Executive Summary

This report analyzes Airbnb listing descriptions to uncover how language supports pricing strategies and influences perceived property value. Listings were categorized as cheap, normal, or expensive based on a price-per-guest metric. Key findings show that expensive listings use more positive sentiments and luxurious vocabulary to justify premium pricing, while cheaper listings emphasize location and functionality. Expensive listings highlight service, prestige, and exclusivity; cheap listings often include non-English words and terms that suggest practical benefits or safety cues. Sentiment analysis confirms positive sentiments across all listings, with luxury properties using the most optimistic language. Visualizations like TF-IDF, correlograms, sentiment scoring, and bigram networks support these insights in a dashboard. These findings offer actionable strategies for hosts, like tailoring descriptions to signal appropriate value for each pricing tier, balancing positivity with practical information to optimize listing performance.

TF-IDF Analysis

The first visualization in this study uses a TF-IDF analysis (Figure 1) to uncover the most distinctive words for each property type. This visualization clearly illustrates how language varies with price. Cheap listings are often filled with Chinese characters, other nationalities like ‘Mexican’, or time spams that maybe are to places potentially close to famous sightings, and even the word ‘police’, potentially indicating that cheap properties are in dangerous areas, and police is a token to make customers think that the area is safer. In contrast, expensive listings use words like ‘kapalua’ and other Hawaiian words, suggesting that most expensive listings are in Hawaii, ‘generous’, and ‘management’, which are service-related, and ‘gold’, indicating luxury and expensive things. These words not only indicate exclusivity but also prestige and status. Normal listings use functional but less emotionally powerful words like ‘toddler’, ‘lodge’, and ‘vacation’. This tier appears to focus on practical descriptions over showing luxury, and also indicates more like properties more made for families and tourists.

Correlogram analysis

The correlogram (Figure 2) compares word frequency proportions between cheap, normal, and expensive listings. We can see some clear outliers like ‘villa’, ‘resort’, ‘pool’ for expensive properties, suggesting more exclusivity. For cheap, we can see ‘basement’ or ‘24h’, which may indicate properties that are in a house and that are for the short term. And in normal we can find ‘Astoria’ which indicates possibly a neighborhood with many normal price properties or ‘cute’ which may be a good adjective to identify the property type, in contrast to expensive that might be looking for adjectives like luxurious or prestigious. Both cheap and normal have a high frequency in tokens with minutes, potentially indicating that tourists choose those properties because they want to be near famous landmarks.

Sentiment Analysis AFINN, Bing, and NRC

The sentiment analysis (Figure 3) applies three lexicons to evaluate the emotional tone of listings. All three price categories display positive sentiment across all lexicons, affirming that Airbnb descriptions are written with the intent to persuade and attract. Expensive listings, however, exhibit the highest sentiment scores, particularly in the AFINN lexicon, where numeric intensity plays a key role. The dominance of positive words like ‘beautiful’, ‘enjoy’, and ‘quite ’ in all listing types reflects a marketing orientation. Having more emotion in expensive listings suggests an effort to match premium pricing with premium appeal.

Bigram Analysis

The bigram network graph (Figure 4) reveals that listings form dense clusters such as ‘short’,’ walk’,’ minutes’, and numbers between 5 and 20 that confirm our initial hypothesis that descriptions that are considered cheap or normal prioritize sinilazing that it is in a location walking distance and nearby tourist places. Another interesting cluster is ‘view’, ‘ocean’ and ‘mountain’, indicating that clients value property with some kind of views, either the ocean if it is on the beach or the mountain. Lastly, an assumption that was 'village' is a token referring to 'East Village', which is a neighborhood in New York, not a property in a village.

Price Setting through Correlation Analysis

Using correlation (Figure 5), it was found that cheap listings had a correlation coefficient of 0.744 between cleaning fee and price, normal listings 0.758, and expensive listings 0.628. These results suggest that for affordable and mid-range listings, cleaning fees are strongly tied to the overall cost. In contrast, luxury listings show a weaker correlation, potentially because pricing is less itemized and more influenced by bundled services or brand prestige.

