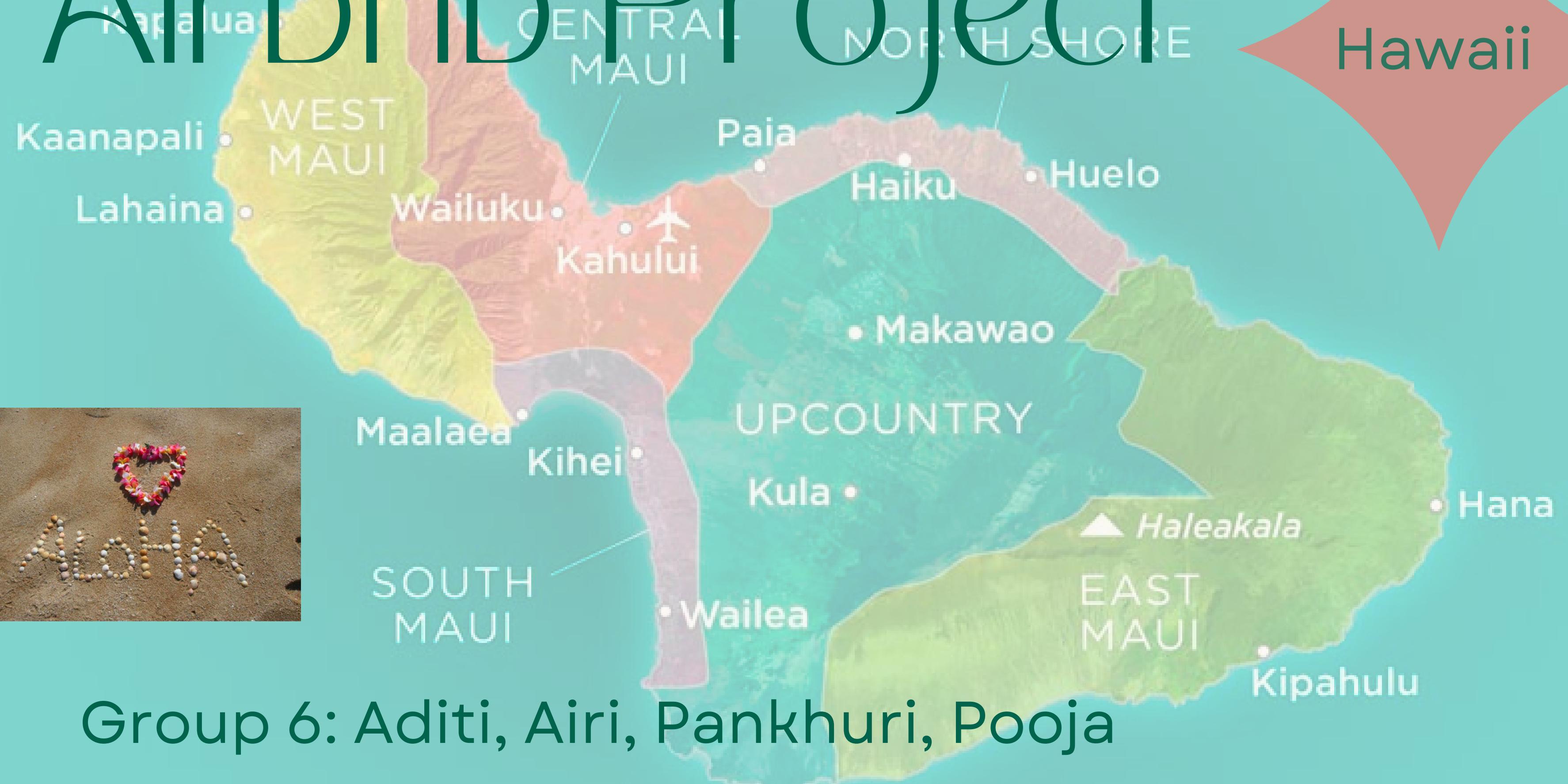
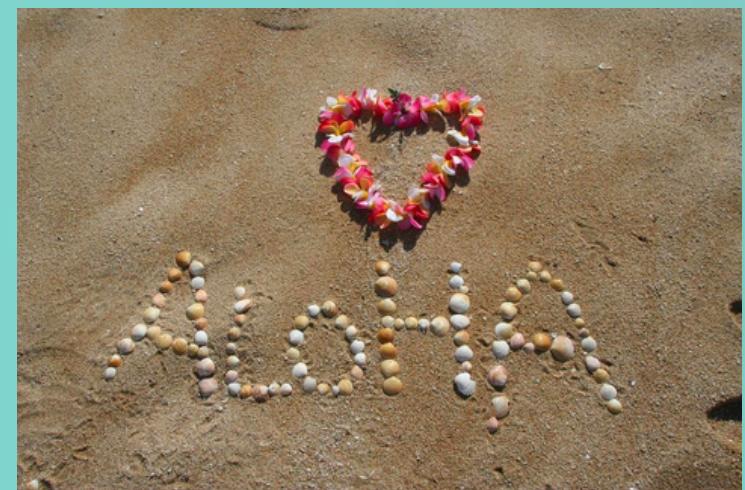


Airbnb Project



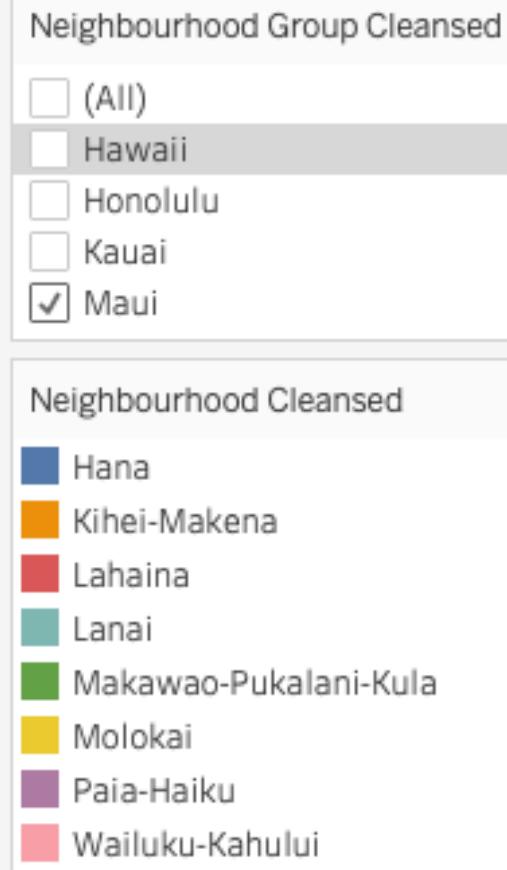
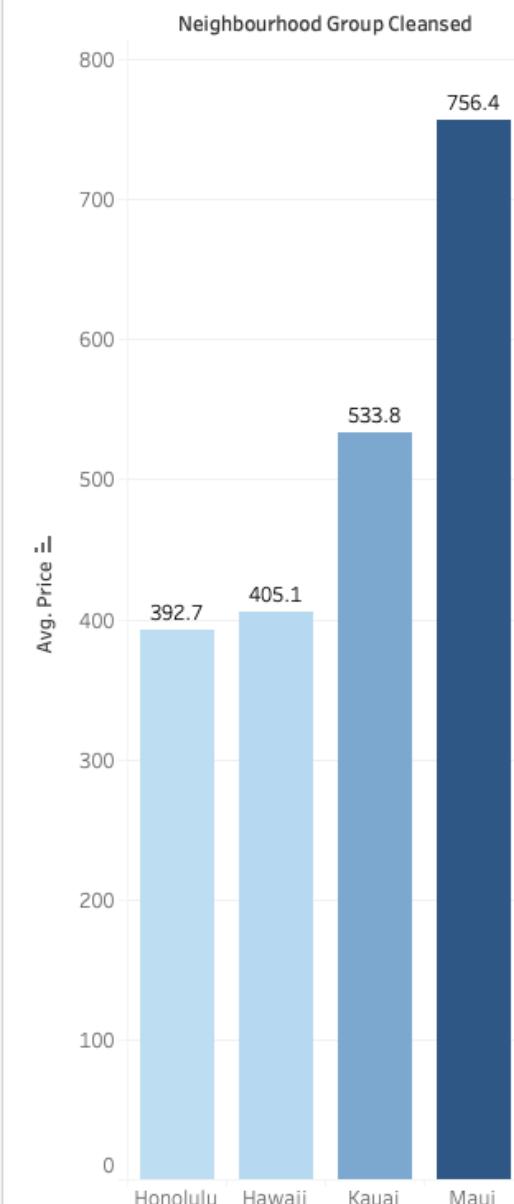
Recap: Airbnb Market in Maui

Maui

- The highest price per night in the neighborhood is Maui.
- There are total of 9,945 listings in Maui.
- In Maui, Lahaina has the highest average price per night, followed by Kihei.



Average Price per Night in Hawaii by Neighbourhood





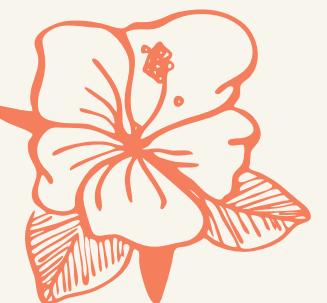
Identify a Problem

- As a guest of Airbnb, we would like to know **Price** per night (Y variable).

<What factors influence price?>

- Number of people?
- Neighborhood?
- Room type?
- Number of bedrooms?
- Which neighborhood has the lowest/highest average price?
- Whether or not the listing is instantly bookable
- **These questions are important in determining and predicting the budget for their stay.**

Recap: Recommendation



- We concluded that all of the variables influence the price, but the neighborhood influences the price the most in our model.
- Kihei and Lahaina are the most visited neighborhoods in Maui and that's where we can find the most Airbnb listings. As a guest, you have to make adjustments to the variables in considering the budget.

By using the profiler, we were able to predict the average price per night in using these variables.

- **Students Vacation Trip:** Private room, Instantly Bookable True, 4 people, Molokai, 2 bedroom - on average, \$270 price per night
- **Couple Vacation:** Private room, Instantly Bookable True, 2 people, Lahaina, 1 bedroom - on average, \$493 price per night
- **Family Get-Together:** Entire Home, Instantly Bookable True, 20 people, Lahaina, 10 bedroom - on average, \$18,033 price per night

Objectives

- Jackknife Analysis - removing outliers
- Unsupervised Learning – PCA, Factor Analysis & Clustering
- Supervised Learning - 3 ML Models & Model Performance
- Text Mining - Word Count and Cloud, Sentiment Analysis, and Topic Modeling
- *Price Prediction profiler using the best ML Model

Data: We used the detailed listing data from Inside Airbnb (Hawaii)

