AirBnB Reviews in Boston

Jianjie Zheng, Daxi Cheng, Nick Scheifler, Reece Wooten

Project Objectives

- Explore how user reviews are related to listings key metrics:
  - Occupancy Rate
  - User Ratings

- User review analysis
  - Lift Analysis
  - Sentiment Analysis
  - Cosine Similarity

Processing the Data

1.

Using stop words,identify language of reviews and select out the English ones

english	64064
french	1285
danish	1173
spanish	630
german	535
dutch	213
italian	192
portuguese	101
russian	27
swedish	18

2.

Tokenize the reviews and host descriptions and remove stop words

68,275 Reviews
3585 Listings

3.

Define Attributes and conduct search and replace in order to calculate lift and cosine similarity

Lift Attributes

location	host	experience	amenities
home	helpful	easy	room
close	friendly	comfortable	bed
neighborhood	welcoming	spacious	kitchen
area	questions	clean	bathroom
restaurants		quiet	parking
walking			apartment
city			house
convenient			
station?			
street			
downtown			
distance			
subway			
short			

Distribution of Key Metrics

Distribution of Key Metrics
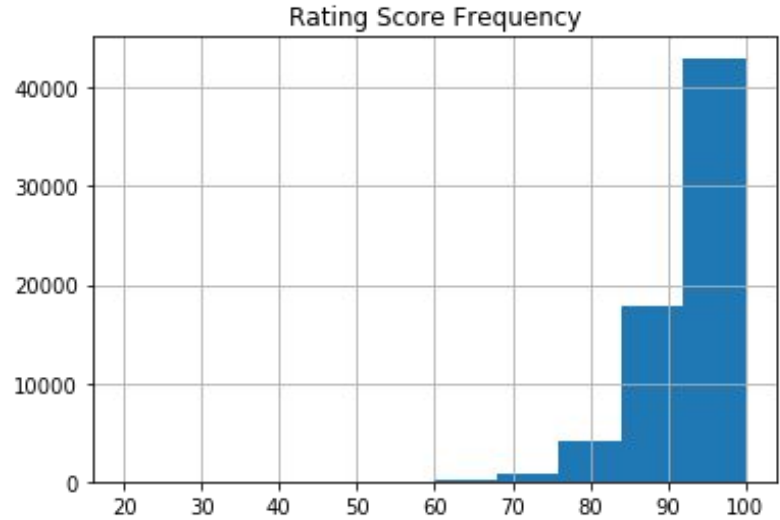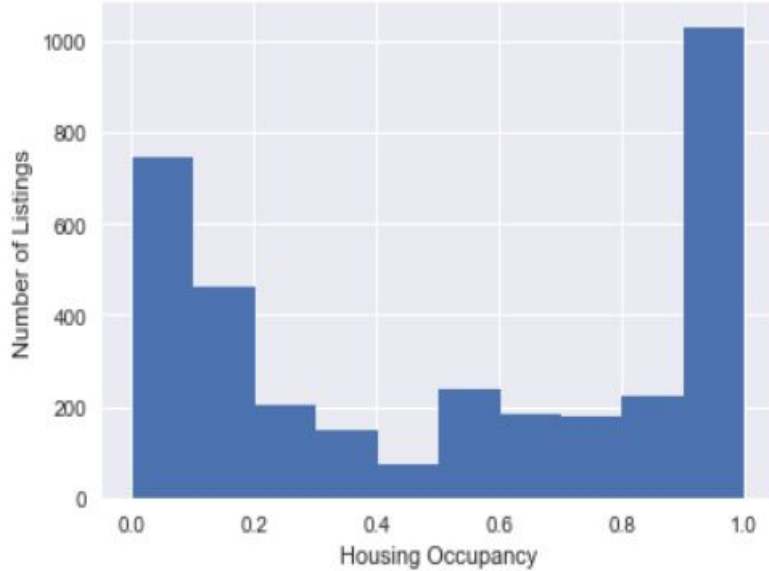
Distribution of Pos/Neg Review Sentiment Scores
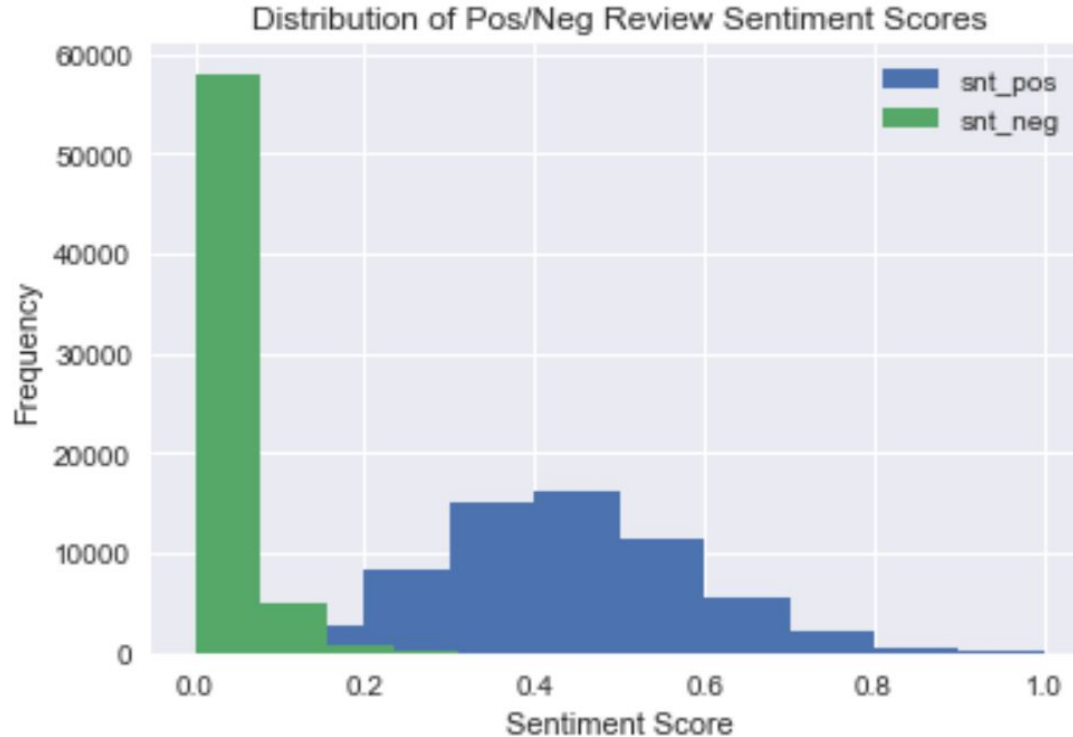
Listings Ratings Consistency

