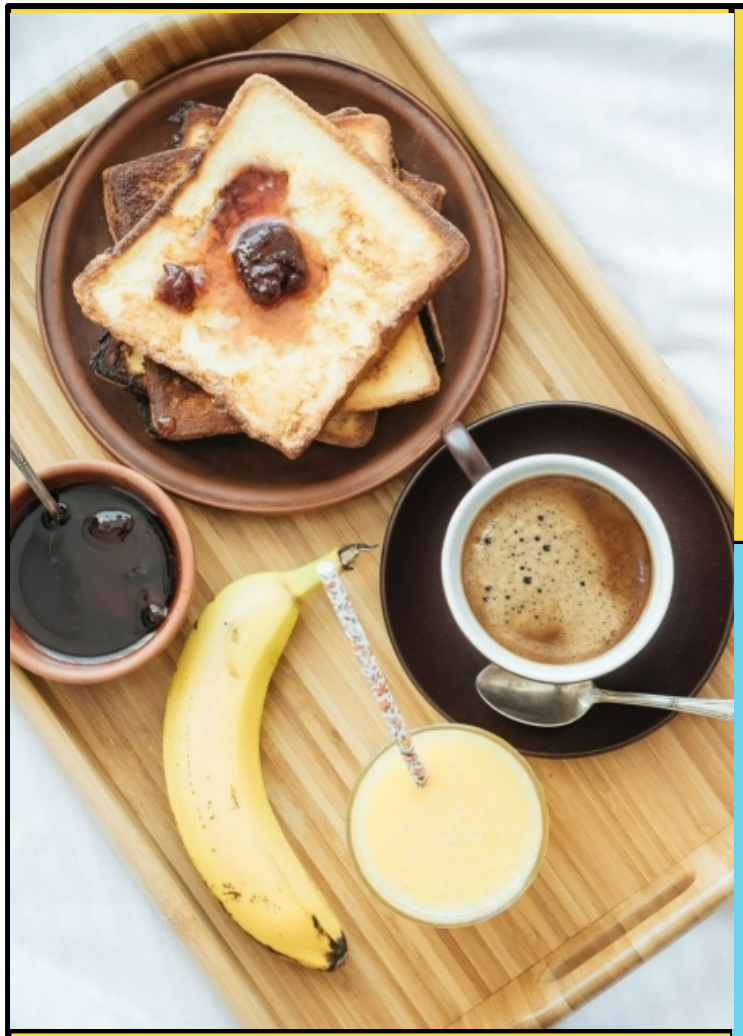
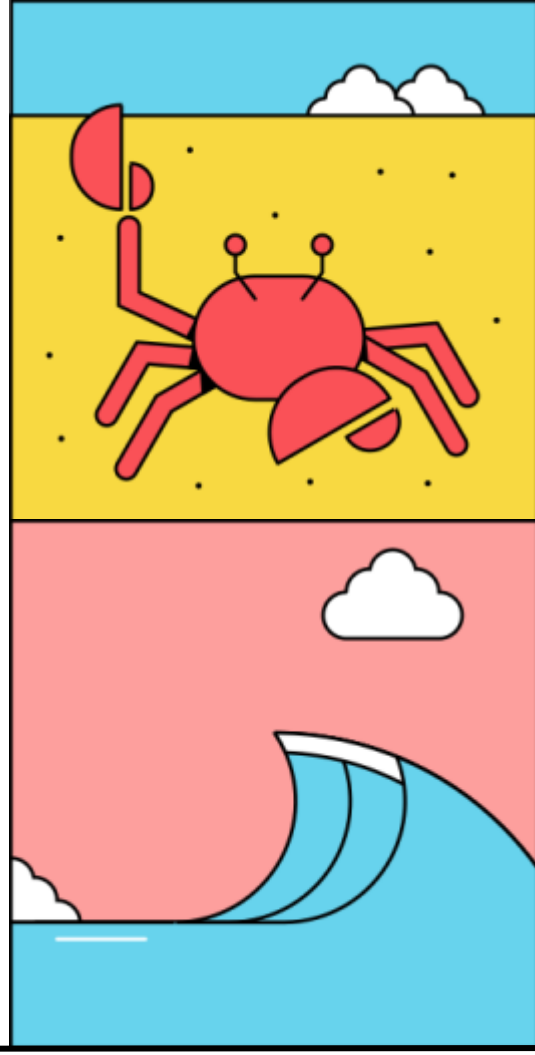


AirBnB

Group 3: James Min, Anh Nguyen, Cam Nguyen, Kim Nguyen, Anh Pham

Quick Recap:

- Our original research question was to help the host predict the optimal price listing for their Airbnb.
- To add on to our research question, we want to help the host figure out the most important factors in receiving good ratings.
- As we previously found, ratings are important for predicting Price and, in general, are beneficial for drawing in customers.

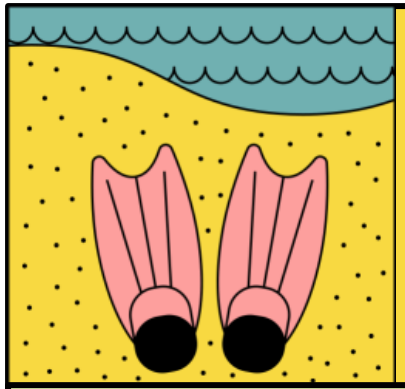


Introduction

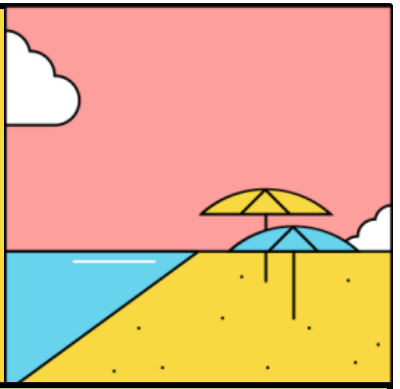
Welcome Back!

01

airbnb



Role:



- Our role for this project is to be in the perspectives of hosts and help them predict whether a guest would leave a 4.9-or higher star review or not.
- This is important for AirBnB hosts because if they are able to have a better idea on what factors/variables take into account in predicting whether the guest will leave a 5-star review, it will help hosts work on factors that are more important.
- By having a higher star rating, hosts will look much more attractive compared to other hosts.

Phenomenon of Interest & Research Question:

Research Question: What are some factors or components that guests consider important when leaving a 4.9 or higher -star review?

Phenomenon of Interest: We want to understand how to help hosts be more attractive by helping them gain more 4.9 or higher star reviews. This could be done by perfecting the factors that guests find to be most important when leaving a 4.9 or higher -star review and by using that, hosts are able to be better prepared.

02

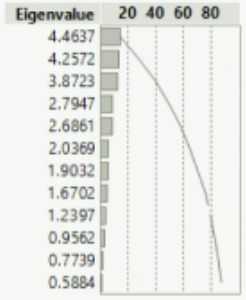
Unsupervised Learning

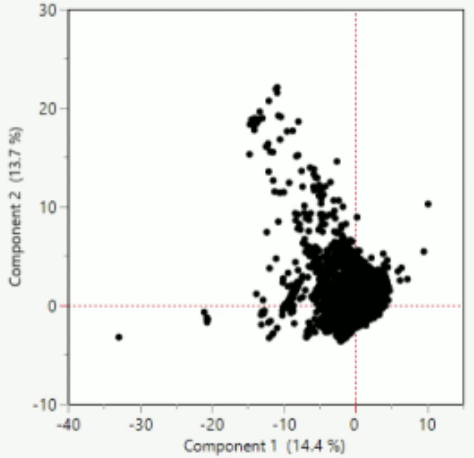
airbnb

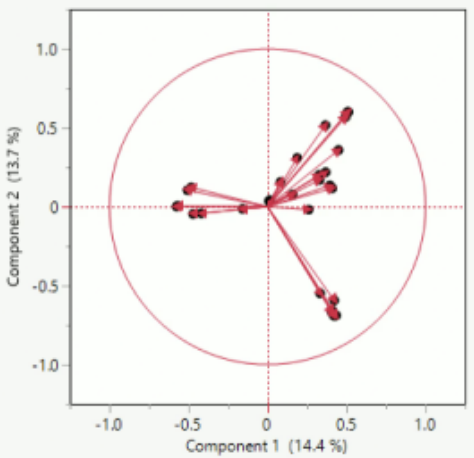
PCA/Factor Analysis

Principal Components: on Correlations

Summary Plots







Principal Components: on Correlations

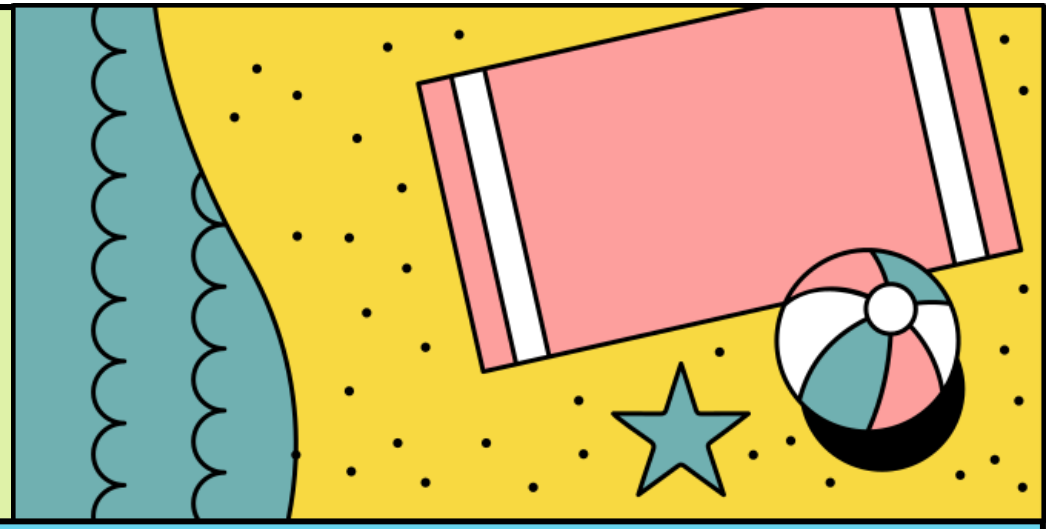
Factor Analysis: Maximum Likelihood / Varimax

Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
host_listings_count	0.04983	-0.04742	0.00513	0.09434	0.98654	-0.05616	-0.09467	0.03088
accommodates	0.11593	-0.00020	0.00236	-0.05416	-0.03249	0.90661	0.01025	0.00859
bedrooms	0.07235	0.02834	-0.01918	-0.00555	-0.02835	0.92484	-0.02327	0.00987
beds	0.08864	-0.00583	-0.00374	-0.02764	-0.01784	0.84725	0.02161	0.01010
price	0.07039	0.02813	0.00232	-0.00426	0.00179	0.38040	-0.04767	-0.01808
minimum_nights	-0.00042	0.01352	-0.00638	0.96624	0.02651	-0.02812	-0.05809	-0.11465
maximum_nights	-0.06255	-0.02881	-0.01063	0.06296	0.14691	0.01228	0.02737	0.01254
minimum_minimum_nights	-0.01080	0.01390	-0.00500	0.90147	0.14823	-0.03353	-0.05564	-0.03002
maximum_minimum_nights	0.04472	0.02530	-0.00143	0.69396	0.56848	-0.05125	-0.05269	0.42009
minimum_maximum_nights	-0.00759	-0.00422	0.97390	-0.00363	-0.00781	-0.00287	-0.02049	-0.00059
maximum_maximum_nights	0.00367	-0.00667	0.90265	-0.00647	-0.00801	-0.00772	-0.01686	-0.00180
minimum_nights_avg_ntm	0.04144	0.02509	-0.00136	0.70878	0.56449	-0.05655	-0.05602	0.41280
maximum_nights_avg_ntm	-0.00772	-0.00215	1.01675	-0.00412	-0.00840	-0.00437	-0.02133	-0.00066
availability_30	0.96476	-0.03195	-0.01256	-0.01307	-0.12752	0.04766	-0.02604	-0.01260
availability_60	1.02705	-0.02738	-0.00545	-0.00008	-0.09768	0.03137	-0.00923	-0.00489
availability_90	1.00124	-0.01834	-0.00183	0.00644	-0.07700	0.03631	0.01630	0.00164
availability_365	0.69273	-0.02269	0.00742	0.04967	0.04580	0.13151	0.00891	0.02578
calendar_last_scraped	0.08888	-0.01895	-0.00760	0.01387	0.02548	0.02567	0.04495	0.02031
number_of_reviews	0.04785	0.03985	-0.00418	-0.01817	0.01641	-0.03228	0.98961	-0.04688
number_of_reviews_ltm	0.16691	0.05132	-0.00715	-0.08553	-0.03153	0.00573	0.67340	-0.01005
number_of_reviews_l30d	0.18764	0.06155	-0.01201	-0.09578	-0.05188	-0.02128	0.40033	0.00880
first_review	0.27036	-0.01964	0.02907	-0.09738	0.04003	0.05554	-0.42238	-0.03427
last_review	0.45262	0.06130	0.01673	-0.06427	0.04372	0.09732	0.15463	-0.06005
review_scores_accuracy	-0.00383	0.89982	0.00096	0.00032	-0.01019	0.01044	0.03085	-0.00473
review_scores_cleanliness	0.03025	0.79106	0.00917	-0.01239	0.02353	0.03202	0.03294	-0.02465
review_scores_checkin	-0.02330	0.78557	-0.01656	0.01560	-0.06720	0.04364	0.03637	0.02388
review_scores_communication	-0.03220	0.79961	-0.00941	0.01555	-0.08193	0.01514	0.02958	0.03156
review_scores_location	-0.00073	0.64674	0.00773	0.00890	0.00673	-0.01210	0.02370	-0.03348
review_scores_value	-0.04927	0.83794	-0.00687	-0.00060	-0.06750	-0.00572	0.04030	0.01389