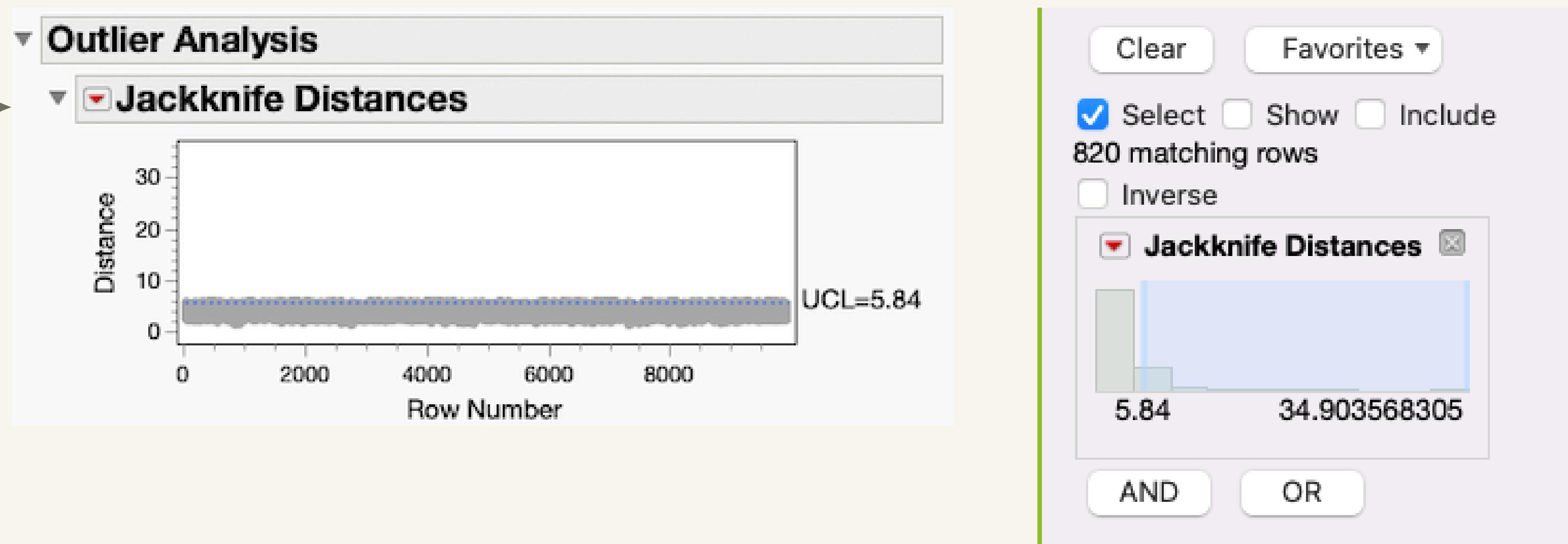
- **Listings_detailed**
- **Reviews**



Jackknife Analysis: Removing Outliers

Potential continuous x-variables

- Since our data contains outliers, we used jackknife analysis to remove them - 820 listings/values





Unsupervised Learning (Principal Component Analysis)

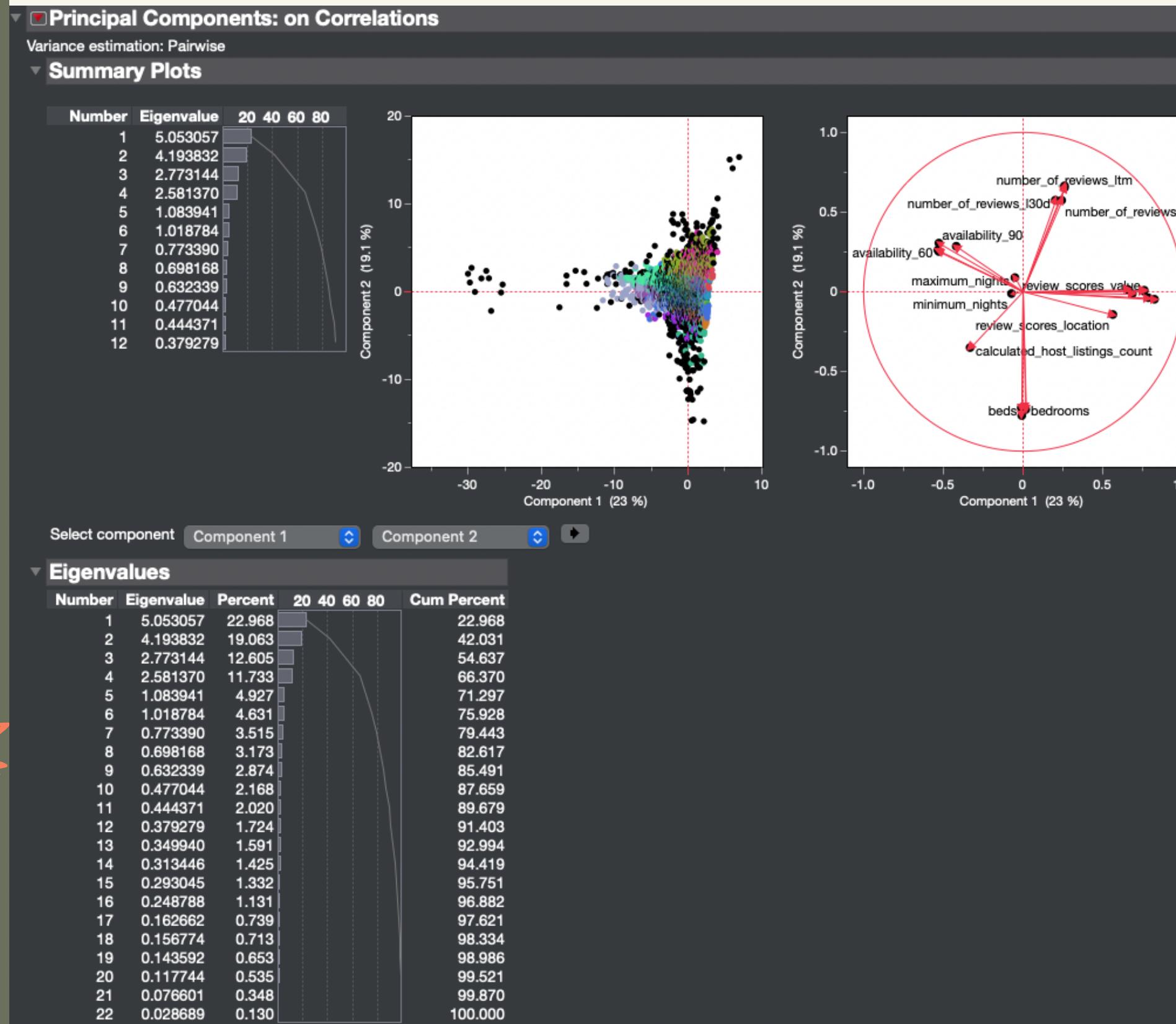
Potential continuous x-variables - 2

accommodates
number_of_bathroom
bedrooms
beds
minimum_nights
maximum_nights
availability_30
availability_60
availability_90
availability_365
number_of_reviews
number_of_reviews_ltm
number_of_reviews_l30d
review_scores_rating
review_scores_accuracy
review_scores_cleanliness
review_scores_checkin
review_scores_communication
review_scores_location
review_scores_value
reviews_per_month
calculated_host_listings_count

- Some of these variables are highly correlated (# of bedrooms, # of bathrooms, availability, and # of reviews).
 - These may result in multicollinearity - variables have a higher correlation with one another and may undermines the statistical significant of an independent variable.

	✓ Multivariate										
	▼ Correlations										
accommodates	1.0000	0.7688	0.8404	0.8397	-0.0140	-0.0386	-0.1092	-0.0689	-0.1040	-0.0749	-0.2031
number of bathroom	0.7688	1.0000	0.8195	0.6986	-0.0097	-0.0149	-0.1176	-0.0704	-0.0929	-0.0515	-0.1713
bedrooms	0.8404	0.8195	1.0000	0.8388	0.0317	-0.0081	-0.1127	-0.0609	-0.0912	-0.0386	-0.1196
beds	0.8397	0.6986	0.8388	1.0000	-0.0161	-0.0174	-0.1100	-0.0602	-0.0906	-0.0517	-0.1590
minimum_nights	-0.0140	-0.0097	0.0317	-0.0161	1.0000	0.0410	0.1032	0.0761	0.0593	0.0003	-0.0553
maximum_nights	-0.0386	-0.0149	-0.0081	-0.0174	0.0410	1.0000	0.0223	0.0357	0.0520	0.0527	0.1483
availability_30	-0.1092	-0.1176	-0.1127	-0.1100	0.1032	0.0223	1.0000	0.8864	0.8121	0.5228	-0.0991
availability_60	-0.0689	-0.0704	-0.0609	-0.0602	0.0761	0.0357	0.8864	1.0000	0.9556	0.6491	-0.0727
availability_90	-0.1040	-0.0929	-0.0912	-0.0906	0.0593	0.0520	0.8121	0.9556	1.0000	0.7150	-0.0423
availability_365	-0.0749	-0.0515	-0.0386	-0.0517	0.0003	0.0527	0.5228	0.6491	0.7150	1.0000	0.0642
number_of_reviews	-0.2031	-0.1713	-0.1196	-0.1590	-0.0553	0.1483	-0.0991	-0.0727	-0.0423	0.0642	1.0000
number_of_reviews_ltm	-0.2250	-0.2003	-0.1376	-0.1673	-0.0770	0.0968	-0.0836	-0.0405	0.0029	0.1176	0.8060
number_of_reviews_l30d	-0.1697	-0.1608	-0.1008	-0.1253	-0.0587	0.0378	-0.0232	0.0261	0.0594	0.1224	0.5416
review_scores_rating	0.0137	0.0519	0.0299	0.0082	-0.1125	-0.0468	-0.1841	-0.1747	-0.1770	-0.1551	0.0423
review_scores_accuracy	0.0130	0.0471	0.0193	0.0015	-0.0283	-0.0334	-0.1848	-0.1729	-0.1747	-0.1357	0.0674
review_scores_cleanliness	-0.0073	0.0253	0.0096	-0.0115	0.0097	-0.0261	-0.1388	-0.1221	-0.1296	-0.1291	0.0446
review_scores_checkin	0.0024	0.0287	0.0161	-0.0058	0.0006	-0.0395	-0.1824	-0.1586	-0.1499	-0.1106	0.1068
review_scores_communication	0.0182	0.0184	0.0303	0.0107	-0.0185	-0.0379	-0.2139	-0.1886	-0.1900	-0.1580	0.1083
review_scores_location	0.0618	0.0756	0.0350	0.0411	-0.0151	-0.0340	-0.1357	-0.1356	-0.1467	-0.1322	-0.0337
review_scores_value	-0.0227	-0.0054	-0.0239	-0.0328	0.0197	-0.0218	-0.1998	-0.2086	-0.2133	-0.1878	0.1046
calculated_host_listings_count	0.2183	0.2211	0.1529	0.1976	-0.0765	0.0178	0.0937	0.0885	0.0973	0.1415	-0.2719
reviews_per_month	-0.2214	-0.2367	-0.1442	-0.1741	-0.0617	-0.0054	-0.0592	-0.0187	0.0202	0.0832	0.6451

Principle Components Analysis



- To avoid the multicollinearity, we conducted the PCA to summarize large number of similar/redundant variables into several factors.
- As a result, we got 6 components with their Eigenvalues > 1.
- These components capture approximately 76% of information contained in our 22 variables.

Factor Analysis:

Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
accommodates	0.007794	0.922750	-0.136175	-0.052945	-0.034504	-0.025898
number of bathroom	0.042029	0.876468	-0.131813	-0.043635	-0.046955	0.019968
bedrooms	0.012372	0.949861	-0.037465	-0.036982	0.032434	0.001633
beds	-0.006712	0.910700	-0.077337	-0.043158	-0.031461	-0.008906
minimum_nights	-0.036499	0.028149	-0.108391	0.071383	0.886612	0.102190
maximum_nights	-0.027394	-0.014899	0.069877	0.026947	0.031047	0.967874
availability_30	-0.132037	-0.093516	-0.095504	0.880361	0.084134	-0.017685
availability_60	-0.111580	-0.029640	-0.034667	0.962429	0.047110	-0.008976
availability_90	-0.117411	-0.056001	0.008321	0.956844	0.011204	0.011691
availability_365	-0.111777	-0.000901	0.129111	0.776761	-0.125187	0.054511
number_of_reviews	0.034503	-0.095313	0.823457	-0.053504	0.016509	0.168393
number_of_reviews_ltm	0.044146	-0.098766	0.950465	-0.001119	-0.013331	0.057688
number_of_reviews_l30d	0.042312	-0.055621	0.831912	0.067546	-0.020793	-0.038274
review_scores_rating	0.918502	0.011144	0.029902	-0.067110	-0.102113	-0.018720
review_scores_accuracy	0.869607	0.010516	0.034495	-0.067857	-0.029294	0.007418
review_scores_cleanliness	0.772957	-0.006508	0.030150	-0.032900	0.001988	0.017202
review_scores_checkin	0.686948	0.017649	0.111269	-0.080890	0.101978	-0.069910
review_scores_communication	0.786605	0.023701	0.121381	-0.108199	0.067272	-0.057394
review_scores_location	0.658771	0.032043	-0.124470	-0.066315	-0.034254	0.032544
review_scores_value	0.798870	-0.038780	0.062120	-0.128807	0.080772	0.017887
reviews_per_month	0.056986	-0.115745	0.906731	0.014075	0.022823	-0.086848
calculated_host_listings_count	-0.196797	0.207940		0.127849	-0.490102	0.177434

Rotated Factor Loading

We renamed these 6 Factors as:

- Ratings (Reviews Scores)
- Room (# of bedrooms, bathrooms, beds, and accommodation)
- Number of Reviews
- Instant Availability
- Minimum # of Nights
- Maximum # of Nights

Also, Multivariate data shows that these factors are not correlated.

