

Beyond Yelp Stars: Better Review Insights for Businesses & Customers ☒

- by K. K. Hong

Leveraging sentiment analysis on hundreds of thousands of Yelp reviews, this project provides deeper emotional insights, helping businesses understand customer sentiment and enabling consumers to make more informed decisions.

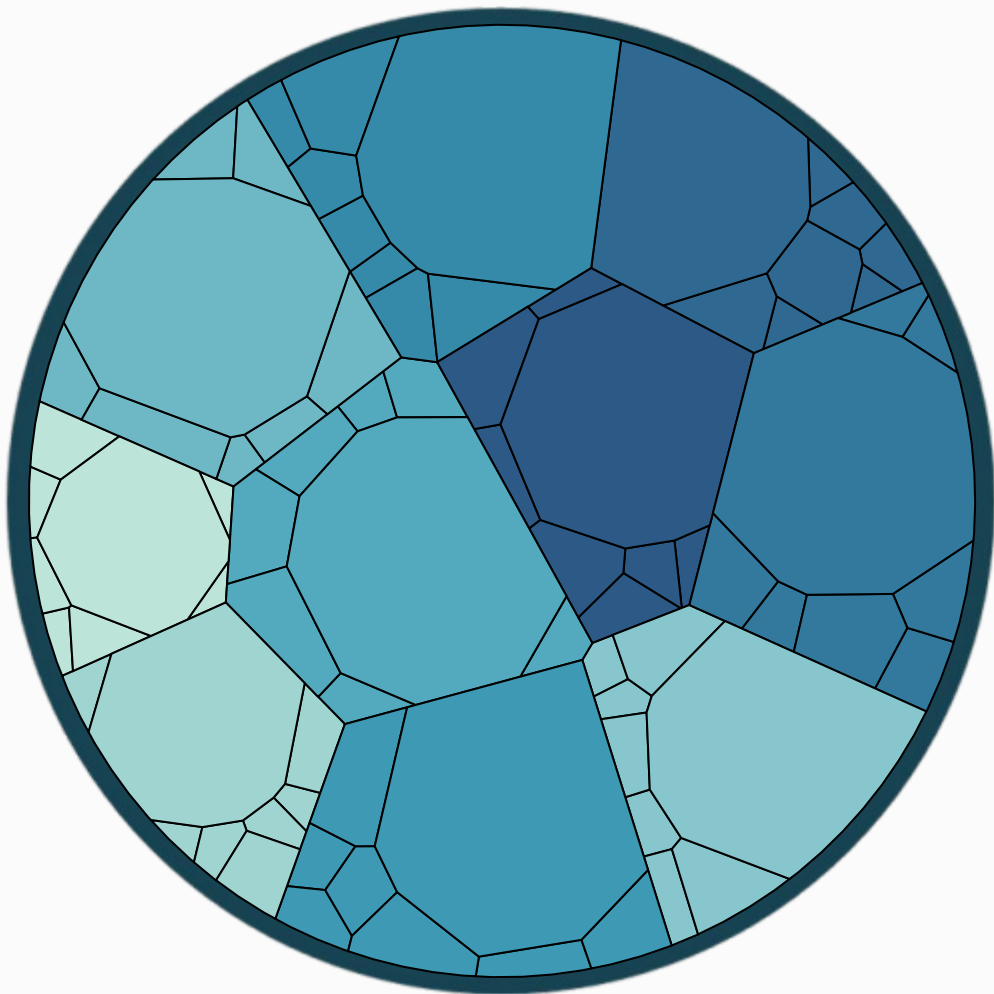
| | | | | |
|--------------|-----------|-----------------|----------|------------|
| Introduction | Sentiment | Timing & Trends | Rankings | Conclusion |
|--------------|-----------|-----------------|----------|------------|

Between 2012 and 2021

120,359 users reviewed 3,772 businesses in Santa Barbara, totaling 242,490 reviews.

For this project, I built a dashboard that goes beyond Yelp's star ratings and text reviews by applying sentiment analysis. I used a pretrained RoBERTa model, fine-tuned on the GoEmotions dataset, to process 242,490 reviews from 120,359 users across 3,772 Santa Barbara businesses (2012 to 2021). The result is an interactive dashbaord that highlights customer emotions, uncovers market shifts, and tracks business performance.

The biggest challenge was transforming raw model outputs into clear charts that serve both business owners and customers. After extracting sentiment labels, I analyzed their distributions and year to year trends to shape the dashboard's design. Now business owners can monitor sentiment trends and key metrics in real time, while customers get an emotion driven way to compare local merchants. By moving beyond traditional NLP, this deep learning approach delivers a richer and more nuanced view of what customers truly feel and how businesses are performing.



DATA SOURCE:

Yelp Open Dataset

TOOLS:

Python, Tableau

KEYWORDS:

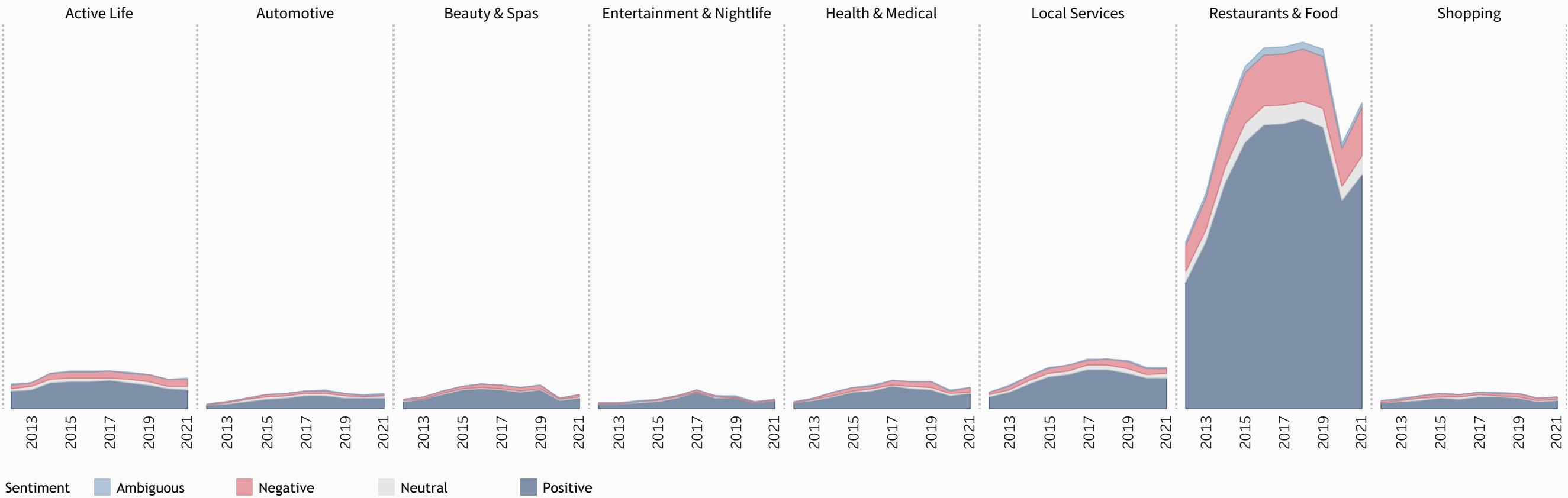
NLP, Fine-Tuning, Sentiment Analysis, Tableau, Data Storytelling, Ranking & Performance Measurement

REFERENCE PAPERS:

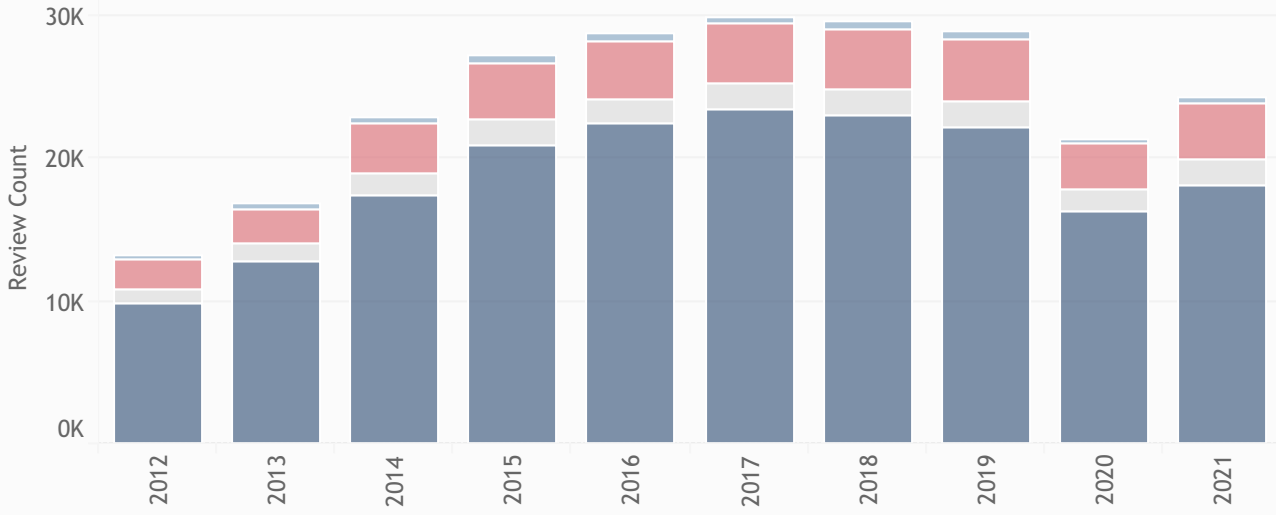
- BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
- RoBERTa: A Robustly Optimized BERT Pretraining Approach
- GoEmotions: A Dataset of Fine-Grained Emotions



