11/30/2024

An Analysis of Airbnb Reviews in Austin, TX

Presenter: Alessandra Bielli

Agenda

1. Introduction
2. Methodology
3. Analysis

Section 1

Research Questions

Purpose: to analyze Airbnb reviews from the Austin, TX area and answer…

1. How do Airbnb reviews reflect customers' attitudes towards diversity and inclusion in Austin?
2. What are the key factors that influence customer satisfaction on the platform?
3. How do hosts practices or characteristics impact guests' experiences and reviews?

Section 1

Data Collection

Title: Austin_Reviews

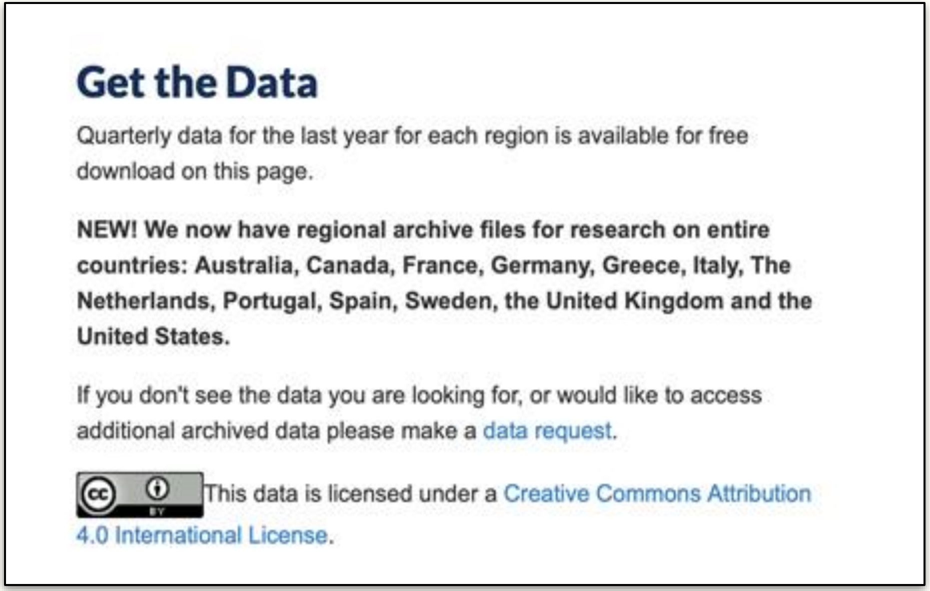
Source: Inside Airbnb -

https://insideairbnb.com/get-the-data/

Dimensions: 5 columns and 633,196 rows

Variables:

- listing_id (unique identifier for listings)
- date (date the review was posted)
- reviewer_id (unique identifier for reviewer)
- reviewer_name (name of the reviewer)
- comments (the raw text of the review)



Get the Data

Quarterly data for the last year for each region is available for free download on this page.

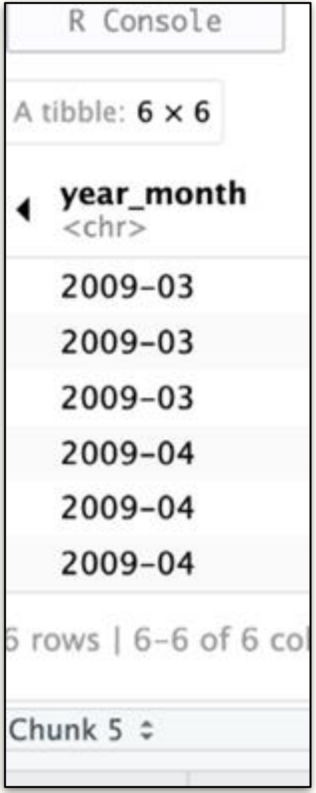
NEW! We now have regional archive files for research on entire countries: Australia, Canada, France, Germany, Greece, Italy, The Netherlands, Portugal, Spain, Sweden, the United Kingdom and the United States.

If you don't see the data you are looking for, or would like to access additional archived data please make a data request.

Exploratory Data Analysis

Data Preparation

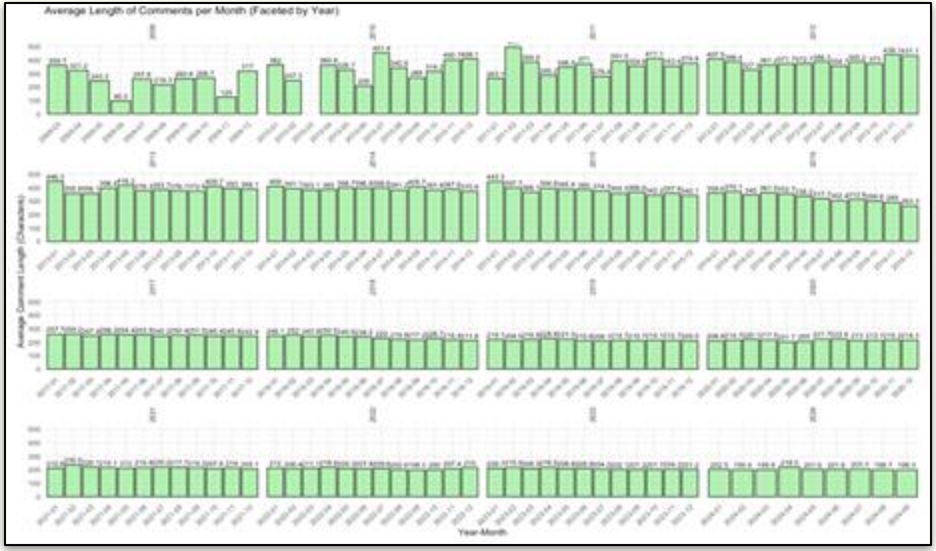
- Used colSums to count the number of missing rows in each columns
- Removed 32 rows with NA in the comment column using the filter function
- Converted the date column to a date object and extracted Year-Month

