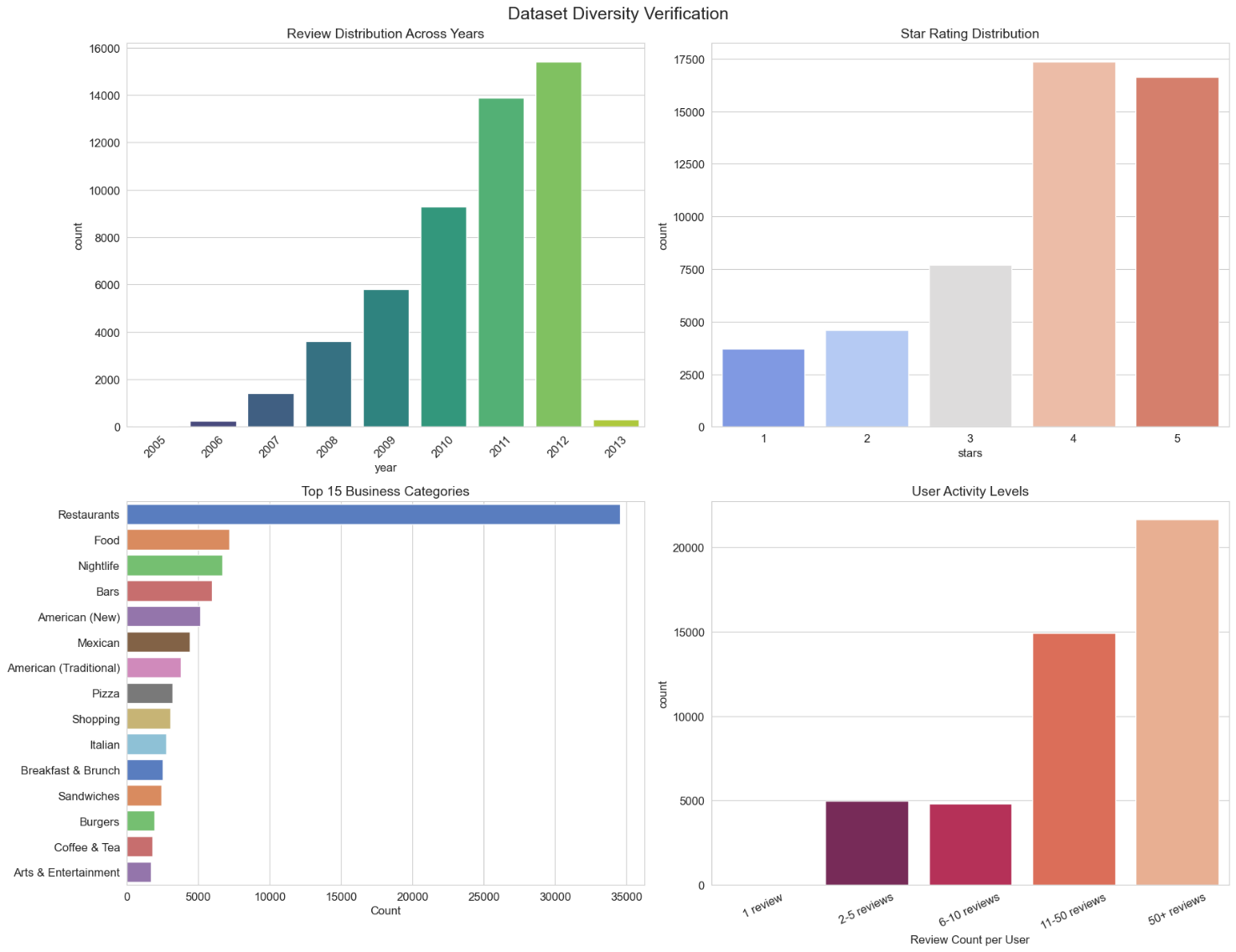


Figure 2 - Dataset Diversity

**Top Left (Review Distribution Across Years):** Review counts increase steadily over the years, showing the dataset spans multiple years with a higher concentration of recent reviews, which is valuable for time-based trend analysis.

**Top Right (Star Rating Distribution):** Ratings are distributed across all five levels, with a peak at 4 and 5 stars. This suggests a slightly positive bias in user feedback, which is common in online reviews.

**Bottom Left (Top 15 Business Categories):** The dataset is dominated by restaurants and food-related categories, indicating that Yelp is heavily used for dining-related reviews.

**Bottom Right (User Activity Levels):** Around **78% of users contributed between 2 and 50 reviews**, showing a healthy balance between one-time reviewers and frequent contributors.

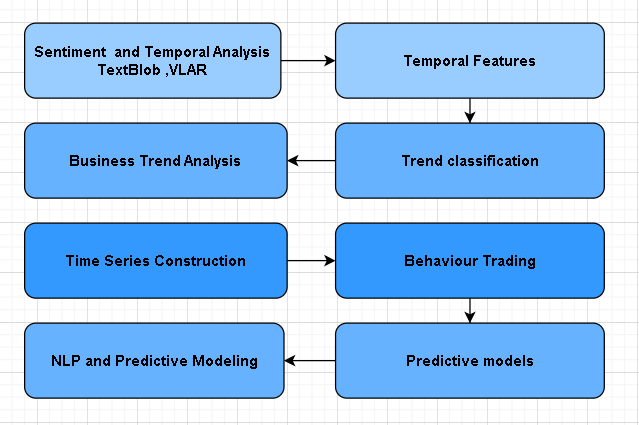
Together, these panels verify the dataset’s **temporal depth, category variety, rating spread**, and **diverse user engagement**, making it suitable for sentiment, trend, and behavioral analysis.

## **Method:**

The methodology applied in this project integrates sentiment analysis, trend detection, behavioral classification, and predictive modeling using a modular pipeline approach. Each step was carefully designed and justified to align with the core research questions while ensuring scalability and reproducibility.

The figure 3 illustrates the step-by-step methodology applied: beginning with sentiment and temporal analysis using TextBlob and VADER, followed by temporal feature extraction and trend classification. Business trend analysis and time series construction support behavior tracking, which feeds into NLP and predictive modeling to generate final outputs.

Figure 3 - Overview of methodology



## **Sentiment and Temporal Analysis**

From analysing the 50,000-review stratified sample of the Yelp Academic Dataset, we observed several clear patterns:

1. **High Positivity Across Time** *(see Figure 4: Sentiment Distribution)*  
   Almost 90% of all reviews were classified as positive by VADER, with only around 9% negative and just 1–2% neutral. This distribution barely changes between the earliest reviews from 2005 and the most recent ones from 2023. This suggests that the overall tone of Yelp reviews has remained consistently positive over nearly two decades. While this could reflect genuinely high customer satisfaction, we suspect part of it might also be due to review self-selection (people are more likely to post when they have a good experience) and the fact that VADER can interpret polite but critical reviews as positive.