- Average sentiment rating scores from customers vs listing review rating scores provided.
- Not a linear relationship between them.
- Listings rating score do not reflect customers' reviews.
- Work on consistency of reviews and reviews score.
- Next, explore consistency bet. Rental description and user's reviews.
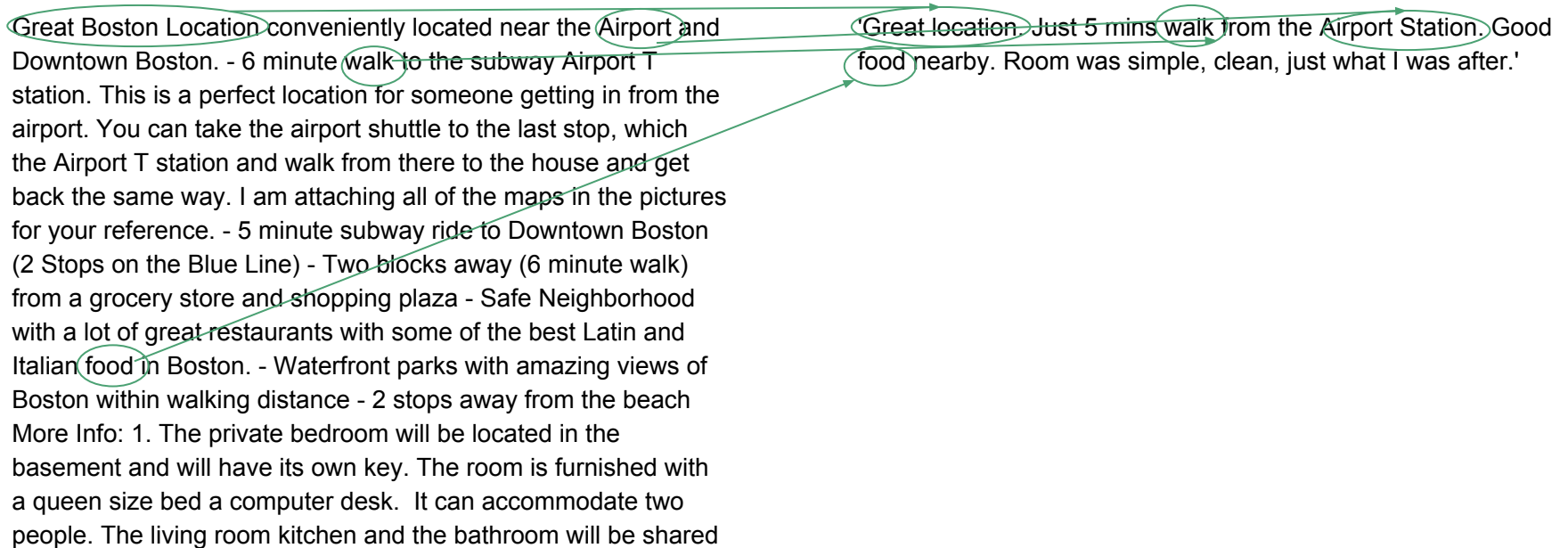