Additional Analytical Insights

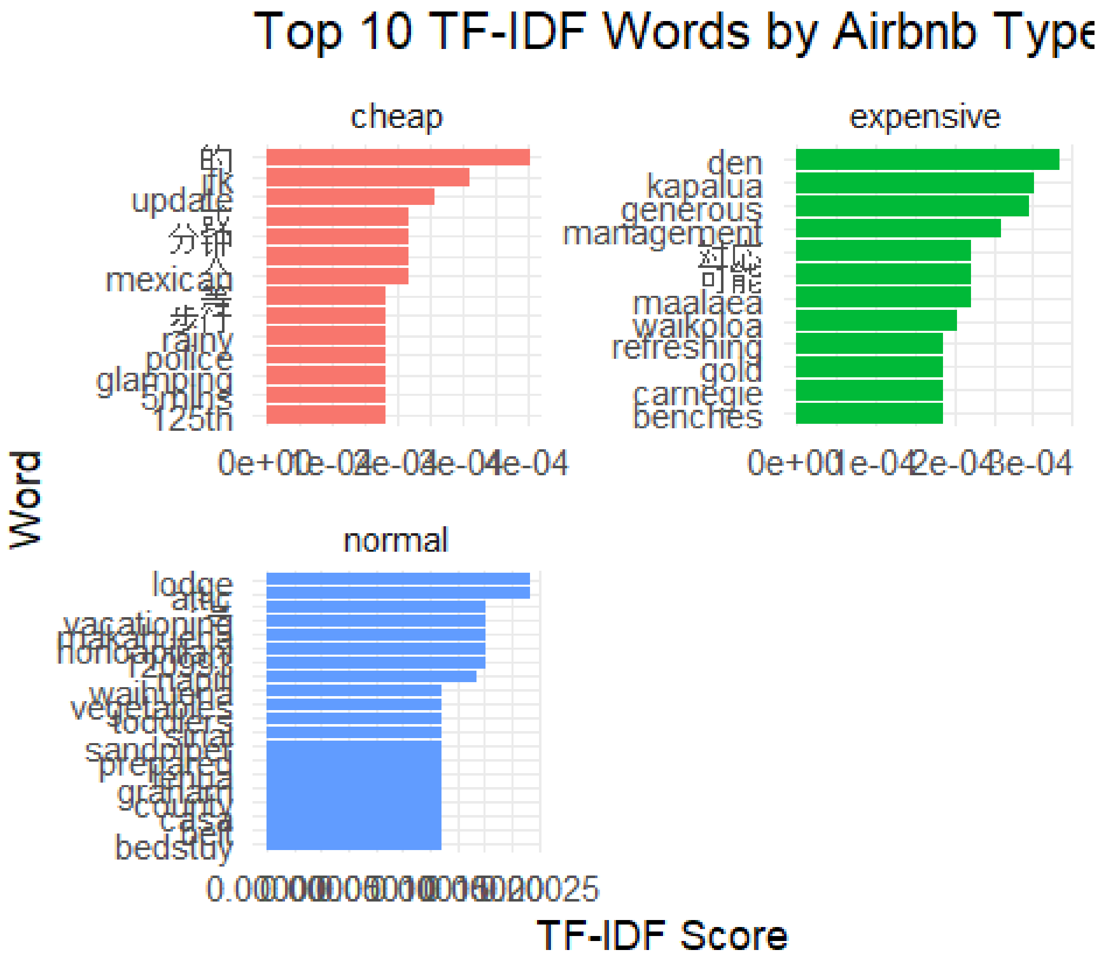
A Zipf's Law plot (Figure 6) confirms that all Airbnb listing types follow a common linguistic distribution, but normal ones have more unique tokens. Additionally, topic modeling through LDA (Figure 7), I found topics like nature, functional space, and convenience. These themes show that Airbnb listings focus on common user needs but are weighted differently according to the listing type.