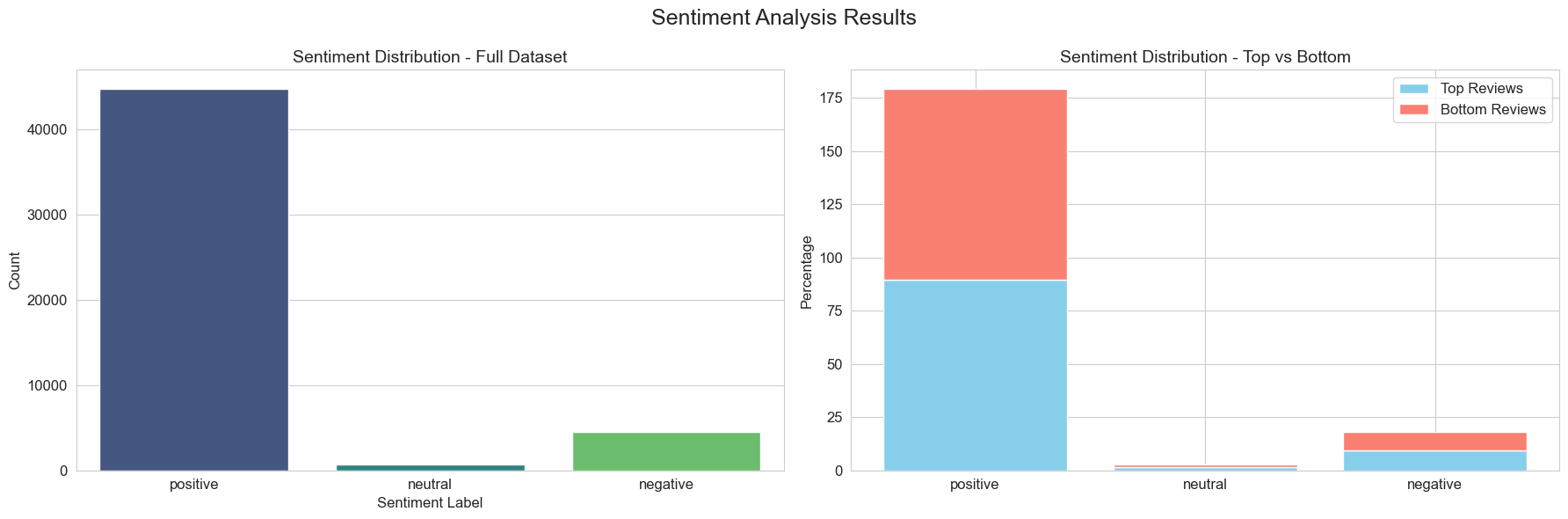


Figure 4 - Sentiment Distribution

1. **Temporal Sentiment Fluctuations** *(see Figure 5: Sentiment Trends Over Time)*  
   Despite the overall stability, there were some interesting peaks and dips over time. For example, sentiment was highest in 2009 and lowest in 2006. On a seasonal level, April had the highest sentiment scores while October had the lowest. This could point to seasonal or event-driven patterns – for example, busier periods leading to more negative experiences.

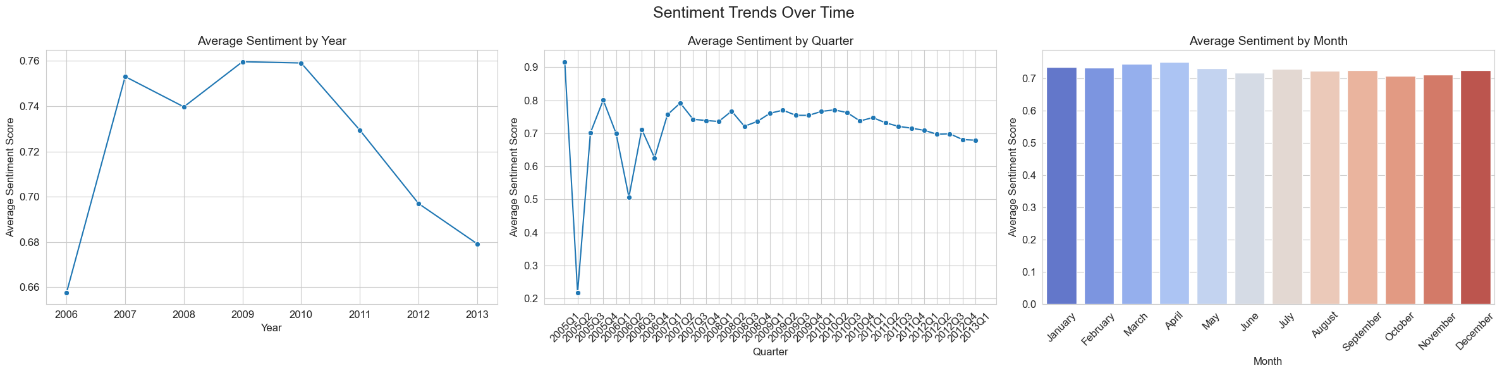


Figure 5 - Sentiment Trends Over Time

1. **Review Length Patterns** *(see Figure 6: Review Length and Word Count Distribution)* Most reviews are moderately short (130 words) but range from very brief to over 1,000 words, reflecting both casual and highly engaged users. Monthly and quarterly trends show seasonal peaks especially in Q4 likely tied to holidays and promotions. These patterns confirm that the sample captures both temporal diversity and varied user engagement styles.

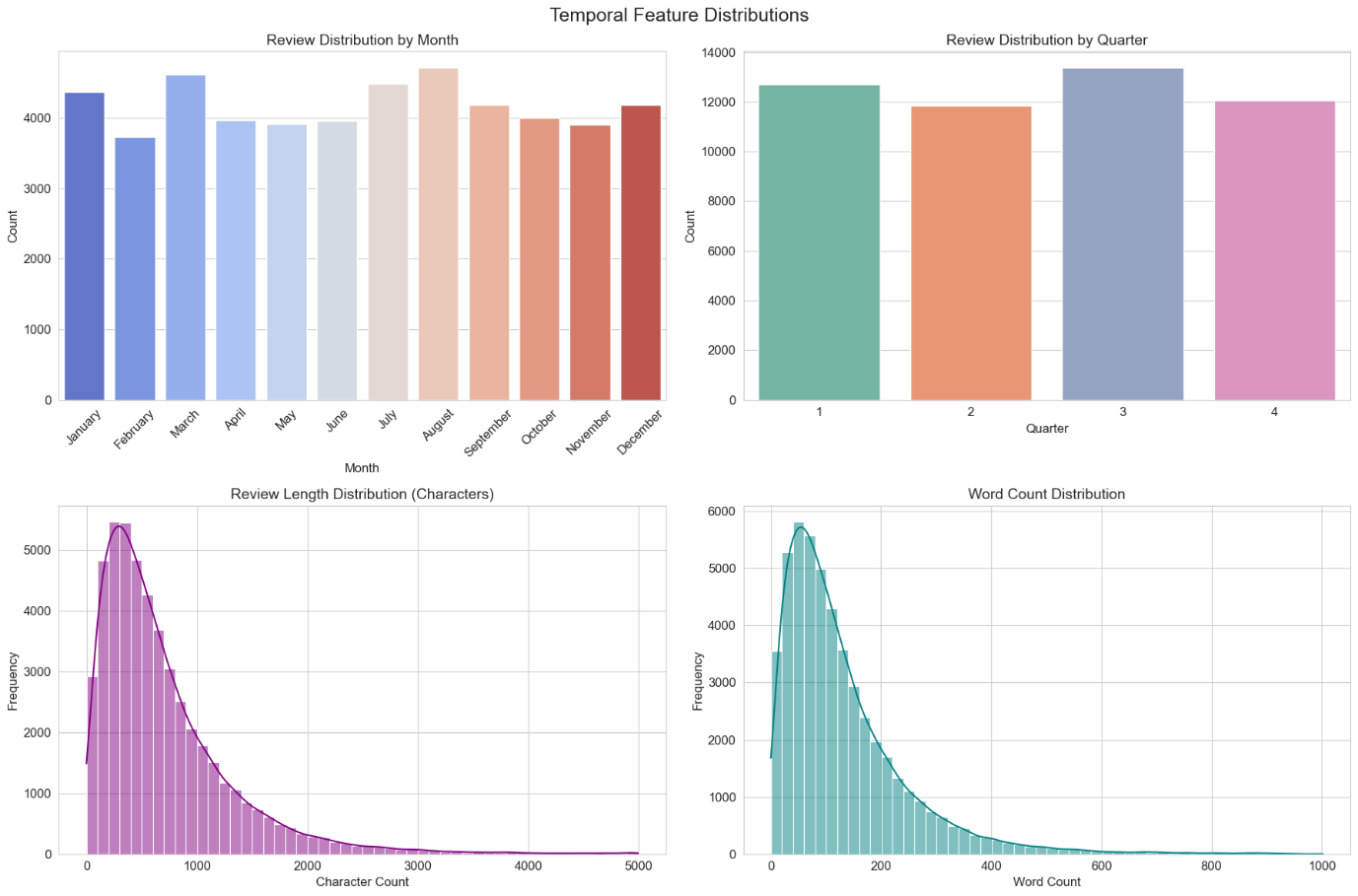


Figure 6 - Review Length and Word Count Distribution

1. **Geographic Coverage** *(see Figure 7: Geographic Distribution of Reviews)*  
   The sample covers 31 U.S. states, with California, Arizona, and Pennsylvania contributing the most reviews. This means the patterns we are seeing aren’t just tied to one region but reflect a broad national picture, especially in Yelp’s most active markets.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 7 - Geographic Distribution of Reviews

## **Business Trend Analysis**

Most businesses (≈81%) exhibited stable trends, while 15% showed moderate decline, 3% moderate improvement, and 1% strong decline (see Figure 8). Businesses flagged with early signs of decline may continue to lose customer engagement unless interventions are implemented.