▼ Multivariate

▼ Correlations

	Ratings	Number of Reviews	Instant Availability	Room	Minimum Nights	Maximum Nights
Ratings	1.0000	0.0027	-0.0298	-0.0098	0.1818	-0.0240
Number of Reviews	0.0027	1.0000	-0.0209	0.0333	0.1780	-0.0142
Instant Availability	-0.0298	-0.0209	1.0000	0.0224	-0.2551	-0.0148
Room	-0.0098	0.0333	0.0224	1.0000	-0.0396	-0.0071
Minimum Nights	0.1818	0.1780	-0.2551	-0.0396	1.0000	-0.1673
Maximum Nights	-0.0240	-0.0142	-0.0148	-0.0071	-0.1673	1.0000



Linear Regression Model

Response price

Effect Summary

Source	Logworth	PValue
Room	95.802	0.00000
Number of Reviews	28.633	0.00000
Ratings	25.656	0.00000
Minimum Nights	4.196	0.00006
Maximum Nights	1.788	0.01630
Instant Availability	0.083	0.82632

Remove Add Edit FDR

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	6184	4426204779	715751	.
Pure Error	3	0	0	Prob > F
Total Error	6187	4426204779	.	.

Max RSq
1.0000

Summary of Fit

RSquare	0.103672
RSquare Adj	0.102802
Root Mean Square Error	845.8156
Mean of Response	531.1723
Observations (or Sum Wgts)	6194

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	511945890	85324315	119.2673
Error	6187	4426204779	715404.04	Prob > F
C. Total	6193	4938150669	.	<.0001*

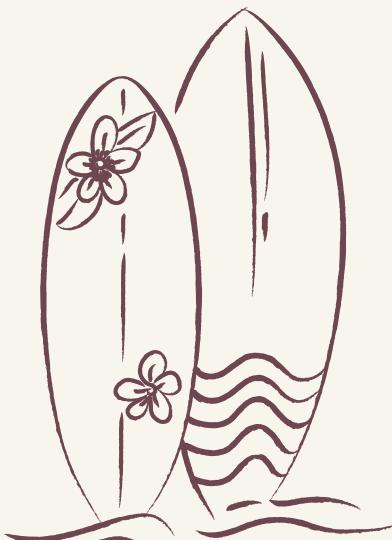
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	542.90167	11.0715	49.04	<.0001*
Ratings	136.53365	12.78748	10.68	<.0001*
Number of Reviews	-135.4464	11.97808	-11.31	<.0001*
Instant Availability	-2.658862	12.11686	-0.22	0.8263
Room	293.30376	13.82148	21.22	<.0001*
Minimum Nights	-89.46983	22.35814	-4.00	<.0001*
Maximum Nights	26.031577	10.83433	2.40	0.0163*

Effect Tests

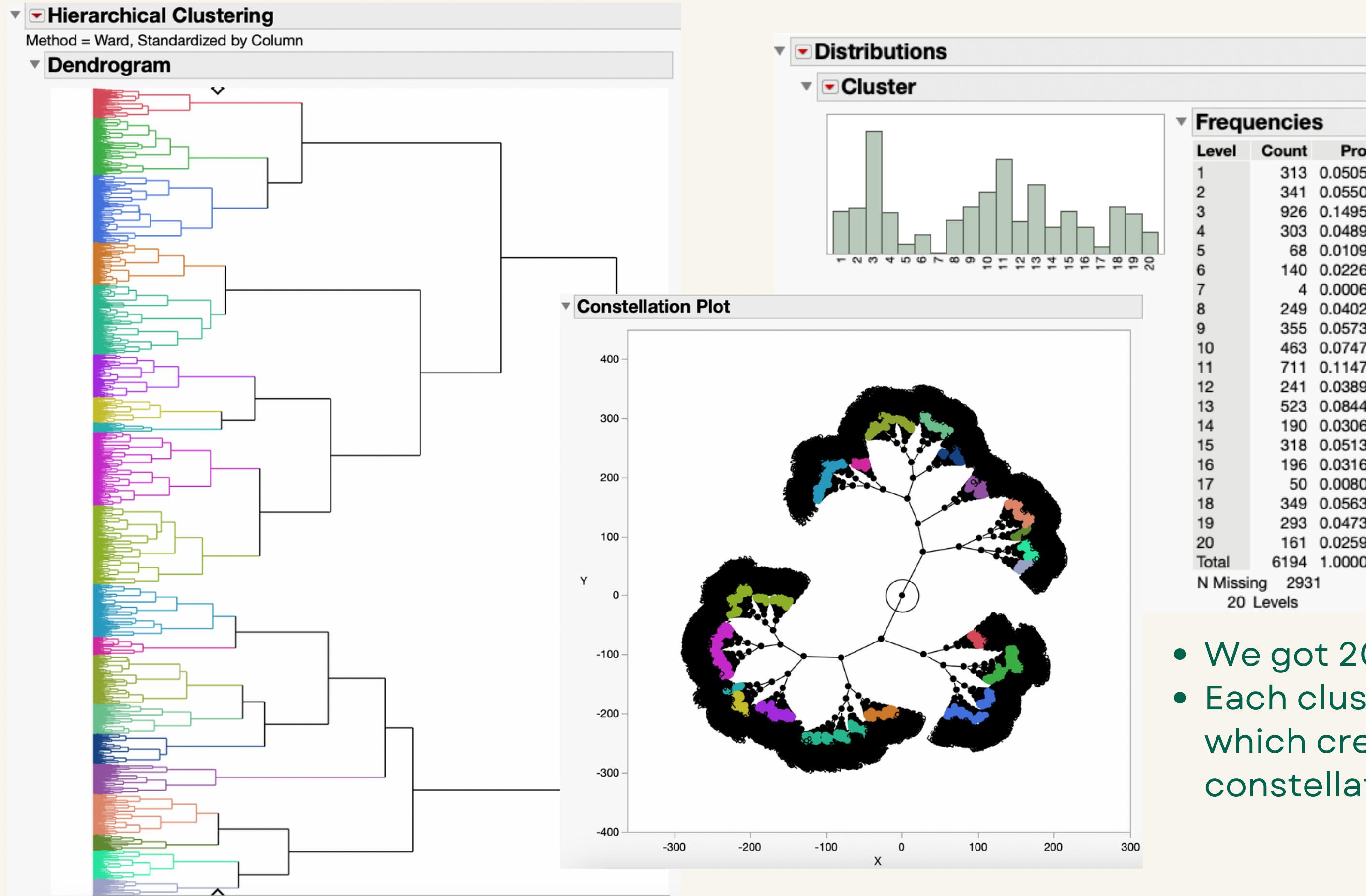
Effect Details

- The most important factor is the Room, followed by Number of reviews, Ratings, which matter in predicting our y-variable (price per night).
- These are all statistically significant with our y-variable, as T-value indicates that all values have probability of less than 0.05, except instant availability.



Clustering

6 Factors + Room type + neighborhoods_cleansed + host_is_superhost

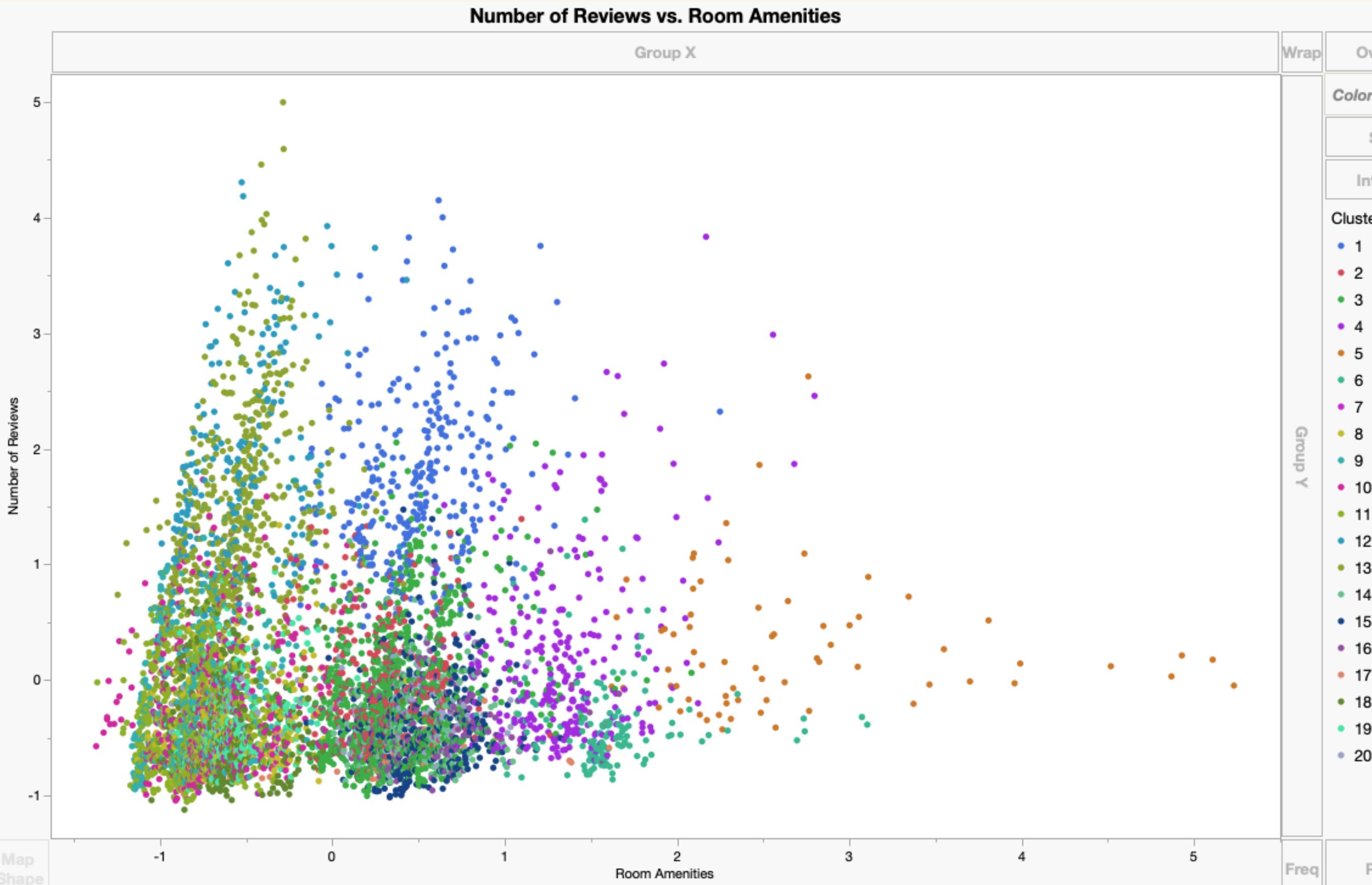


- We got 20 clusters
- Each cluster contains many listings, which created condensed constellation plot.

