



Topic Modelling for Reviews, Sources Affecting Guests' Positive and Negative Experiences

Why reviews matter?

- Reviews **helps** guests choose their travel plans wisely and enables hosts to open their homes with confidence and **attract guests**.

The screenshot shows a modal window for Airbnb reviews. At the top, it displays a star rating of 4.71 and 7 reviews. Below this, a table lists various categories with their respective ratings. The reviews themselves are listed below, each with a user profile picture, name, date, and the review text.

Category	Rating
Cleanliness	5.0
Accuracy	4.7
Communication	4.7
Location	4.7
Check-in	4.7
Value	4.4

Reviews:

- Shane** (November 2021): Amazing location, true value!
- Chris** (October 2021): Wow. This place is just awesome. From walking down the alley to the amazing people at the front desk, this was a 5-star experience top to bottom. I come in to the City regularly, and I have definitely now found a new go-to. If you're coming from out of town, you'll really feel like you're getting a true NYC experience. For those of us familiar with the City, how could it be anything but awesome to stay at this super chill spot on the lower east side. Spring St, Prince St, Little Italy, the F up to Broadway, so many vibes 🌟
- Tristan** (October 2021): Very clean small space. Nice location
- Fred** (September 2021): Great location, modern and comfortable facility.
- Jace** (September 2021): Awesome spot!
- Sergey** (October 2021): well-designed room, quiet space
- Clara** (September 2021):

Why reviews matter?

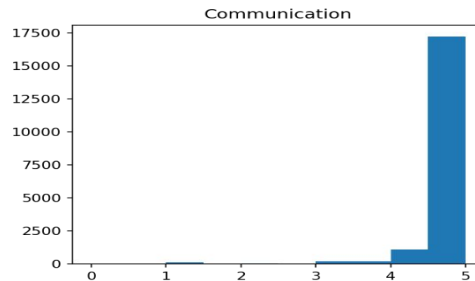
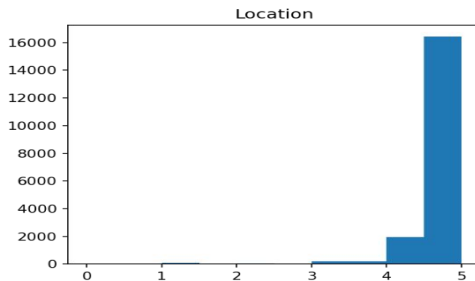
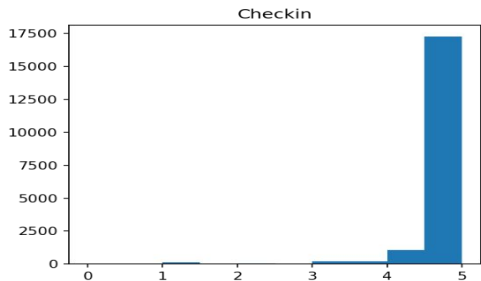
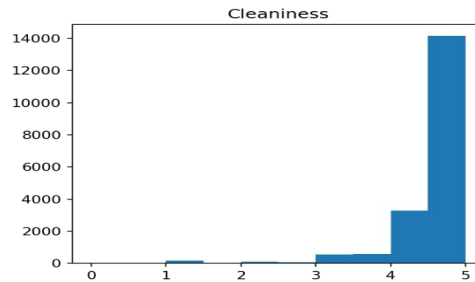
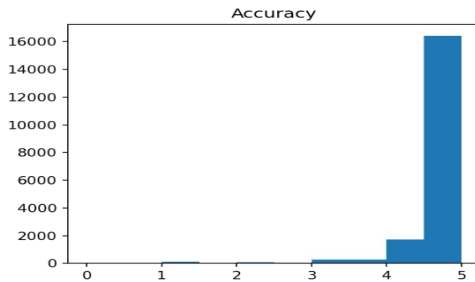
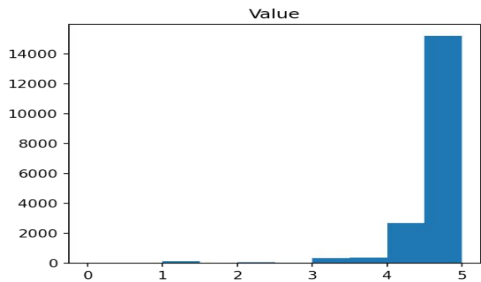
- The first thing to avoid when booking an Airbnb:

Little or No Reviews



What about ratings?