Conclusion and dashboard

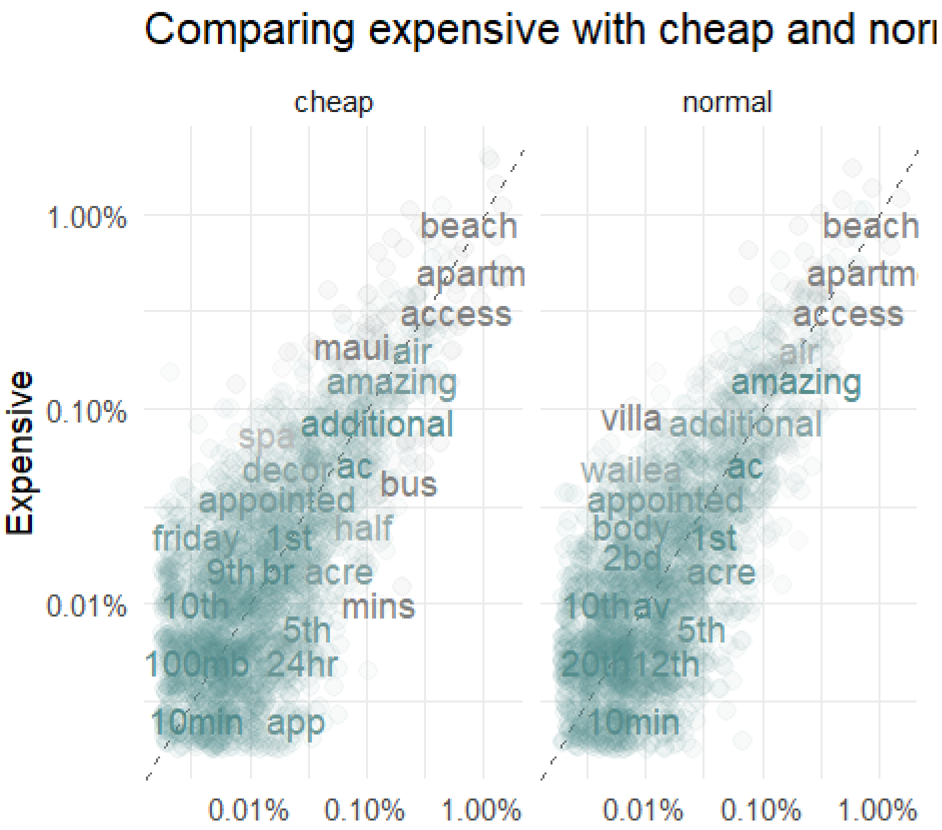
The Airbnb dashboard (Figure 8) brings together key visualizations that reveal how listing language aligns with pricing strategy. The TF-IDF/Quadrogram panel shows that expensive listings use exclusive, luxury-related tokens. The correlogram compares word proportions. The sentiment score chart shows that expensive listings are written in more emotionally positive tones. Finally, the bigram network shows clusters around walkability, scenic views, and convenience. The dashboard supports the conclusion that expensive listings benefit from persuasive, optimistic language, while value-tier listings are best served by emphasizing location and practical benefits.

Appendix

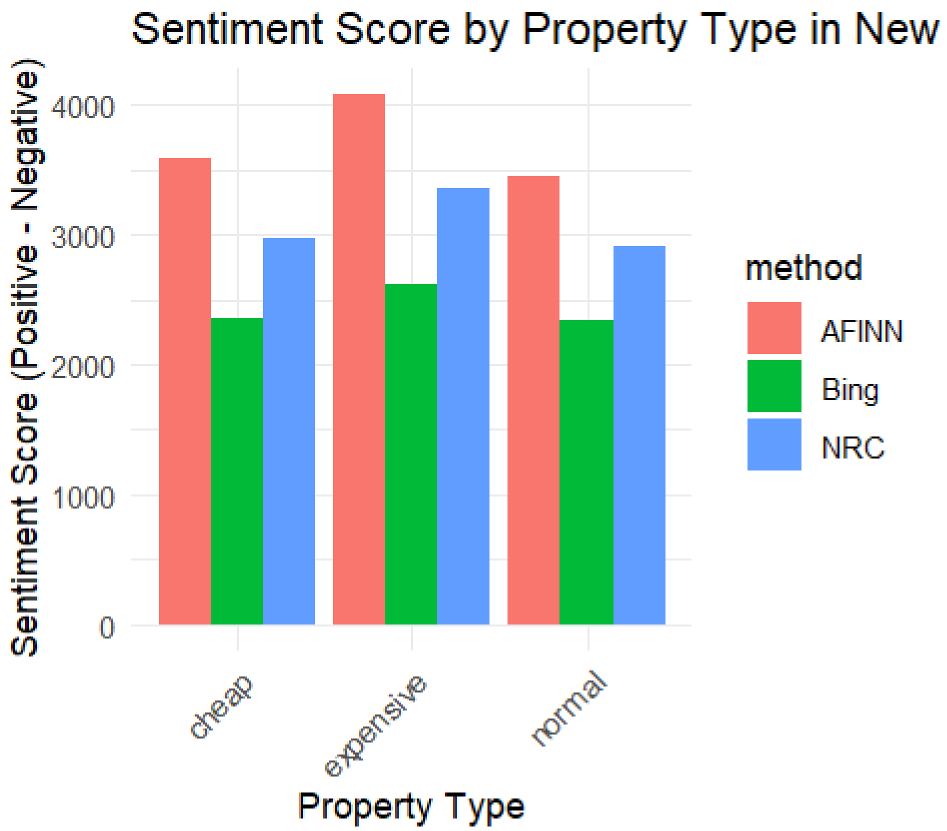
(1)