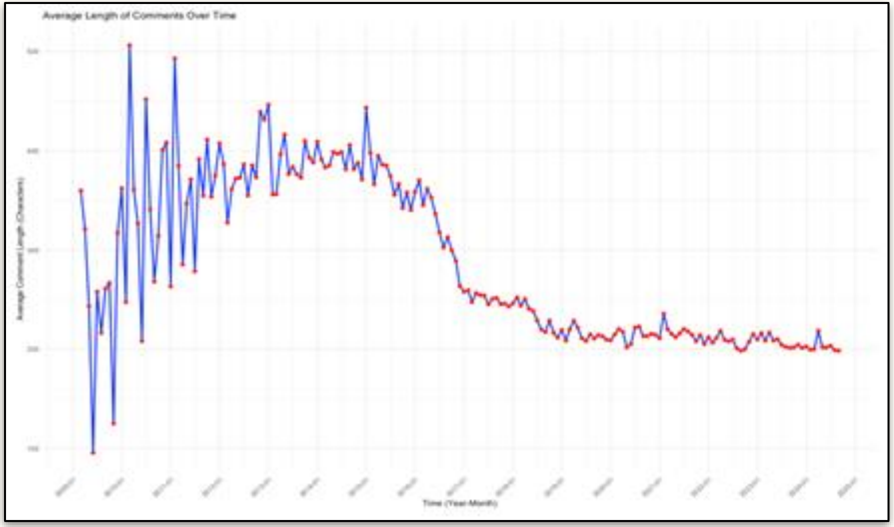


Section 2

Nchar Analysis

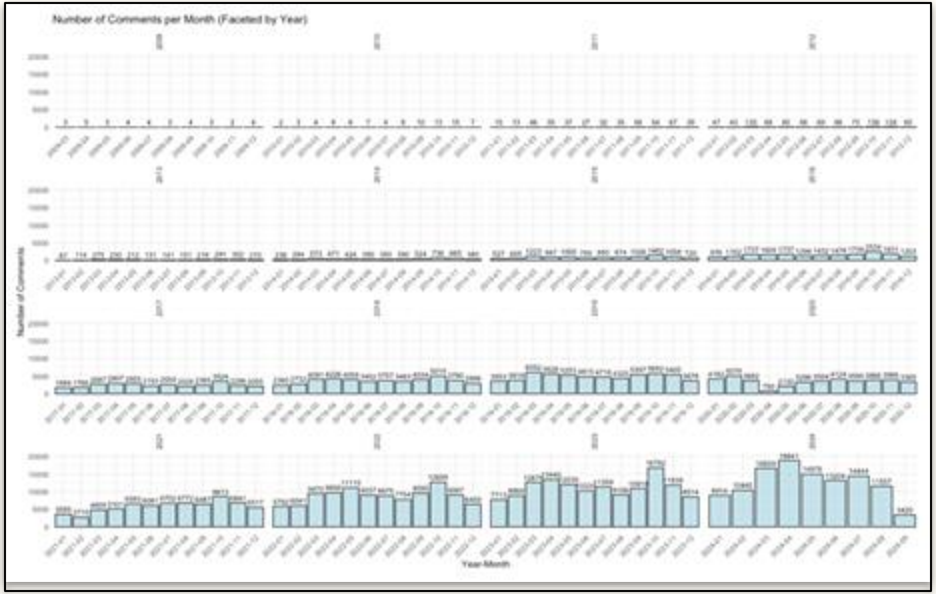
- **Methodology:** nchar, ggplot2



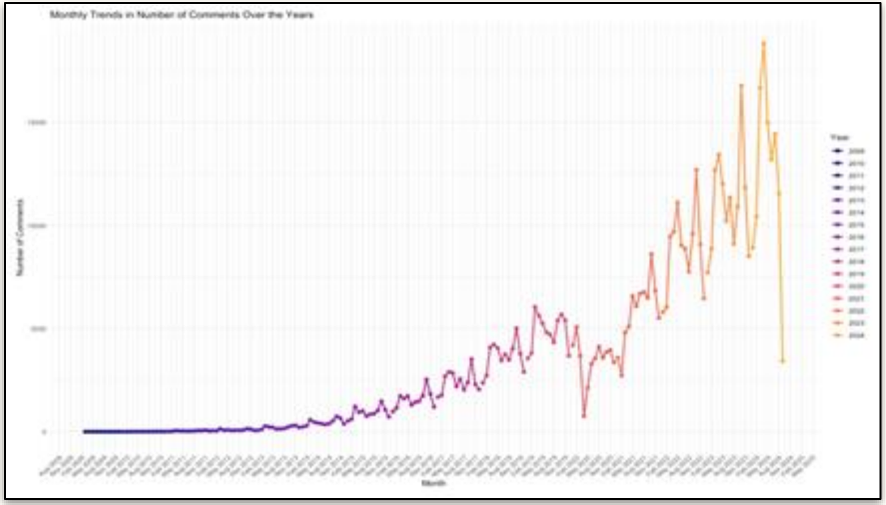


Length of Comments per Month by Year, Average Comment Length: 221.3008

Numerical Analysis



- **Methodology:** num, ggplot2

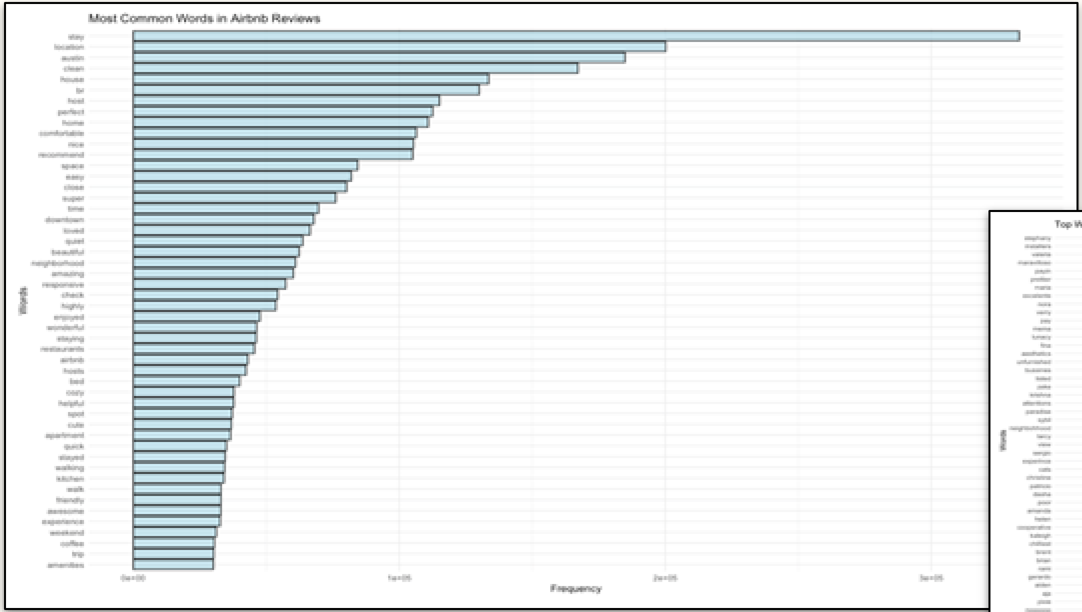


Number of Comments per Month by Year

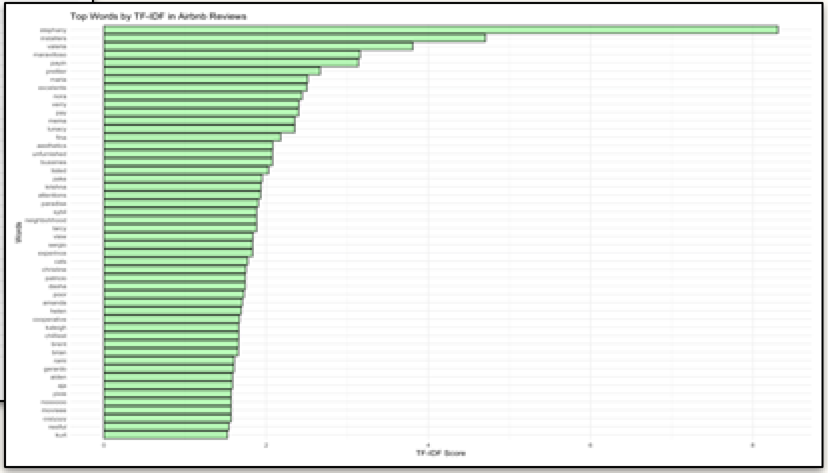
Frequency Analysis

STUDENT 2

- **Methodology:** removed stop words, tokenized comments

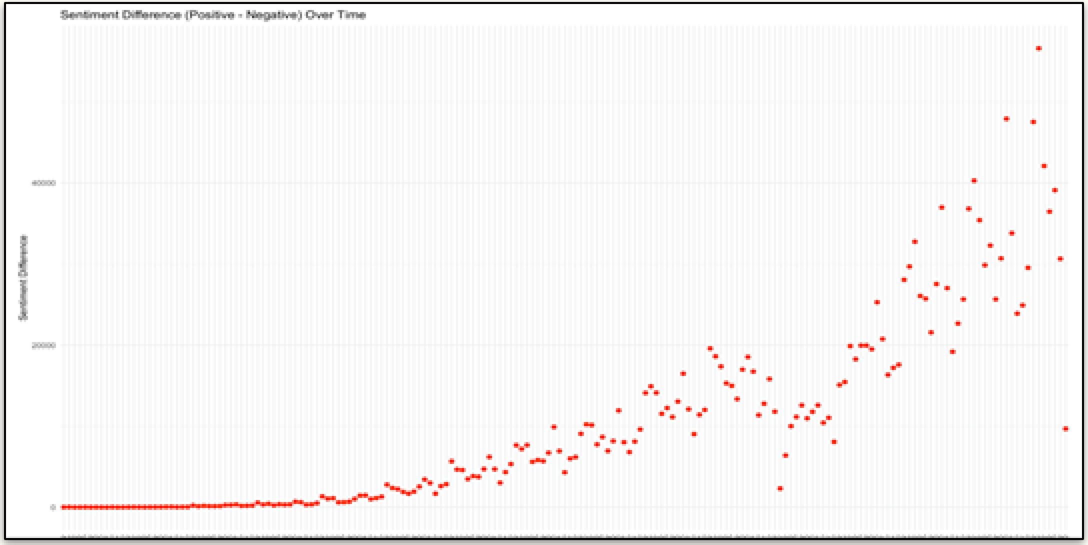