- Guests can provide star ratings with 1–5 stars on specific aspects of their experience.
- The problem:
 - All review scores are highly positive scores.
 - There is no scores less than 4.5 out of 5(see below graphs).



Methodology

Data

- NYC Listings Dataset, InsideAirbnb
- More than 80K Reviews

Data Cleaning & EDA

- Remove numbers, capital letters and punctuations
- Eliminate non-English reviews
- Lemmatize
- Tools: Pandas, Numpy, langdetect, NLTK, Matplotlib, Seaborn

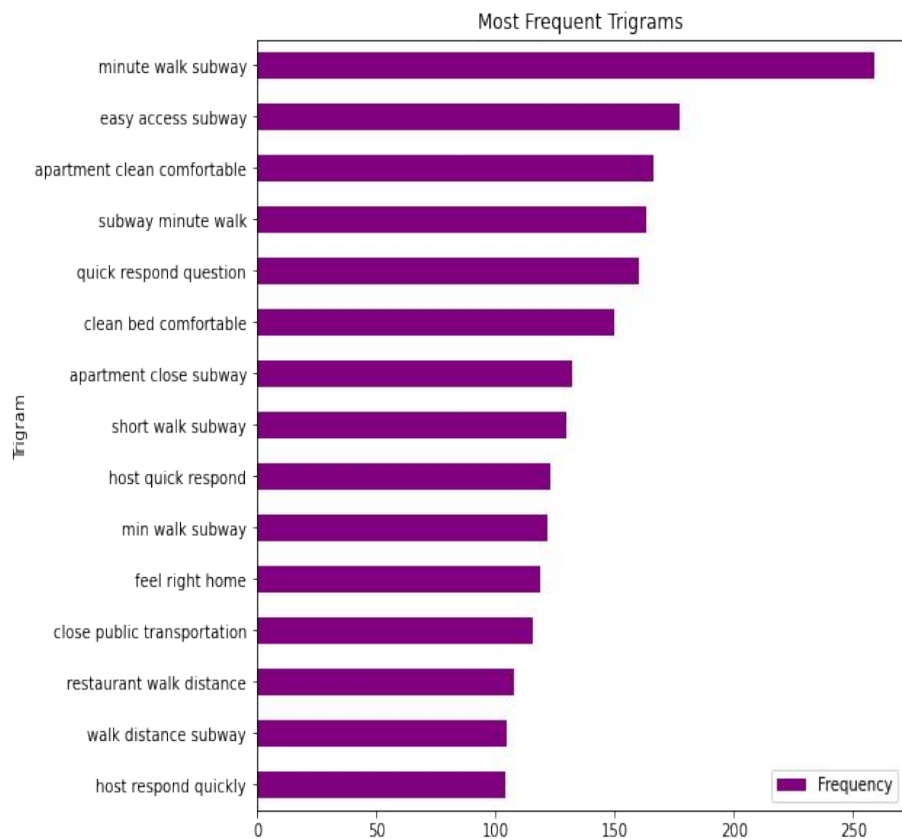
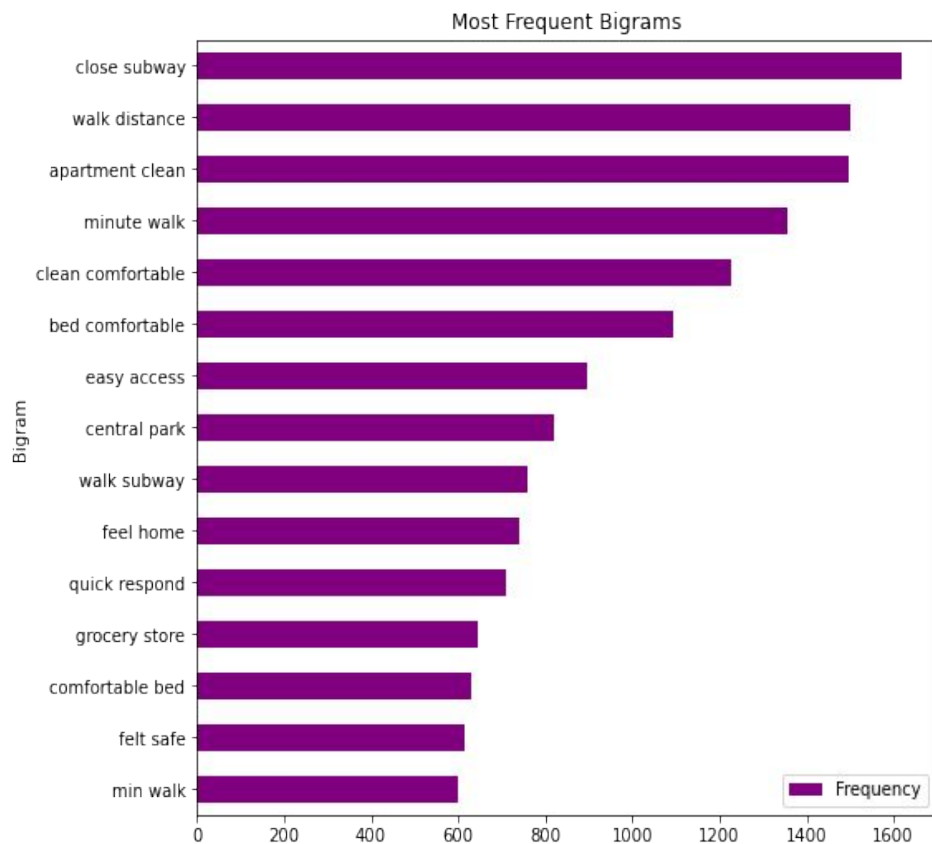
Topic Modeling