Top 10 TF-IDF Words by Airbnb Type
cheap
的
jfk
update
致
分钟
人
mexican
等
步行
rainy
police
glamping
5mins
125th
expensive
den
kapalua
generous
management
对应
可能
maalaea
waikoloa
refreshing
gold
carnegie
benches
normal
lodge
bedstuy
Word
TF-IDF Score

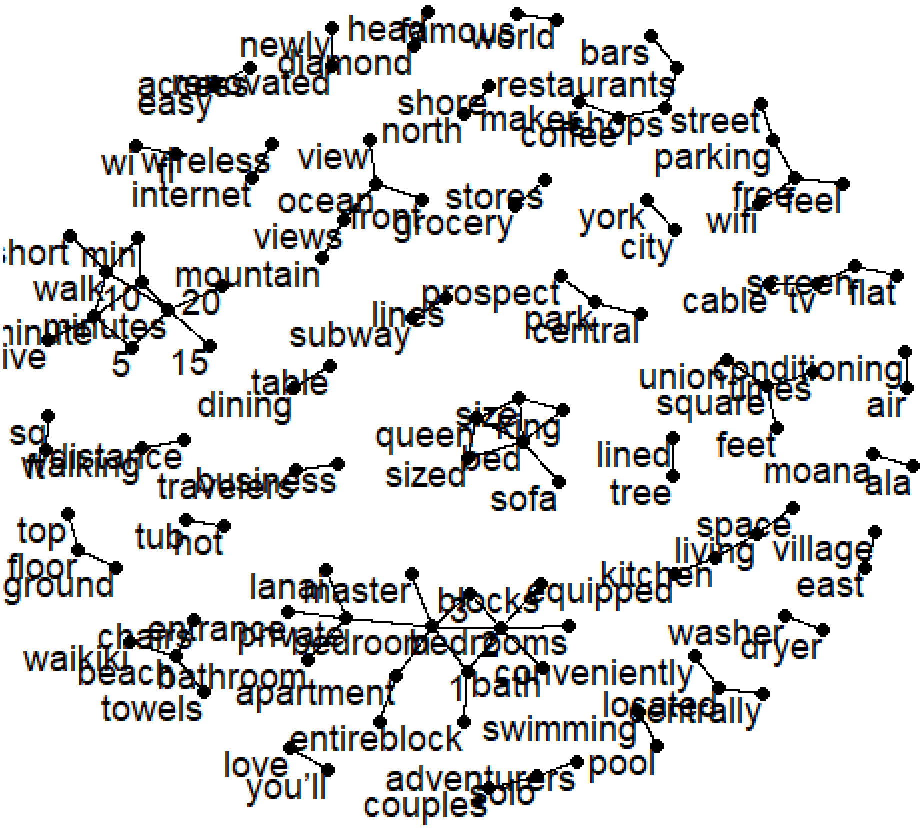
(2)


Comparing expensive with cheap and nor
cheap
normal
Expensive
1.00%
0.10%
0.01%
0.01%
0.10%
1.00%
beach
apartm
access
maui
air
amazing
additional
spa
decor
ac
bus
appointed
half
friday
1st
9th
br
acre
10th
mins
5th
100mb
24hr
10min
app
villa
wailea
body
2bd
10th
av
20th
12th