Most Common Words



TF-IDF Analysis

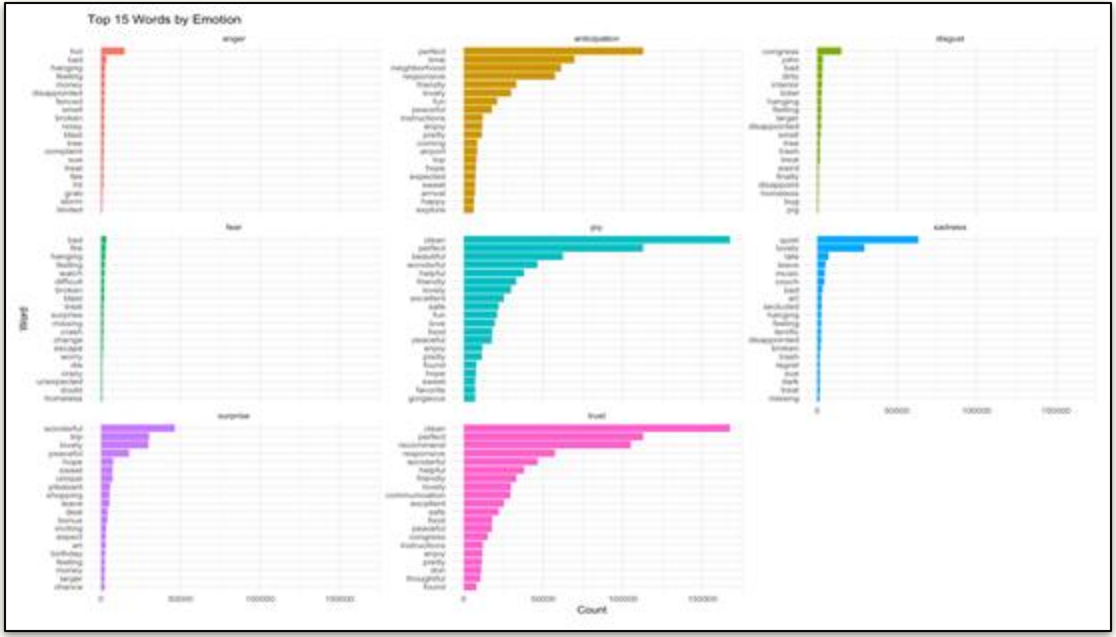
Sentiment Analysis



Comment Sentiment Score Over Time

- **Methodology:** functionget_sentiments and Bing lexicon for scoring
- **Results**: positive relationship between sentiment score and time; increased variability

Sentiment Analysis



Top 15 Words per Emotion

- **Methodology:** function get_sentiments and NRC lexicon
- **Results**: large number of words associated with positive emotions; interesting words associated with negative emotions