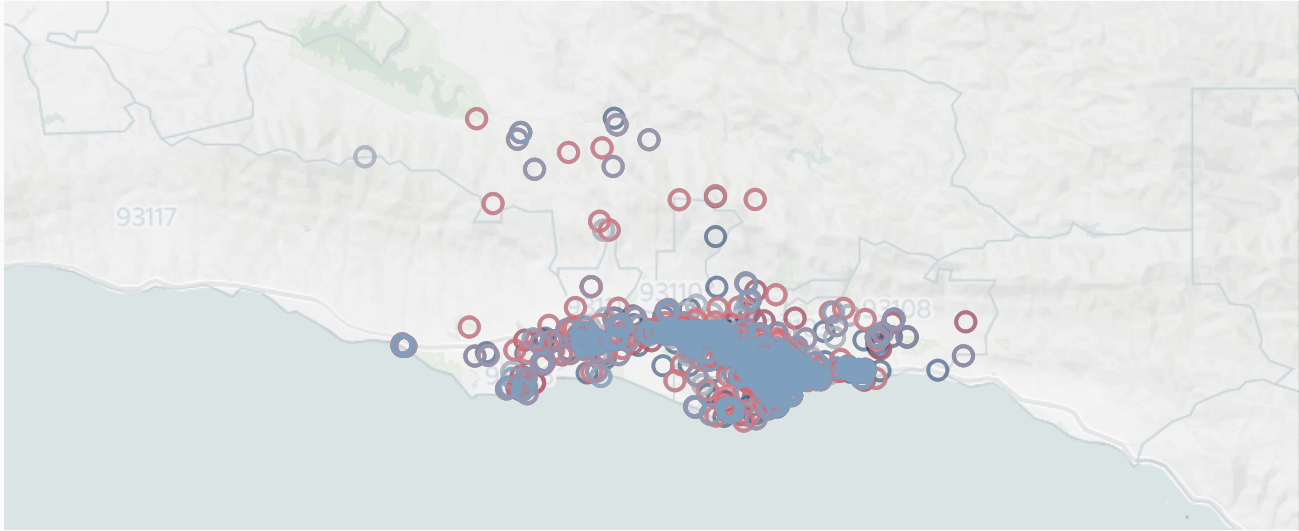
This area chart shows the distribution of sentiments across key industries each year. The four sentiment categories—positive, negative, neutral and ambiguous—are plotted by total review counts per year, enabling a clear comparison of sentiment trends. This visualization offers an in depth look at how customer perceptions and emotional responses evolve over time across different industry sectors.



Changes in Sentiment Over Time



Geographic Distribution of Reviews Across Santa Barbara



On this page I turn Yelp's star and text format into an interactive dashboard for Santa Barbara. It begins with a Voronoi treemap that shows how review counts across eight key industries have changed over time. Below that, an area chart overlays positive, negative and ambiguous sentiment to reveal year to year shifts in customer emotions.

Next, a bar chart highlights which industries saw the largest increases in review volume and sentiment intensity, making it easy to spot peak seasons or turning points. For example, before COVID review counts increased steadily each year; in 2020 they dropped sharply, and the Restaurants and Food industry was the hardest hit, showing its vulnerability to major events. Finally, a GIS based heat map highlights areas with the highest concentration of reviews, particularly downtown Santa Barbara, suggesting prime business opportunities in the most popular neighborhoods.

Beyond Yelp Stars: Better Review Insights for Businesses & Customers

- by K. K. Hong

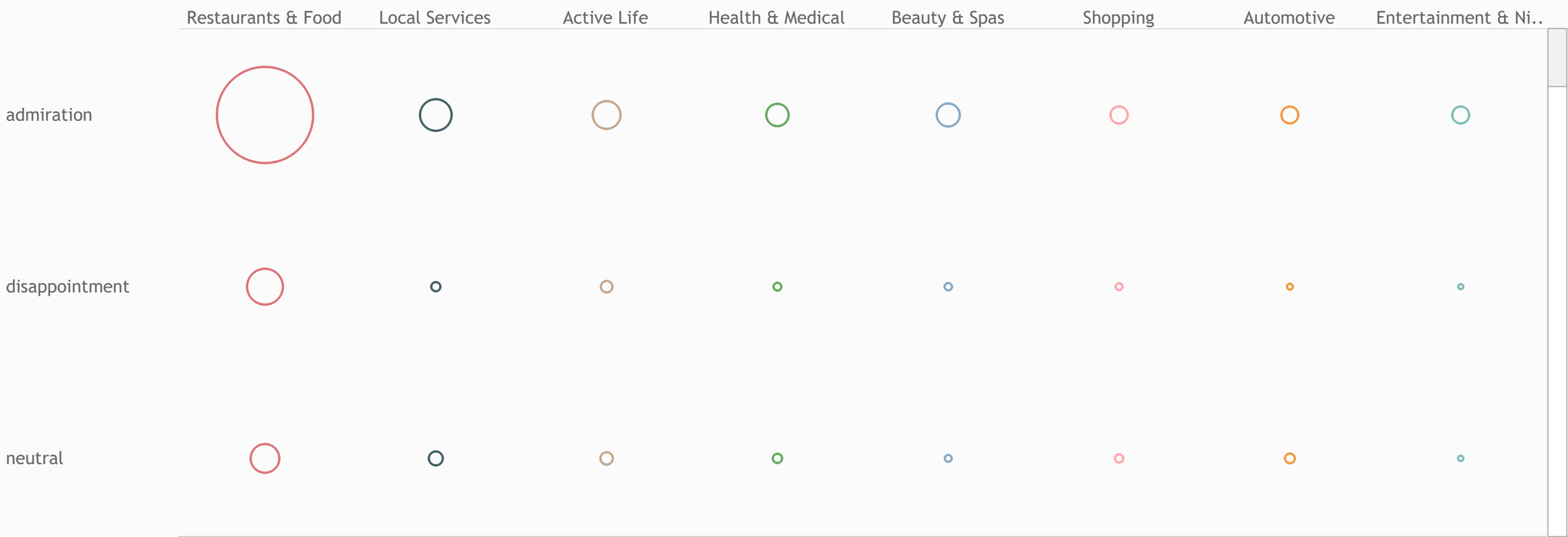
Leveraging sentiment analysis on hundreds of thousands of Yelp reviews, this project provides deeper emotional insights, helping businesses understand customer sentiment and enabling consumers to make more informed decisions.

| Introduction | Sentiment | Timing & Trends | Rankings | Conclusion |
|--------------|-----------|-----------------|----------|------------|
|--------------|-----------|-----------------|----------|------------|