airbnb

Linear Regression



Response review_scores_rating

Effect Summary

Source	LogWorth	PValue
Review Score	3907.560	0.00000
Accomodate	20.185	0.00000
Host Listing Count	9.512	0.00000
Number of reviews	8.190	0.00000
Availability	1.428	0.03729
Minimum Nights	0.536	0.29106
Average Min Max nights	0.406	0.39277
Maximum Nights	0.339	0.45782

Remove Add Edit ☐ FDR

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	10556	264.17171	0.025026	.
Pure Error	1	0.00000	0.000000	Prob > F
Total Error	10557	264.17171		.
				Max RSq
				1.0000

Summary of Fit

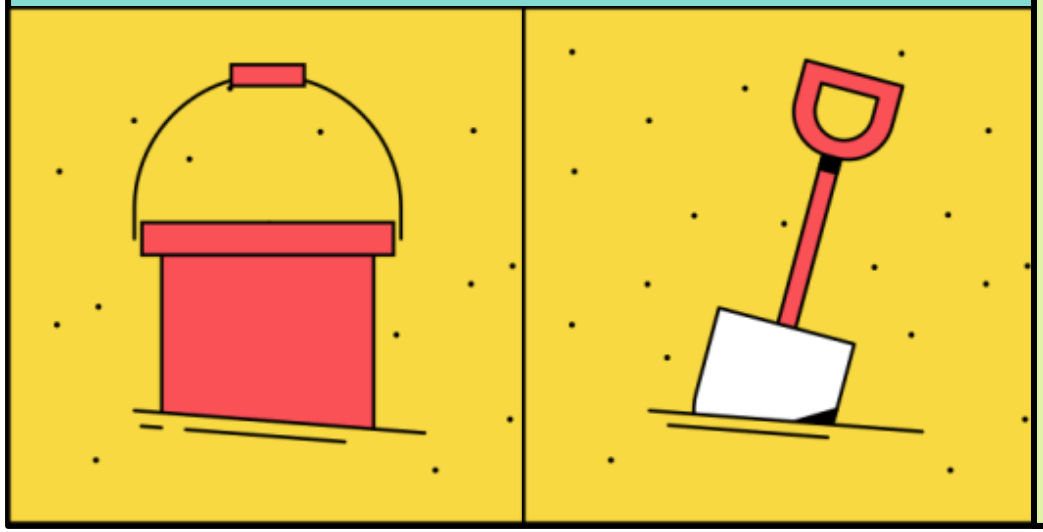
RSquare	0.820001
RSquare Adj	0.819864
Root Mean Square Error	0.158188
Mean of Response	4.816994
Observations (or Sum Wgts)	10566

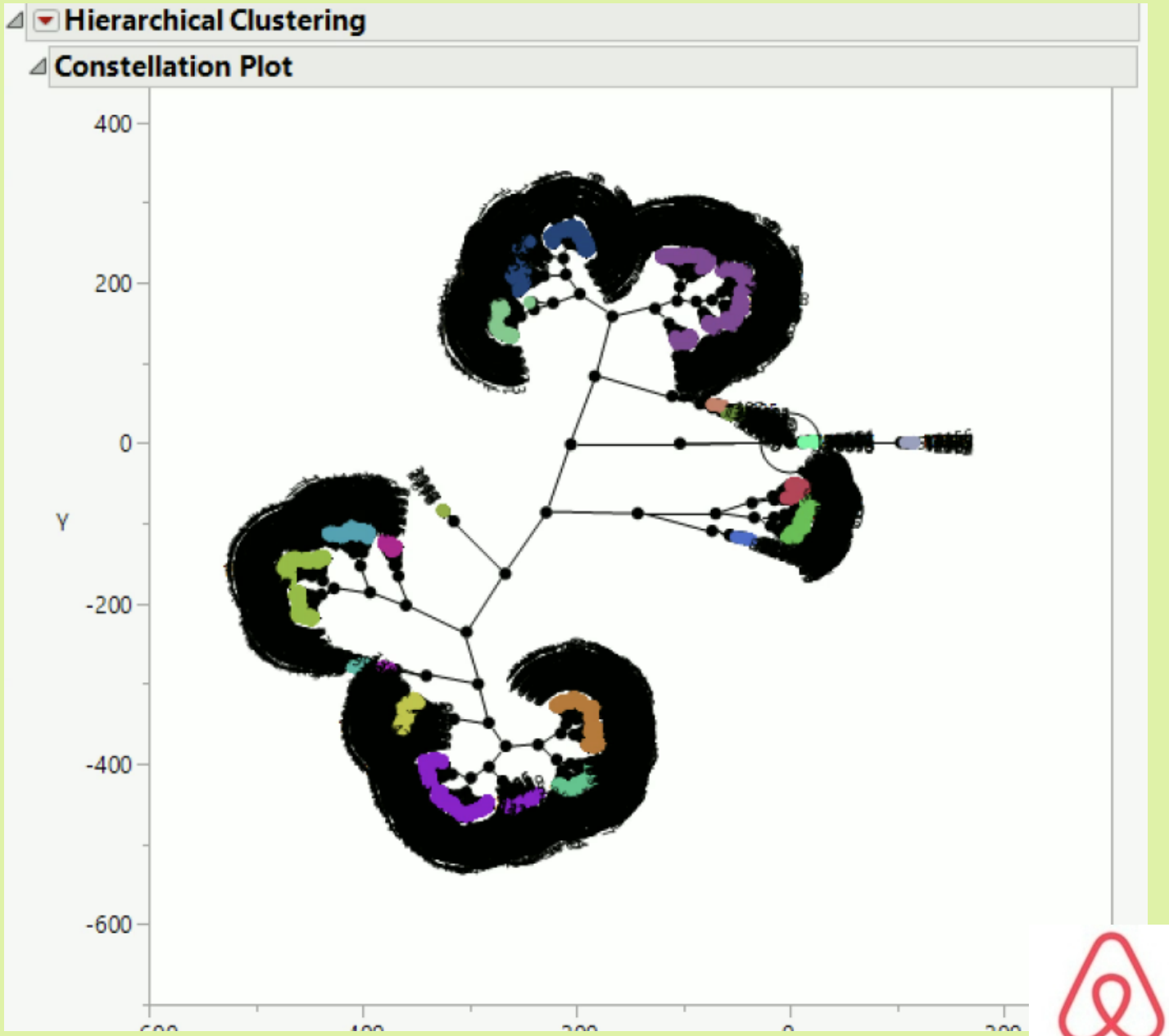
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.8127819	0.001637	2940.5	<.0001*
Availability	-0.003238	0.001555	-2.08	0.0373*
Review Score	0.360218	0.001654	217.81	<.0001*
Maximum Nights	-0.001146	0.001544	-0.74	0.4578
Minimum Nights	0.0016754	0.001587	1.06	0.2911
Host Listing Count	-0.022305	0.00354	-6.30	<.0001*
Accomodate	0.0148519	0.00158	9.40	<.0001*
Number of reviews	0.0088286	0.00152	5.81	<.0001*
Average Min Max nights	0.0030744	0.003597	0.85	0.3928