- TF-IDF Vectorizer
- CountVectorizer
- Models :
 - NMF
 - SVD
 - LDA
 - CorEx
- Tools : sklearn, pyLDAvis

Final Model

- Count Vectorizer (ngram_range=(1,2), max_df = 0.7, min_df = 10)
- LDA
- VaderSentimentAnalysis
- Tools : sklearn, pyLDAvis

Reviews: Most common words in reviews



Topic Modelling

- Vectorizer :
 - CountVectorizer
- Topic Modeler :
 - LDA
- Number of Topic 15

Topics

-  Rental interior issues
-  Kitchen Experience
-  Overall Airbnb Experience
-  Neighborhood- accessibility to transportation
-  Home-like comfort/experience
-  Neighborhood/ accessibility to dining, social attractions
-  Cleanliness
-  Host-hospitality
-  Location- safety/family friendly
-  Bed/Bathroom
-  Overall trip experience
-  Host-responsiveness
-  Convenience (check in/out, comfort, hotel like)
-  Comfort/Value
-  Listing Accuracy

Topic Visualization

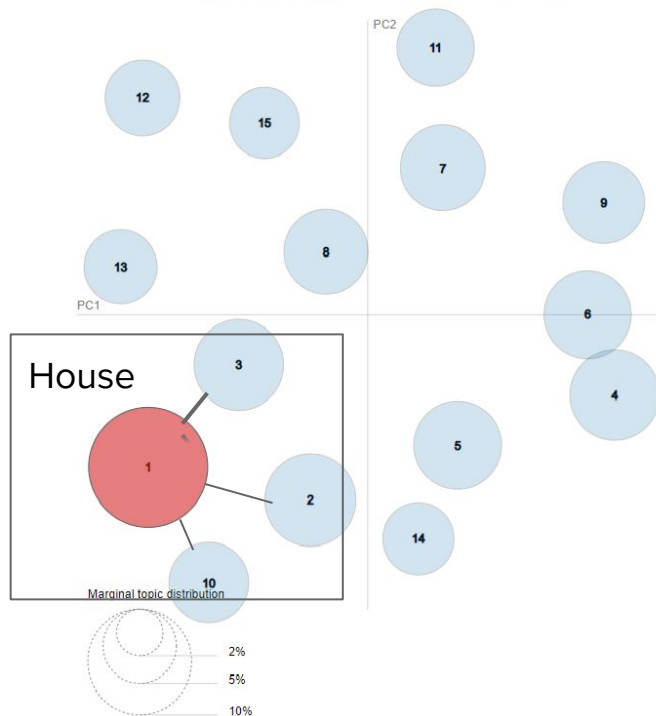
Selected Topic:

Slide to adjust relevance metric:⁽²⁾

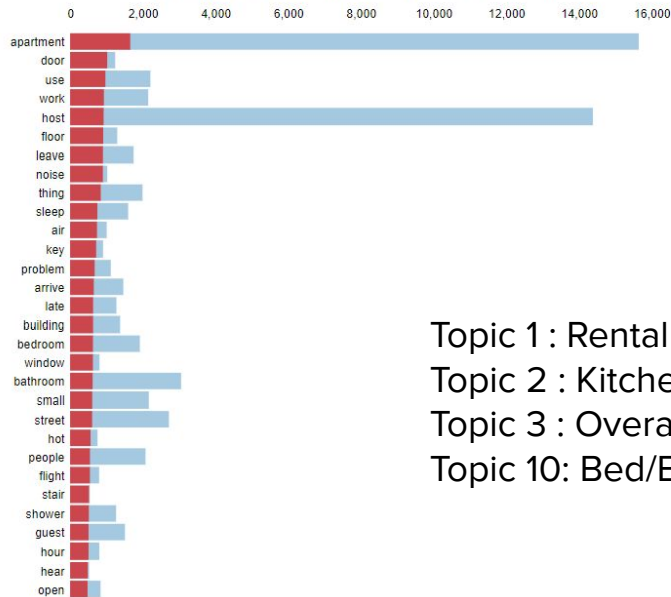
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 1 (13.2% of tokens)



Topic 1 : Rental interior
Topic 2 : Kitchen
Topic 3 : Overall experience
Topic 10: Bed/Bathroom

Overall term frequency

Estimated term frequency within the selected topic

1. $\text{saliency}(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t ; see Chuang et. al (2012)

2. $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Topic Visualization

Selected Topic: Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾

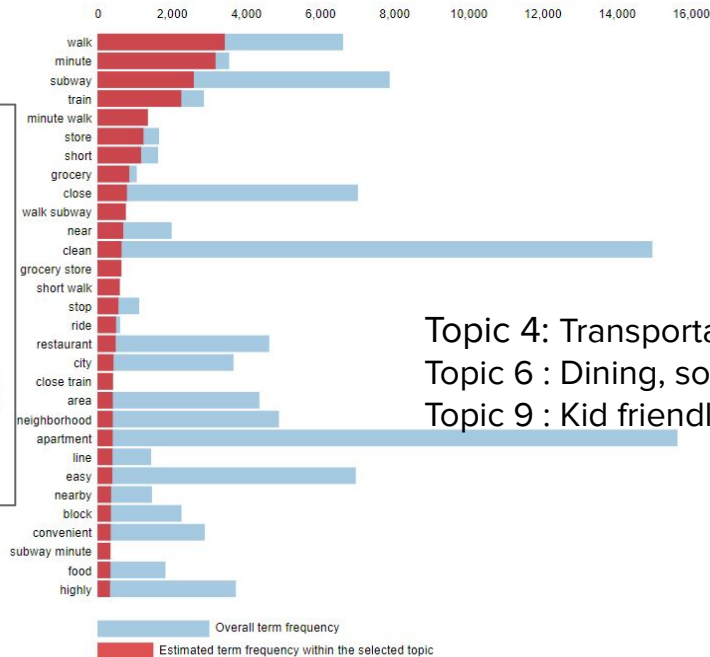
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 4 (7.4% of tokens)