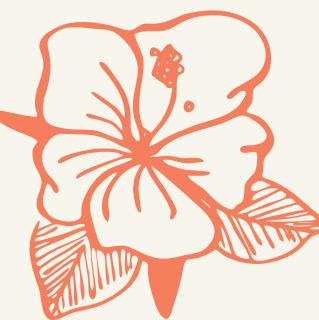


Clustering

Perceptual Map



- Same factors (colors) are located in same/similar area.
- Orange cluster tends to have higher number in room (# of bathrooms, bedrooms, beds, and accommodation).
- Light blue cluster seems to have higher number of reviews

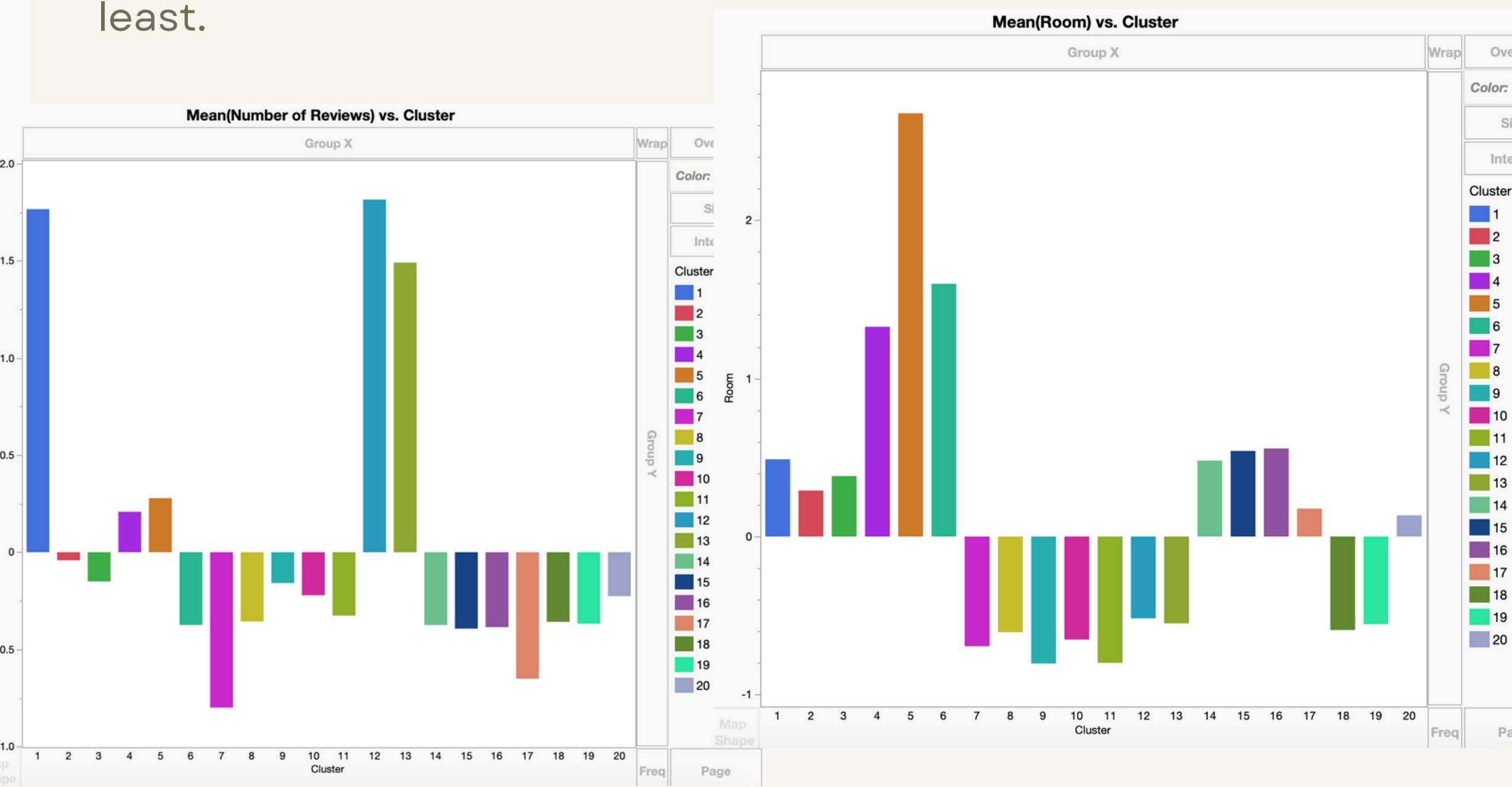
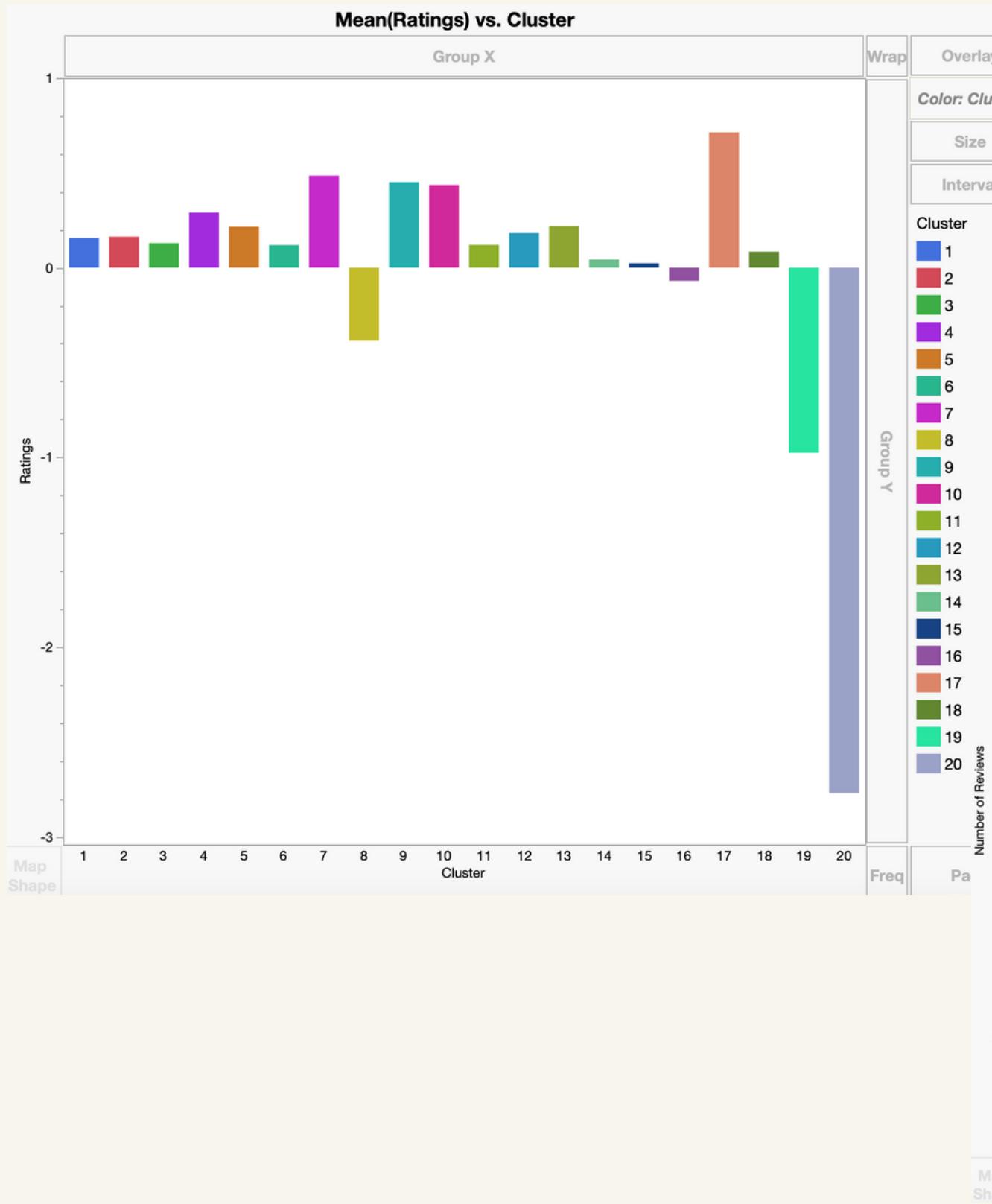




Clustering

Comparing Clusters

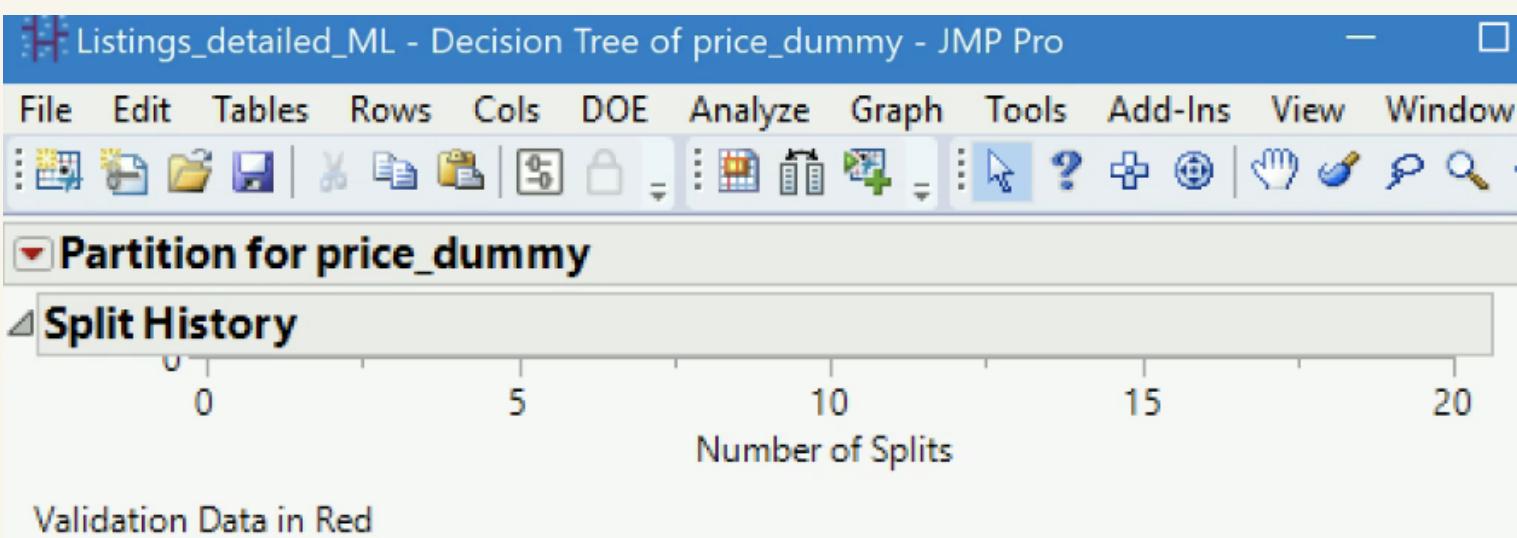
- Cluster 17 does the best in terms of average rating and cluster 20 has the lowest average rating.
- Cluster 12 has the highest average number of reviews and cluster 7 has the lowest.
- Cluster 5 has the highest average room, meaning more # of bedrooms, accommodation, and bathrooms, and cluster 9 has the least.