airbnb

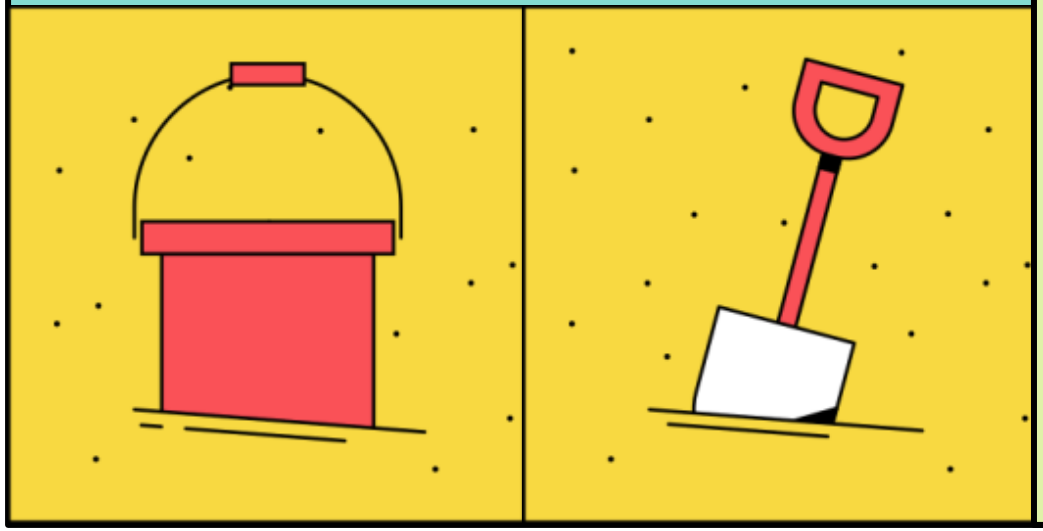
Constellation Plot

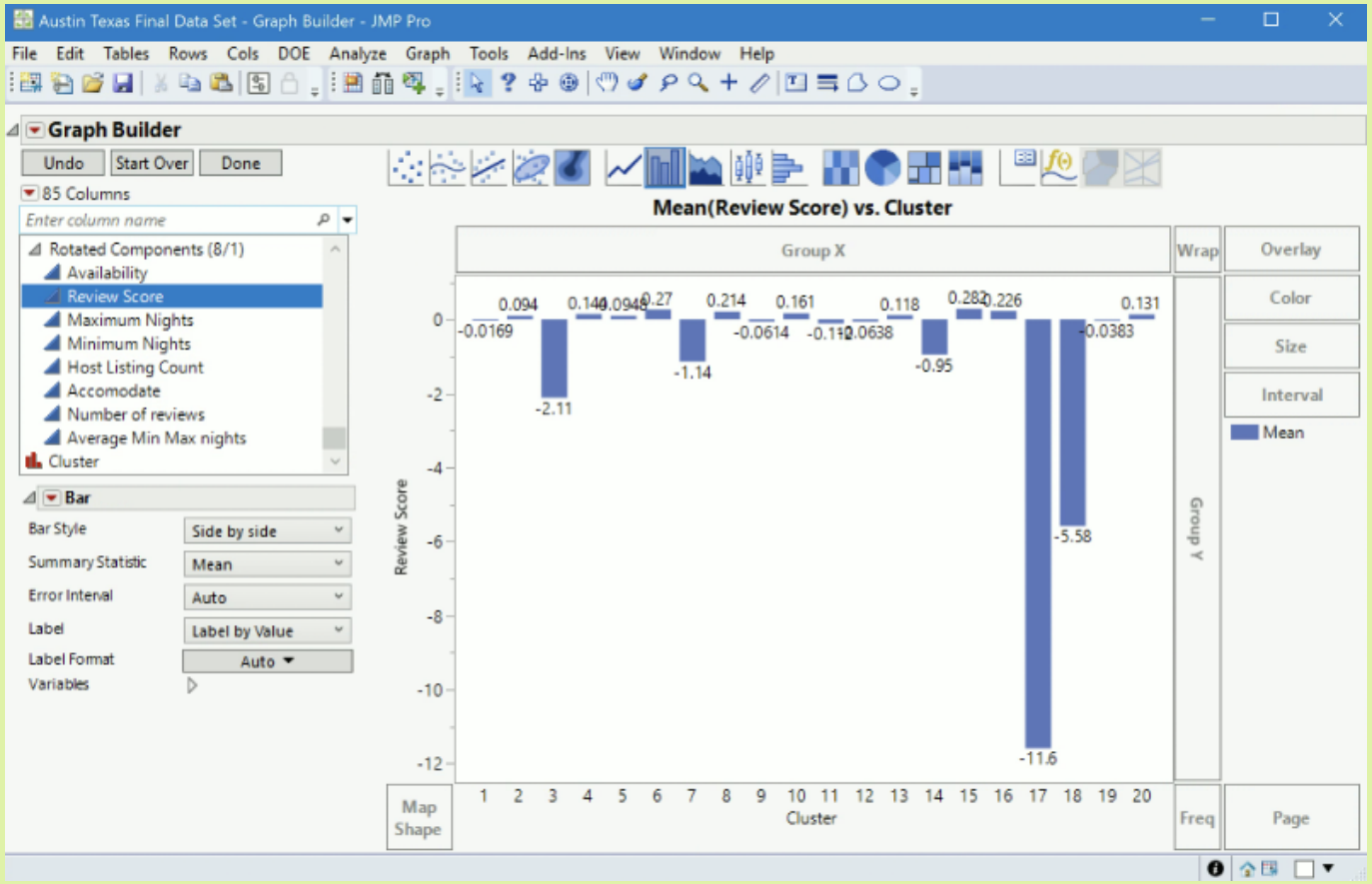




airbnb

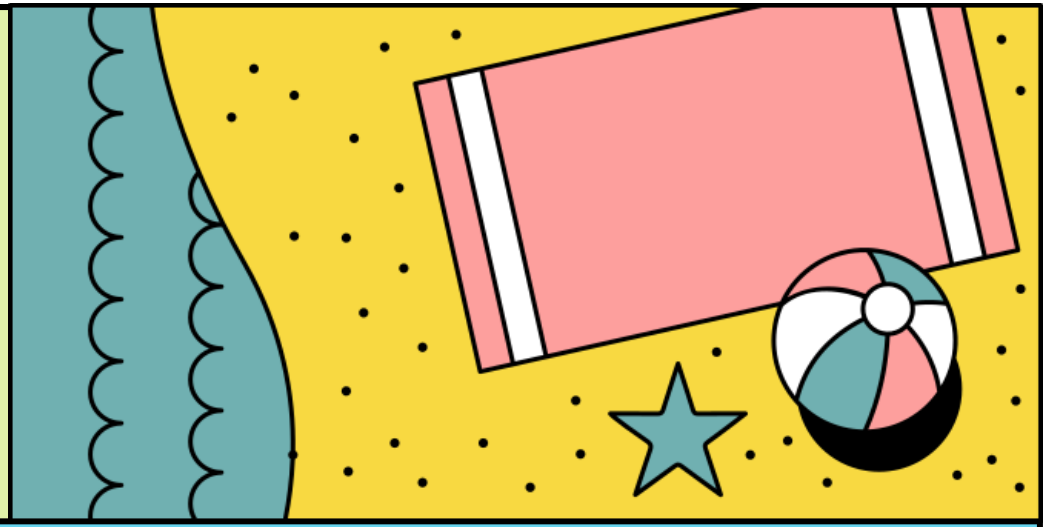
Clustering

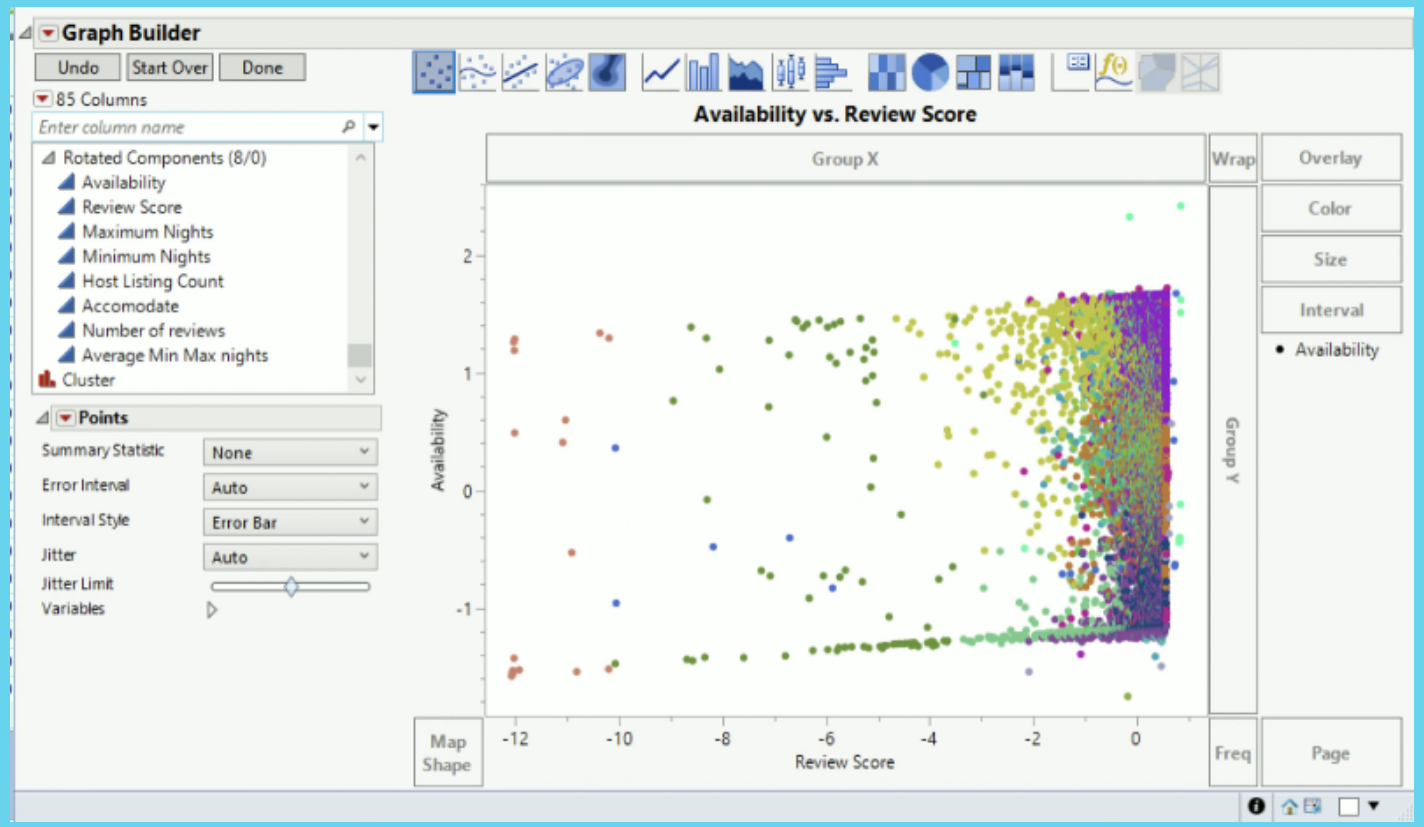




20 clusters

Perceptual Map





The two most important factors:

- Review Score
- Availability

It's concentrated from point 1 to –2 and scattered until –12.

Supervised Machine Learning

- Logit Regression Model & Neural Network
- Which model would be best to use when predicting how our guests would rate their stays
- Transformed review star ratings to our dummy variable
- Star rating < 4.9 = 0
- Star rating > 4.9 = 1
- Going to analyze misclassification rate and AUC (ROC Chart)

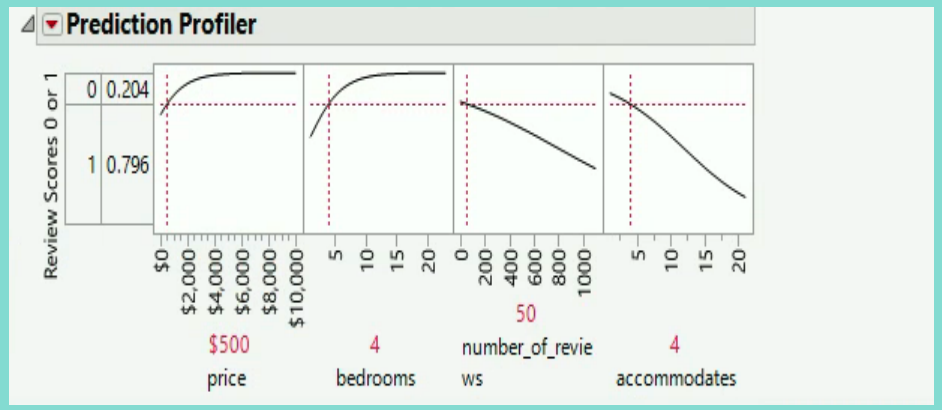
- Used 60% for training set
- 40% for validation set
- 10715 listings
- Accommodates was the most important factor
- Bedrooms and price had similar importance
- Misclassification rate = 40.57%

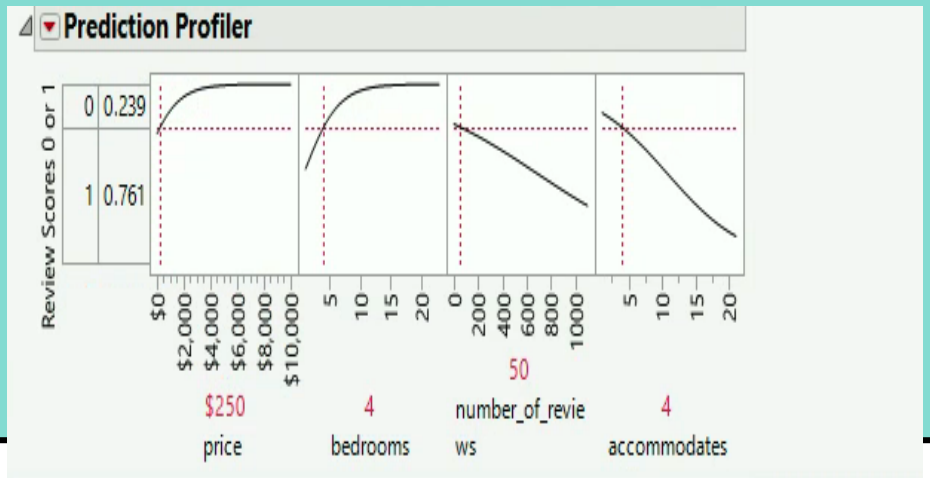
Fit Details

Measure	Training	Validation	Definition
Entropy RSquare	0.0231	0.0187	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0416	0.0339	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.6645	0.6696	∑ -Log(ρ[j])/n
RASE	0.4851	0.4869	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.4716	0.4727	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.4028	0.4057	∑ (ρ[j]≠ρMax)/n
N	6475	4240	n