Listings Rating Consistency

Cosine Similarity

Rental Description

Great Boston Location conveniently located near the Airport and Downtown Boston. - 6 minute walk to the subway Airport T station. This is a perfect location for someone getting in from the airport. You can take the airport shuttle to the last stop, which the Airport T station and walk from there to the house and get back the same way. I am attaching all of the maps in the pictures for your reference. - 5 minute subway ride to Downtown Boston (2 Stops on the Blue Line) - Two blocks away (6 minute walk) from a grocery store and shopping plaza - Safe Neighborhood with a lot of great restaurants with some of the best Latin and Italian food in Boston. - Waterfront parks with amazing views of Boston within walking distance - 2 stops away from the beach More Info: 1. The private bedroom will be located in the basement and will have its own key. The room is furnished with a queen size bed a computer desk. It can accommodate two people. The living room kitchen and the bathroom will be shared
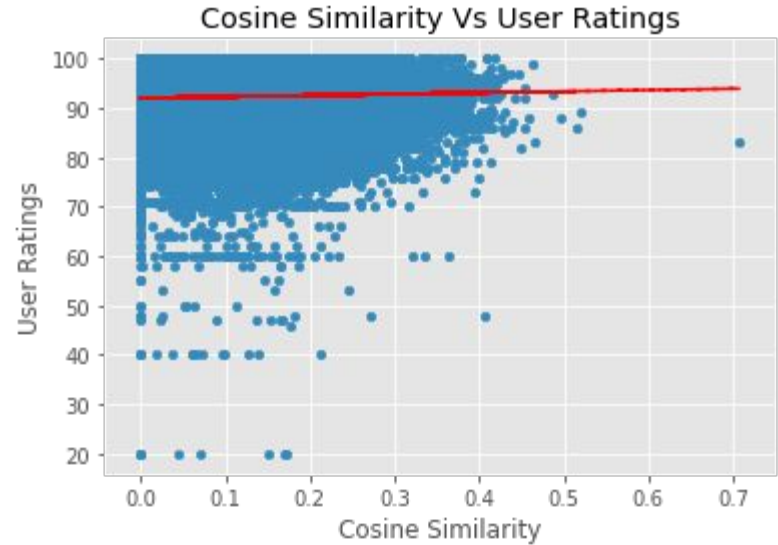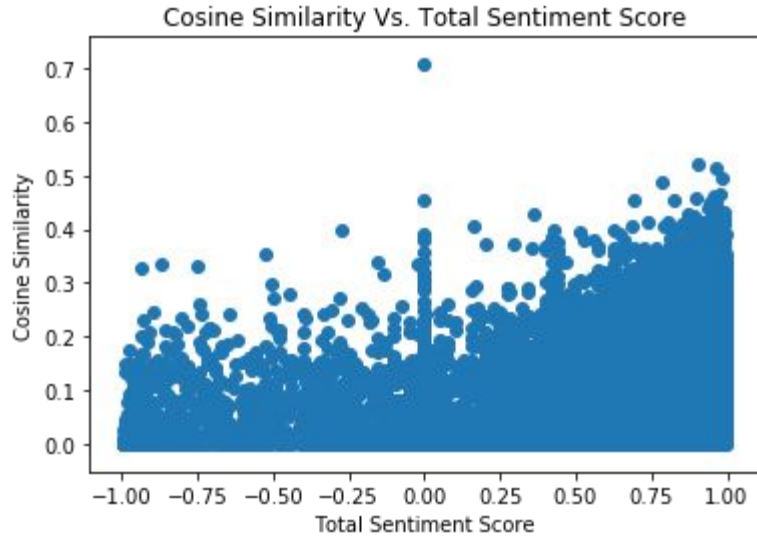
User Review

'Great location. Just 5 mins walk from the Airport Station. Good food nearby. Room was simple, clean, just what I was after.'