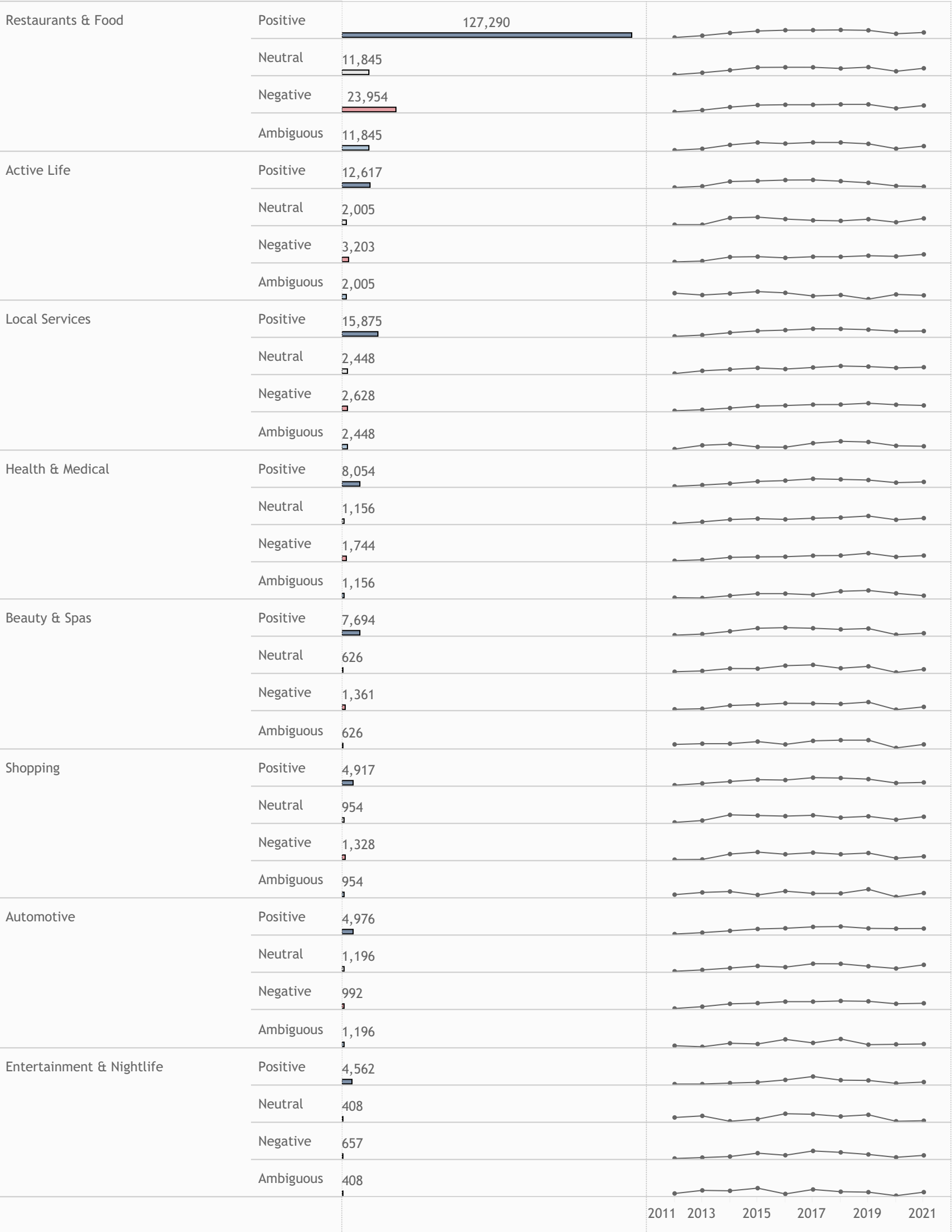
This page maps emotional associations across eight industries. The horizontal axis lists detailed emotions from positive feelings like Admiration and Joy to negative responses such as Disappointment and Disgust. These labels come from a RoBERTa model fine tuned on the GoEmotions dataset. Each bubble's size reflects how often that emotion appears in an industry's reviews.

For example, Restaurants & Food not only leads in review volume but also has the largest Admiration bubble, underlining its power to create delight and connection. Other industries show a more balanced mix of emotions, although Admiration still dominates overall.

It is important to note that these patterns can reflect model and data biases. Since the GoEmotions dataset was originally labeled using Reddit comments, its categories may not line up perfectly with Yelp reviews. Reviewers also tend to start with praise before pointing out problems, which can push the model toward more positive labels. Even so, this matrix delivers each industry's unique emotional fingerprint, offering clear insights businesses can use to shape customer experiences and sharpen their marketing.



Trends in Customer Sentiment (2010-2022): Category-Specific Review Volumes and Yearly Patterns



The combined chart uses bar charts to show the distribution and volume of positive, negative, neutral and ambiguous sentiments, while a line chart tracks the growth of positive reviews from 2010 to 2021. Restaurants and Food leads with 127,290 positive reviews, far outpacing 23,954 negative and 11,845 ambiguous mentions. Over that period, total reviews increased steadily, reflecting the growing reliance on Yelp for service selection.

When we compare Restaurants and Food to other industries, it clearly outperforms in both review count and positive sentiment, confirming that it drives the most engagement on Yelp. Although negative reviews remain low overall, Shopping and Beauty and Spa show the highest share of negative feedback, signalling opportunities for quality improvement.

The line chart trends highlight consistent growth in review volumes, especially for positive sentiment, suggesting rising customer engagement and satisfaction. Lower-volume sectors like Health and Medical, Automotive and Beauty and Spa follow similar sentiment patterns. Overall, food and service industries dominate user feedback, while smaller industries maintain a steady niche presence. These insights offer a clear roadmap for targeted improvements and ongoing sentiment monitoring.