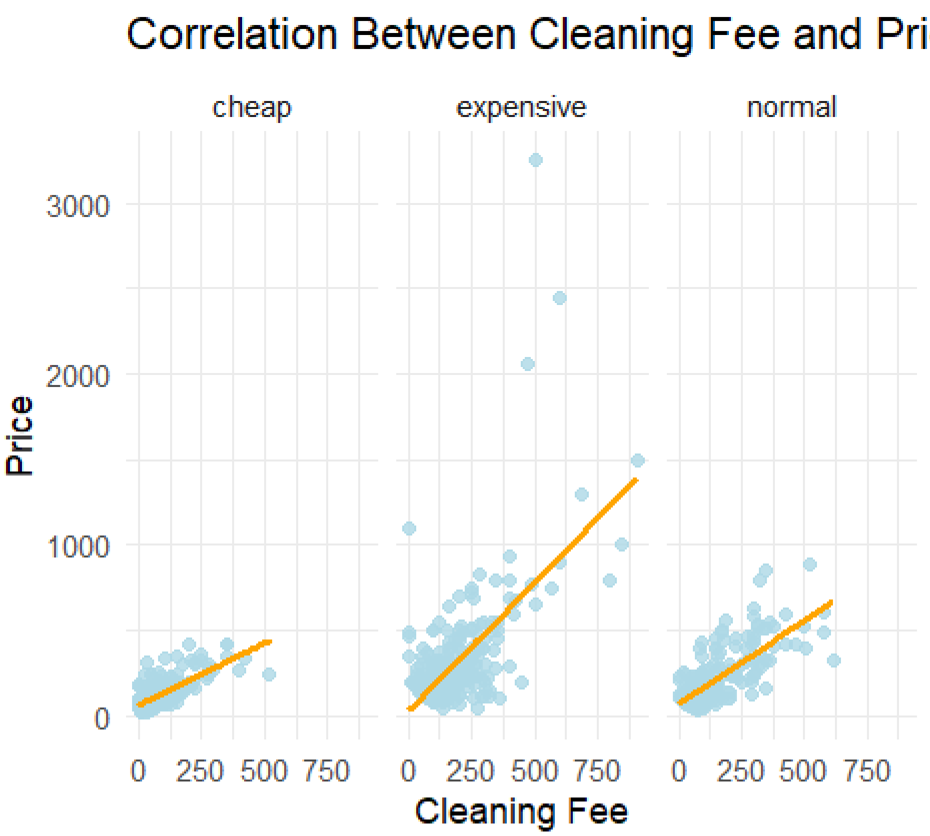
(3)


Sentiment Score by Property Type in New
Sentiment Score (Positive - Negative)
4000
3000
2000
1000
0
cheap
expensive
normal
Property Type
method
AFINN
Bing
NRC

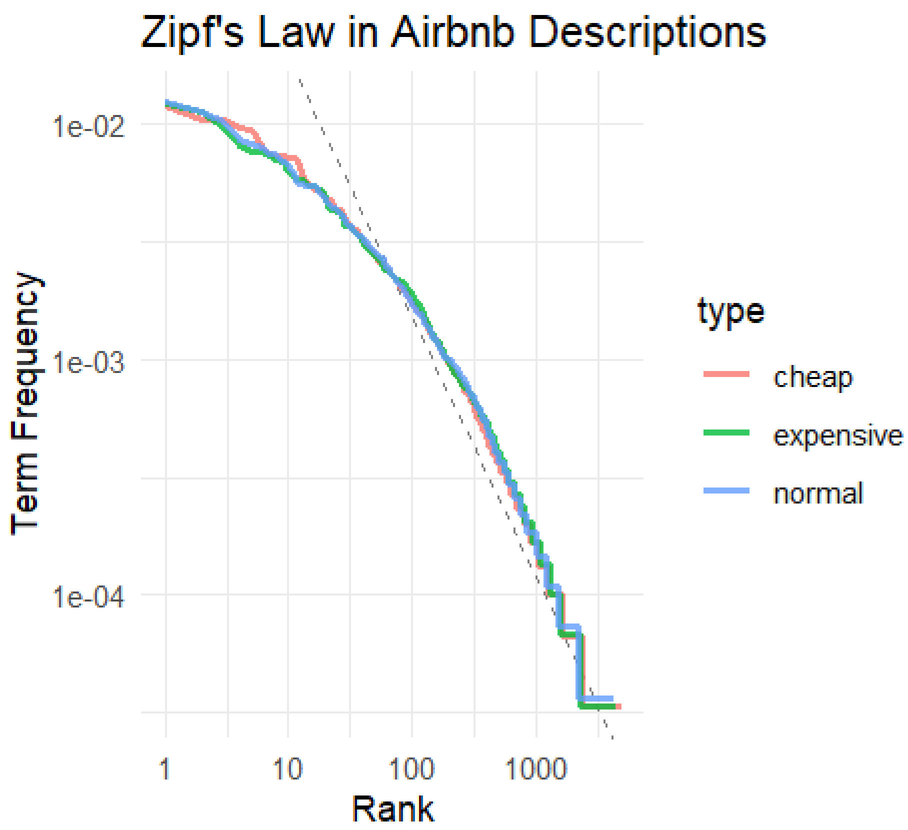
(4)


bars
restaurants
street
parking
wifi
york
city
view
north
ocean
views
stores
grocery
internet
mountain
subway
prospect
park
central
cable
tv
flat
union
square
air
dining
table
queen
sized
sofa
king
bed
lined
tree
feet
moana
ala
business
top
floor
ground
tub
hot
space
living
village
east
kitchen
equipped
master
washer
dryer
waikiki
beach
towels
apartment
bathroom
conveniently
swimming
pool
love
you'll
couples
entire
block