Cosine Similarity

Cosine Similarity Vs. Total Sentiment Score

Cosine Similarity Vs User Ratings

Lift Scores for High and Low Occupancy Listings
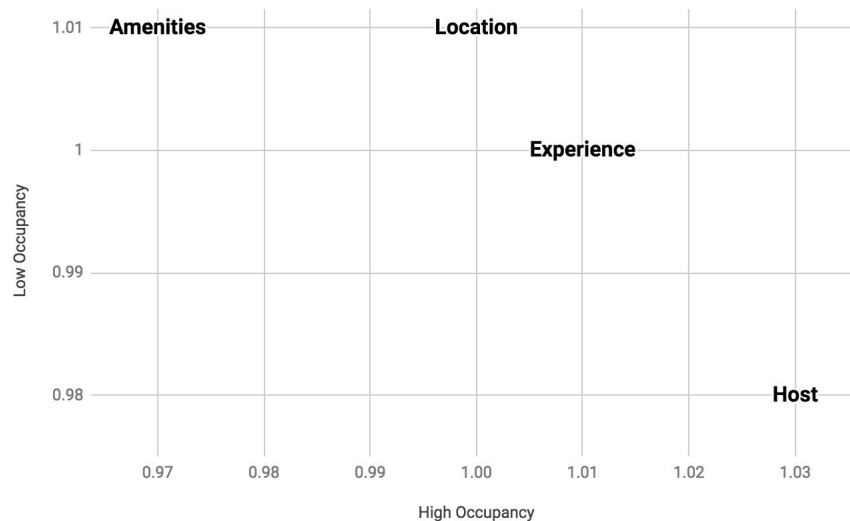
Reviews for listings with high rental rates tend to talk more about the host and experience more than average

Reviews for listings with low rental rates tend to talk about amenities more than average

High and Low

Amenities   Location   1.01

Experience   1

0.99

Host   0.98

Low Occupancy

0.97   0.98   0.99   1.00   1.01   1.02   1.03

High Occupancy

All positive reviews resemble one another;

Each negative review is negative in its own way.

	POS	NEG
neighborhood	1.0057	0.3303
restaurants	1.011	0.1747
station	1.0031	0.4581
street	0.9955	0.6247
downtown	1.0067	0.3403
subway	1.0016	0.5105
helpful	1.0118	0.1307
friendly	1.0119	0.132
welcoming	1.0117	0.1323
clean	1.001	0.4265
quiet	1.0077	0.2903
parking	0.9958	0.6008