Beyond Yelp Stars: Better Review Insights for Businesses & Customers

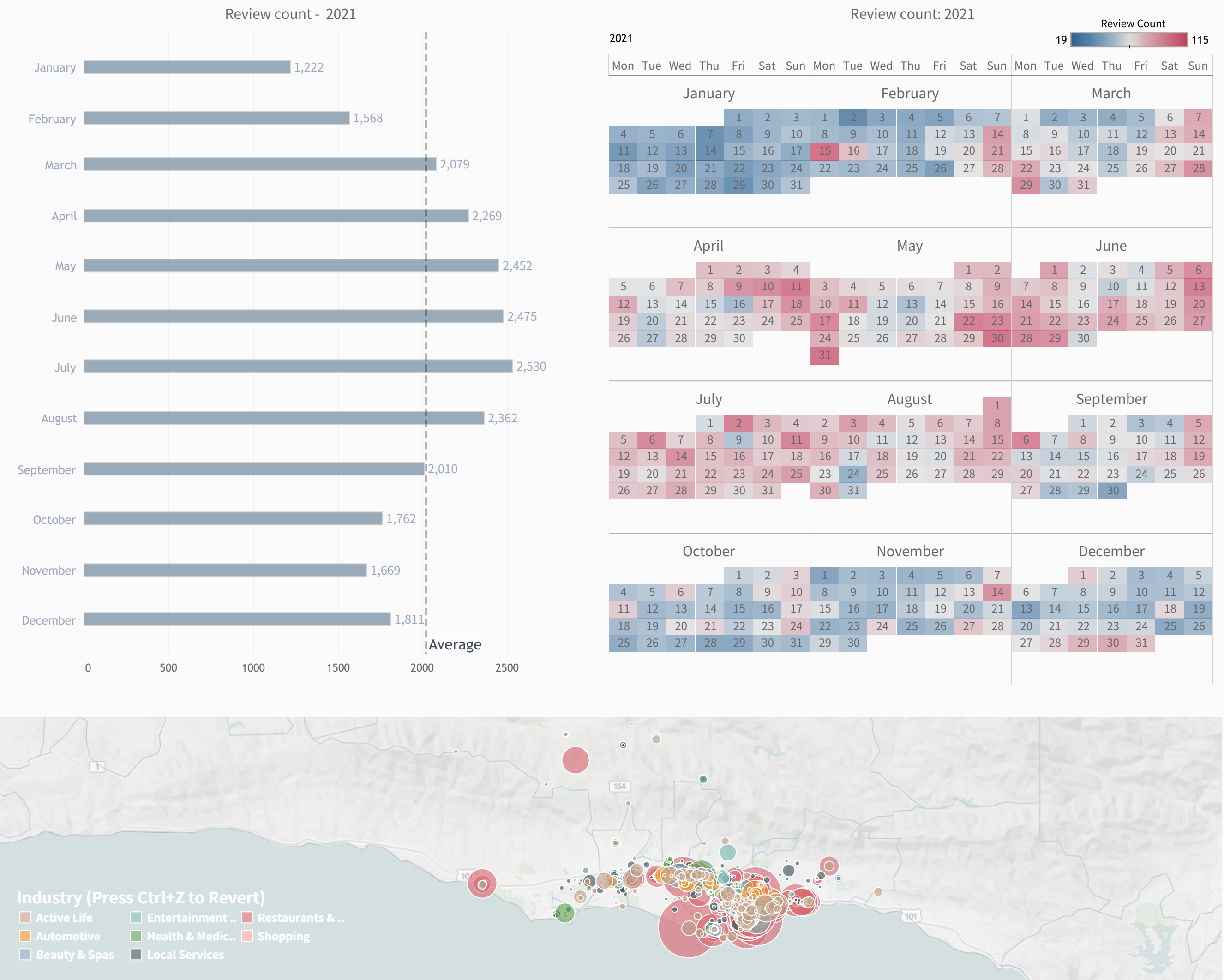
- by K. K. Hong

Leveraging sentiment analysis on hundreds of thousands of Yelp reviews, this project provides deeper emotional insights, helping businesses understand customer sentiment and enabling consumers to make more informed decisions.