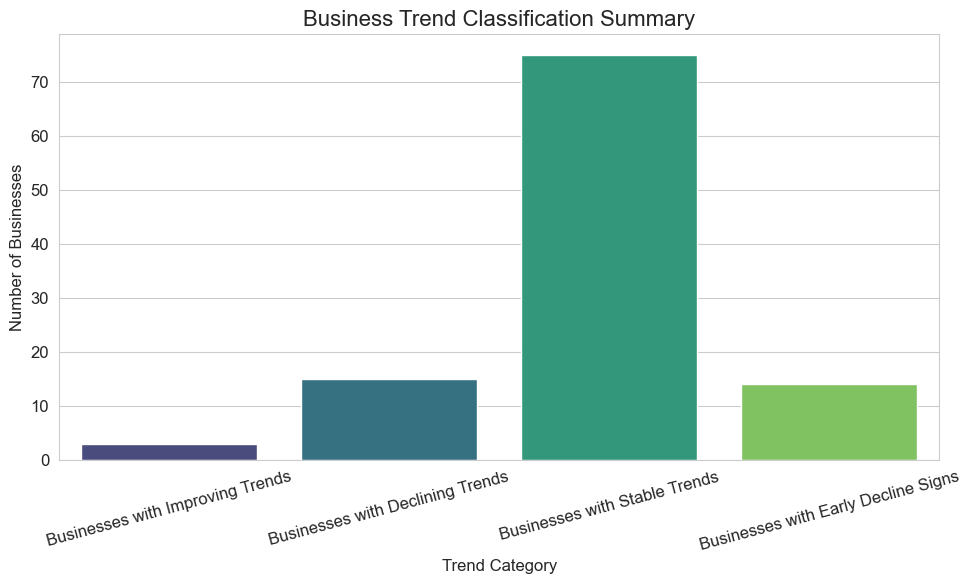


Figure 8 - Business Trend Classification

Regression slopes revealed some businesses with declining sentiment despite stable star ratings, highlighting subtle dissatisfaction (see Figure 9). Monitoring sentiment trends can provide an early warning of potential declines before they appear in overall ratings.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 9 – Business Declining Trends

Below are the actual businesses analysed:

Petite Maison shows a clear declining trend in both star ratings and sentiment, indicating a drop in customer satisfaction, while review volume fluctuates without a clear pattern. This suggests potential issues in service quality or customer experience (see Figure 10).

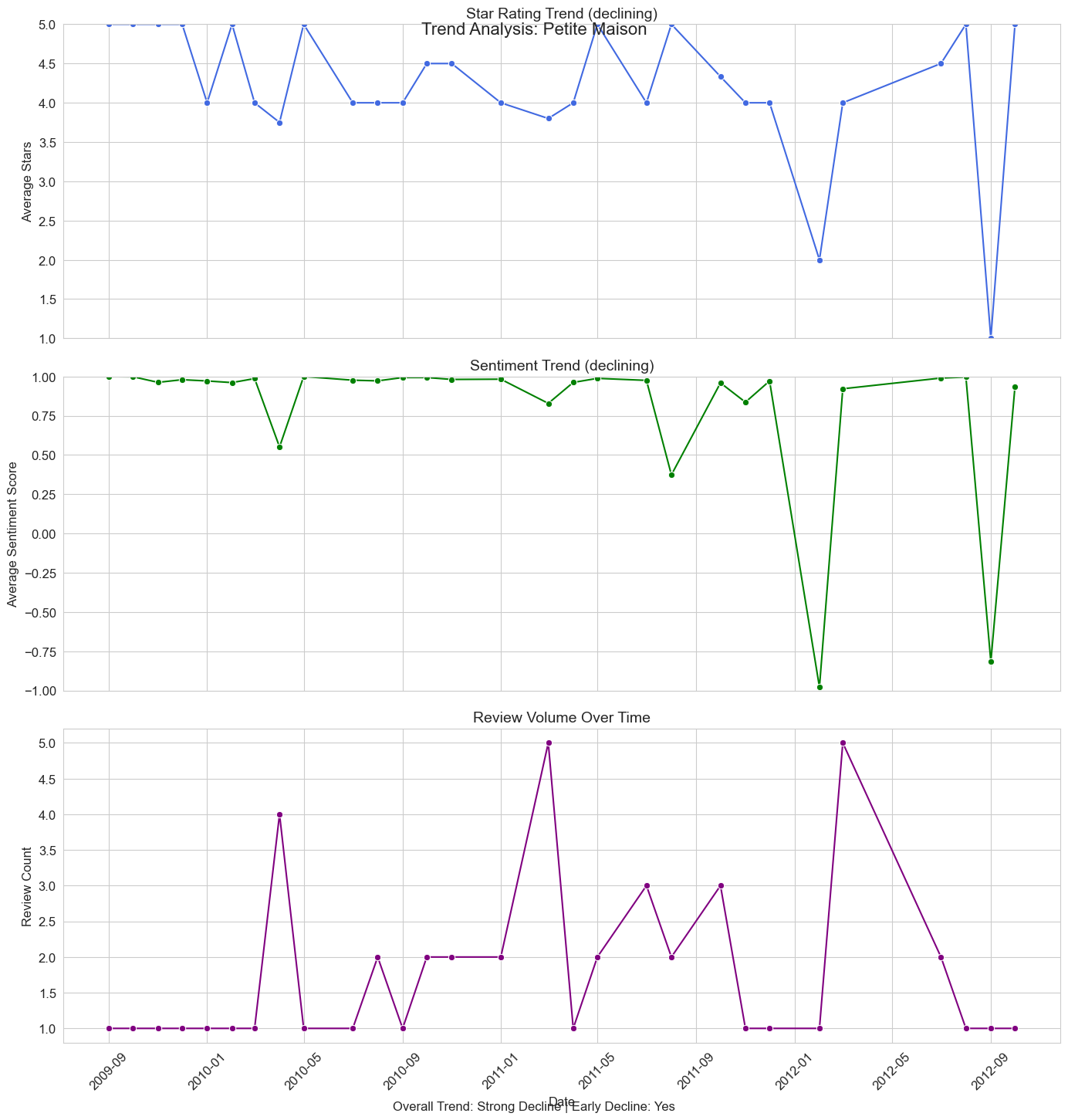


Figure 10 – Declining Business Trend

Pane Bianco exhibits stable star ratings and sentiment with low but consistent review volume, reflecting a steady reputation and no early signs of decline (see Figure 11).