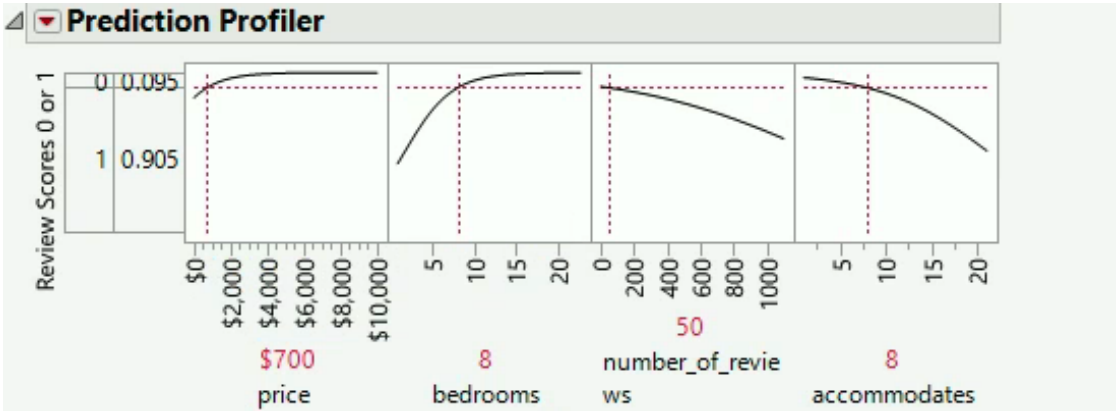
Source	LogWorth	PValue
accommodates	23.946	0.00000
bedrooms	14.895	0.00000
price	12.560	0.00000
number_of_reviews	7.595	0.00000

- Probability of leaving a 4.9+ star rating (top model) = 80%
- Probability of leaving a 4.9+ star rating (bottom model) = 76%
- Higher priced AirBnB listings had a higher probability of guests leaving a 4.9-star rating or higher
- Price may reflect true quality of the listing
- Therefore, quality may differentiate how customers may leave reviews

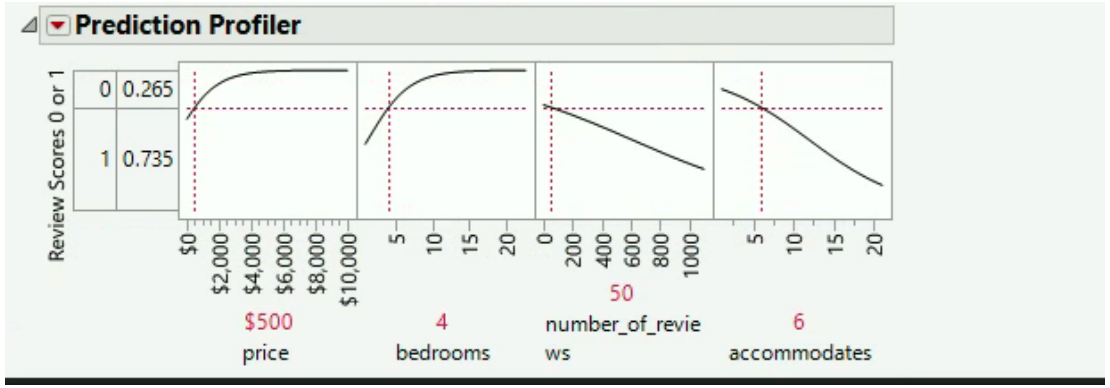




- $700, 8 bedrooms, 50 reviews, 8 accommodates = 90% probability of guest leaving a 4.9-star rating or higher

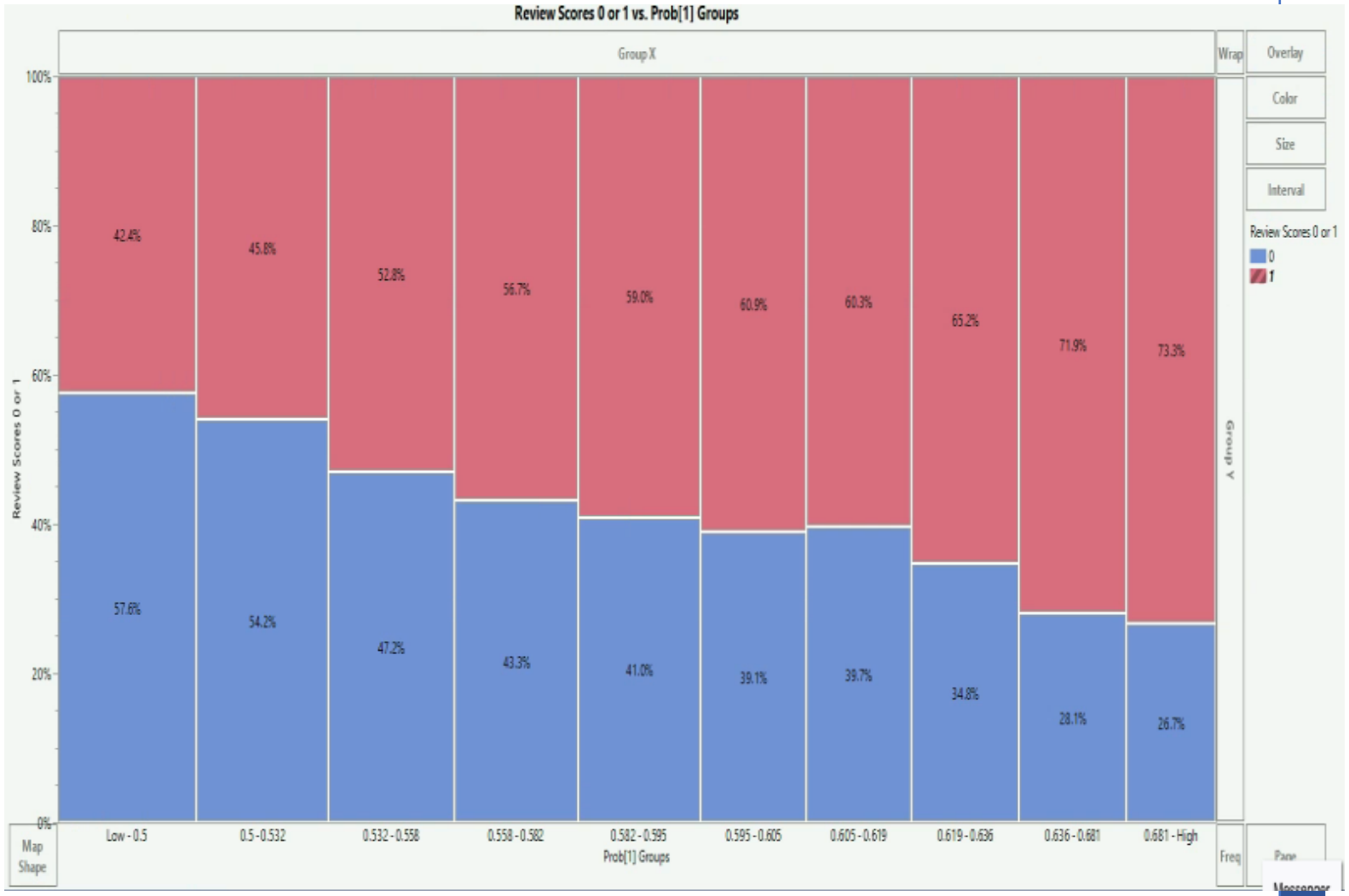


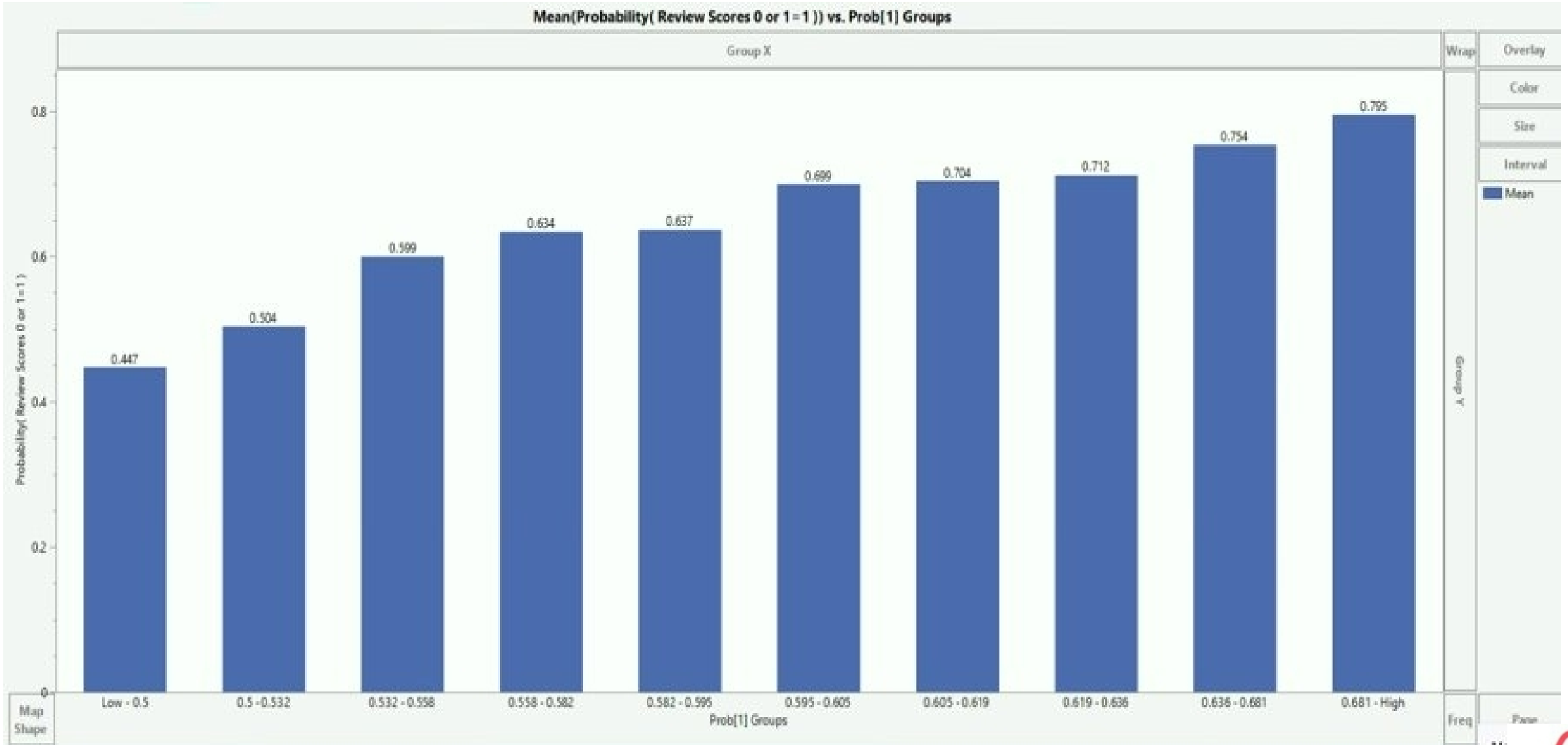
- $500, 4 bedrooms, 50 reviews, accommodates = 73.5% probability of guest leaving a 4.9-star rating or higher