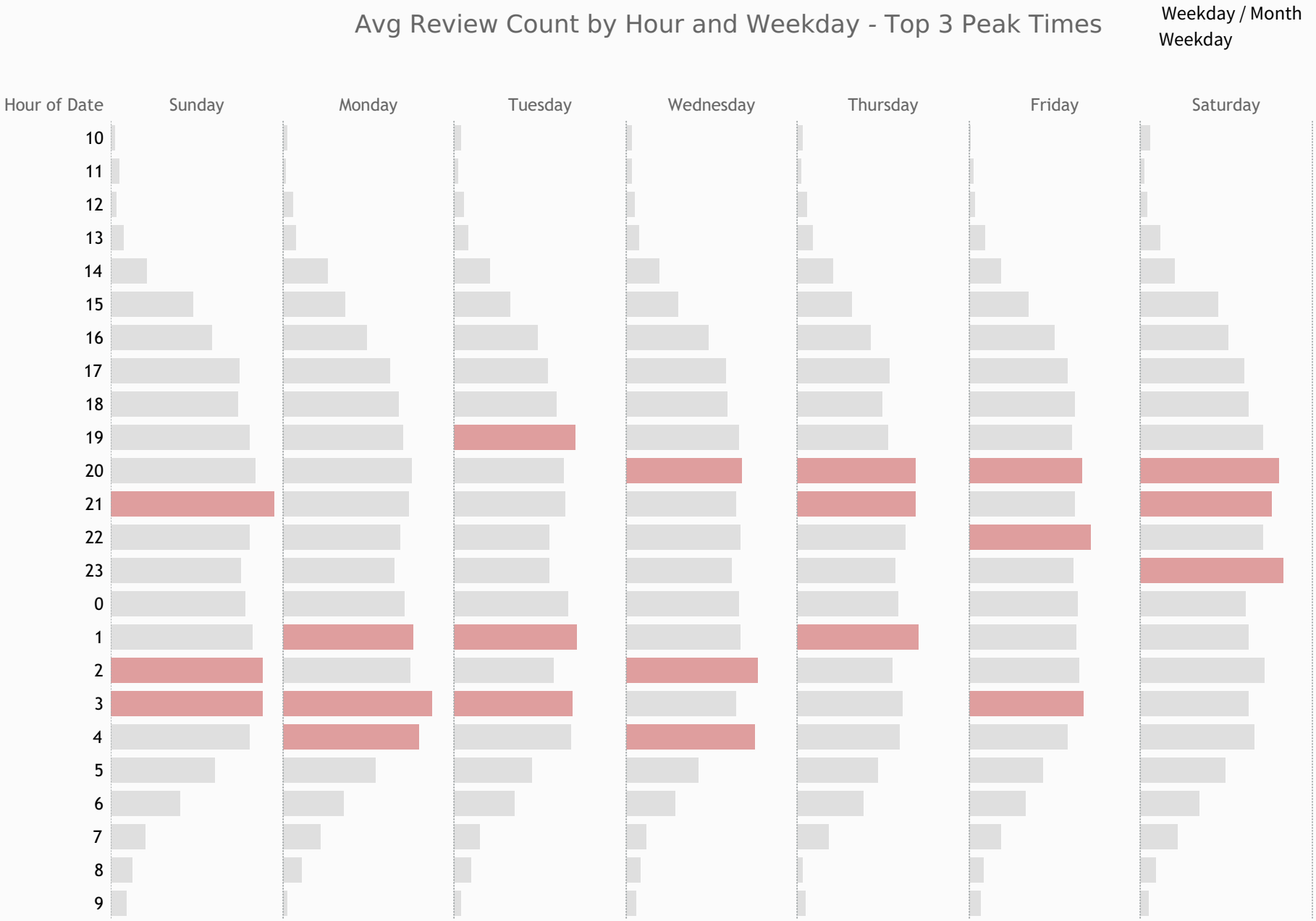
| | | | | |
|--------------|-----------|-----------------|----------|------------|
| Introduction | Sentiment | Timing & Trends | Rankings | Conclusion |
|--------------|-----------|-----------------|----------|------------|

This page breaks down review activity on a weekly and daily basis using 2021 as an example. A heat map calendar highlights monthly, weekly, and daily review volumes. Beside it, a bar chart plots monthly review volume, showing the highest counts in June and July, likely reflecting seasonal demand shifts. The GIS map below displays the distribution and volume of reviews by industry, using a color coded legend.

In the bottom right, a bar chart links review activity to hourly, daily, and monthly patterns, with the top three peak times highlighted in red (for example, late afternoons and weekends). This matches the calendar view, where weekends and midweek days like Wednesday and Thursday stand out. The color coded design enhances readability, while the map adds geographic context. Overall, this page reveals 2021's cyclical trends, regional engagement patterns, and operational benchmarks, offering actionable insights for resource planning and strategic adjustments.



To the right, a bar chart shows customer review volume by hour across weekdays, with the three busiest hours highlighted in red. This makes it easy to spot peak activity times and understand when customers are most engaged. The chart helps link daily patterns to broader trends, supporting informed decisions around staffing, business activity planning, and resource allocation. An interactive filter also allows users to switch to a monthly view for quick comparisons, making it easier to track how customer behavior shifts from month to month.



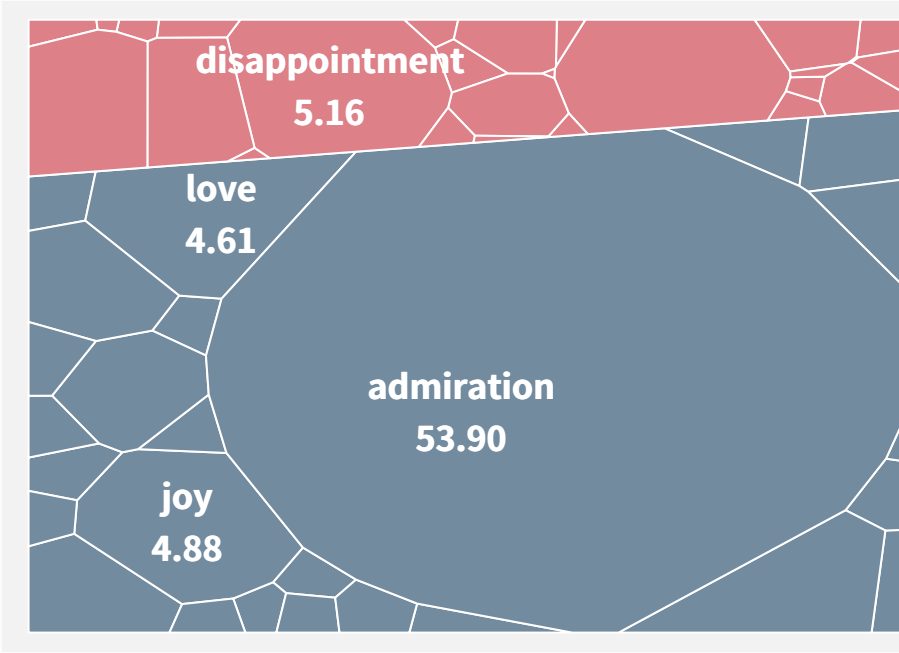
Beyond Yelp Stars: Better Review Insights for Businesses & Customers

- by K. K. Hong

Leveraging sentiment analysis on hundreds of thousands of Yelp reviews, this project provides deeper emotional insights, helping businesses understand customer sentiment and enabling consumers to make more informed decisions.