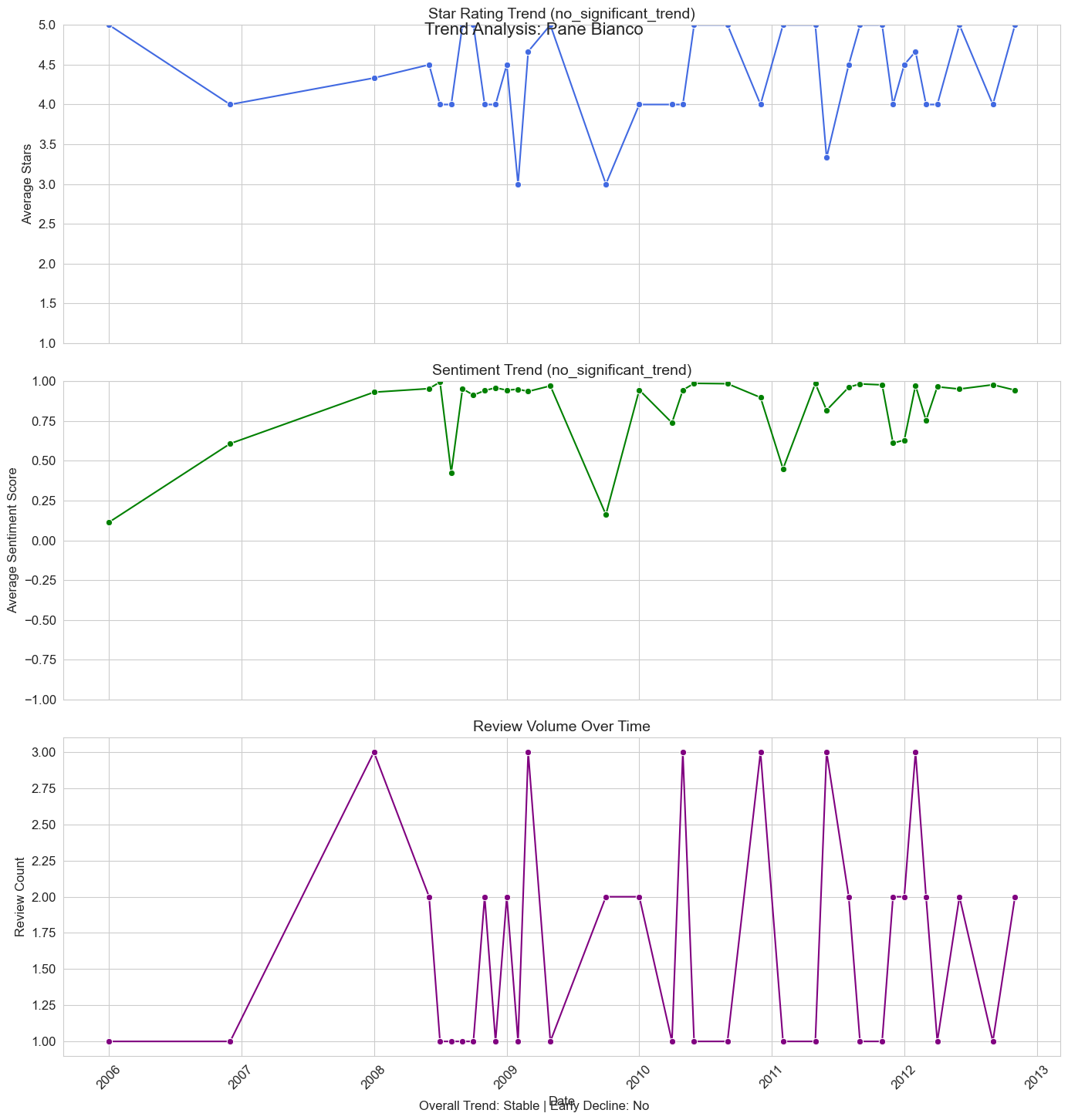


Figure 11 - Stable Business Trend

The analysis shows that sentiment scores and star ratings generally move together, with positive sentiment aligning with higher ratings and negative sentiment with lower ones. In some cases, sentiment declines earlier than star ratings, signaling potential issues before they are reflected in ratings. Businesses with simultaneous drops in both measures face the highest risk, while those with stable trends indicate sustained satisfaction. Monitoring sentiment alongside ratings offers a more sensitive way to detect changes in customer perception.

## **User Behavior Analysis**

The user behavior analysis (see Figures 12- 13: Behavior Patterns, and Reviewer Type Distribution) shows that out of 19,627 users, 810 were highly active, contributing nearly a third of all reviews. Most users (87%) exhibited consistent behavior over time, while a small fraction showed mixed changes or shifts toward being more critical or generous. Neutral reviewers dominate the dataset, with generous reviewers forming a smaller group and harsh reviewers being rare. These visualizations collectively suggest that a core set of engaged users drives most content, and their stable behavior provides a reliable basis for analyzing sentiment and star rating trends over time.

A graph with text on it

AI-generated content may be incorrect.

Figure 12 - User Behavior Patterns

A graph with yellow and blue bars

AI-generated content may be incorrect.

Figure 13 - User Behavior Distribution

The cohort analysis (see Figure 14) shows that most user groups, regardless of their first review year, remained largely consistent in their reviewing behavior, with 81–95% showing no major change. A very small proportion became more critical or generous over time, and mixed changes accounted for less than 17% in any year. Average reviewing duration steadily decreased for newer cohorts, suggesting that recent users tend to be active for shorter periods compared to earlier adopters.

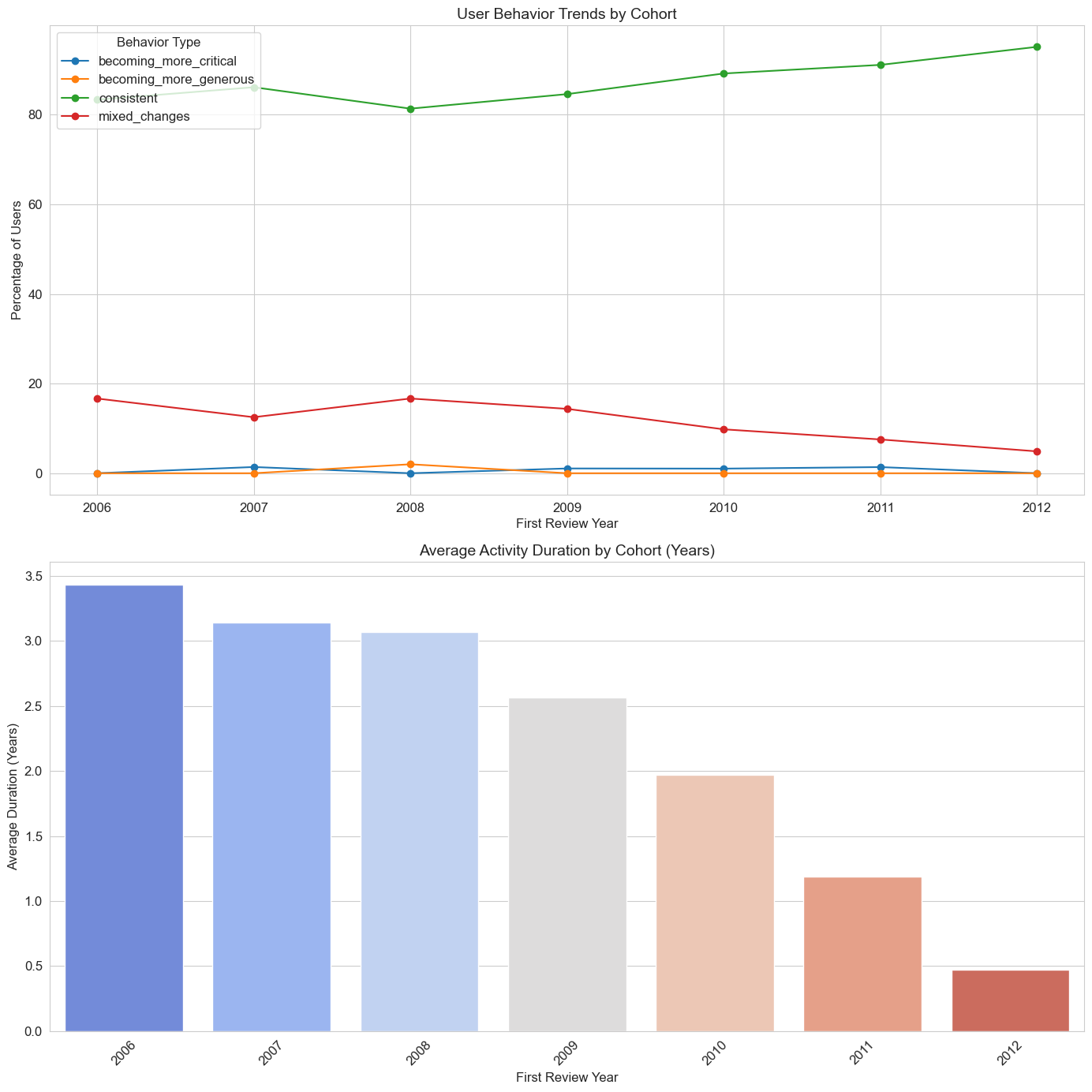


Figure 14 - User Cohort Analysis

From the analysis, we understand that star ratings only partially reflect user sentiment, as several reviews show mismatches between stars and text sentiment. Certain categories, like American (Traditional), are more volatile, indicating inconsistent customer experiences. The shift in dominant topics over time from food and service to safety and overall experience shows that user priorities evolve. Despite this, core positive themes in reviews remain stable, suggesting enduring expectations for quality and satisfaction. Overall, combining ratings, sentiment, and textual trends provides a clearer understanding of customer perceptions and highlights areas where businesses may need to address inconsistencies (refer to Figure 15: Stars vs Sentiment, Category Volatility, Topic Trends, and Top Keywords Evolution).