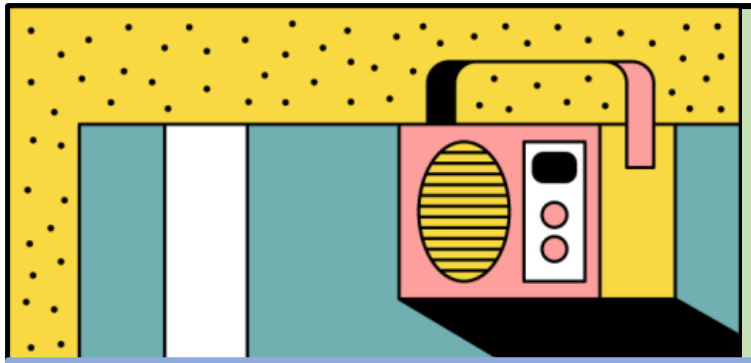


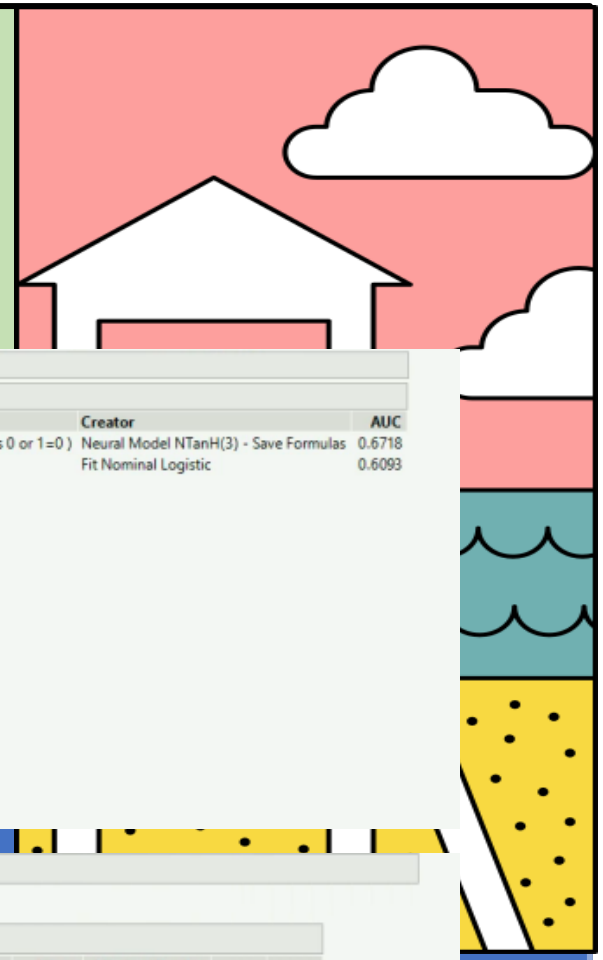
- Higher price but more rooms and accommodates had a very positive effect on guests' rating
- When accommodates went up while bedrooms remained the same, the probability also went down
- This may be due to guests' feeling undervalued when some may have to stay in a non-bedroom

- The 1st percentile group of consumers who had left a review rating of 4.9 or higher= 42.4%
- The top percentile group of consumers had left a review rating of 4.9 or higher= 73.3%
- While as hosts we want the top percentile groups; we want to find out why the 1st percentile group leaves the review they do
- Ask AirBnB for a follow up feedback on their stay which may help hosts make positive changes
- Ask guests who are leaving more positive reviews what makes their stay enjoyable
- Then we can use those to even better improve our greatness

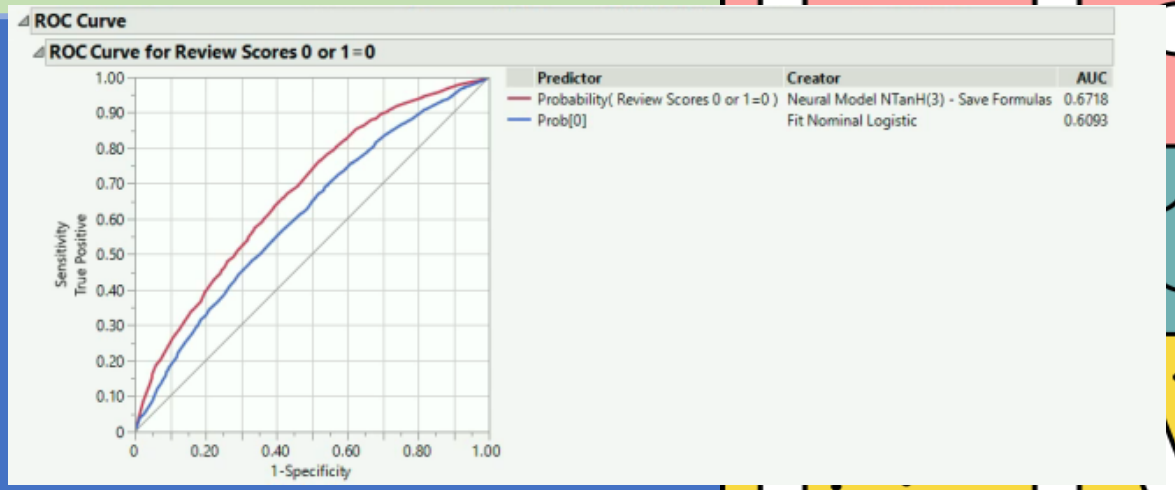


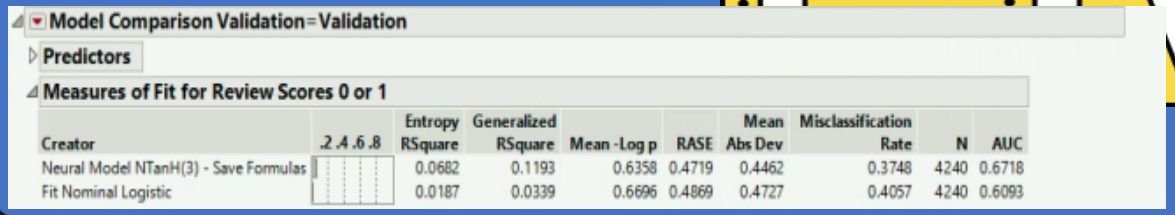

Mean(Probability(Review Scores 0 or 1 = 1)) vs. Prob[1] Groups
Group X
Wrap
Overlay
Color
Size
Interval
Mean
Group Y
Freq
Page
Map Shape
Probability(Review Scores 0 or 1 = 1)
0
0.2
0.4
0.6
0.8
0.447
0.504
0.599
0.634
0.637
0.699
0.704
0.712
0.754
0.795
Low - 0.5
0.5 - 0.532
0.532 - 0.558
0.558 - 0.582
0.582 - 0.595
0.595 - 0.605
0.605 - 0.619
0.619 - 0.636
0.636 - 0.681
0.681 - High
Prob[1] Groups



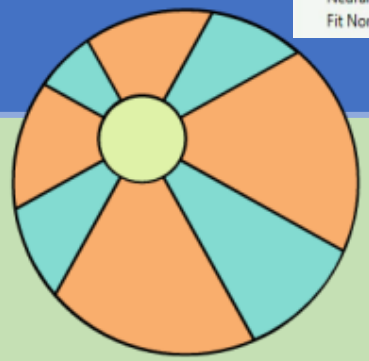


- Neural Model was the better model; because...
- Neural Model had a lower misclassification rate
- Neural Model had a lower misclassification rate
- Neural model also had a higher AUC
- Neural model has a higher true positive rate
- Neural model also has a lower false positive rate
- Both may have a high misclassification due to the range of prediction being so low

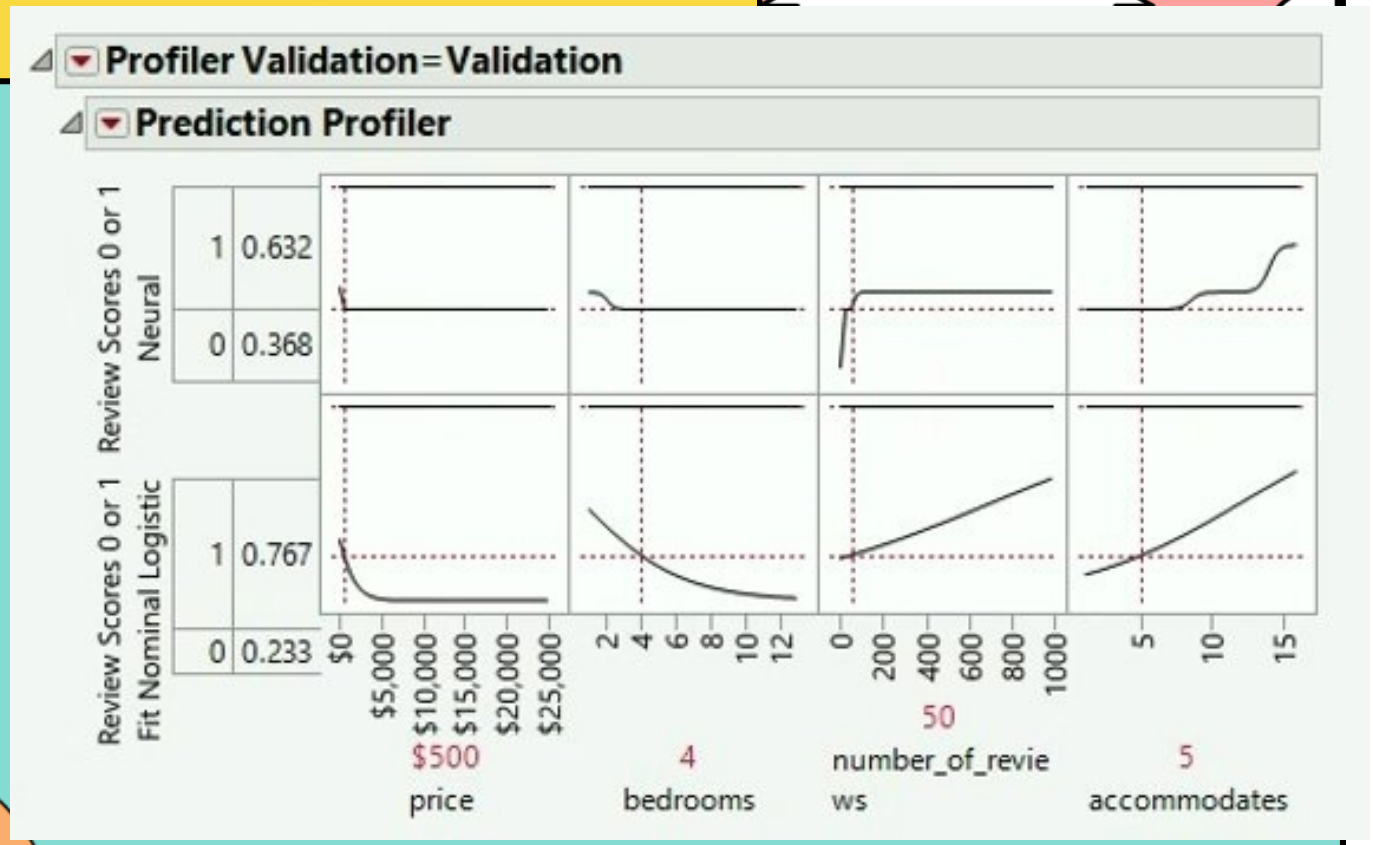




- Misclassification Neural: 37.48%
- AUC Neural: .6718
- Misclassification Logit: 40.57%
- AUC Logit: .6093

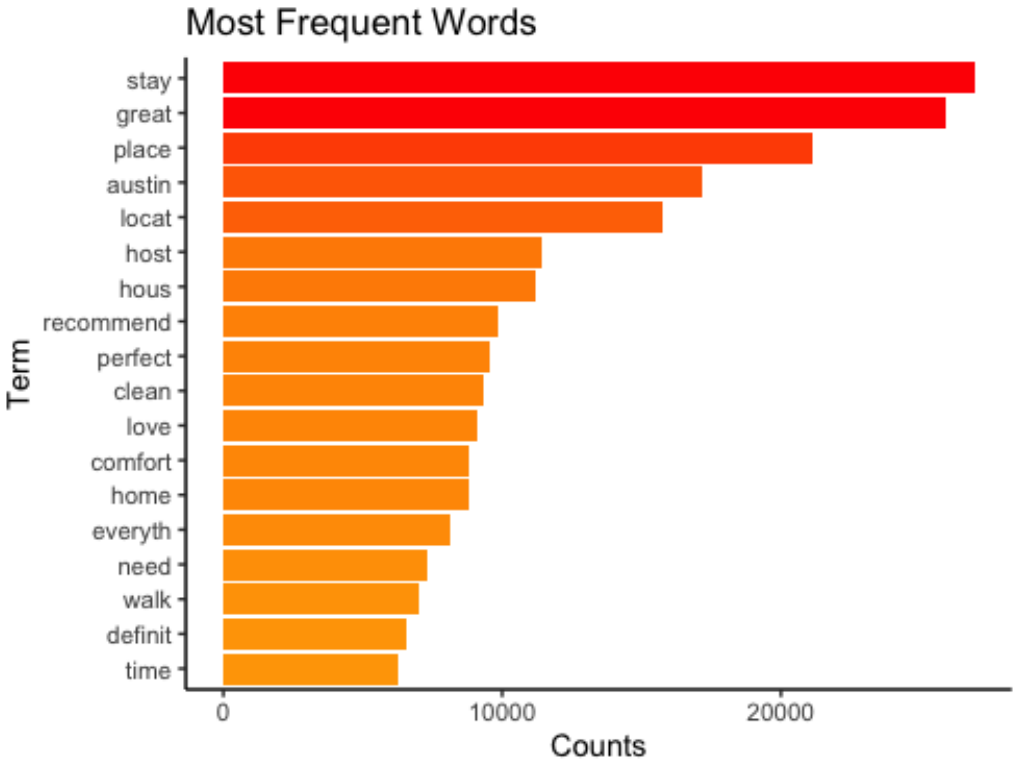


- When predicting using these two models...
- Neural = 63.2%
- Logit: 76.7%
- Price $500, 4 bedrooms, 50 reviews, 5 accommodates
- Neural had a lower probability of guests leaving a 4.9 or + higher reviews
- This may be because logit model has a higher misclassification rate and therefore it may be over predicting the reviews
- Vice versa, to have a low misclassification rate, you must never over predict