Diversity and Inclusion Analysis

Data Dictionary

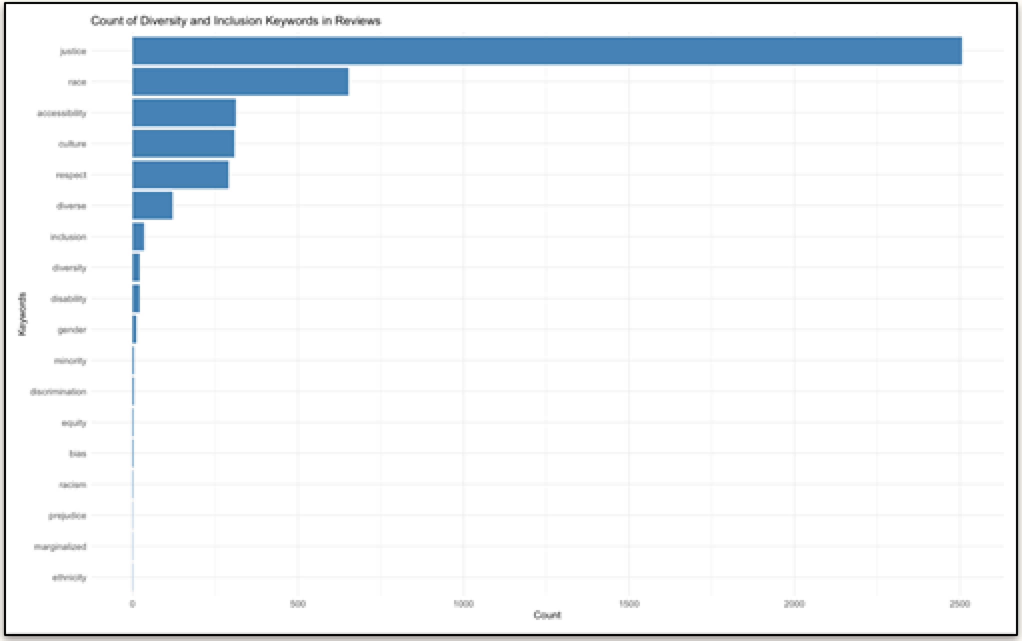
Custom List of Words: keywords_di
"diverse", "inclusion", "culture", "accessibility", "respect", "ethnicity", "race", "racism", "sexism", "homophobia", "transphobia", "discrimination", "equality", "justice", "prejudice", "bias", "marginalized", "minority", "equity", "gender", "sexuality", "disability", "intersectionality", "diversity", "inclusive"