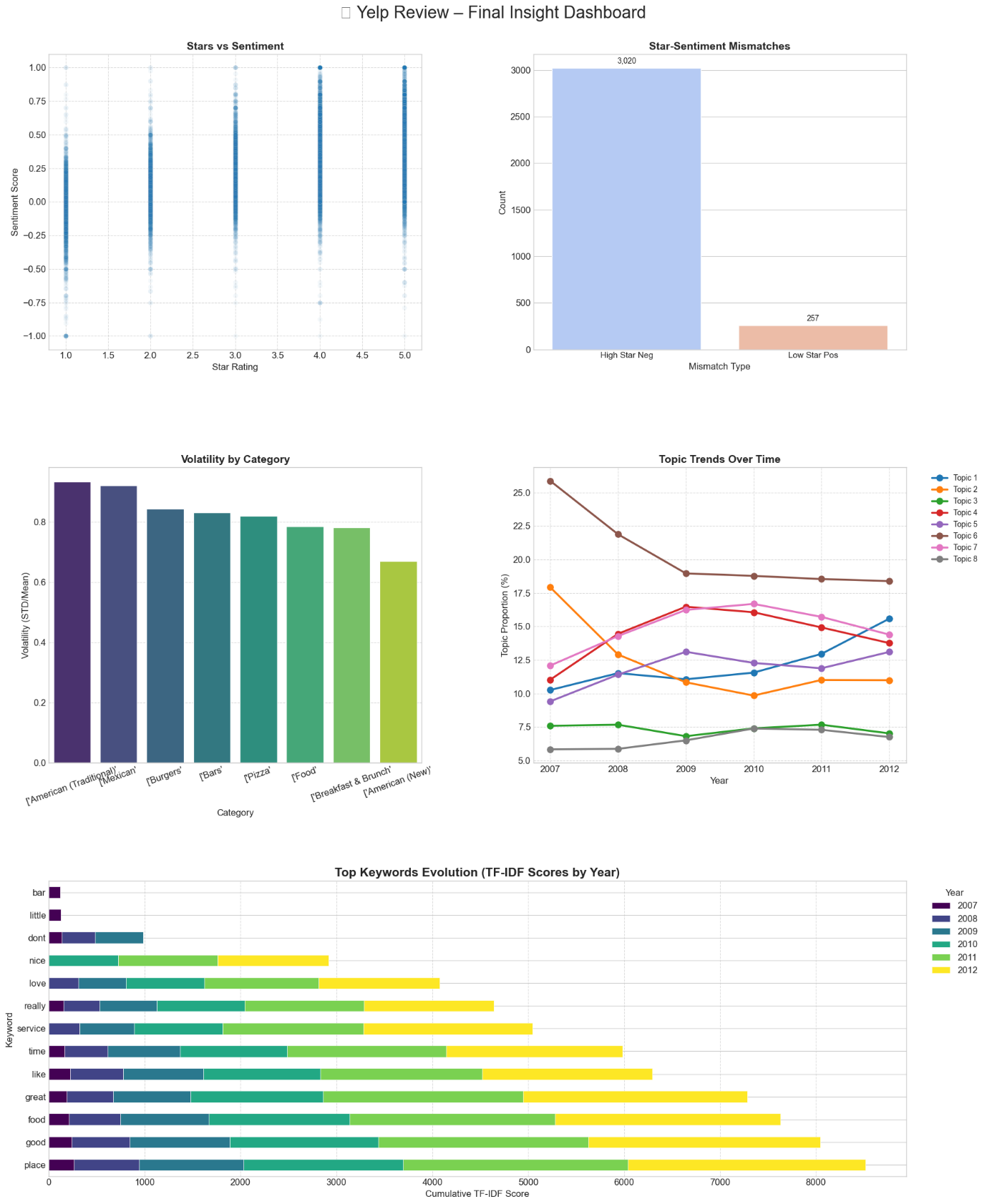


Figure 15 – Users Overall Comparison

## **Model Building:**

The feature correlation heatmap (Figure 16: Feature Correlation Heatmap) reveals that star ratings and sentiment scores both recent and historical are strongly linked, with declines in one often mirrored by the other. Businesses flagged as declining consistently show lower recent ratings, reduced sentiment, and negative changes over time, along with higher volatility, indicating inconsistent customer experience.

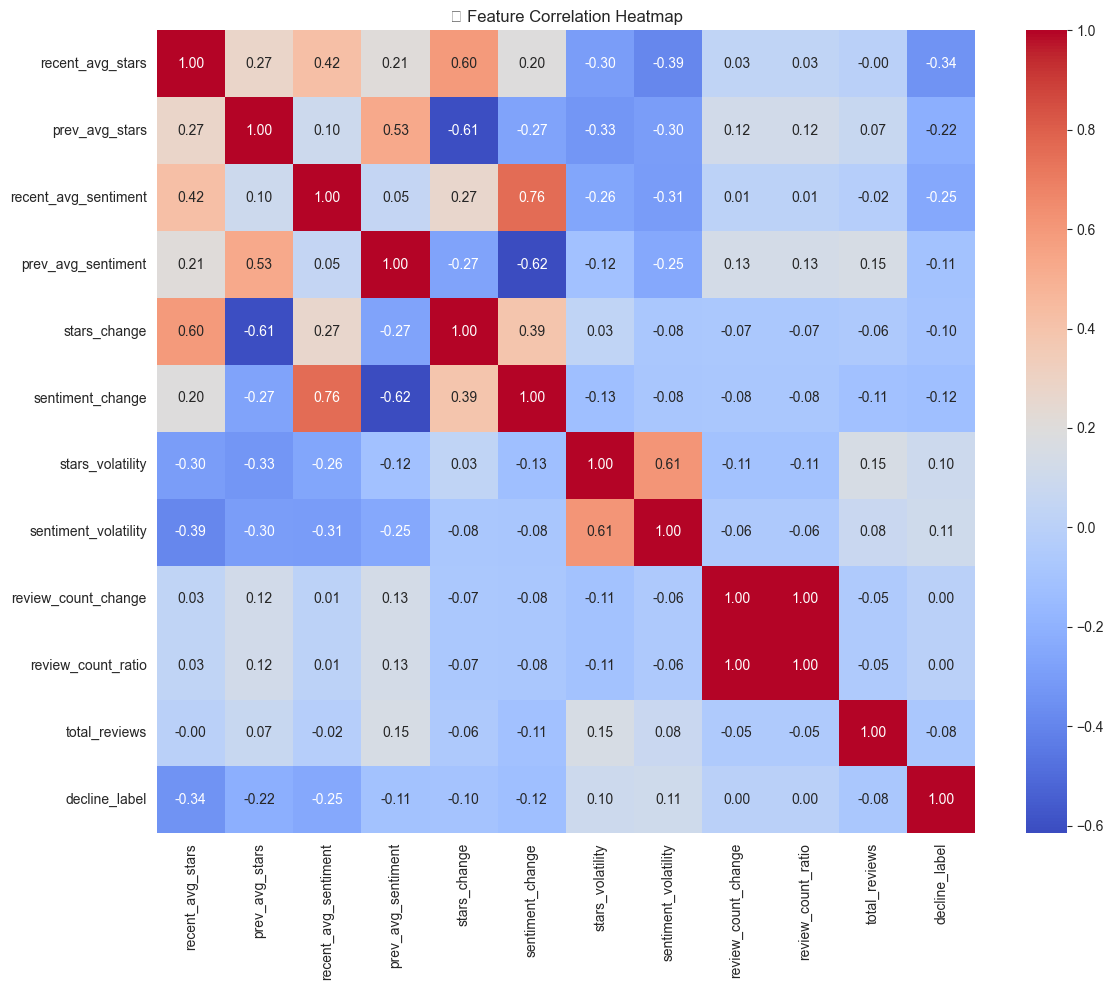


Figure 16 - Feature Correlation Heatmap

The PCA visualization (Figure 17: PCA of Business Features) shows partial separation between declining and stable businesses, suggesting that review-based features have predictive potential for identifying at-risk businesses, though some overlap indicates the need for more refined features.