Topic 4: Transportation

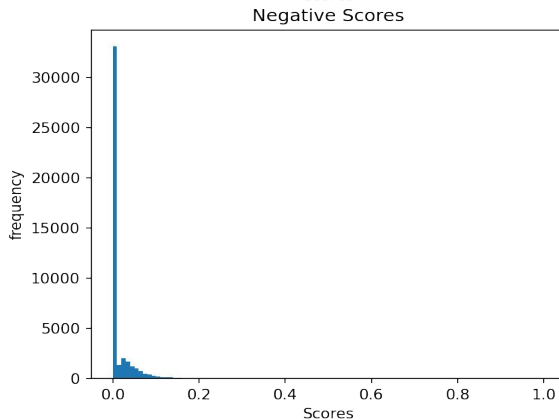
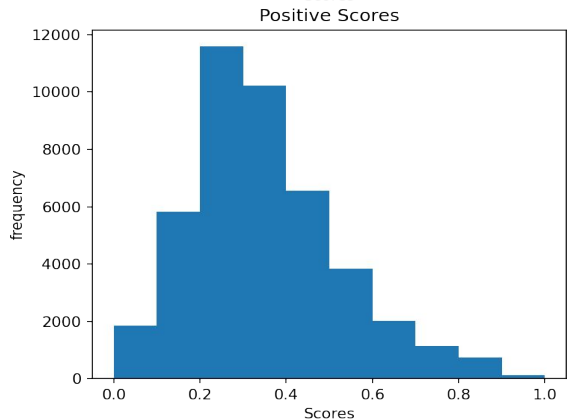
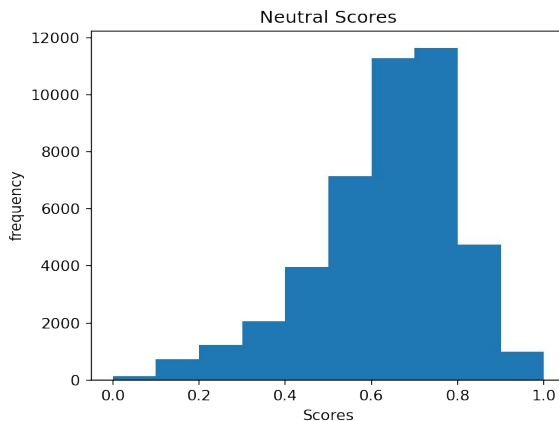
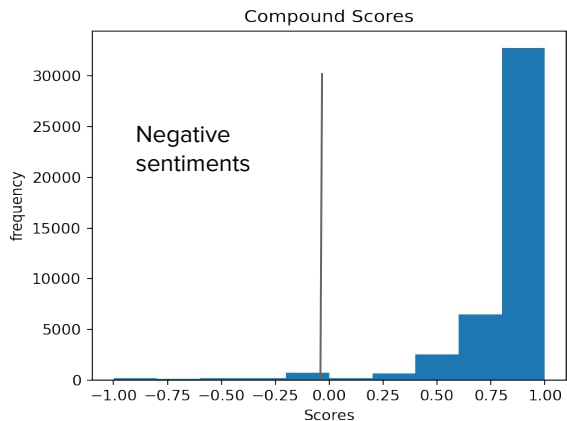
Topic 6 : Dining, social attractions

Topic 9 : Kid friendly, safety

1. $sallency(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t ; see Chuang et. al (2012)