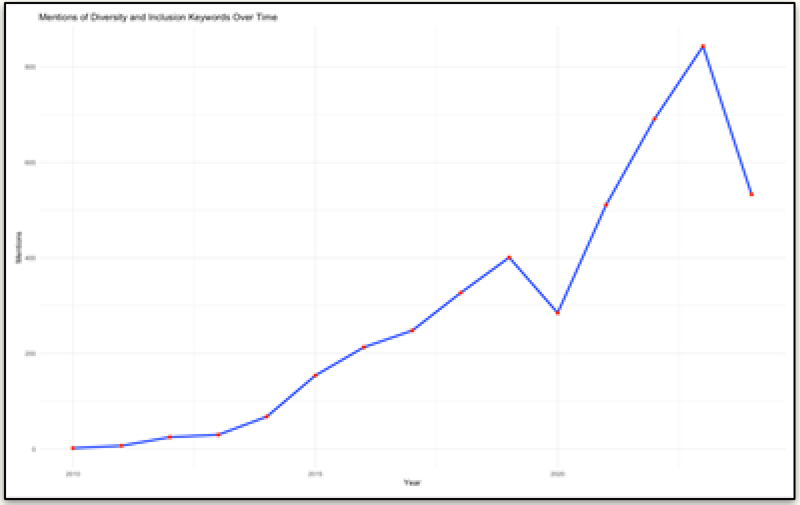
Section 2

Section 2

D&I Count

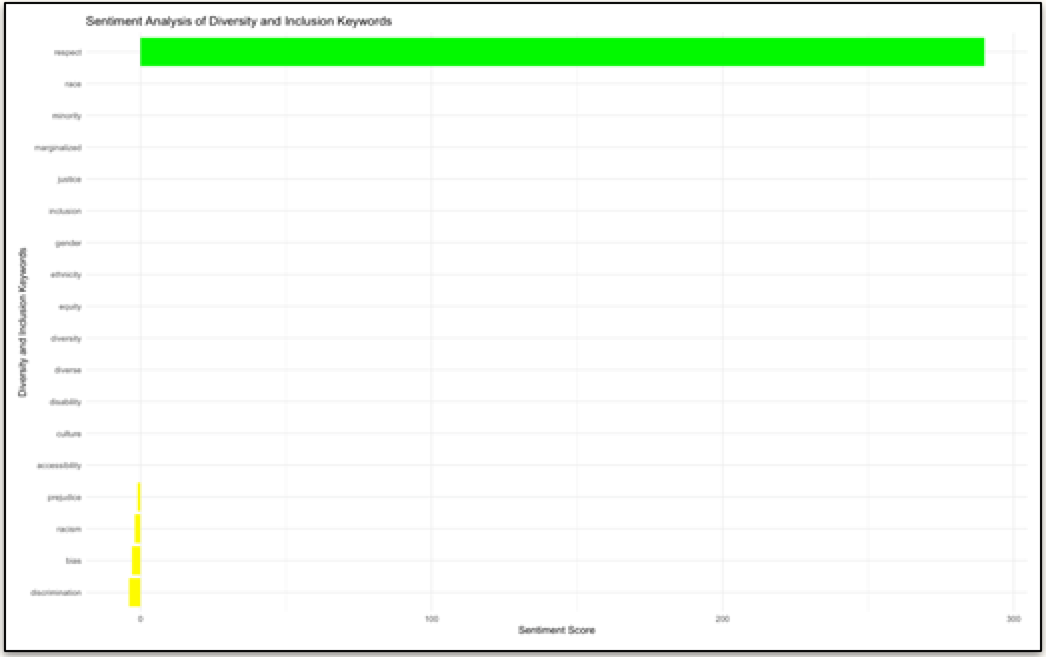
- **Methodology:** count, ggplot2



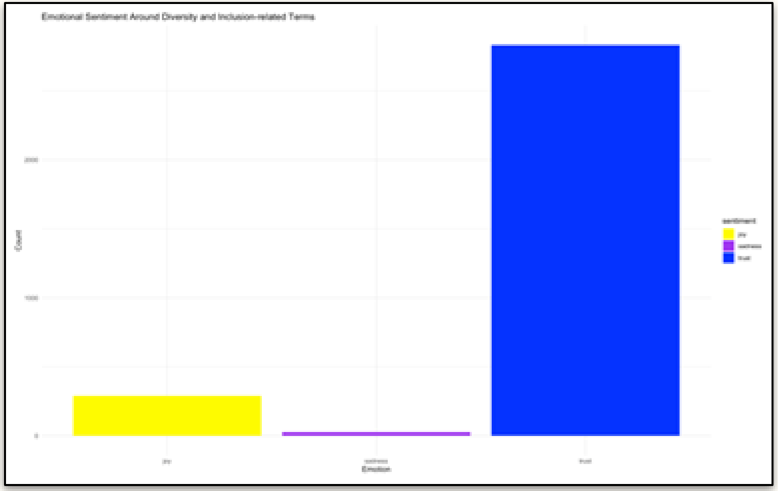


Frequency of Diversity Terms

D&I Sentiment

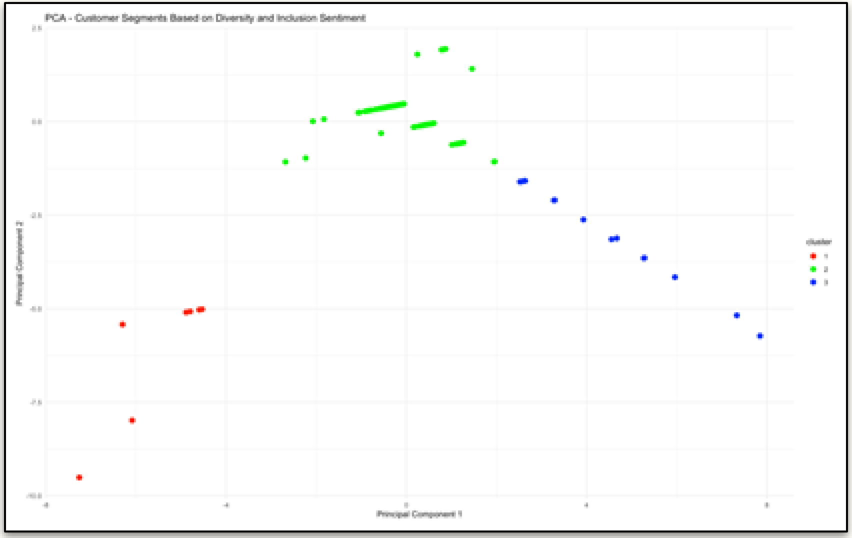


- **Methodology:** get_sentiments, nrc lexicon, ggplot2



Sentiment of Diversity Terms

D&I K Means



K Means Segmentation Using Diversity Terms

- **Methodology:** k means,, ggplot2

- **C1 -** negative sentiment, longer reviews, low number of positive reviews
- **C2** - positive sentiment, moderately long reviews, some positive reviews
- **C3** - slightly positive sentiment, short reviews, many positive reviews

A tibble: 3 × 6

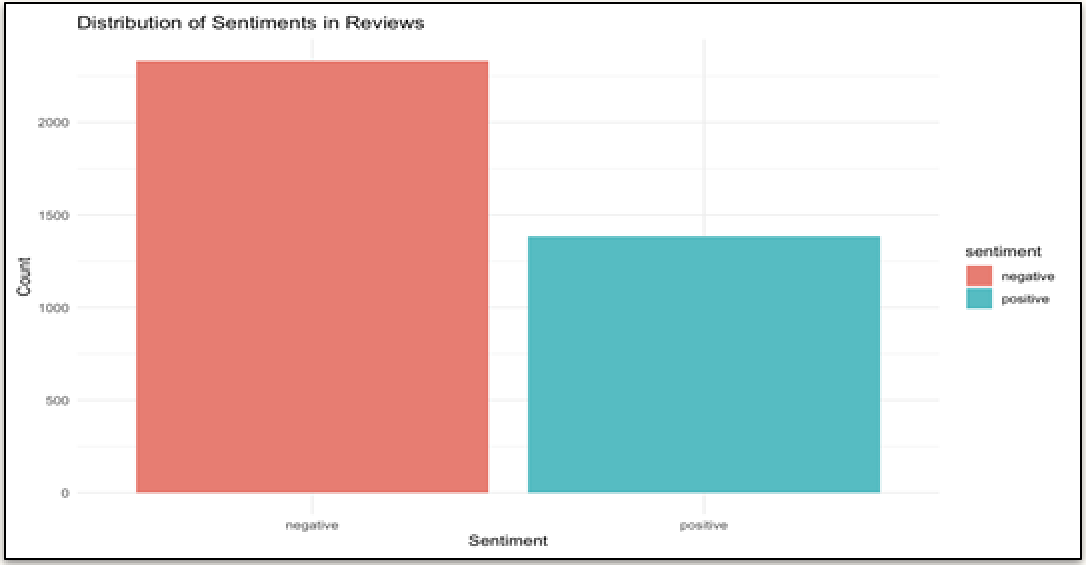
cluster <int>	avg_sentiment <dbl>	avg_review_length <dbl>	avg_positive_reviews <dbl>	avg_negative_reviews <dbl>	total_customers_in_cluster <int>
1	-1.142857	1381.8571	0.000000	1.285714	7
2	1.018727	635.6451	1.220974	0.000000	267
3	1.000000	261.8431	7.437500	0.000000	16