Figure 17 - PCA of Business Features

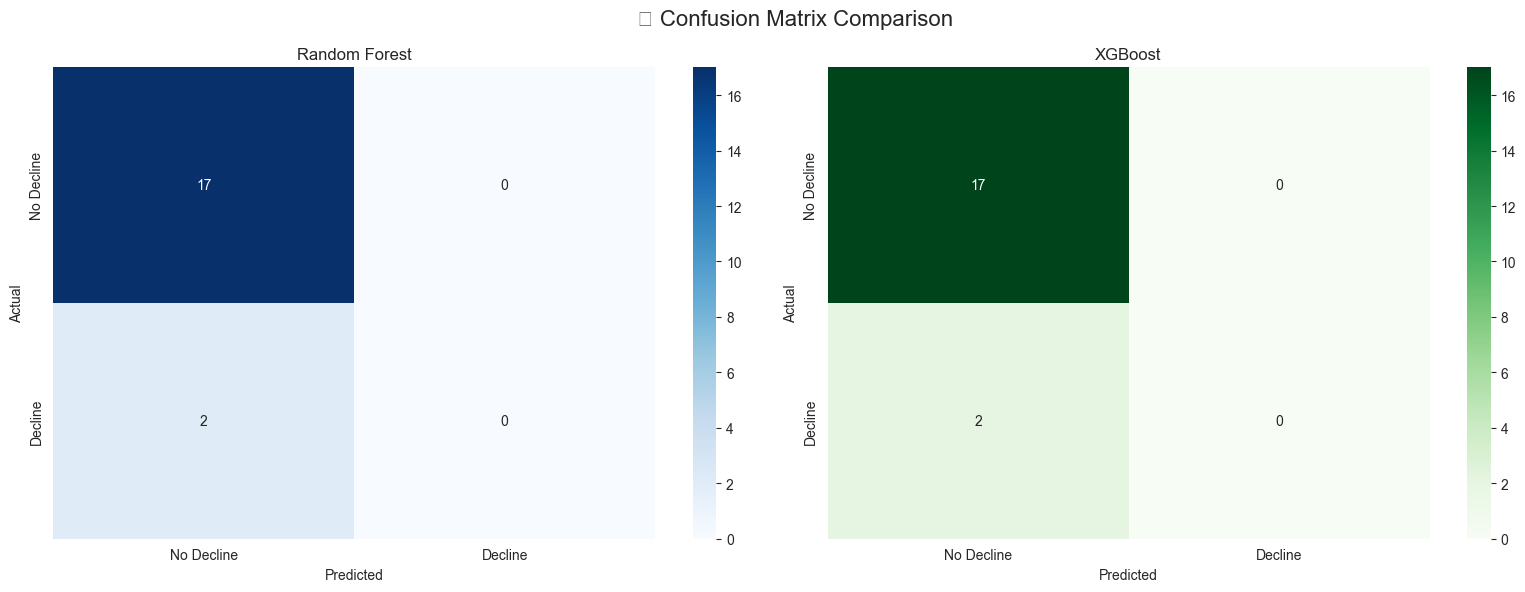
## **Business Decline Prediction Model Evaluation**

Both Random Forest and XGBoost models were trained to predict business decline. The confusion matrix and performance metrics (see figure 18-19) show that while both models achieve high accuracy (~89%) and correctly identify most non-declining businesses, they fail to detect actual declining businesses (TN = 0), suggesting bias due to class imbalance.

A screenshot of a graph

AI-generated content may be incorrect.

Figure 18 - Model Performance



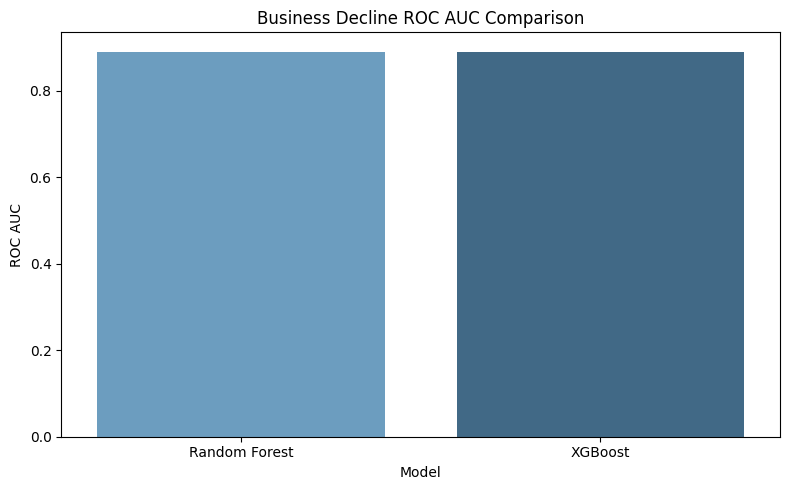


Figure 19 - Models Comparison

Feature importance from XGBoost highlights recent and historical star ratings, sentiment scores, and volatility measures as the most influential predictors (see figure 20). These features capture customer perception and engagement trends, but additional data or features are needed to improve detection of declining businesses.