Supervised Learning

Decision Tree



Measure	Training	Validation	Definition
Entropy RSquare	0.1853	0.1695	$1 - \text{Loglike(model)}/\text{Loglike(0)}$
Generalized RSquare	0.3020	0.2792	$(1 - (L(0)/L(\text{model})))^{(2/n)} / (1 - L(0))^{(2/n)}$
Mean -Log p	0.5647	0.5756	$\sum -\text{Log}(p[j])/n$
RASE	0.4391	0.4436	$\sqrt{\sum (y[j] - p[j])^2/n}$
Mean Abs Dev	0.3858	0.3903	$\sum y[j] - p[j] /n$
Misclassification Rate	0.2964	0.3039	$(\sum (p[j] \neq p_{\text{Max}}))/n$
N	6838	2287	n

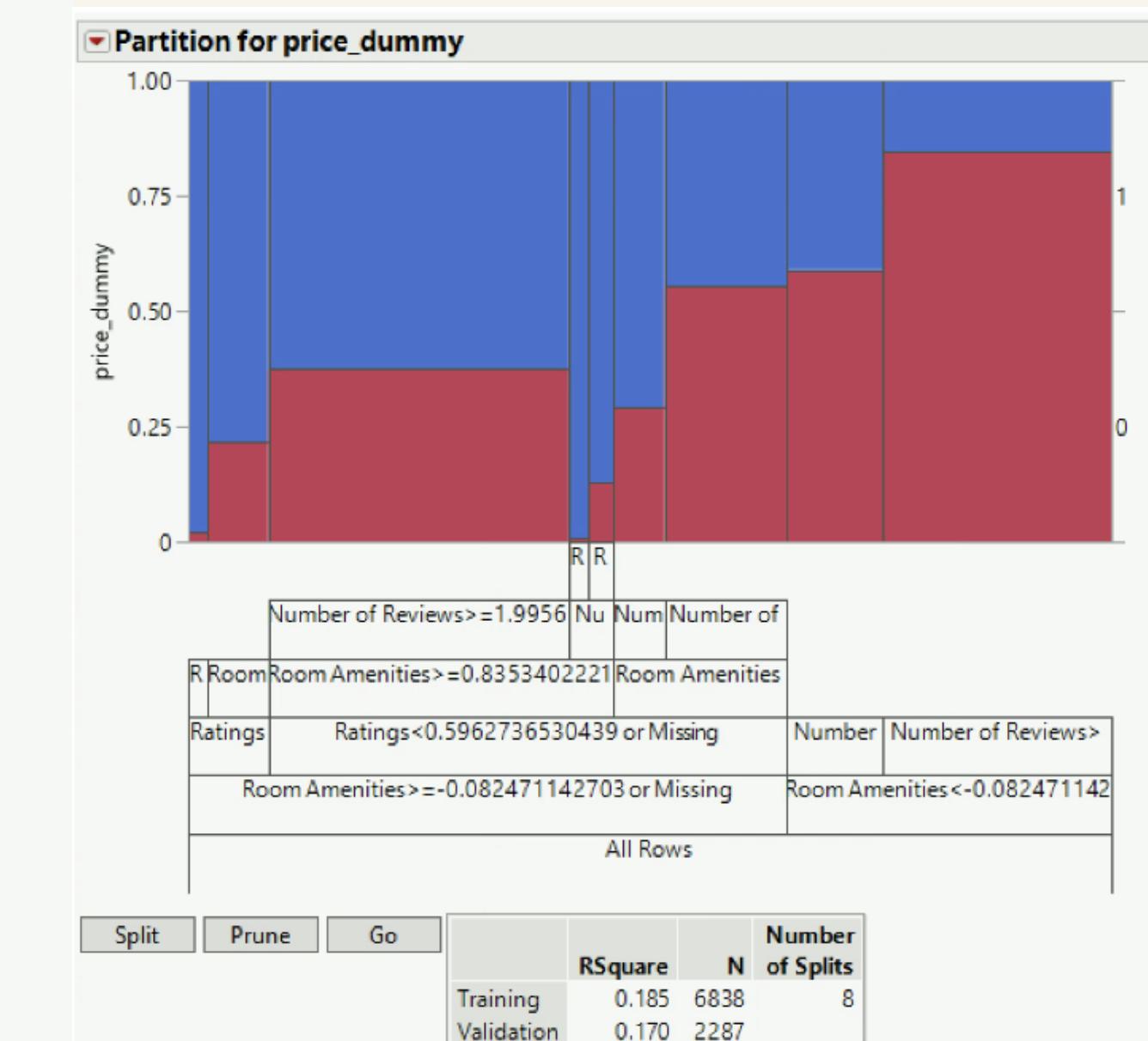
Confusion Matrix

Training		Validation	
Actual	Predicted Count	Actual	Predicted Count
price_dummy	0 1	price_dummy	0 1
0	2350 1066	0	789 348
1	961 2461	1	347 803

Actual	Predicted Rate	
price_dummy	0	1
0	0.688	0.312
1	0.281	0.719

Actual	Predicted Rate	
price_dummy	0	1
0	0.694	0.306
1	0.302	0.698

- There are 8 Splits in our decision tree.
- Of the 2,287 data on validation, 30.39% are misclassified.



Supervised Learning



Bootstrap Forest

Listings.detailed_ML - Bootstrap Forest of price_dummy - JM... - X

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View
Window Help

Bootstrap Forest for price_dummy

Specifications

Target	price_dummy	Training Rows:	6838
Validation Column:	Validation	Validation Rows:	2287
Number of Trees in the Forest:	100	Test Rows:	0
Number of Terms Sampled per Split:	5	Number of Terms:	6
		Bootstrap Samples:	6838
		Minimum Splits per Tree:	10
		Minimum Size Split:	9

Overall Statistics

Measure	Training	Validation	Definition
Entropy RSquare	0.3892	0.2131	$1 - \text{Loglike}(\text{model})/\text{Loglike}(0)$
Generalized RSquare	0.5560	0.3410	$(1 - (L(0)/L(\text{model}))^{(2/n)})/(1 - L(0)^{(2/n)})$
Mean -Log p	0.4234	0.5454	$\sum -\text{Log}(p[j])/n$
RASE	0.3694	0.4309	$\sqrt{\sum (y[j] - p[j])^2/n}$
Mean Abs Dev	0.3146	0.3729	$\sum y[j] - p[j] /n$
Misclassification Rate	0.1925	0.2798	$\sum (p[j] \neq p_{\text{Max}})/n$
N	6838	2287	n

Confusion Matrix

Training		Validation	
Actual	Predicted Count	Actual	Predicted Count
price_dummy	0 1	price_dummy	0 1
0	2418 998	0	726 411
1	318 3104	1	229 921

Actual	Predicted Rate
price_dummy	0 1
0	0.708 0.292
1	0.093 0.907

Actual	Predicted Rate
price_dummy	0 1
0	0.639 0.361
1	0.199 0.801

- Misclassification Rate: 27.98% (0.2798)
- Of the 2,287 data on validation, 27.98% are misclassified.

Supervised Learning



Neural Network

Listings_detailed_ML - Neural of price_dummy - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help

Validation Column: Validation
Informative Missing

Model Launch

Model NTanH(3)

Training

price_dummy

Measures	Value
Generalized RSquare	0.3425409
Entropy RSquare	0.2141915
RASE	0.4307987
Mean Abs Dev	0.3705252
Misclassification Rate	0.2860486
-LogLikelihood	3724.5265
Sum Freq	6838

Validation

price_dummy

Measures	Value
Generalized RSquare	0.3250907
Entropy RSquare	0.2015995
RASE	0.4339492
Mean Abs Dev	0.3743613
Misclassification Rate	0.2833406
-LogLikelihood	1265.617
Sum Freq	2287

Confusion Matrix

		Predicted
		Count
Actual	Count	
price_dummy	0 1	
0	2118 1298	
1	658 2764	

Confusion Rates

		Predicted
		Rate
Actual	Rate	
price_dummy	0 1	
0	0.620 0.380	
1	0.192 0.808	

Confusion Matrix

		Predicted
		Count
Actual	Count	
price_dummy	0 1	
0	716 421	
1	227 923	

Confusion Rates

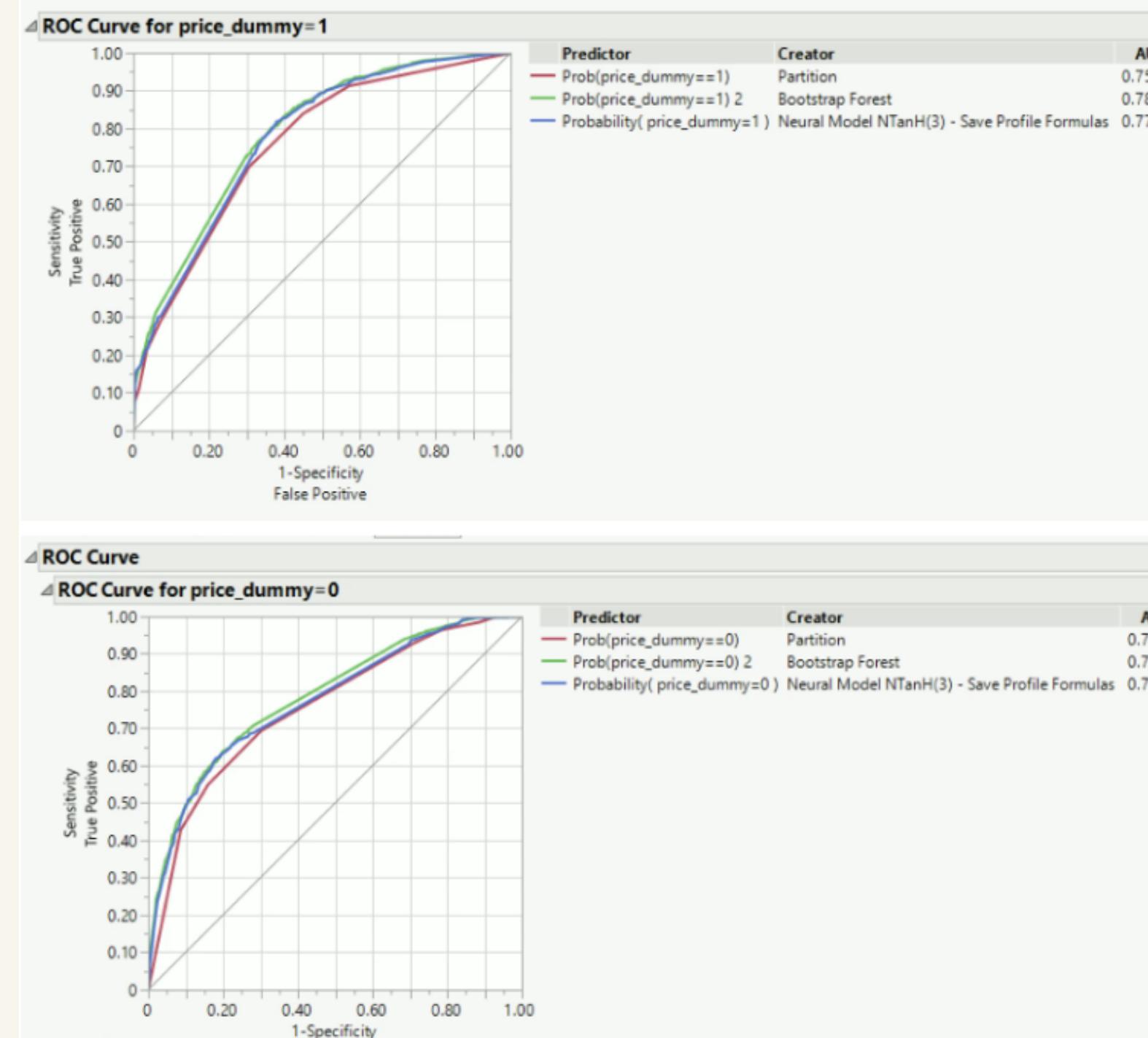
		Predicted
		Rate
Actual	Rate	
price_dummy	0 1	
0	0.630 0.370	
1	0.197 0.803	

- Misclassification rate is 28.33% (0.2833)
- Of the 2,287 data on validation, 28.33% are misclassified.