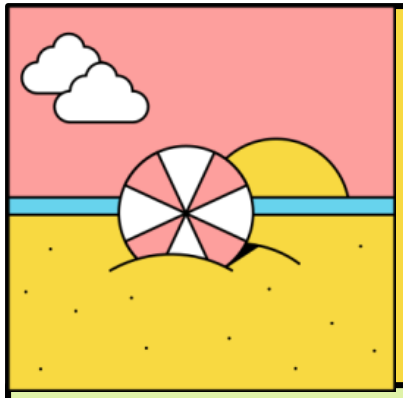


Text Mining

Word Frequency for Reviews 4.9 and above

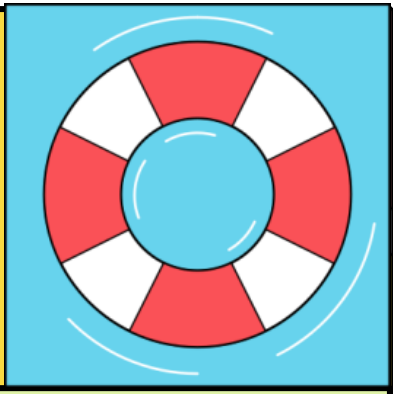


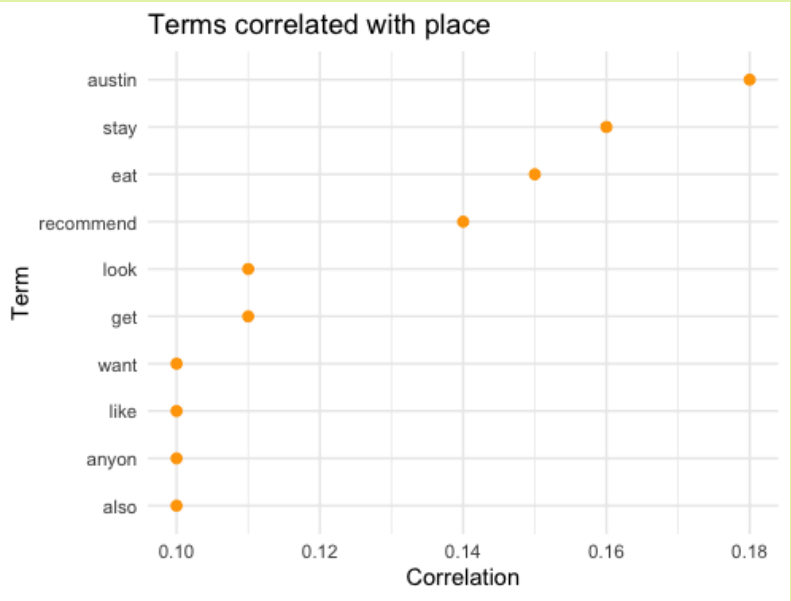
airbnb

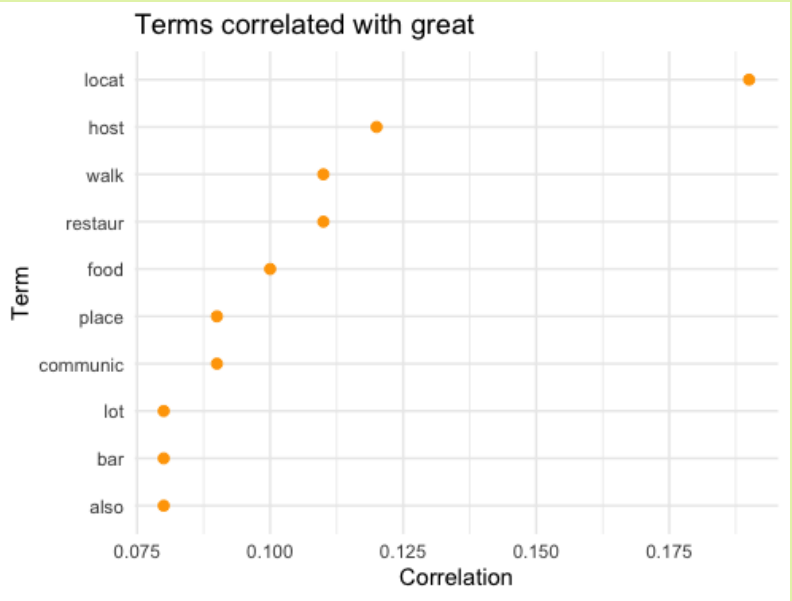


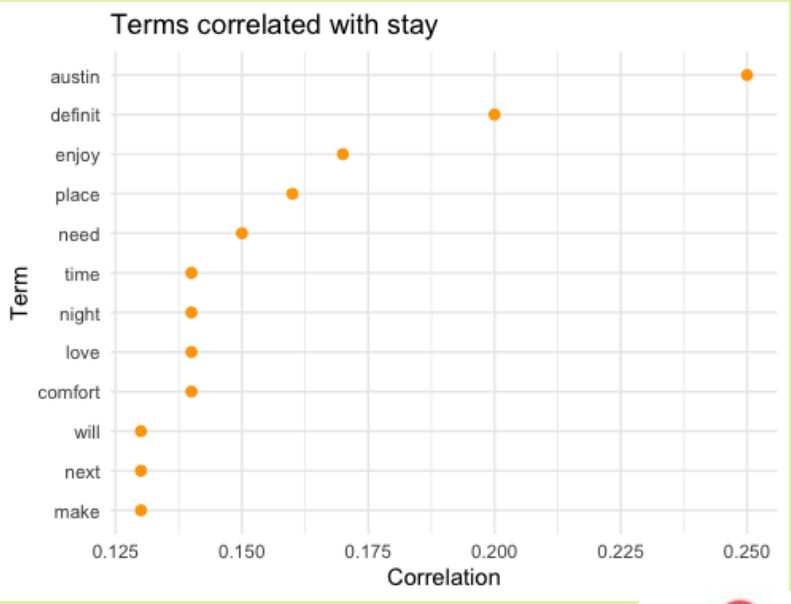
Text Mining

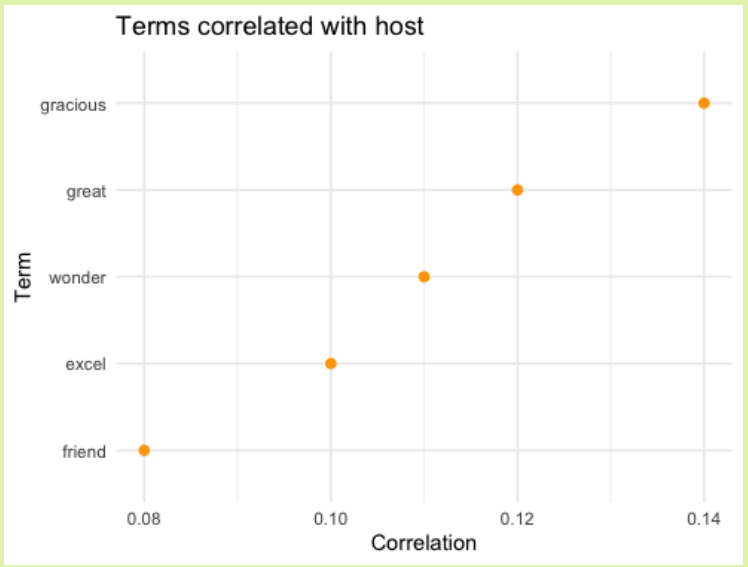
Word Association

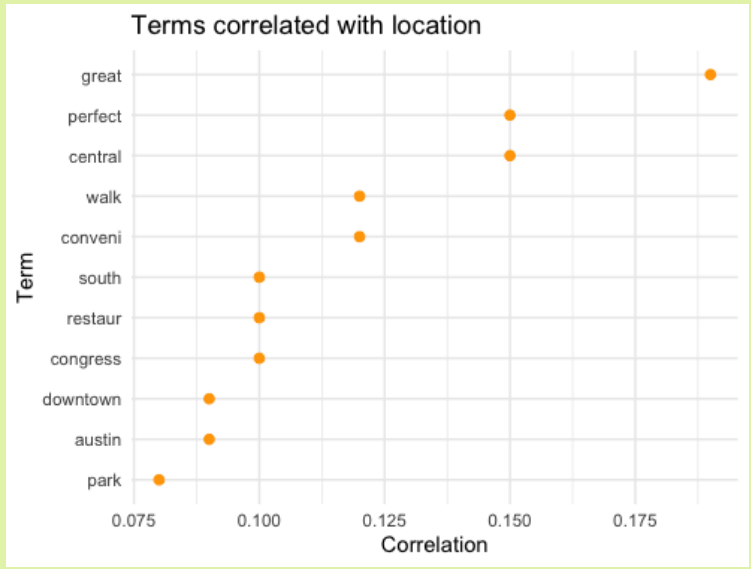


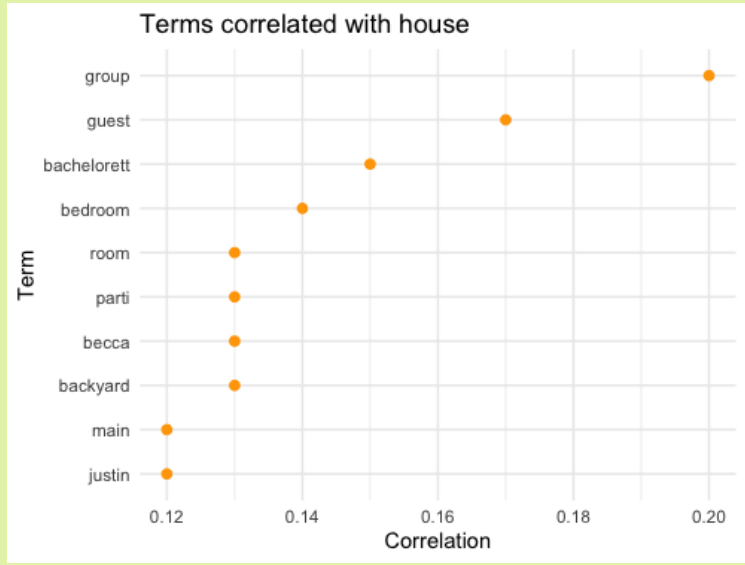




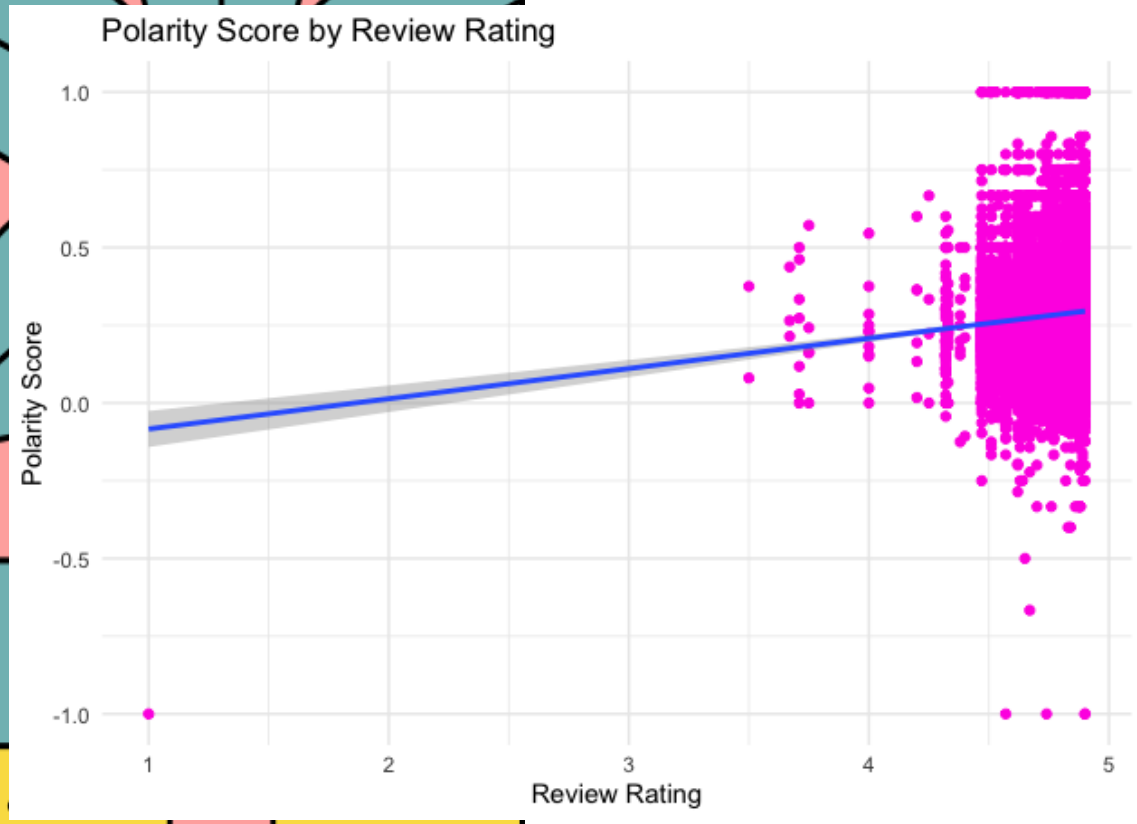