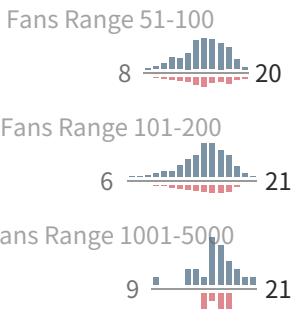
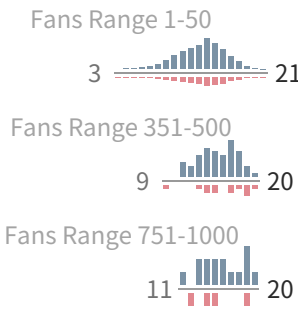
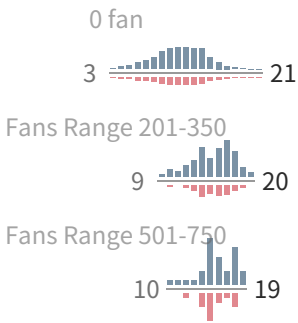
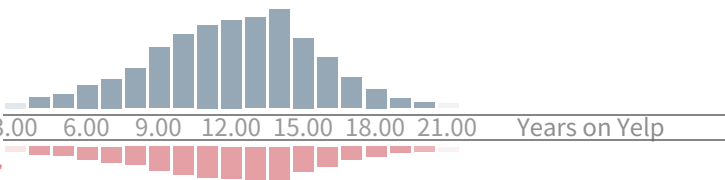
| | | | | |
|--------------|-----------|-----------------|----------|------------|
| Introduction | Sentiment | Timing & Trends | Rankings | Conclusion |
|--------------|-----------|-----------------|----------|------------|

This page uses customer attributes to build a segmentation metric that divides users into Supporters and Complainers based on their positive review rate (Supporters exceed 50 percent). Of 120,359 users, 74.49 percent are Supporters and 25.51 percent are Complainers. To compare sentiment by fan count and Yelp tenure, I created a bar chart showing each group's share of Supporters and Complainers. Customers with more fans or more years on Yelp have a higher share of Complainers, revealing stricter reviewing standards. Finally, I combined business attributes and user sentiment metrics into a single ranking score to identify the top 10 businesses for recommendations.



ComplainerSupporter

Total Users:
130,346 Users
74.49 % Supporter | 25.51 % Complainer



Sentiment Value Mapping:

- Favourable (2): caring, gratitude, joy, love, optimism, excitement
- Somewhat Favourable (1): admiration, amusement, approval, desire, pride, relief
- Neutral (0): confusion, curiosity, realization, surprise, neutral
- Somewhat Unfavourable (-1): annoyance, disappointment, disapproval, embarrassment, nervousness, remorse
- Unfavourable (-2): anger, disgust, fear, grief, sadness

Fans Range and Score:

- 0 fans: 0
- 1 to 50 fans: 2
- 51 to 100 fans: 4
- 101 to 200 fans: 5
- 201 to 350 fans: 6
- 351 to 500 fans: 7
- 501 to 750 fans: 8
- 751 to 1000 fans: 9
- 1001 to 5000 fans: 10

Years on Yelp Score:

- 2.5 to 5 years: 2
- 5 to 7.5 years: 3
- 7.5 to 10 years: 5
- 10 to 12.5 years: 6
- 12.5 to 15 years: 7
- 15 to 17.5 years: 8
- 17.5 to 20 years: 9

The Review Score combines sentiment strength with reviewer credibility.

It is calculated as:

Review Score :

$$\begin{aligned} & (\text{Primary sentiment value} \times \\ & \quad \text{Primary sentiment score} \times 0.45) + \\ & (\text{Secondary sentiment value} \times \\ & \quad \text{Secondary sentiment score} \times 0.20) + \\ & (\text{Years on Yelp Score} \times 0.20) + \\ & (\text{Fan score} \times 0.15) \end{aligned}$$

Overall Ranking Score:

$$\begin{aligned} & (\text{Current stars} \times 0.58) + \\ & (\text{Current review count} \times 0.002) + \\ & (\text{Average of latest 50 review scores} \times 0.4) \end{aligned}$$

This formula balances positive emotion, reviewer authority, star ratings, and recent engagement. All weights reflect my business selection priorities and can be refined via A/B testing, user surveys, focus groups, or advanced recommendation techniques such as collaborative filtering and matrix completion.