Supervised Learning



Model Performance Comparison using the Misclassification Rate and AUC (ROC Chart)



- The lowest Misclassification is 27.98% and the highest AUC % is at 78.66%

Bootstrap Forest

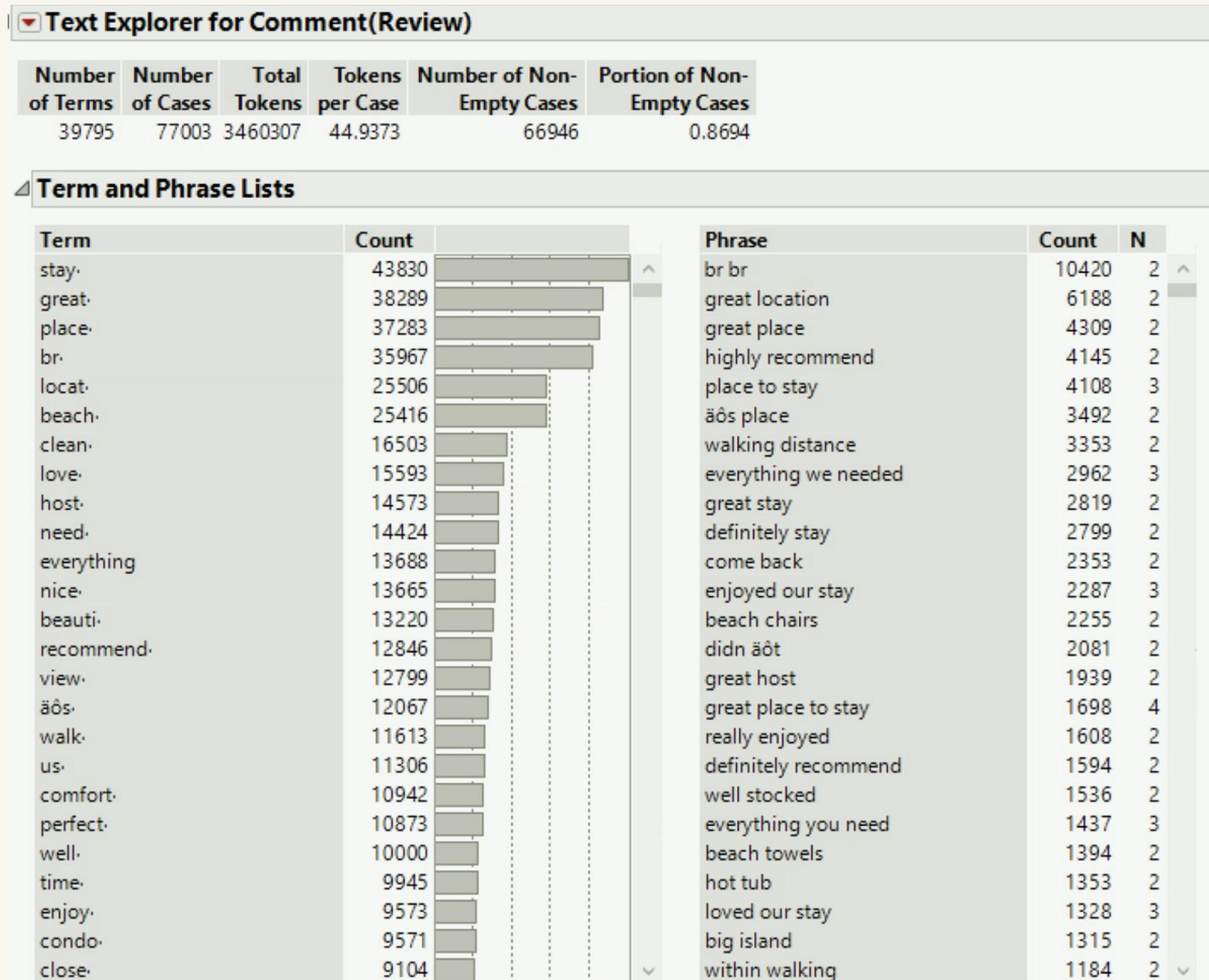
The screenshot shows the JMP Pro software interface with the title bar 'Listings_detailed_ML - Model Comparison - JMP Pro'. The main window displays 'Model Comparison Validation=Training' and 'Model Comparison Validation=Validation' sections. Under 'Predictors' and 'Measures of Fit for price_dummy', data is presented in tables for three models: Partition, Bootstrap Forest, and Neural Model NTanh(3) - Save Profile Formulas. The tables include columns for Creator, Entropy RSquare, Generalized RSquare, Mean -Log p, RASE, Abs Dev, Misclassification Rate, and N.

Creator	.2	.4	.6	.8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Abs Dev	Misclassification Rate	N
Partition					0.1853	0.3020	0.5647	0.4391	0.3858	0.2964	6838
Bootstrap Forest					0.3892	0.5560	0.4234	0.3694	0.3146	0.1925	6838
Neural Model NTanh(3) - Save Profile Formulas					0.2142	0.3425	0.5447	0.4308	0.3705	0.2860	6838

Creator	.2	.4	.6	.8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Abs Dev	Misclassification Rate	N
Partition					0.1695	0.2792	0.5756	0.4436	0.3903	0.3039	2287
Bootstrap Forest					0.2131	0.3410	0.5454	0.4309	0.3729	0.2798	2287
Neural Model NTanh(3) - Save Profile Formulas					0.2016	0.3251	0.5534	0.4339	0.3744	0.2833	2287

Text Mining

Word Count



- As shown below, most of the reviews consist of terms and phrases, such as “have stay, great, place, locat, beach, clean, love, etc.

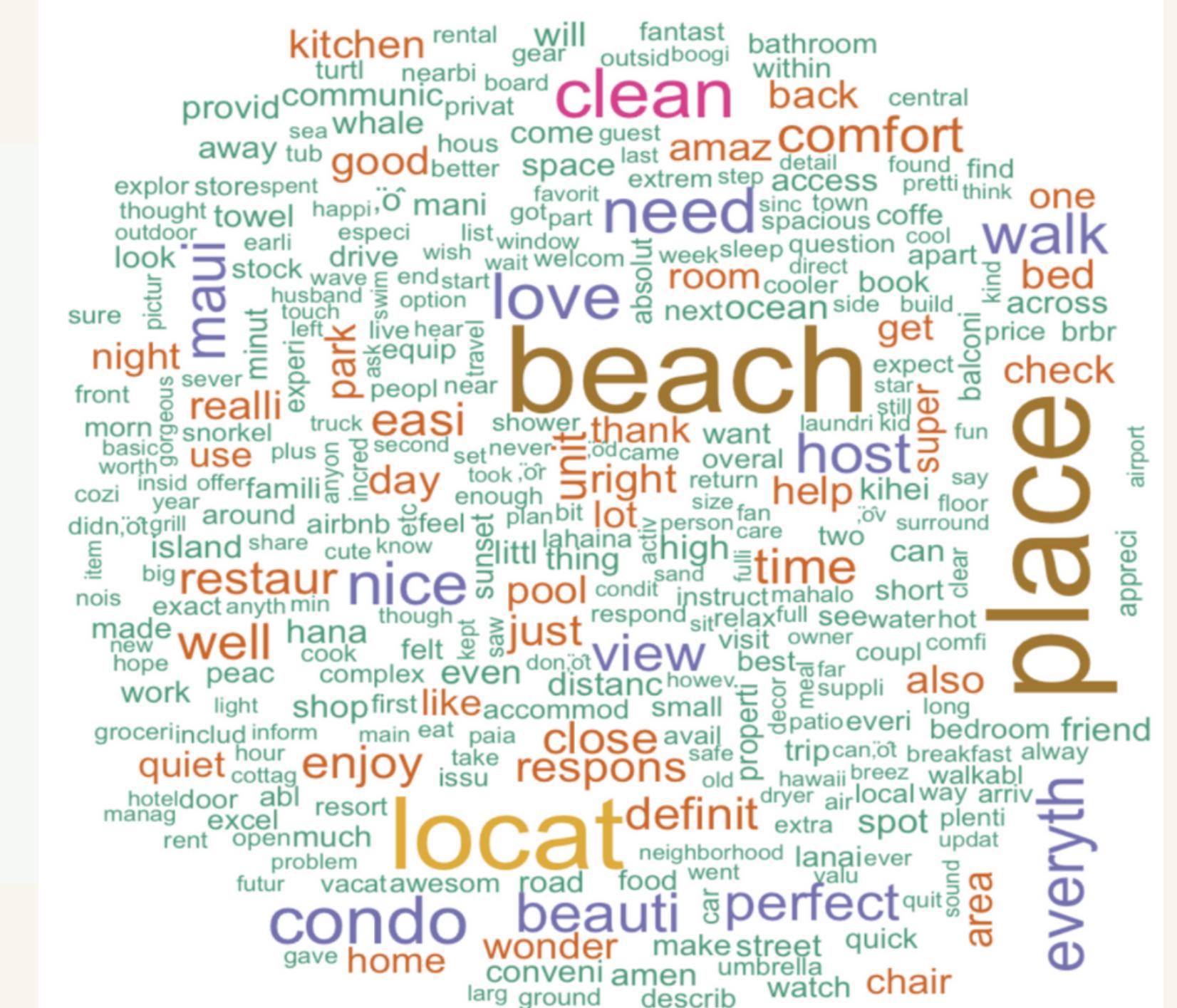
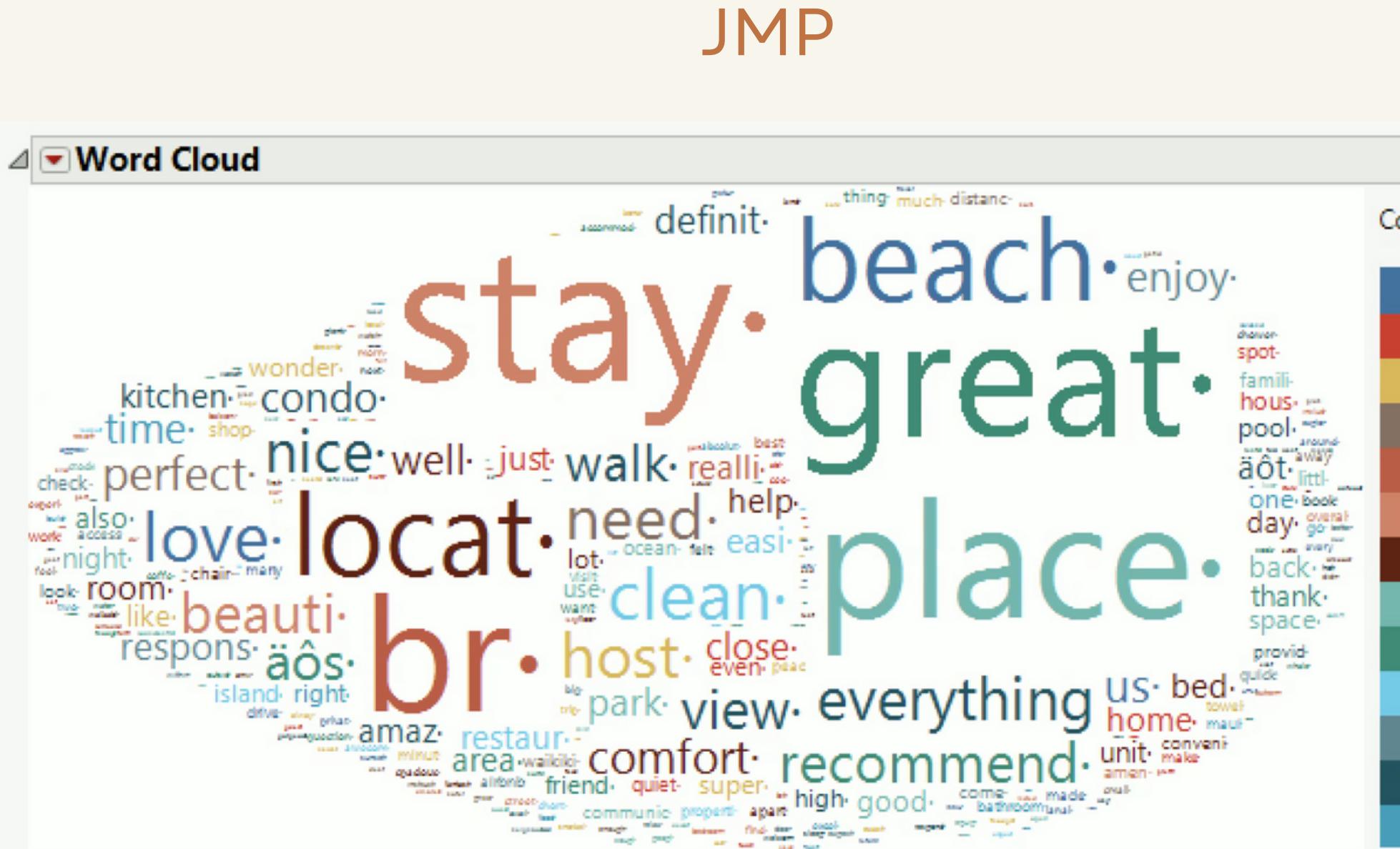