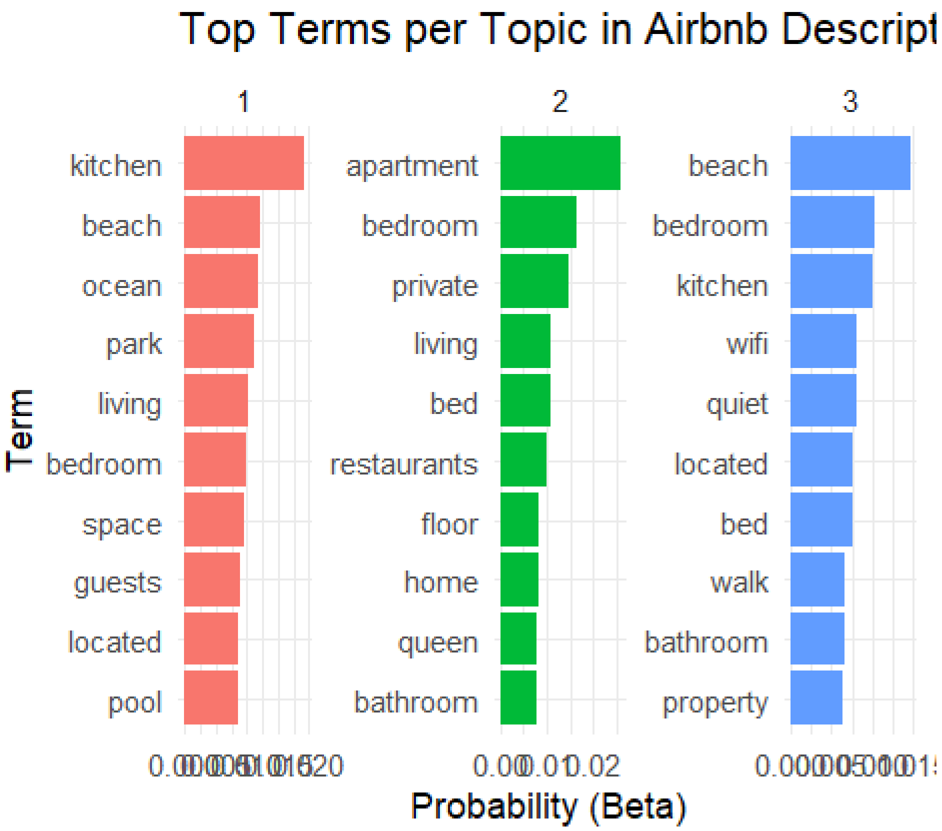
(5)


Correlation Between Cleaning Fee and Pric
cheap
expensive
normal
Price
3000
2000
1000
0
0 250 500 750
0 250 500 750
0 250 500 750
Cleaning Fee

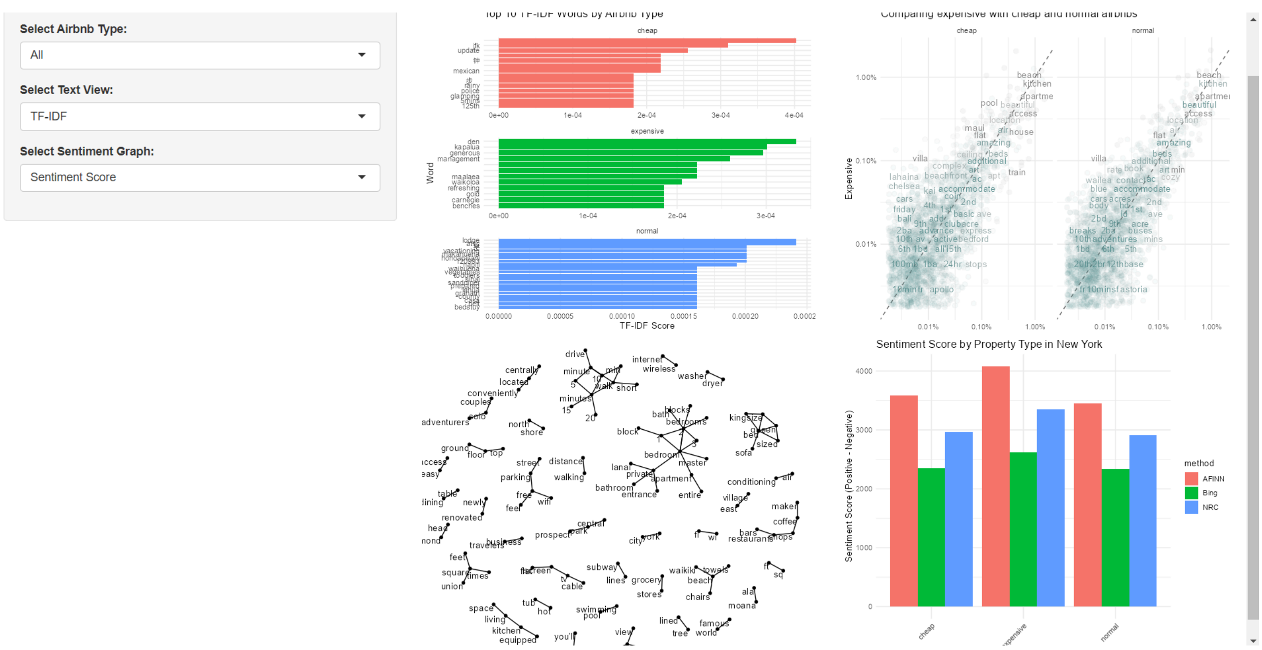
(6)