Ranking based on the Ranking Score and Industry

| | | | | | | | | | | |
|---------------------------|---------------------------------------|--|-------------------------------------|------------------------------------|--|------------------------------------|--|--|---------------------------------------|--------------------------------------|
| Beauty & Spas | Lux Nails & Spa Rank: 1 | Salon Patine Rank: 2 | Champagne Nail Bar Rank: 3 | Angels Nail & Spa Rank: 4 | Modern Nails Rank: 5 | Carlyle Salon and Style Rank: 6 | Energy Tattoo Rank: 7 | Cielo Spa Boutique Rank: 8 | Santa Barbara Salon Rank: 9 | Underground Hair Artists Rank: 10 |
| Nightlife & Entertainment | The Good Lion Rank: 1 | EOS Lounge Rank: 2 | Santa Barbara Bowl Rank: 3 | Celebration Cruises Rank: 4 | Test Pilot Rank: 5 | Wildcat Lounge Rank: 6 | Santa Barbara Wine Tasting Rank: 7 | The Red Piano Rank: 8 | Burgundy Blue Photography Rank: 9 | Tonic Nightclub Rank: 10 |
| Automotive | Big Brand Tire & Service Rank: 1 | Lexus Santa Barbara Rank: 2 | Ally Detail Rank: 3 | Bunnin Chevrolet Rank: 4 | Jan's Tire and Auto Repair Rank: 5 | Super Value Smog Rank: 6 | Educated Car Wash Rank: 7 | American Tire Depot Rank: 8 | Bob Holzer Towing Rank: 9 | Team Tire Pros Rank: 10 |
| Health & Medical | Salt Cave Santa Barbara Rank: 1 | Santa Barbara Chicken Rank: 2 | Royal Thai Massage Rank: 3 | Float Luxury Spa Rank: 4 | Massage Green SPA - Santa Barbara Rank: 5 | Zen Massage Place Rank: 6 | Urban Optics Optometry Rank: 7 | Evolutions Medical & Dental Rank: 8 | Trevor Denny, DDS Rank: 9 | Johnson Family Dental Rank: 10 |
| Restaurants & Food | Los Agaves Rank: 1 | Brophy Bros - Santa Barbara Rank: 2 | Boathouse at Hendry's Rank: 3 | Santa Barbara Shellfish Rank: 4 | Mesa Verde Rank: 5 | La Super-Rica Taqueria Rank: 6 | McConnell's Fine Ice Cream Rank: 7 | Sandbar Rank: 8 | The Lark Rank: 9 | The Palace Grill Rank: 10 |
| Shopping | The Closet Trading Company Rank: 1 | Chaucer's Bookstore Rank: 2 | Apple State Street Rank: 3 | Alpha Floral Rank: 4 | Cat Therapy Rank: 5 | Nordstrom Paseo Nuevo Rank: 6 | Riley's Flowers Rank: 7 | Lush Rank: 8 | The Dress Rank: 9 | Plum Goods Rank: 10 |
| Local Services | Santa Barbara Art Glass Rank: 1 | Nick's Appliance Service Rank: 2 | Segway of Santa Barbara Rank: 3 | iGeeks iPhone Repair Rank: 4 | Paradise Retreats Rank: 5 | GadgetFix Rank: 6 | Best Western Plus Santa Barbara Rank: 7 | Golden Eagle Tattoo Rank: 8 | Channel Islands Outfitters Rank: 9 | Movegreen Rank: 10 |
| Active Life | Hotel Milo Santa Barbara Rank: 1 | Santa Barbara Zoo Rank: 2 | Fess Parker's DoubleTree Rank: 3 | Kimpton Canary Hotel Rank: 4 | The Eagle Inn Rank: 5 | Harbor House Inn Rank: 6 | Harbor View Inn Rank: 7 | Santa Barbara County Courthouse Rank: 8 | Arroyo Burro Beach Rank: 9 | SB Buggie Rank: 10 |

Beyond Yelp Stars: Better Review Insights for Businesses & Customers

- by K. K. Hong

Leveraging sentiment analysis on hundreds of thousands of Yelp reviews, this project provides deeper emotional insights, helping businesses understand customer sentiment and enabling consumers to make more informed decisions.

| | | | | |
|--------------|-----------|-----------------|----------|------------|
| Introduction | Sentiment | Timing & Trends | Rankings | Conclusion |
|--------------|-----------|-----------------|----------|------------|

Conclusion:

This project successfully transforms traditional Yelp review data into a rich, emotion-driven, and actionable analytics tool. By leveraging a fine-tuned RoBERTa model trained on the GoEmotions dataset, I was able to extract detailed emotional signals from over 240,000 Yelp reviews across 3,700+ businesses in Santa Barbara.

Through a combination of deep learning, geographic visualization, and structured sentiment mapping, the dashboard provides new dimensions of insights into customer behavior, emotional trends, industry dynamics, and business performance over time.

The analysis revealed strong, steady growth in review volume and positive sentiment from 2012 to 2021, particularly driven by the Restaurants & Food sector. However, it also uncovered important nuances, such as higher proportions of negative sentiment in industries like Shopping and Beauty & Spas, highlighting areas for service improvement. Seasonal and regional trends were captured through calendar heatmaps and GIS maps, offering strategic operational insights.

Customer segmentation into "Supporters" and "Complainers" based on positivity rates added further layers of understanding, with findings showing that more experienced Yelp users or those with larger fan bases tend to have stricter review standards. Finally, a custom business ranking model was developed by weighting sentiment strength, reviewer tenure, fan following, and business star ratings, offering a refined, multi-factor metric for business evaluation and recommendation.

While the dashboard and ranking system provide a comprehensive view, future enhancements could include A/B testing, user surveys, or collaborative filtering techniques to refine the business recommendation engine further.

Overall, this project demonstrates how advanced natural language processing and thoughtful data visualization can unlock deeper customer insights, inform business strategies, and enhance the decision-making experience for users on platforms like Yelp.