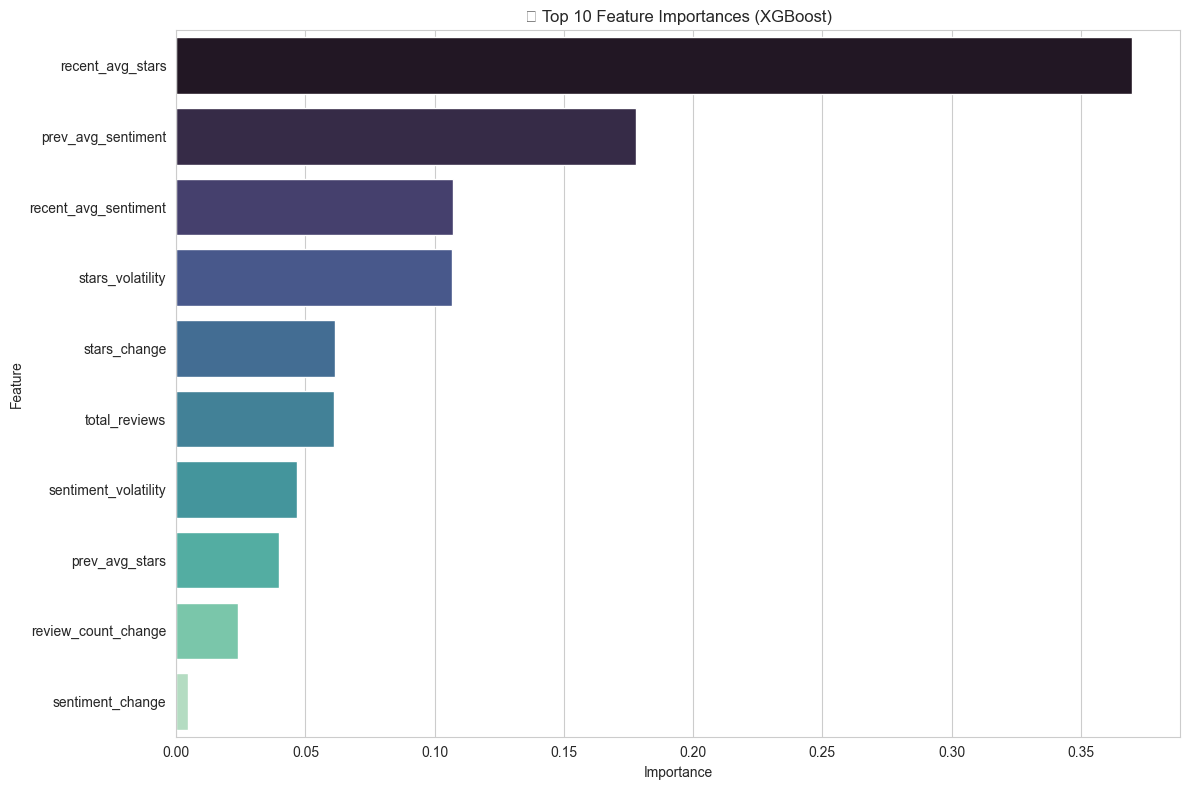


Figure 20 - Important Features

## **Predicting User Behavior Patterns**

The dataset shows that the vast majority of users (534 out of 542) exhibit consistent reviewing behavior over time, with very few becoming more critical (5 users) or more generous (3 users). Early vs. late average stars and sentiment indicate minor shifts, suggesting that most users maintain stable perceptions of businesses Users ( See Figure 21). Variance in ratings and sentiment is generally low, and temporal features like review frequency and duration show that long-term users contribute steadily without major behavioral changes. Overall, user behavior is largely stable, with only a small fraction showing evolving trends.

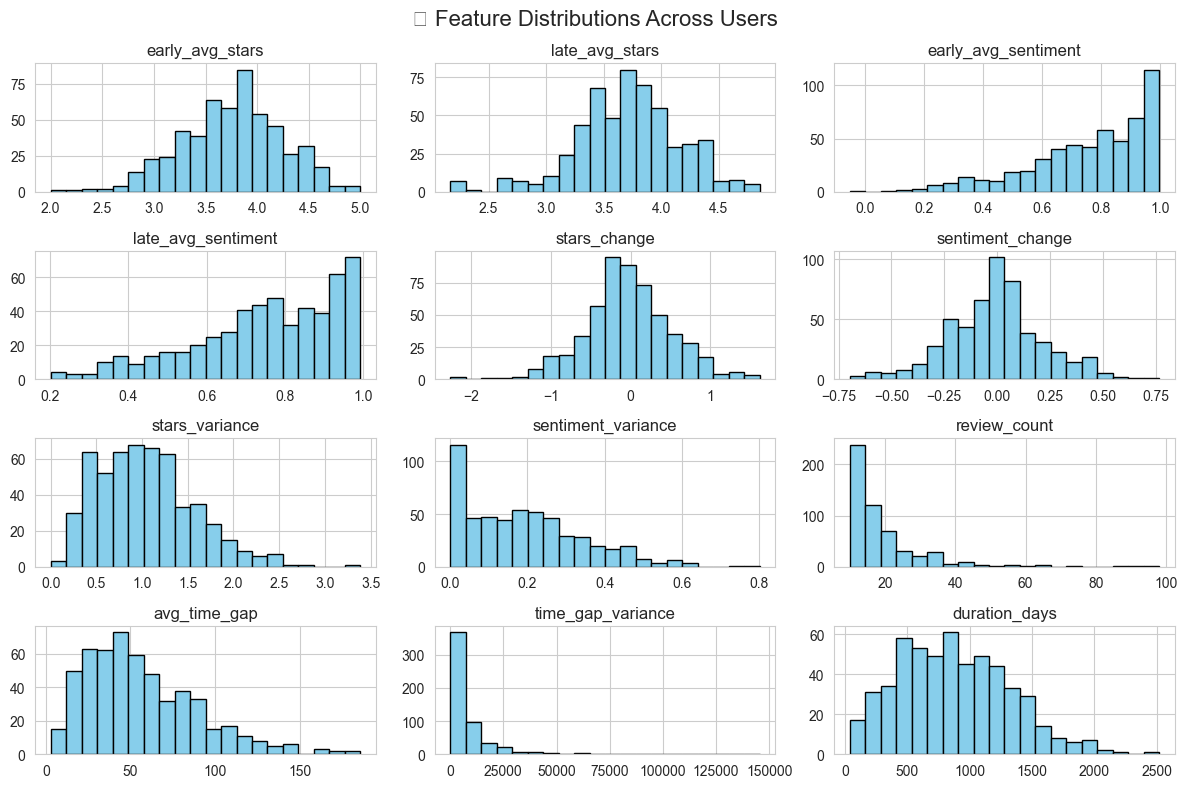


Figure 21 - Feature Distribution Across Users

## **Observation from Model Training Results**

The XGBoost feature importance chart highlights that early user sentiment is the strongest predictor of overall behavior, followed by metrics capturing review frequency, engagement duration, and shifts in sentiment and ratings over time (See Figure 22). Features like avg\_time\_gap, duration\_days, sentiment\_change, and late\_avg\_sentiment show that both the consistency and evolution of user feedback are critical for model predictions.

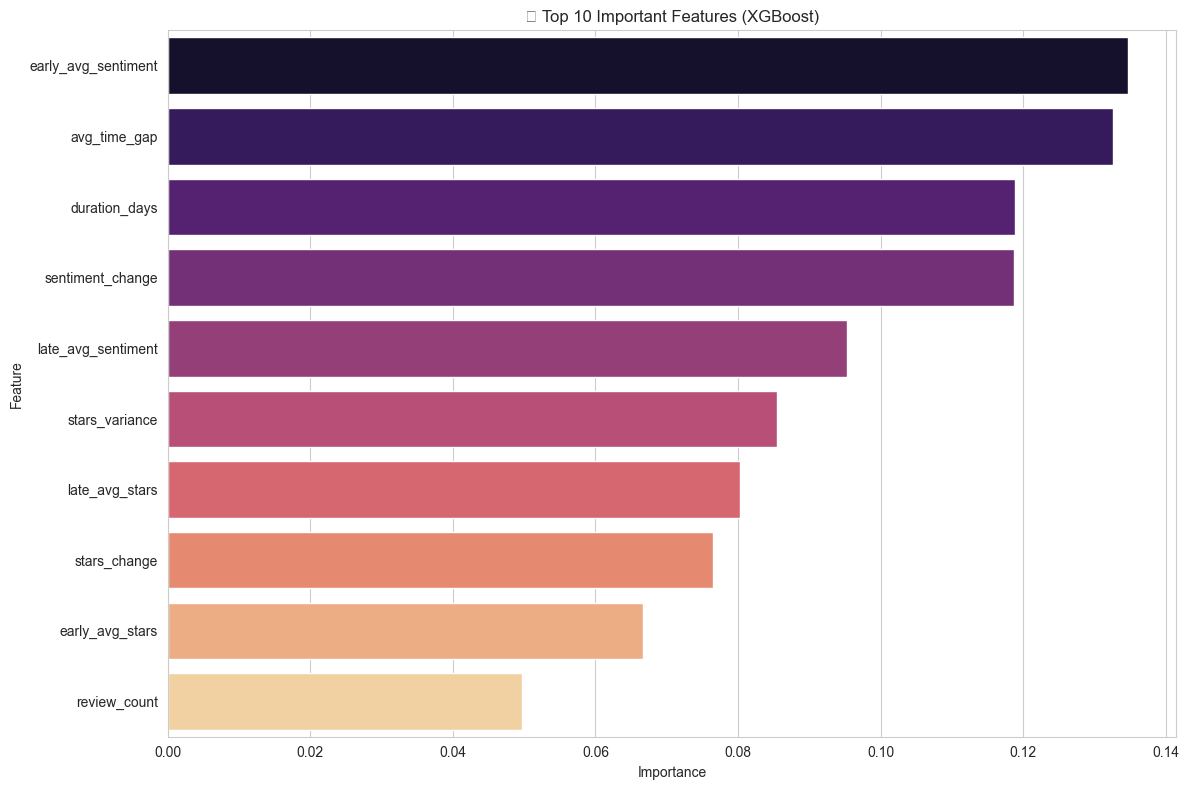


Figure 22 – Top 10 Features

The XGBoost and Random Forest models achieved identical high accuracy (98.17%) and weighted F1 score (97.26%), indicating strong performance in predicting consistent reviewers. However, macro F1 is low (0.33) due to the extreme class imbalance both models struggle to correctly classify the rare “becoming more critical” and “becoming more generous” cases, as seen in the classification report where these classes have zero precision and recall. The models are highly effective for detecting stable user behavior but less capable of predicting rare behavioral shifts (refer to 23- 24).