Text Mining

Word Cloud

R - Cleaned data

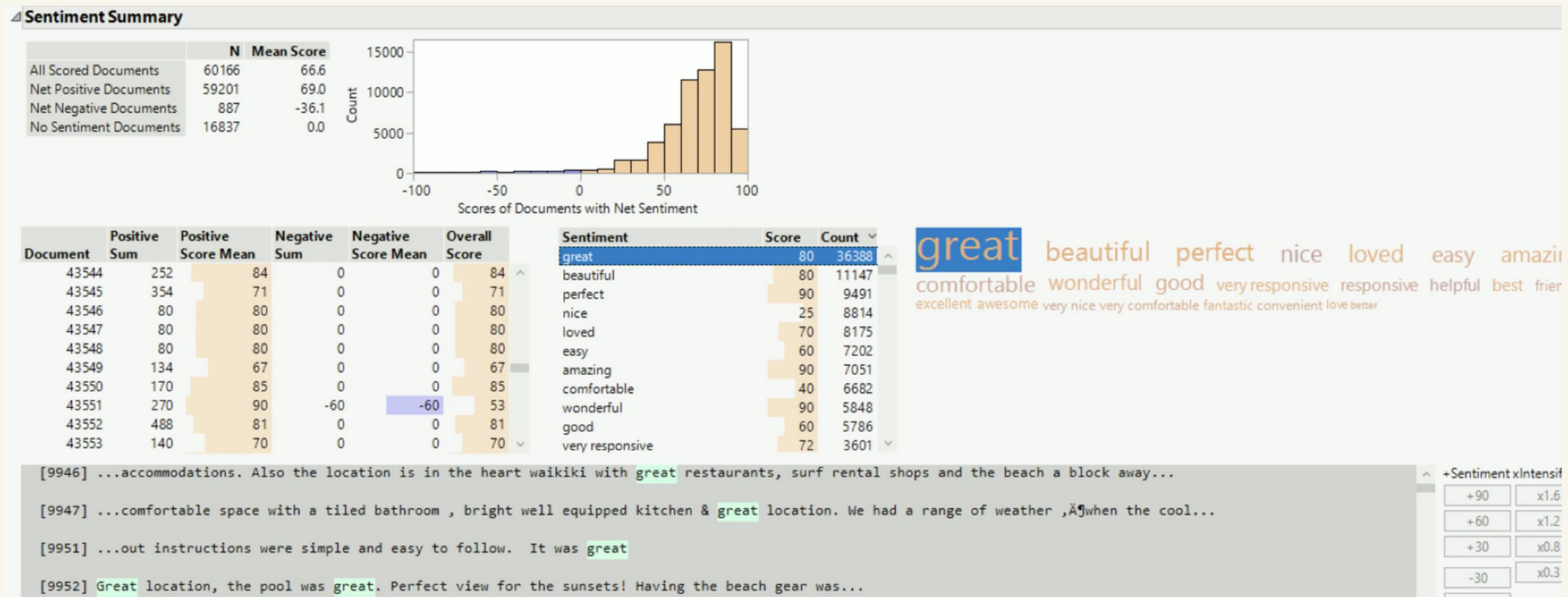


Text Mining



Sentiment Analysis

- As we explained previously, these terms and phrases are very positive, it is proven by the sentiment analysis that almost 70% are positive, including the word, “great, beautiful, perfect, loved, easy, amazing, comfortable, wonderful, etc.”





Text Mining

Topic Modeling (JMP)

Top Loadings by Topic									
Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
Term	Loading	Term	Loading	Term	Loading	Term	Loading	Term	Loading
åæ·	0.94361	und·	0.86780	et·	0.85140	µ·	0.99052	y·	0.810
éº·	0.90982	die·	0.84462	le·	0.76915	á·	0.98894	el·	0.800
åß·	0.90733	der·	0.74897	est·	0.72907	å·	0.98840	muy·	0.700
åü·	0.90537	ist·	0.73139	pour·	0.72051	æ·	0.98492	todo·	0.670
åñ·	0.90040	das·	0.72653	vt·	0.71535	ω·	0.98100	lo·	0.640
åó·	0.89744	ein·	0.72034	de·	0.71488	≤·	0.98095	una·	0.640
åñ·	0.88591	war·	0.69604	nous·	0.68387	¥·	0.97814	para·	0.600
åå·	0.87830	sehr·	0.69481	les·	0.67725	ç·	0.97698	las·	0.590
å·	0.87449	wir·	0.68151	trv @ s·	0.67159	Π·	0.97510	en·	0.580
çå·	0.87316	mit·	0.65893	pas·	0.64631	è·	0.97309	que·	0.570
äç·	0.86564	zu·	0.65506	il·	0.60831	ää·	0.97292	la·	0.570
é≥·	0.85628	nicht·	0.65389	des·	0.60202	oo·	0.97223	tien·	0.560
åää·	0.85070	f/vºr·	0.63303	tout·	0.59736	j·	0.94761	playa·	0.540
åç·	0.84624	auch·	0.60403	avon·	0.59070	á·	0.94656	lugar·	0.520
çä·	0.83951	man·	0.59288	la·	0.58836	ø·	0.94528	pero·	0.510
åø·	0.83923	auf·	0.57522	avec·	0.56940	º·	0.94065	los·	0.500
åô·	0.83152					π·	0.93717		
éé·	0.82804					ð·	0.93581		
						é·	0.92873		

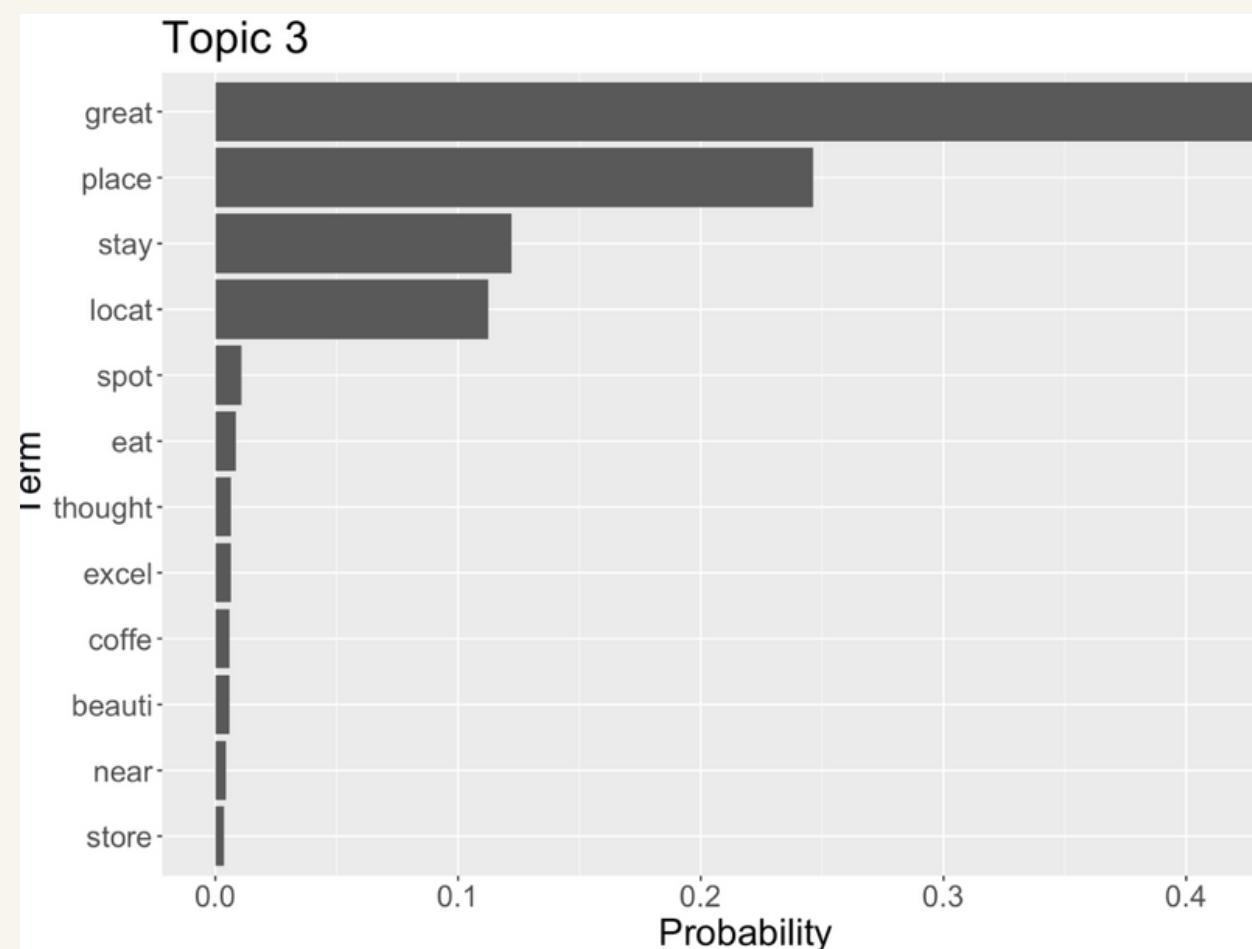
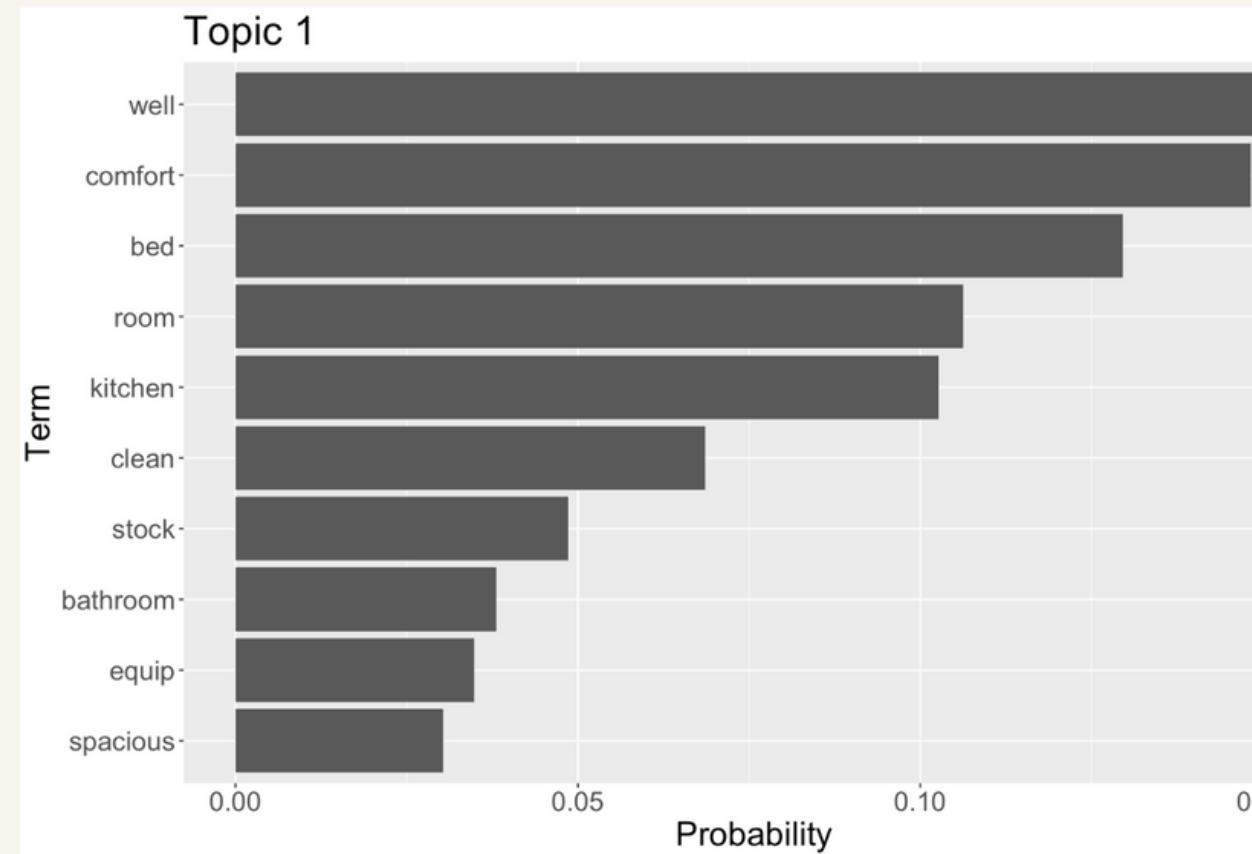
- With the topic modeling, Topic 1 to 5 has many unknown symbols where the system cannot translate into English.
 - This means that the reviews are left with other foreign languages. (Topic 2 contains German and Topic 3 contains French, Topic 5 and 9 contains Spanish or Portuguese).
 - Topic 6 contains amenities
 - Topic 8 contains words with location and amenities and nearby activities).