Zipf's Law in Airbnb Descriptions
Term Frequency
1e-02
1e-03
1e-04
1
10
100
1000
Rank
type
cheap
expensive
normal

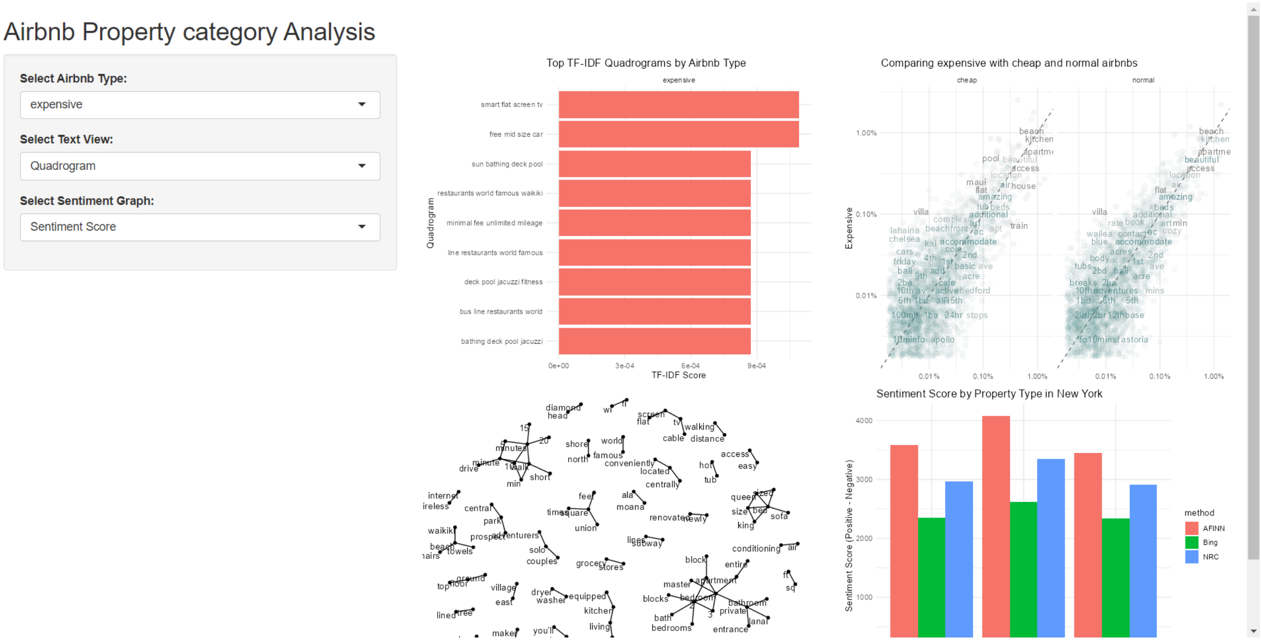
(7)


Top Terms per Topic in Airbnb Descript
1
2
3
Term
kitchen
beach
ocean
park
living
bedroom
space
guests
located
pool
apartment
bedroom
private
living
bed
restaurants
floor
home
queen
bathroom
beach
bedroom
kitchen
wifi
quiet
located
bed
walk
bathroom
property
0.00 0.01 0.02
Probability (Beta)