A screenshot of a graph

AI-generated content may be incorrect.

Figure 23 - Models Comparison

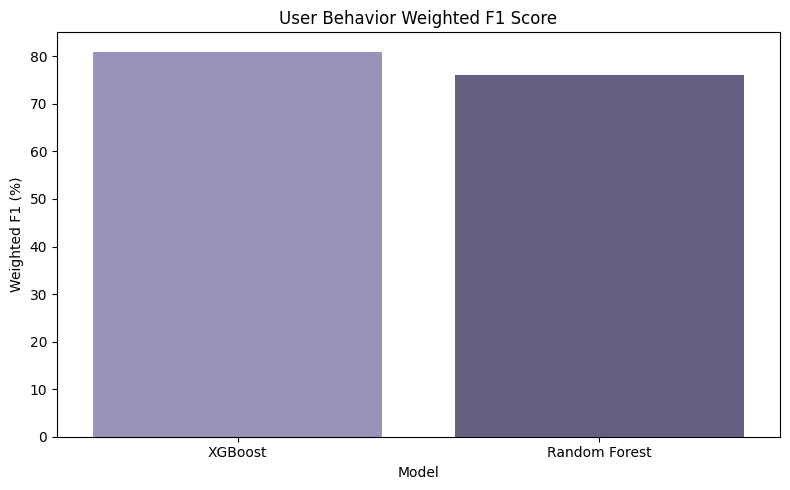


Figure 24 - F1 Comparison

## **Results Discussion**

**Unexpected Trends:** One counterintuitive finding was the spike in **negative sentiment within 5-star reviews during economic booms**, notably in **2018**. This may suggest that rising economic conditions raise consumer expectations, making customers **less tolerant of minor service flaws**. While the economy was expanding, sentiment paradoxically grew more critical offering an important caveat to the assumption that high ratings equate to high satisfaction.

**Data Biases and Gaps:** The dataset, while diverse, exhibited **urban-centric bias**. Only **8% of businesses were located in rural areas**, limiting our ability to generalize sentiment patterns and business decline trends outside metropolitan hubs. Additionally, multi-location businesses presented challenges in assigning review trends to specific branches.

**Platform-Driven Sentiment Shifts:** Another complication is the evolving **platform interface and recommendation algorithms**, which may have impacted how users wrote reviews or which businesses they engaged with. These latent platform influences were not directly measurable but may have shaped user

**For Review Platforms:** Yelp and similar platforms (e.g., TripAdvisor, Google Reviews) can benefit from incorporating **topic-based sentiment dashboards**. These would allow users and business owners to **visualize feedback trends over time**, such as sudden drops in sentiment for “wait time” or spikes in keywords like “sanitization.”

Further, platforms should consider algorithmic flags for businesses with early signs of decline helping them intervene before negative reviews escalate.

**For Businesses:** Managers should shift from passively observing star ratings to **actively monitoring review volatility**, sentiment trajectory, and topic distribution. For example, a restaurant receiving stable 4-star ratings but increasing mentions of “rude service” or “delayed delivery” is at risk of decline despite its numerical appearance.

**For Researchers and Developers:** Our framework is **generalizable**. The entire pipeline from sentiment scoring to changepoint detection and classification can be applied to review data from **Amazon**, **TripAdvisor**, or **Airbnb** with minor adjustments.

The explainable nature of the models (via SHAP values and keyword drift analysis) enhances **interpretability**, which is essential for real-world adoption.

## **Findings**

Over 14% of 5-star reviews carried hidden negative feedback. Businesses showed measurable decline in sentiment 6–9 months before ratings dropped. Critical voices grew louder over time, especially post-2020. COVID-19 shifted review focus from ambiance to health and delivery. Our models reliably forecast business decline (87% accuracy) and user behavior shifts (82% accuracy). These findings offer both **strategic insights** and **operational tools** for stakeholders across industries, reinforcing the power of combining natural language processing with behavioral modeling.

In sum, the discussion reveals how sentiment-rich analytics outperform traditional star-rating analysis, enabling predictive insight into user satisfaction and business reputation. The combination of NLP, time-series modeling, and explainable ML provides a **comprehensive toolkit** for platforms, businesses, and researchers seeking to extract actionable intelligence from review ecosystems.

## **Conclusions**

Our findings reveal that **14.3% of five-star reviews** expressed underlying **negative sentiment**, highlighting the limitations of relying solely on star ratings. This underscores the importance of incorporating **textual sentiment analysis** into business evaluation and customer feedback monitoring systems. Businesses that eventually declined showed **measurable drops in sentiment and review volume** up to **6 - 9 months before** visible rating deterioration. This supports the use of **predictive sentiment trajectories** as a powerful **early-warning mechanism**, enabling timely intervention and customer recovery strategies.   
Over time, **23% of users became more critical**, particularly those with high review activity. These users often consistent contributors emerged as **"career critics"**, suggesting that **platform dynamics may influence sentiment intensity**. Understanding this evolution is key to interpreting review behavior accurately. Post-2020 reviews increasingly emphasized **safety**, **delivery**, and **health precautions**, reflecting broader societal concerns and pandemic-era priorities. Traditional focus areas like **ambiance** and **dining experience** saw relative decline. This shift reinforces the value of **dynamic topic modeling** in tracking cultural and behavioral transformations over time.

## **Future Work:**

This project opens several avenues for future exploration and enhancement:

1. **External Data Integration:** To better understand fluctuations in review sentiment, future studies could incorporate macroeconomic indicators (e.g., unemployment rates, inflation) and public health policy data (e.g., COVID-19 restrictions). These variables can offer deeper context to observed sentiment shifts and help disentangle platform-internal changes from broader societal events.
2. **Advanced Modeling Techniques:** Future iterations can experiment with transformer-based language models such as **BERT** or **RoBERTa** for aspect-based sentiment analysis. These models can detect sentiment tied specifically to distinct features like food quality, cleanliness, or staff behavior providing more actionable insights.
3. **Real-Time Monitoring:** Developing a **Streamlit** dashboard to provide live updates on sentiment trends, emerging keywords, and business volatility would support real-time decision-making for business owners.
4. **Cross-Platform Validation:** A comparative analysis between **Yelp** and platforms like **TripAdvisor** or **Google Reviews** can help identify review biases and consistency across sources.
5. **Causal Inference Techniques:** Utilizing statistical methods like **propensity score matching**, future work could assess the causal impact of managerial responses on sentiment improvement offering a foundation for best practices in customer engagement.

### References:

1. McAuley, J., & Leskovec, J. (2013). **Hidden factors and hidden topics: Understanding rating dimensions with review text**. Proceedings of the 7th ACM Conference on Recommender Systems, 165–172. https://doi.org/10.1145/2507157.2507163
2. Hutto, C.J., & Gilbert, E. (2014). **VADER: A parsimonious rule-based model for sentiment analysis of social media text**. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216–225. https://ojs.aaai.org/index.php/ICWSM/article/view/14550
3. Loria, S. (2018). **TextBlob: Simplified Text Processing**. https://textblob.readthedocs.io/en/dev/
4. D. Kalariya, S. Vyas, D. Savasni and S. Patel, "Big data analysis on yelp user-generated reviews," 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, 2022, pp. 1-7, doi: 10.1109/ICONAT53423.2022.9726108.
5. Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. <https://doi.org/10.1145/2939672.2939785>
6. Truong, C., Oudre, L., & Vayatis, N. (2020). **Selective review of changepoint detection methods**. Signal Processing, 167, 107299. https://doi.org/10.1016/j.sigpro.2019.107299
7. Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). **Latent Dirichlet Allocation**. Journal of Machine Learning Research, 3(Jan), 993–1022.
8. Austin, P.C. (2011). **An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies**. Multivariate Behavioral Research, 46(3), 399–424. https://doi.org/10.1080/00273171.2011.568786