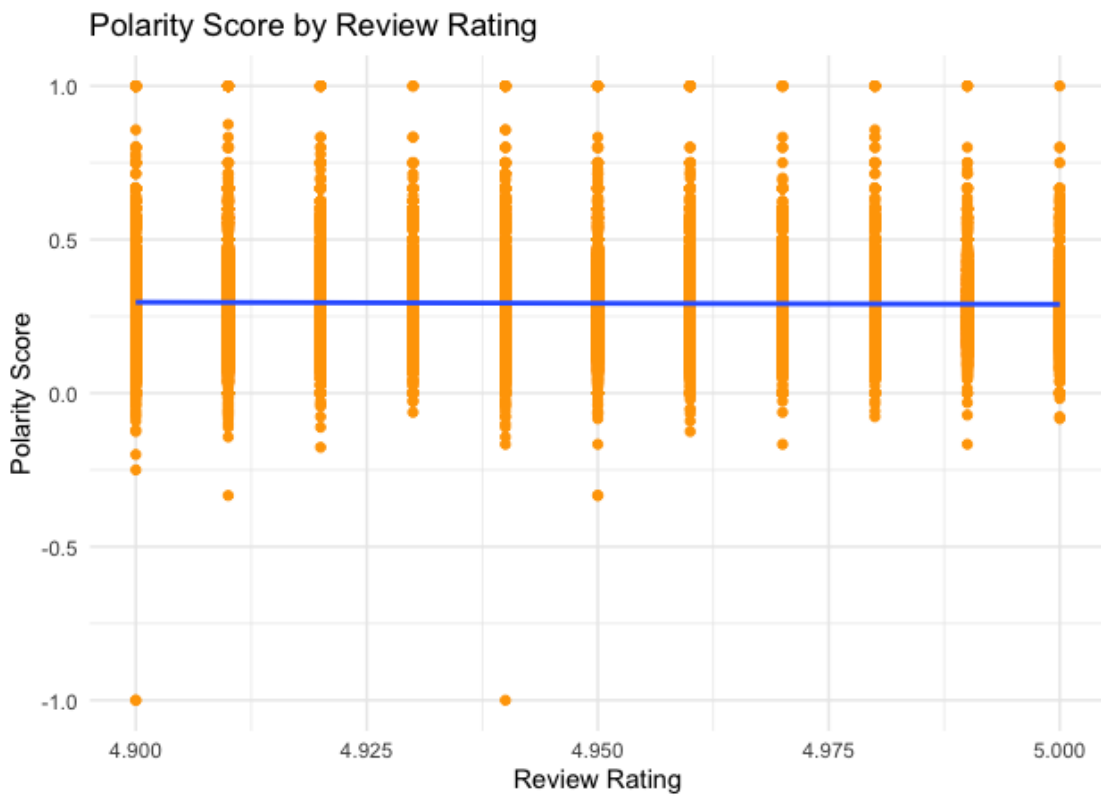

Terms correlated with host
gracious
great
wonder
excel
friend
Term
0.08
0.10
0.12
0.14
Correlation


Terms correlated with location
great
perfect
central
walk
conveni
south
restaur
congress
downtown
austin
park
Term
0.075
0.100
0.125
0.150
0.175
Correlation

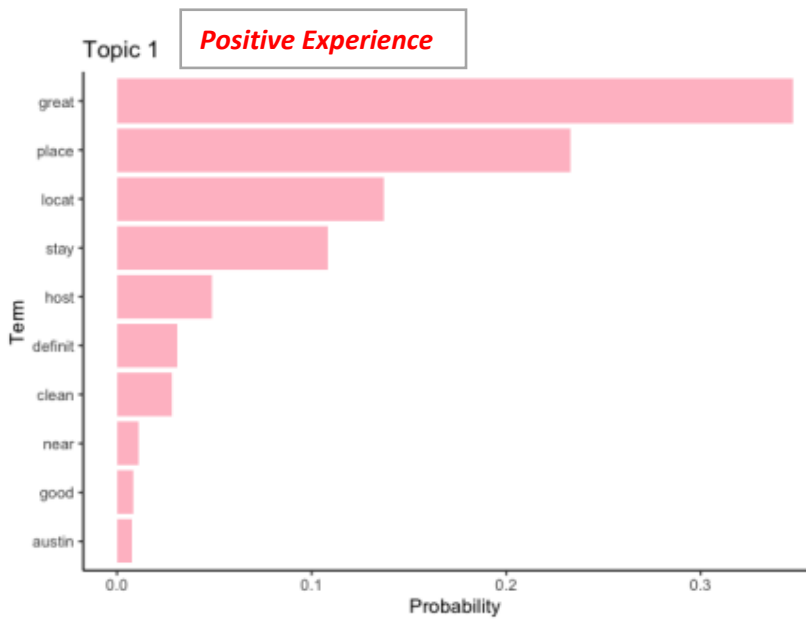

Terms correlated with house
group
guest
bachelorett
bedroom
room
parti
becca
backyard
main
justin
Term
0.12
0.14
0.16
0.18
0.20
Correlation

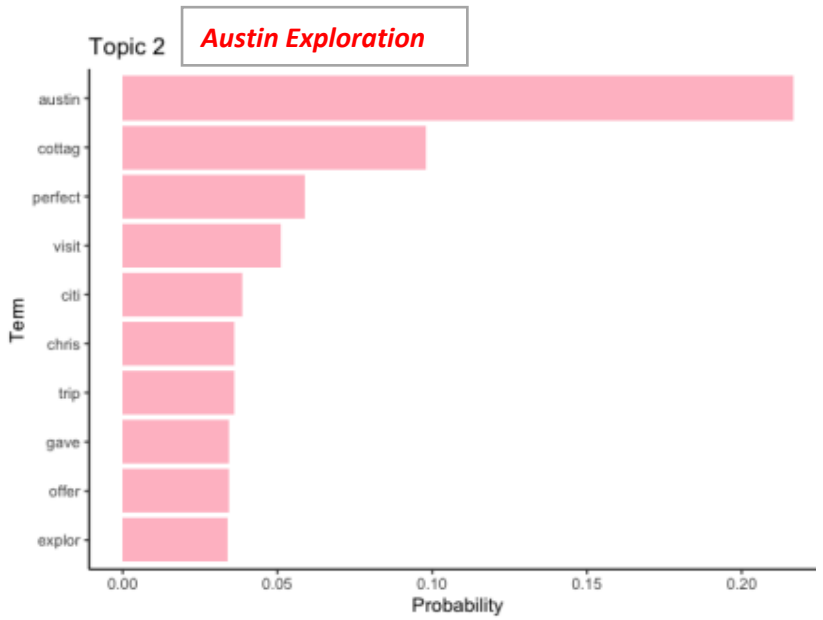
Sentiment Analysis

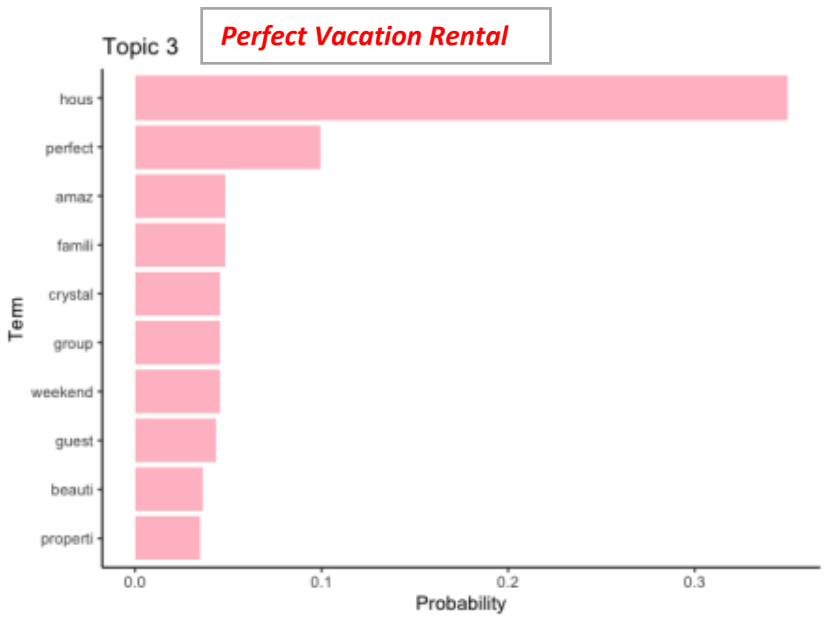


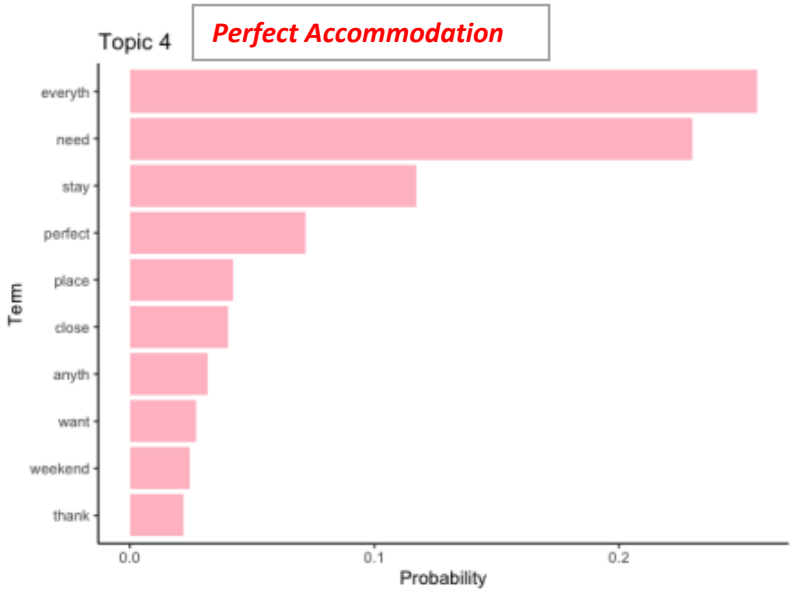


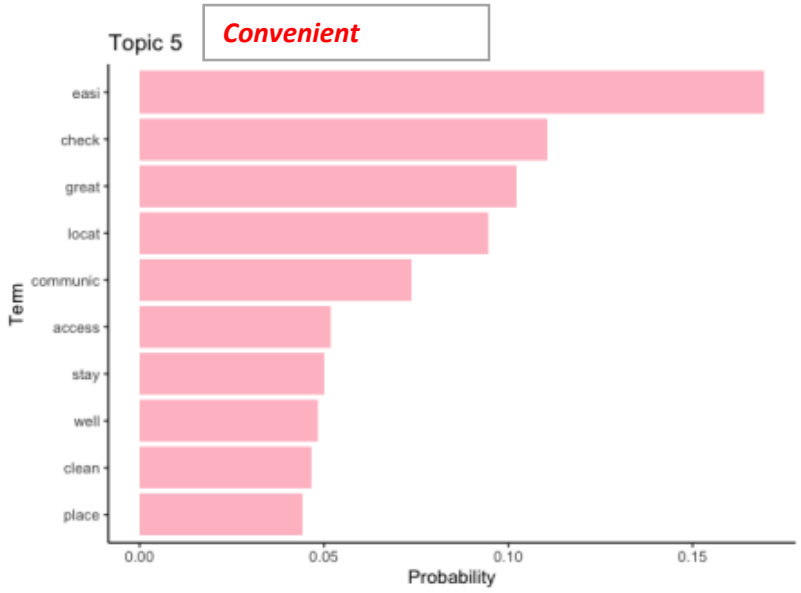
Topic Word Distribution







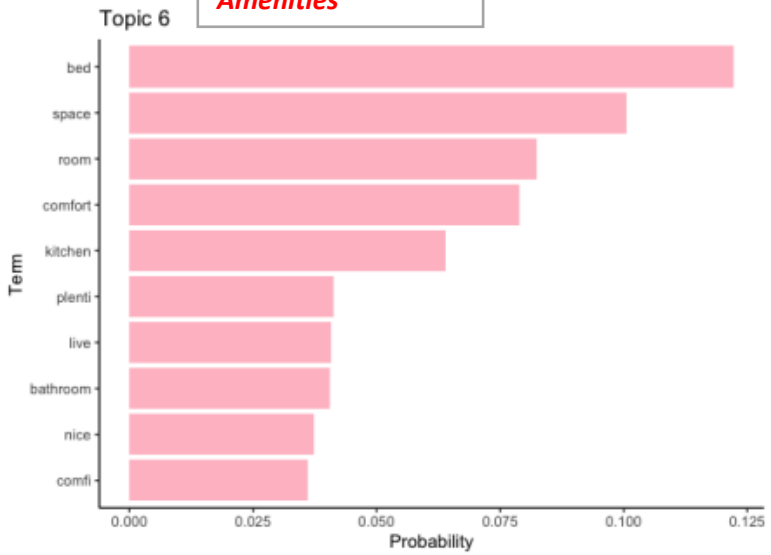




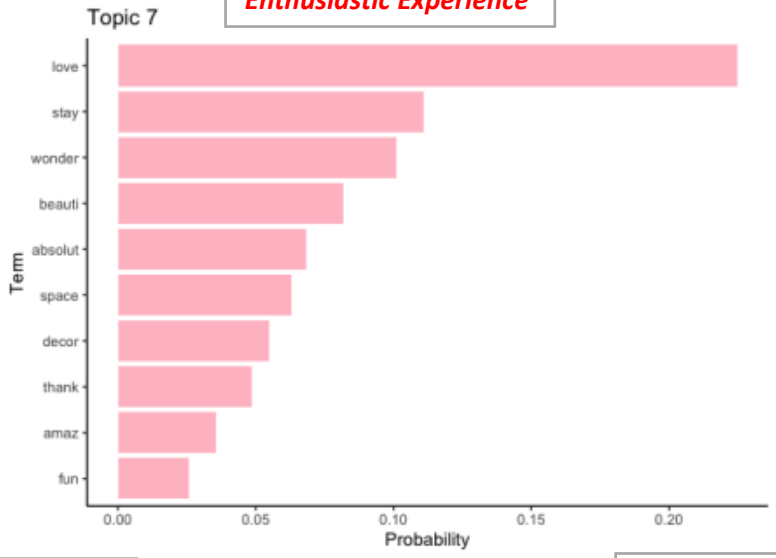
airbnb

Topic Word Distribution

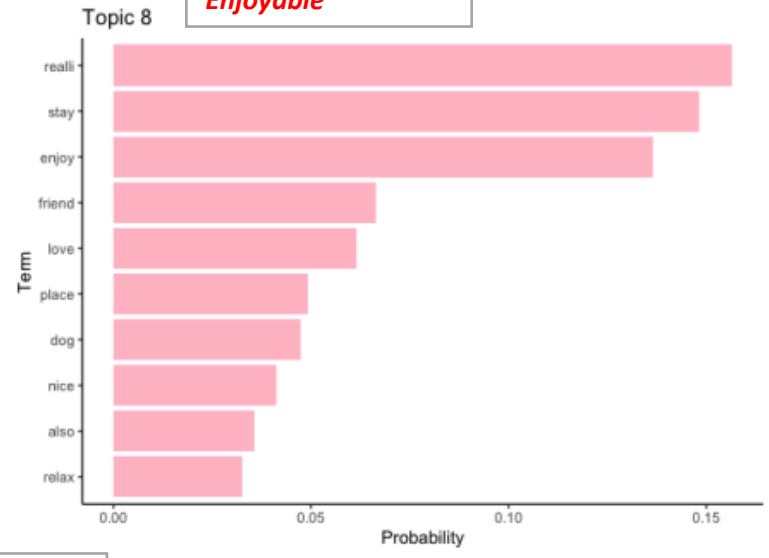
Amenities



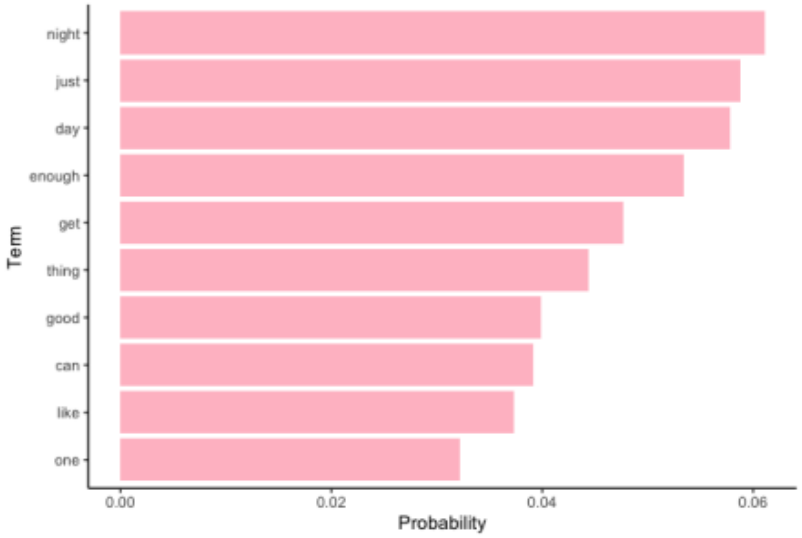
Enthusiastic Experience



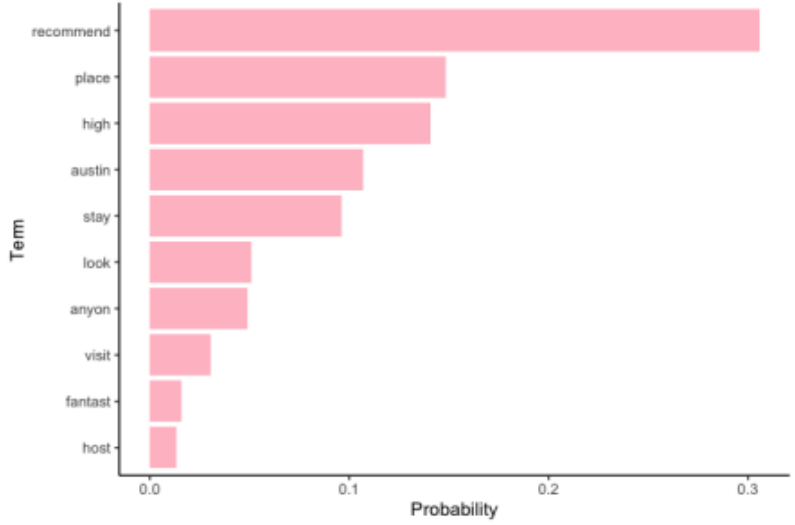
Enjoyable



Pet-Friendly Accommodations

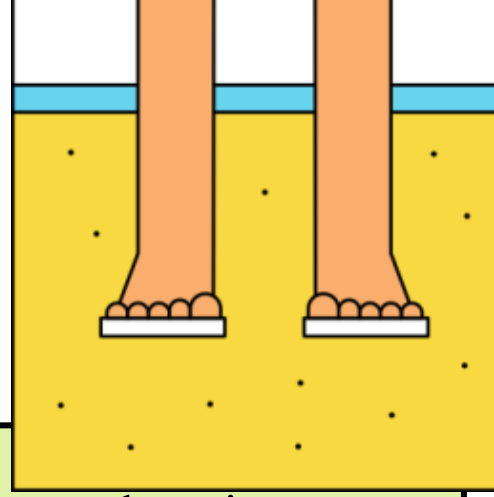


Fantastic Hosts



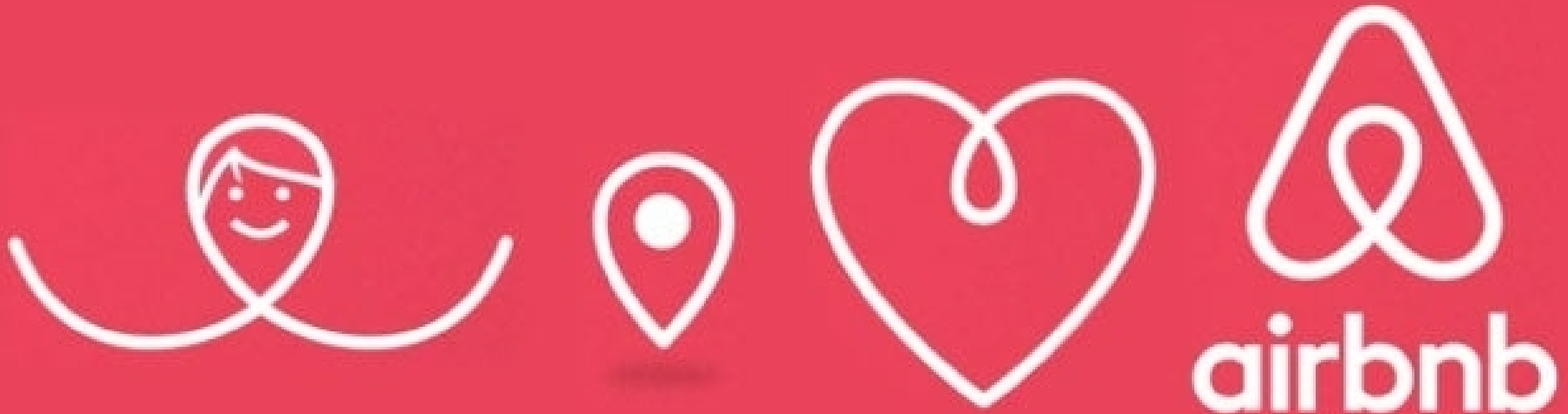
airbnb

Summary Recommendations



- Focus on these qualities which impact review scores (cleanliness, accuracy, location, value, communication) and availability
- Try to list your listing so that the accomodates do not exceed bedrooms by too much
- Higher priced listings had higher probability of receiving a 4.9 star rating or higher, which means that when guests book their listings, they may not necessarily care about the price if it truly does give them more value
- Ask AirBnB to do special follow ups for an additional feedback to help hosts improve
- Ask guests what we are doing good at so that we can continue to work on them
- Hosts can provide the guest with great places to visit around the Airbnb and tips to get around the city.
- Provide additional amenities

airbnb



Thank you!