3 rows

Customer Satisfaction Analysis

Frequency Analysis

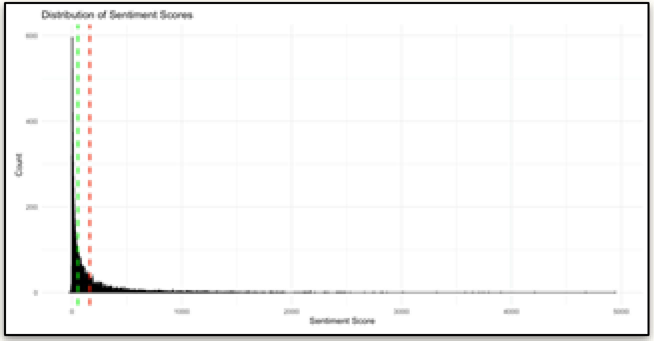
Section 2



A tibble: 1 x 3

min_sentiment <dbl>	max_sentiment <dbl>	avg_sentiment <dbl>
-27	4947	162.3197

1 row

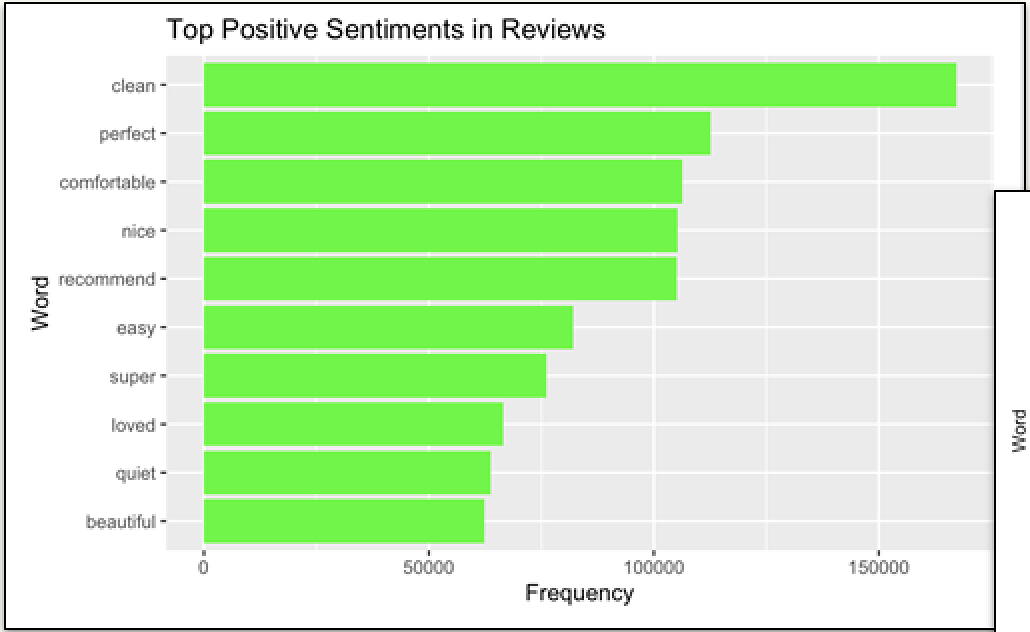


- **Methodology:** get_sentimemts, ggplot2

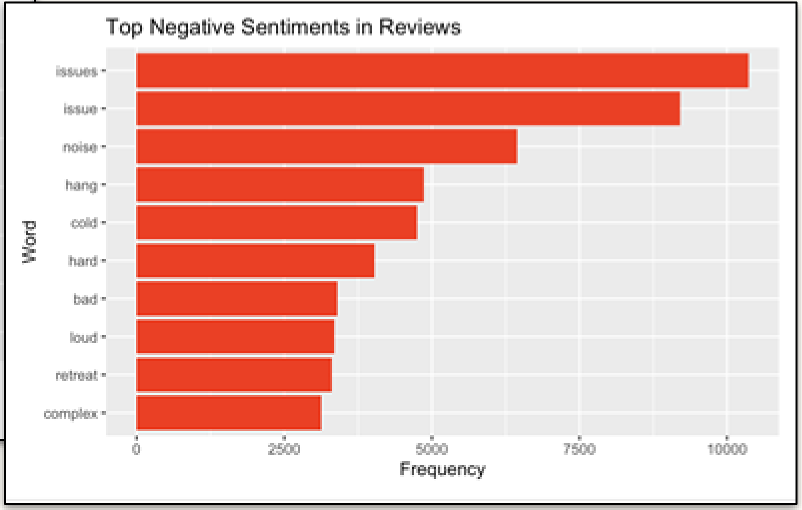
Sentiment Distribution Across Reviews

Frequency Analysis

Section 2



- **Methodology:** get_sentimemts, ggplot2



Top 10 words by Positive and Negative Sentiment

Host and Neighborhood Influence

Count Analysis

A tibble: 12,167 × 3 Groups: listing_id [12,167]

listing_id <chr>	word <chr>	n <int>
1072736146967078528	beautiful	2
1072996921564359040	beautiful	1
1073114645744749056	hugo	2
1073336860615885056	austin	3
10733649	jinny	6
1073377811934184704	bed	4
1073402537429639168	stay	16
1073488432495931392	location	7
1074041615462006912	sandy	12
1074064	harshan	8

651-660 of 12,167 rows Previous 1 ... 64 65 66 67 68 ... 100 Next

Most Common Word per Listing

- **Methodology:** count, group_by
- **Results**: large number of reviews mention the host name, indicating that the host has a strong effect on the guest's experience

Count Analysis

A tibble: 14 x 2

word <chr>	n <int>
location	200180
clean	167176
quiet	63698
responsive	57333
helpful	37776
friendly	32971
convenient	27402
safe	21821
welcoming	13661
central	8525
accessible	4787
organized	3745
hospitable	3513
noisy	1996

14 rows

Host and Neighborhood Keyword Frequency

- **Methodology:** keyword, count
- **Results**: repeated mention of words associated with hosts and neighborhoods

Important Findings

Section 3

Key Insights

1. Comment Trends (2009–2024):

- **Average comment length:** 221.3 characters.
- **Length trend:** High 300s–400s (2010–2016) → Mid-200s (2016–present).
- **Review growth:** From single digits/month (2009–2012) → 18,840 reviews (April 2024).
- No clear link between comment length and satisfaction.

2. Word Frequency Analysis:

- Top words: *stay, location, host, Austin, clean, neighborhood, perfect*.
- Key themes: Importance of hosts and location in guest satisfaction.
- Quality-related words (*clean, comfortable, beautiful*) suggest a strong reputation.

Section 3

Sentiment and Diversity Analysis

1. Sentiment Trends (2009–2024):