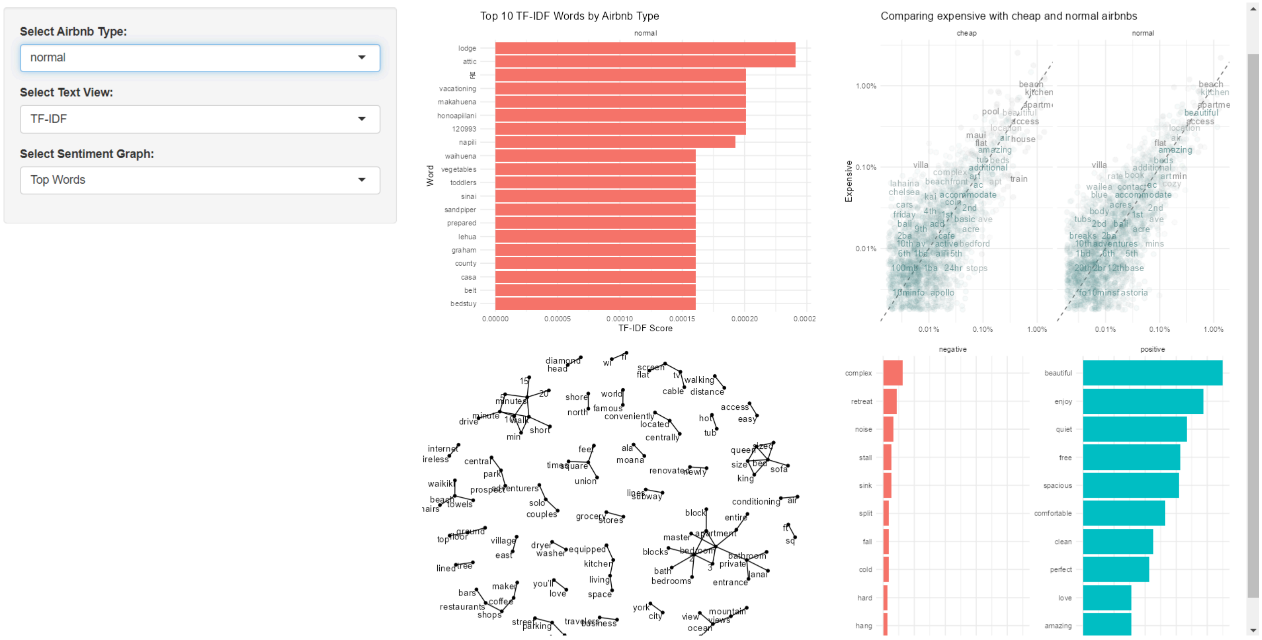
(8)


Select Airbnb Type:
All
Select Text View:
TF-IDF
Select Sentiment Graph:
Sentiment Score
cheap
expensive
normal
Word
TF-IDF Score
Expensive
Sentiment Score by Property Type in New York
Sentiment Score (Positive - Negative)
method
AFINN
Bing
NRC

(9)


Airbnb Property category Analysis
Select Airbnb Type:
expensive
Select Text View:
Quadrogram
Select Sentiment Graph:
Sentiment Score
Top TF-IDF Quadrograms by Airbnb Type
expensive
smart flat screen tv
free mid size car
sun bathing deck pool
restaurants world famous waikiki
minimal fee unlimited mileage
line restaurants world famous
deck pool jacuzzi fitness
bus line restaurants world
bathing deck pool jacuzzi
Quadrogram
TF-IDF Score
Comparing expensive with cheap and normal airbnbs
cheap
normal
Expensive
Sentiment Score by Property Type in New York
Sentiment Score (Positive - Negative)
method
AFINN
Bing
NRC

(10)


Select Airbnb Type:
normal
Select Text View:
TF-IDF
Select Sentiment Graph:
Top Words
Top 10 TF-IDF Words by Airbnb Type
normal
lodge
attic
vacationing
makahuena
honoapiilani
napili
waihuena
vegetables
toddlers
sinai
sandpiper
prepared
lehua
graham
county
casa
belt
bedstuy
Word
TF-IDF Score
Comparing expensive with cheap and normal airbnbs
cheap
normal
Expensive
negative
positive
complex
retreat
noise
stall
sink
split
fall
cold
hard
hang
beautiful
enjoy
quiet
free
spacious
comfortable
clean
perfect
love
amazing

References

OpenAI. (2024). ChatGPT (Apr 12 version) [Large language model]. https://chat.openai.com

Shiny. (2025).User interface (UI). Shiny. https://shiny.posit.co/r/getstarted/build-an-app/hello-shiny/user-interface.html