Positive and Negative Review Attributes

Negative Vs. Positive Attributes

Location

Host

Negative Sentiment

Positive Sentiment

street
parking

subway

station
clean

downtown
neighborhood

quiet

restaurants
welcoming
helpful
friendly
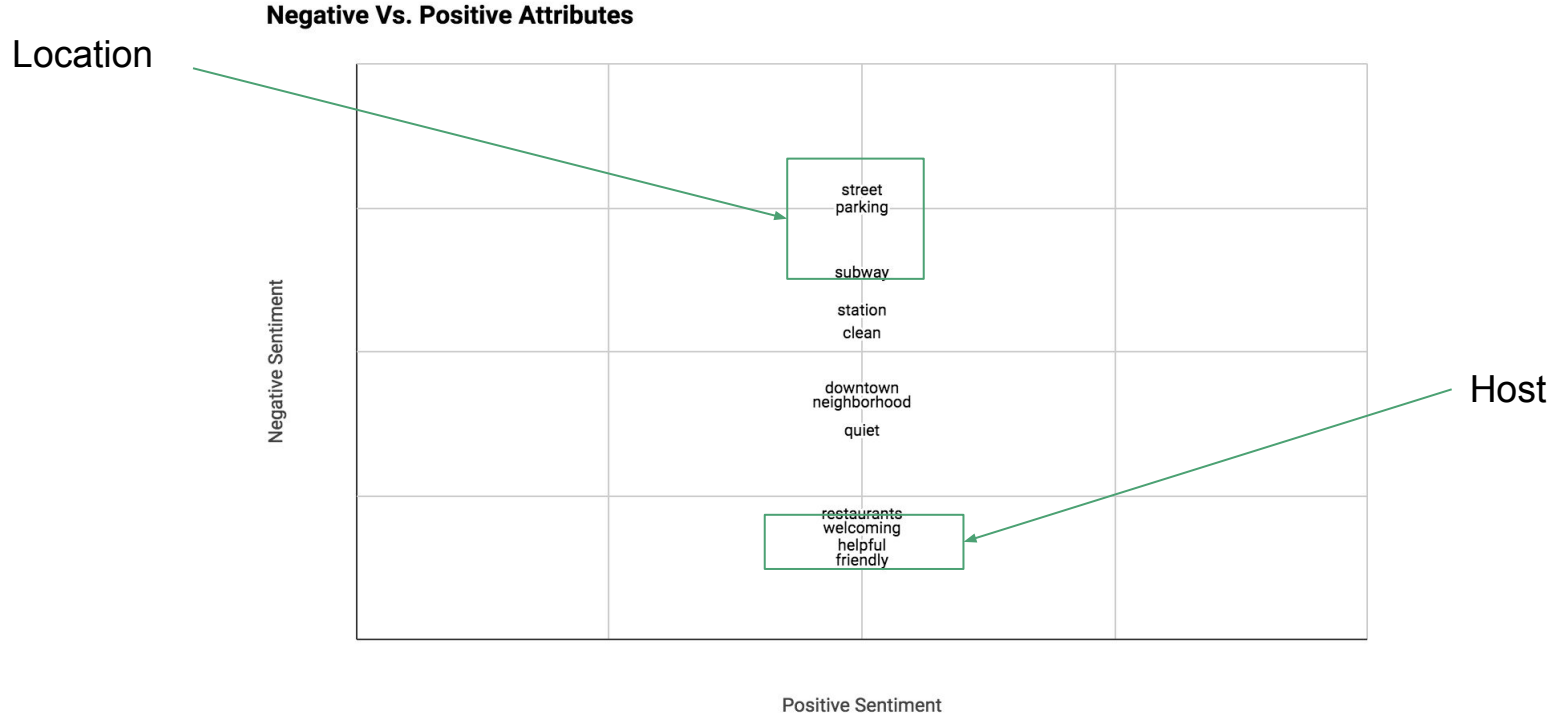
Conclusion

- Listings with the highest occupancy rates have prominent hosts that are at the center of the renter experience

- Hosts should help customers more with location based services to improve location sentiment scores ie. provide parking maps, maps of city, etc.

- Negative experiences are often uniquely negative, except for issues of canceling, unresponsiveness, and lack of cleanliness

Thank You

Appendix and Notes

Lift Scores for Positive and Negative Reviews

If a Review is Negative then they reviewer tends to talk equally about the amenities, location, and experience and a lot about the Host.

If a Review is Positive the review talks mostly about the host.

	POS	NEG
host	1	0.93
amenities	0.24	0.57
location	0.22	0.45
experience	0.17	0.38

Steps for Minimum Viable Product

1. Create master file
2. Create binary variable for high-rental or low-rental
3. Create 5 attributes and merge
4. Lift and cosine similarity.
   a. Description of rental vs what people are saying about it
   b. Tell me what you care about, I'll show you the listings.
      i. Doesn't have to be a fully automated system. Could be prepared.

Business value

1. Comparing description to what people talk about in their review to see if listing owners are in touch with their customers
2. See what attributes affect occupancy rates the most
3. Correlation between review sentiment and average star rating
4. Exploratory analysis on rental attributes
5. See what attributes determine the average review score.
6. Cosine similarity between attributes and rental description for matching
7. Association between high rental and low rental and attributes
8. No more than 5 attributes

Notes from meeting with Barua

We can assume that the owners make no improvements after the reviews. And can still use reviews for predicting. This works because review data is older than calendar data

Cosine similarity analysis between description of the owner and what the users are focusing on.

Make dependent variable high-rental and low-rental. Make a binary of rental rate. Do lift between high-rental and certain attributes.