- Overall sentiment: Increasingly positive with greater variability.
- Key positive terms: *clean, perfect, quiet, loved*.
- Negative themes: *noise, complicated processes, cold, bad*.

2. Diversity & Inclusion:

- Rise in diversity-related mentions: 0 (2010) → 800+ (post-2020).
- Top terms: *justice, respect, accessibility, culture, race*.
- Sentiment clusters (K-Means):
  - **Cluster 2 & 3:** Positive sentiments about inclusivity.
  - **Cluster 1:** Negative sentiments highlighting discrimination concerns.

Conclusions

Key Drivers of Satisfaction:

- Positive factors: Cleanliness, communication ease, quiet environments, and scenic locations.
- Negative factors: Noise, process complexity, and discomfort.

Impact of Hosts and Neighborhoods:

- Top mentions: *location (200K), Austin (185K), host (115K)*.
- Hosts' practices and local features heavily influence reviews.

Section 3

Recommendations for Hosts

Section 3

- **Enhance Communication:**
  - Ensure timely responses and provide clear check-in instructions
- **Focus on Cleanliness and Comfort:**
  - Regularly inspect and clean properties to maintain high standards
- **Address Noise Issues:**
  - Use soundproofing measures where possible
- **Leverage Positive Sentiments:**
  - Highlight amenities like scenic views and a quiet environment in listings

Recommendations for Planners

Section 3

Improve Neighborhood Appeal:

- Enhance accessibility and safety in popular Airbnb locations.
- Collaborate with local businesses to boost the appeal of the area for visitors.

Promote Diversity and Inclusion:

- Support cultural events and local initiatives that emphasize inclusivity.
- Offer resources for creating universally accessible accommodations.

Section 3

Limitations and Future Research

The End

Limitations:

- Data is specific to Austin and may not generalize to other cities.
- Sentiment analysis relies on text, which may not fully capture guest experiences.

Areas for Improvement:

- Analyze additional cities for comparative insights.
- Include more nuanced metrics for diversity and inclusion beyond keywords.

Recommended Future Research Focus:

- Investigate seasonal trends in satisfaction and review counts
- Explore relationships between property types and guest experiences

Questions?