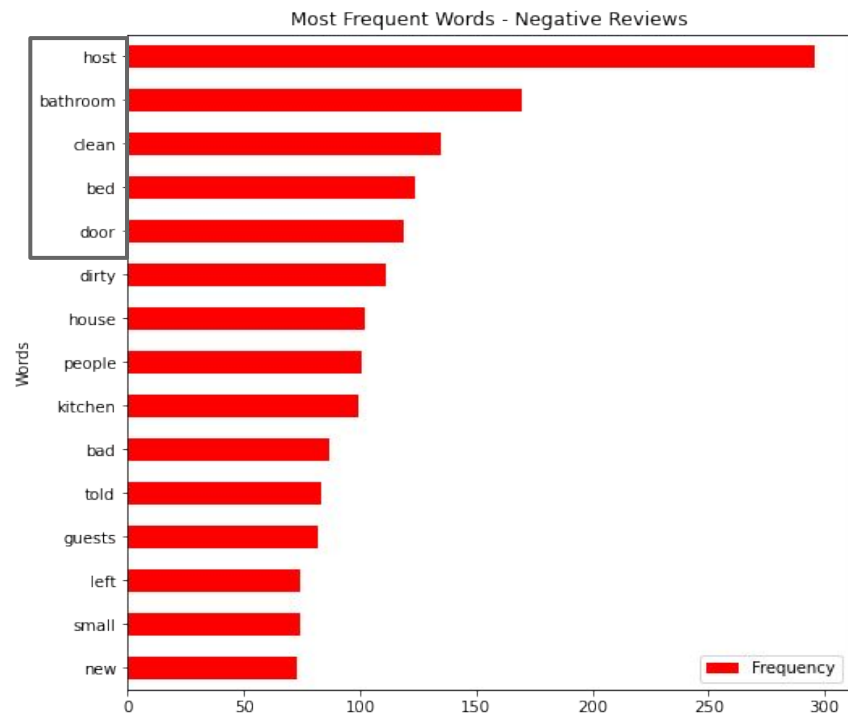
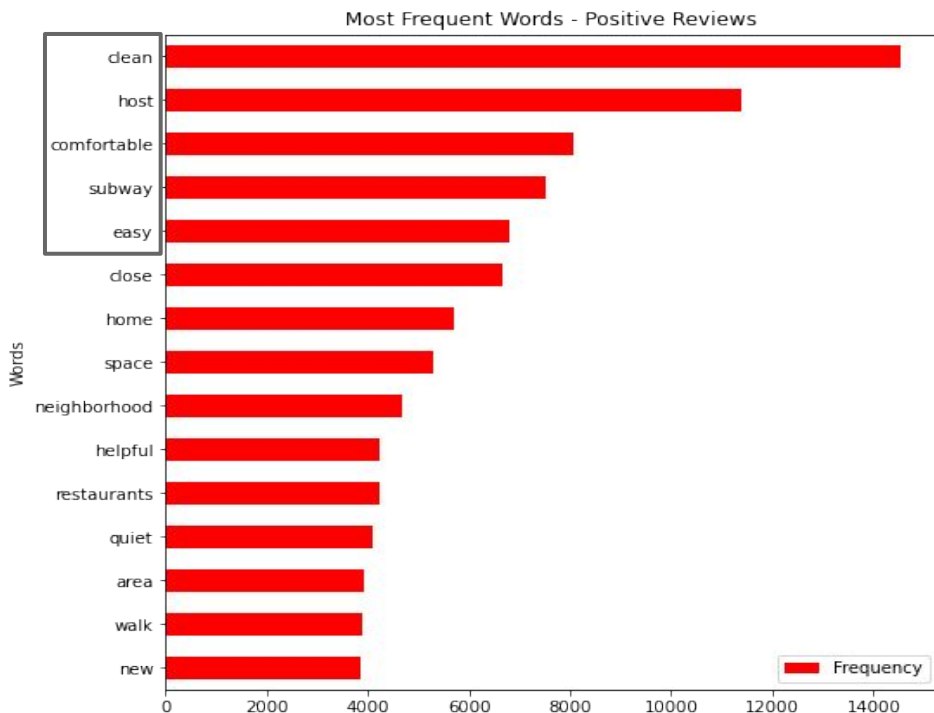
2. $relevance(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Slevert & Shirley (2014)

Reviews: Sentiment Analysis/ VaderSentiment



- Most guest had pleasant experiences during their Airbnb stays.

Positive vs Negative Reviews



Host interaction, Cleanliness → Overall experience

Insights for Hosts

Common topics in reviews

- The condition of the house
 - Kitchen, bedrooms, and bathrooms
- Cleanliness
- Location
- Host responsiveness and hospitality

Next Steps

- Cluster analysis to group topics
- Use keywords as a filtering options for the search tool
- Incorporate reviews as a feature to predict occupancy rates

Questions



Thank you!