Topic 6		Topic 7		Topic 8		Topic 9		Topic 10	
Term	Loading	Term	Loading	Term	Loading	Term	Loading	Term	Loading
not	0.53772	iào·	0.61930	beach·	0.64610	o·	0.80424	us·	0.31623
no	0.40702	iúàïñýñú·	0.50575	walk·	0.44119	✓@·	0.78416	fruit·	0.30630
äöt·	0.35682	iúàïäö·	0.50261	chair·	0.43263	uma·	0.75327	home·	0.28915
br·	0.33159	iñà·ñ·	0.48032	restaur·	0.37231	e·	0.74904	so	0.27440
only	0.33086	iøññ·ú·	0.47011	kitchen·	0.36897	muito·	0.74262	island·	0.27288
dirti·	0.32102	iúàïñç·	0.44988	pool·	0.34977	nñ£o·	0.69338	stay·	0.27086
door·	0.31944	iúò·	0.44599	towel·	0.34203	bem·	0.68972	tree·	0.26307
bathroom·	0.29513	iññ@ññ·	0.44222	condo·	0.32185	com·	0.68753	frog·	0.25603
one·	0.29179	iúàïóàïäµññññ·	0.42943	shop·	0.32133	foi·	0.67559	peac·	0.24899
issu·	0.28791	iàññùäññ·	0.42748	well·	0.31586	mas·	0.65521	feel·	0.24709
if·	0.28539	iúàïäµññññ·	0.42060	board·	0.31524	localiza·/ßñ£o·	0.60706	beautiful	0.24512
room·	0.27292	içññññññññññ·	0.42059	snorkel·	0.31130	✓ñtima·	0.60190	bird·	0.23804
stain·	0.27238	iñññé·	0.41724	locat·	0.30866	poi·	0.60001	hous·	0.23602
so	0.27227	iàññüå·	0.41698	umbrella·	0.29714	praia·	0.53940	fresh·	0.23307
shower·	0.26954	iñññöø·	0.41317	also·	0.28930			recommend·	0.22724
use·	0.26923	içññùå·	0.40824					experi·	0.22597
		iù·	0.39699					natur·	0.22194

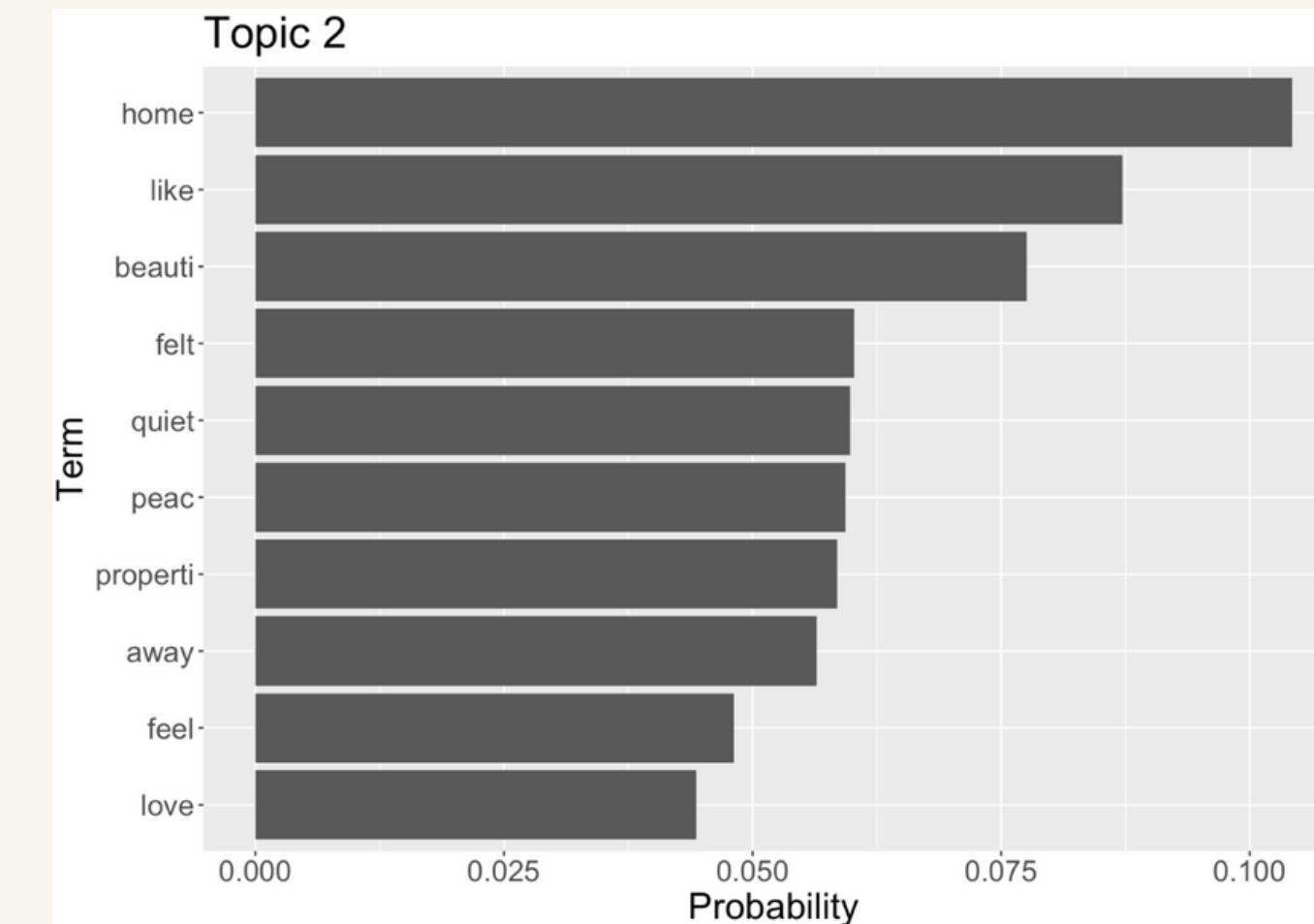


Text Mining

Topic Modeling (R with cleaned data)



- With cleaned data on R, we got related words for each topics.

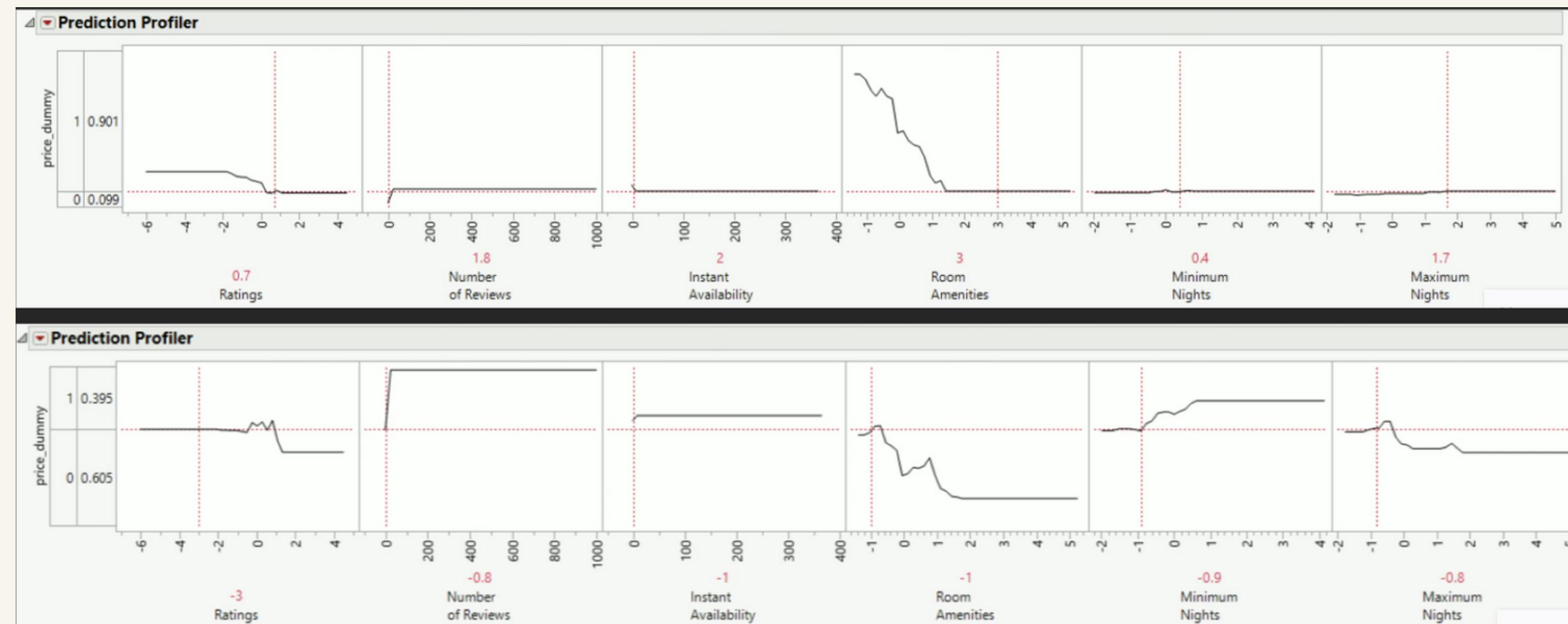


*Price Prediction Profiler - Recommendation

Price 0 = Below \$387, Price 1 = Above \$387 (Boot Strap Forest ML Model)

Scenario 1: More than 90% of our data is above \$387 when we have a high rating, high number of reviews, have more room (# of bathrooms, bedrooms, and beds), has many availabilities, and have higher minimum nights and higher maximum nights.

Scenario 2: More than 60% of data has listings below \$387 when it has the lower ratings, lower number of reviews, less availability, less room (# of bathrooms, bedrooms, and beds), lower minimum nights, and lower maximum nights.





Questions