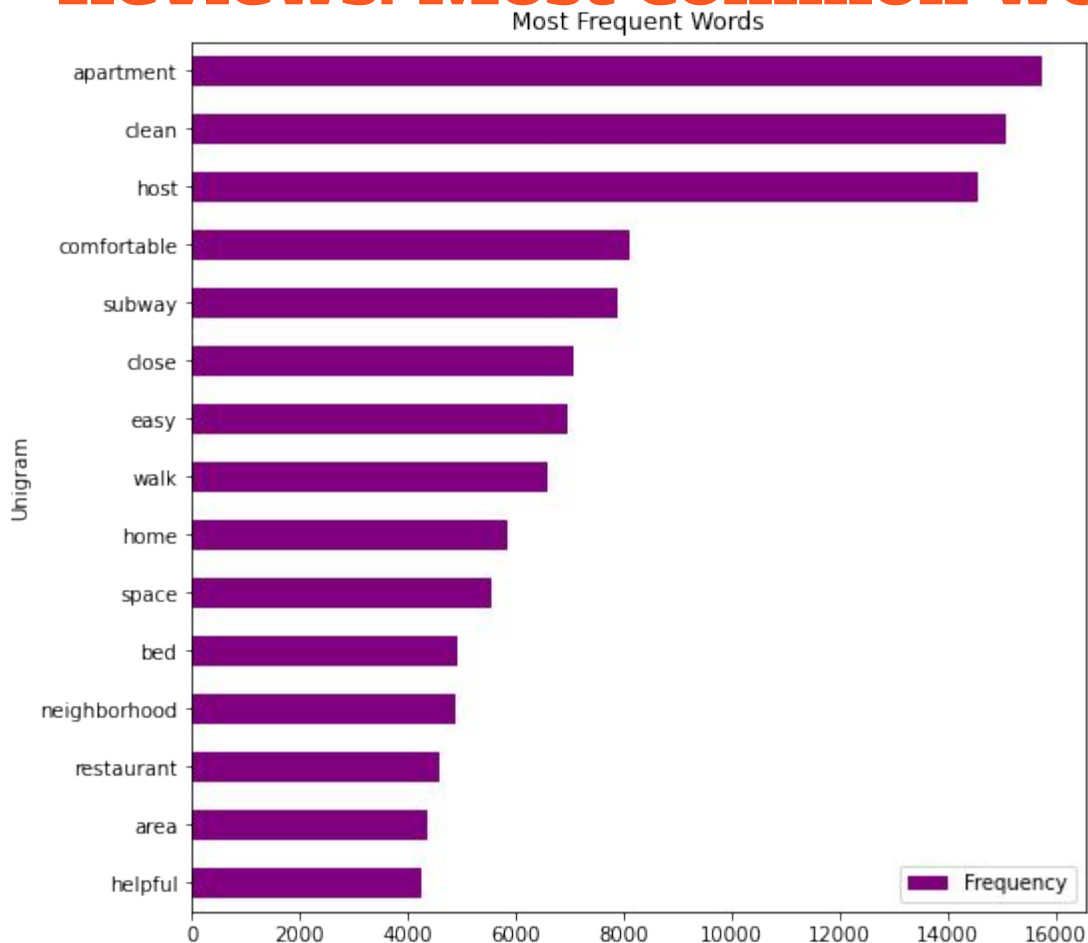
Appendix



Review Topics

- 1 Apartment interior issues
- 2 Kitchen Experience
- 3 Airbnb Experience
- 4 Neighborhood- accessibility to transportation
- 5 Home-like comfort/experience
- 6 Neighborhood/ accessibility to social attractions
- 7 Cleanliness
- 8 Host-hospitality
- 9 Location- safety/family friendly
- 10 Bed/Bathroom
- 11 Overall trip experience
- 12 Host-responsiveness
- 13 Convenience (check in/out, comfort, hotel like)
- 14 Comfort/Value
- 15 Listing Accuracy

Reviews: Most common words in reviews



Negative Topics:

- Topic 1 - Inaccurate listings
- Topic 2 - Check-in/out
- Topic 3 - Bed/Bathroom
- Topic 4 - Dirtiness and smell
- Topic 5 - Uncomfortable sleep conditions
- Topic 6 - Location
- Topic 7 - Poor house maintenance
- Topic 8 - Noise
- Topic 9 - Hot Water/Heater
- Topic 10 - Dirtiness
- Topic 11 - Location
- Topic 12 - Location general
- Topic 13 - Value

Positive Topics Topics: