Joshua Joshua

AIRBNB ANALYSIS FOR ROME: HOST INSIGHTS

RESEARCH QUESTION

- POV: Airbnb Host
- Research Question:
  - What factors contribute most to maximizing occupancy rates for Airbnb listings in Rome?

DATA SELECTION CRITERION

Room Type

Only “Entire home/apt”

CENTRAL ISSUE AND INVESTIGATIVE QUESTION

- Central Issue:
  - Identifying key factors that Airbnb hosts in Rome can leverage to maximize occupancy rates

- Investigative Question:
  - What natural groupings or patterns exist among Airbnb listings in Rome based on property features, host attributes, and reviews?
    - Approach: (PCA, Factor Analysis), Clustering
  - Which property and host characteristics are the most important predictors of occupancy rates
    - Approach: (Random Forest and Neural Networks)
  - What themes emerge in guest reviews, and how do they relate to occupancy rates
    - Approach: (Topic Modelling, Word Cloud and Sentiment Analysis)

UNSUPERVISED LEARNING

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	5.062536	24.107		24.107
2	3.209286	15.282		39.390
3	2.772625	13.203		52.593
4	1.406748	6.699		59.291
5	1.102749	5.251		64.543
6	0.990609	4.717		69.260
7	0.958956	4.566		73.826
8	0.876051	4.172		77.998
9	0.721105	3.434		81.432
10	0.639745	3.046		84.478
11	0.588203	2.801		87.279
12	0.469097	2.234		89.513
13	0.457754	2.180		91.693
14	0.294925	1.404		93.097

Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
host_response_rate	0.086974	-0.003094	0.072799	0.838776	-0.028588
host_acceptance_rate	0.027924	-0.006072	0.183382	0.827771	0.088029
host_total_listings_count	-0.174993	-0.050741	-0.035879	0.127977	0.570539
price	0.056900	0.143949	-0.050086	0.016046	0.639302
accommodates	-0.013051	0.919052	-0.027890	0.032042	0.049036
Number Bathrooms	0.032127	0.804118	-0.026468	0.013902	0.083151
bedrooms	0.006931	0.904045	-0.034647	0.009318	0.041119
beds	-0.002021	0.869412	-0.013443	0.006278	0.010596
minimum_nights	0.018711	-0.023483	-0.013683	-0.274794	0.034713
maximum_nights	-0.078230	0.047005	0.028587	-0.221843	0.555293
number_of_reviews	0.053508	-0.009857	0.741379	-0.015114	0.080474
number_of_reviews_ltm	0.085148	-0.017778	0.926880	0.105003	-0.029434
review_scores_rating	0.929261	0.000033	0.043437	0.028118	-0.091487
review_scores_cleanliness	0.814730	0.001978	0.026068	0.046961	-0.103470
review_scores_communication	0.823799	-0.018930	0.062572	0.004580	-0.086667
review_scores_value	0.890968	0.036681	0.081405	-0.003764	-0.064117
reviews_per_month	0.111666	-0.055985	0.849474	0.154765	-0.087447
number_of_reviews_l30d	0.068010	-0.023144	0.823493	0.086726	-0.053857
review_scores_accuracy	0.890874	-0.025961	0.042349	0.021441	-0.070471
review_scores_checkin	0.782456	0.000533	0.045561	-0.010803	-0.077545
review_scores_location	0.571499	0.043646	0.117401	0.013044	0.242422

PRINCIPAL COMPONENT ANALYSIS (PCA)

- Factor 1: Noise and Quietness
- Factor 2: Arrival and Booking Experience
- Factor 3: Transportation
- Factor 4: Food and Dining
- Factor 5: Luxury and Comfort

This 5 Factors accounts for 64.54% of the variation

LINEAR REGRESSION MODEL

Most Important Factor

- Accommodations
- Instant Bookable
- Property Type

Overall Performance:

Only around 1% of the variation in the y-variable (occupancy_rate_365) can be explained by this model.

Effect Summary

Source	Logworth	PValue
Accomodations	10.125	0.00000
instant_bookable	5.891	0.00000
property_type	4.030	0.00009
Guest Reviews	1.611	0.02449
neighbourhood_cleansed	1.466	0.03417
Guest Activity	1.268	0.05391
Host Engagement	0.884	0.13053
host_is_superhost	0.736	0.18374
Booking Flexibility	0.581	0.26256

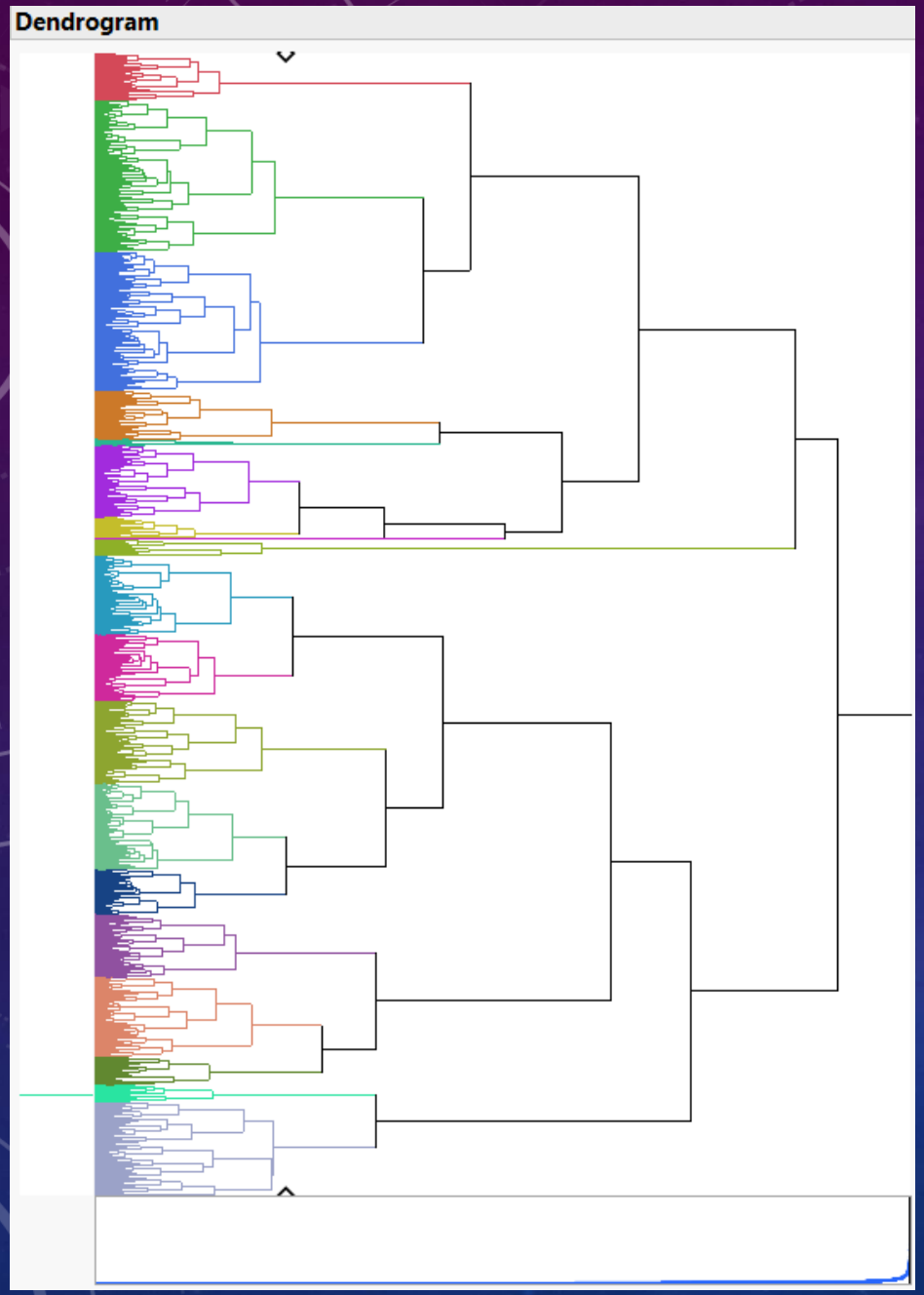
Remove Add Edit Exclude FDR

Summary of Fit

RSquare	0.009878
RSquare Adj	0.007314
Root Mean Square Error	0.282856
Mean of Response	0.413771
Observations (or Sum Wgts)	17810

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.3797643	0.031648	12.00	<.0001*
Guest Reviews	0.0054133	0.002406	2.25	0.0245*
Accomodations	-0.014196	0.002179	-6.51	<.0001*
Guest Activity	0.0044223	0.002294	1.93	0.0539
Host Engagement	0.0042093	0.002784	1.51	0.1305
Booking Flexibility	-0.002901	0.002589	-1.12	0.2626
host_is_superhost[f]	0.0032076	0.002413	1.33	0.1837
neighbourhood_cleansed[I Centro Storico]	0.008483	0.005021	1.69	0.0911
neighbourhood_cleansed[II Parioli/Nomentano]	-0.016141	0.008486	-1.90	0.0572
neighbourhood_cleansed[III Monte Sacro]	0.0308983	0.016036	1.93	0.0540
neighbourhood_cleansed[IV Tiburtina]	0.0166814	0.018243	0.91	0.3605
neighbourhood_cleansed[IX Eur]	-0.01602	0.019102	-0.84	0.4017
neighbourhood_cleansed[V Prenestino/Centocelle]	0.0117162	0.010042	1.17	0.2433
neighbourhood_cleansed[VI Roma delle Torri]	-0.001437	0.024477	-0.06	0.9532
neighbourhood_cleansed[VII San Giovanni/Cinecittà]	0.0147002	0.00834	1.76	0.0780
neighbourhood_cleansed[VIII Appia Antica]	-0.0028	0.01305	-0.21	0.8301
neighbourhood_cleansed[X Ostia/Acilia]	-0.022248	0.012095	-1.84	0.0659
neighbourhood_cleansed[XI Arvalia/Portuense]	-0.000791	0.015016	-0.05	0.9580
neighbourhood_cleansed[XII Monte Verde]	0.0127305	0.009223	1.38	0.1675
neighbourhood_cleansed[XIII Aurelia]	-0.004998	0.008741	-0.57	0.5675
neighbourhood_cleansed[XIV Monte Mario]	-0.00376	0.013873	-0.27	0.7864
property_type[Casa particular]	0.0100664	0.115474	0.09	0.9305
property_type[Castle]	-0.251898	0.27428	-0.92	0.3584
property_type[Cave]	0.2247658	0.273704	0.82	0.4115
property_type[Dome]	-0.149307	0.195309	-0.76	0.4446
property_type[Entire bed and breakfast]	-0.131926	0.115568	-1.14	0.2537
property_type[Entire bungalow]	0.0424849	0.194873	0.22	0.8274
property_type[Entire cabin]	-0.085627	0.160175	-0.53	0.5929
property_type[Entire chalet]	0.3810785	0.194958	1.95	0.0506
property_type[Entire condo]	0.0301766	0.031942	0.94	0.3448
property_type[Entire cottage]	0.0349463	0.115498	0.30	0.7622
property_type[Entire guest suite]	0.1471222	0.05586	2.63	0.0085*
property_type[Entire guesthouse]	0.0276708	0.091715	0.30	0.7629
property_type[Entire home]	0.0094736	0.033436	0.28	0.7769
property_type[Entire home/apt]	0.0207623	0.207373	0.10	0.9202
property_type[Entire loft]	0.0242633	0.034761	0.70	0.4852
property_type[Entire place]	-0.042029	0.081884	-0.51	0.6078
property_type[Entire rental unit]	0.0340405	0.031772	1.07	0.2840
property_type[Entire serviced apartment]	-0.072615	0.037713	-1.93	0.0542
property_type[Entire townhouse]	0.1763618	0.065113	2.71	0.0068*
property_type[Entire vacation home]	0.0284118	0.033211	0.86	0.3923
property_type[Entire villa]	-0.031411	0.044419	-0.71	0.4795
property_type[Farm stay]	-0.157943	0.139779	-1.13	0.2585
property_type[Room in aparthotel]	-0.217241	0.273652	-0.79	0.4273
property_type[Tiny home]	-0.036908	0.047184	-0.78	0.4341
property_type[Tower]	-0.237461	0.273623	-0.87	0.3855
instant_bookable[f]	0.0113102	0.002335	4.84	<.0001*

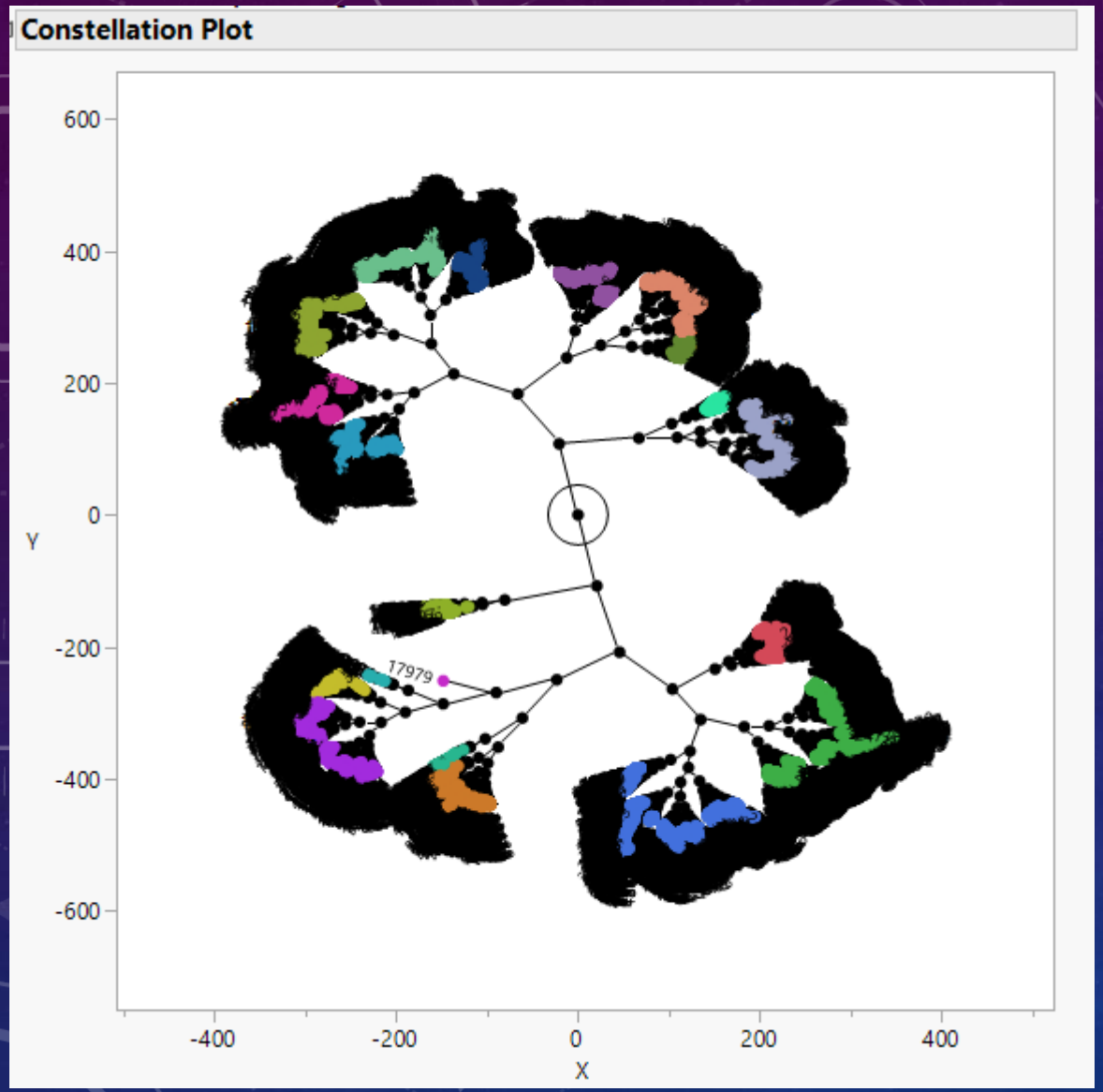


DENDOGRAM

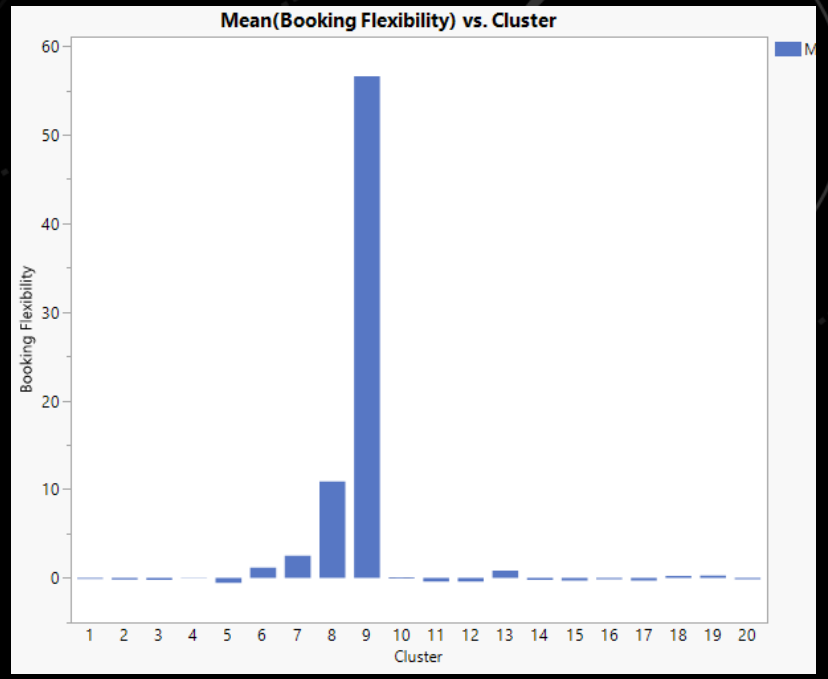
Number of Clusters:

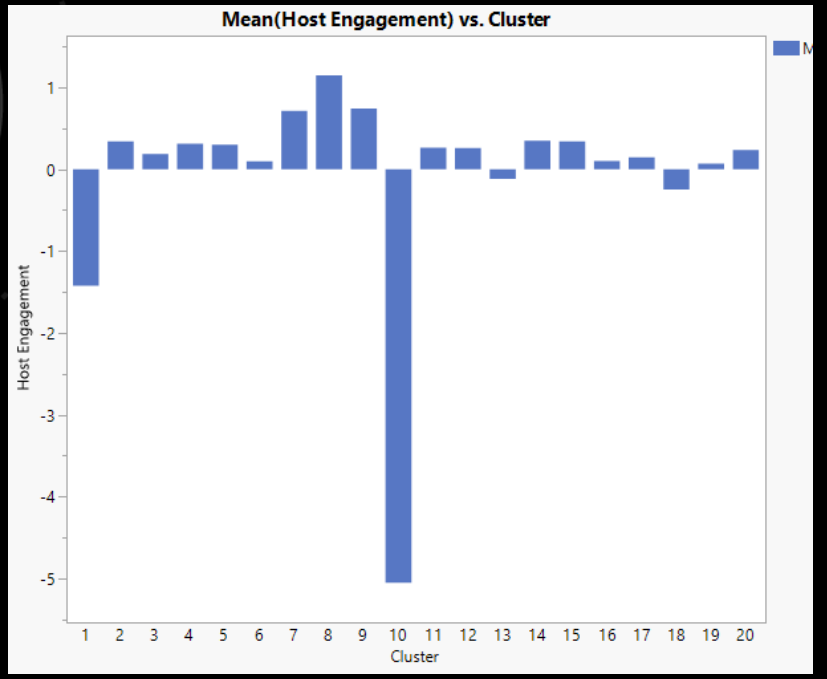
20 Clusters

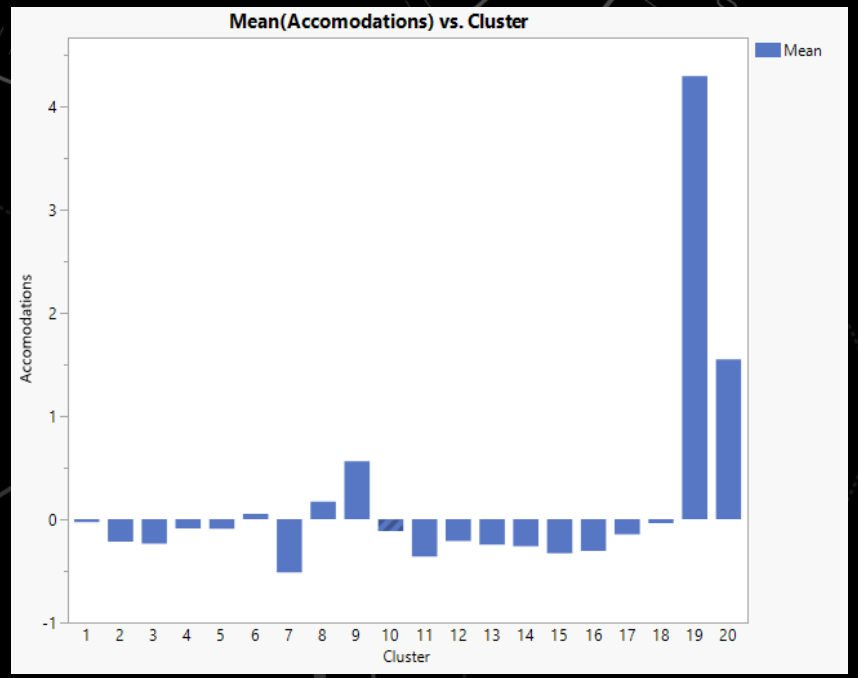
Constellation Plot

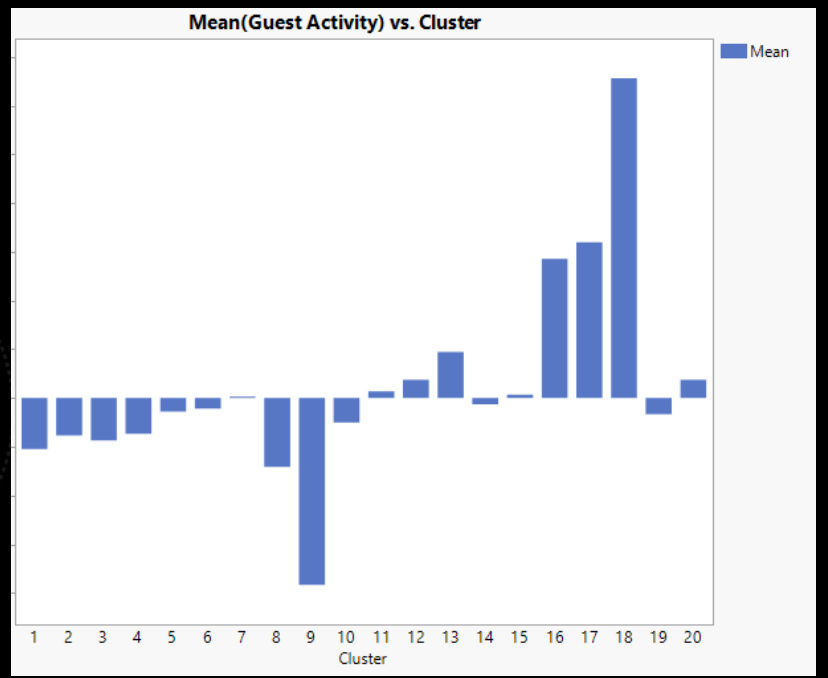


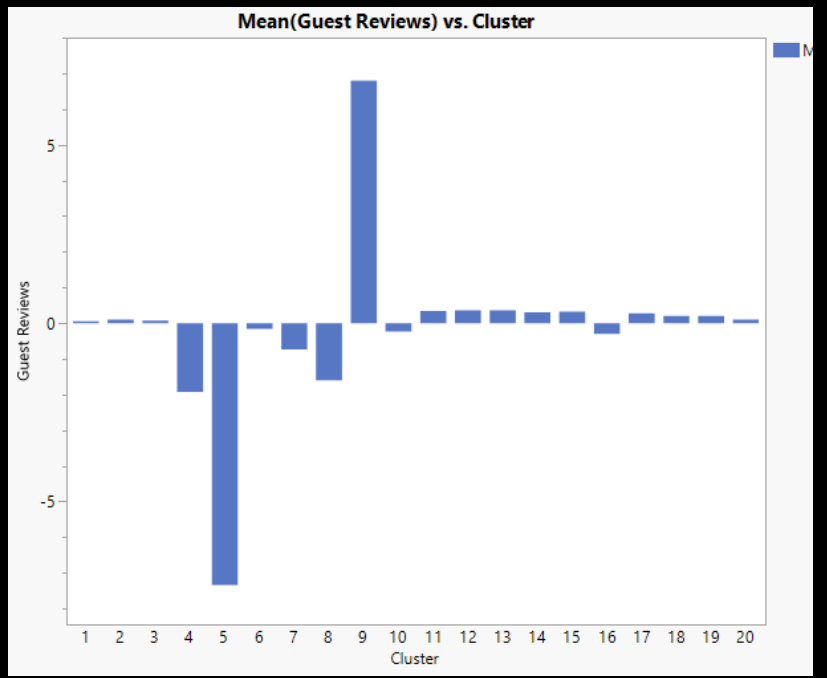
BAR PLOT


Mean(Booking Flexibility) vs. Cluster
Booking Flexibility
Cluster
Mean


Mean(Host Engagement) vs. Cluster
Host Engagement
Cluster

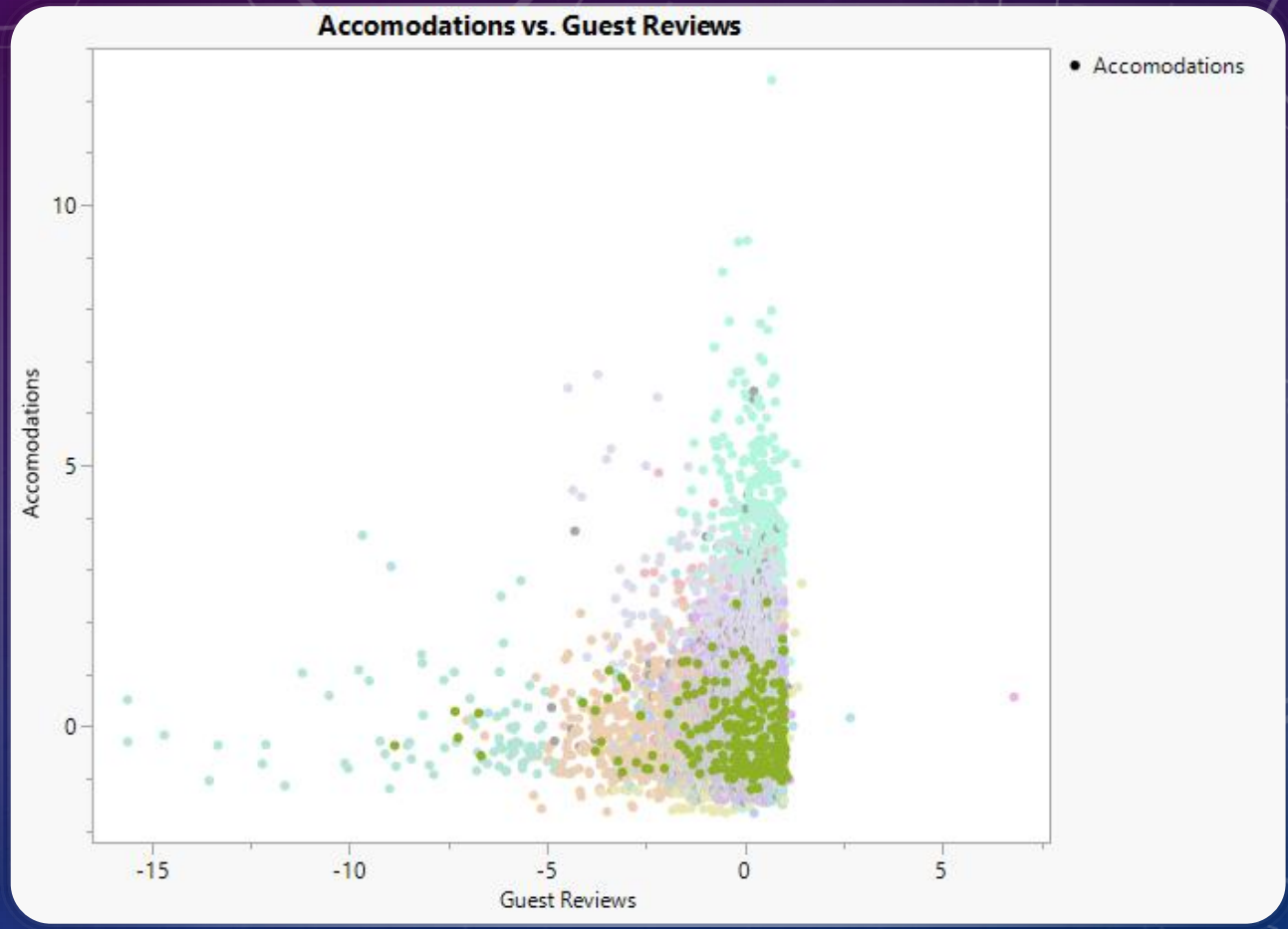

Mean(Accomodations) vs. Cluster
Accomodations
Cluster
Mean


Mean(Guest Activity) vs. Cluster
Cluster
Mean


Mean(Guest Reviews) vs. Cluster
Guest Reviews
Cluster

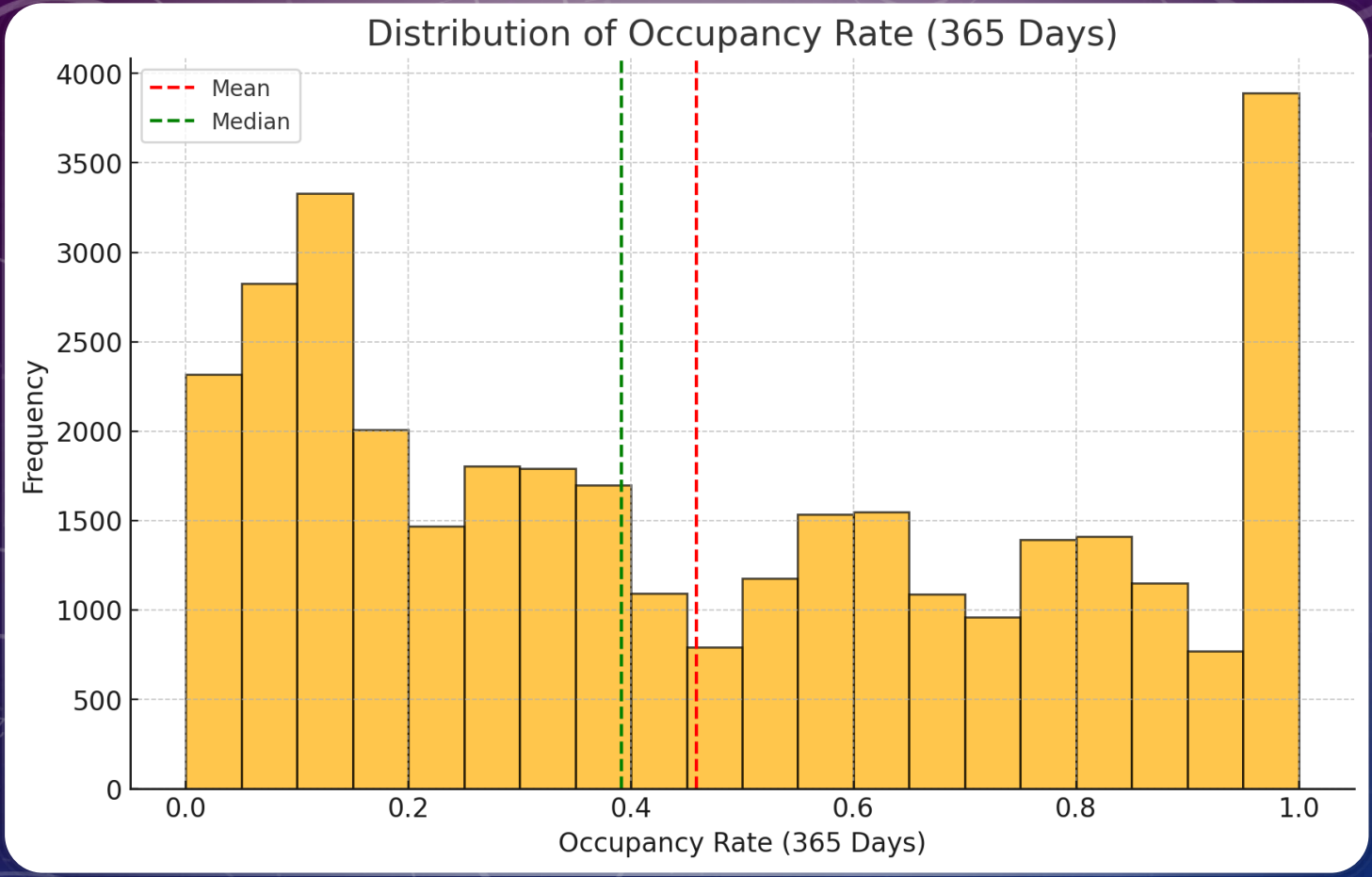
SCATTER PLOT

Listing with higher accommodations tend to have more guest reviews, and potentially indicating higher demand for larger properties.



SUPERVISED LEARNING

BINARY VARIABLE CREATION CRITERIA



- **Mean**: 0.459
- **Median**: 0.392
- **75th Percentile**: 0.753

Decide to go with **0.6 cutoff** that captures well-performing listings without being overly restrictive or lenient

RANDOM FOREST

Column Contributions

Term	Number of Splits	G^2	Portion
price	2887	1111.27473	0.2376
host_total_listings_count	2872	486.234689	0.1040
neighbourhood_cleansed	2243	320.856811	0.0686
maximum_nights	1822	318.507358	0.0681
beds	1151	269.016007	0.0575
review_scores_value	2139	208.930511	0.0447
number_of_reviews_ltm	2230	200.71111	0.0429
reviews_per_month	2057	199.297985	0.0426
number_of_reviews	2060	198.796787	0.0425
host_acceptance_rate	1329	176.229438	0.0377
review_scores_cleanliness	1875	158.883079	0.0340
review_scores_rating	1848	157.244793	0.0336
review_scores_communication	1737	142.388122	0.0304
accommodates	1490	132.85524	0.0284
host_response_rate	927	130.146931	0.0278
minimum_nights	1689	129.127638	0.0276
property_type	981	93.1551132	0.0199
host_is_superhost	1091	65.9060794	0.0141
instant_bookable	1227	61.1762159	0.0131
number_bathrooms	808	60.0615592	0.0128
bedrooms	1037	56.2085448	0.0120

Top Predictors:

- **Price**: The most important variable, with 23.7% contribution to the splits.
- **Host Total Listing Count**: Accounts for 10.4%, reflecting the impact of host experience or portfolio size on occupancy.
- **Neighborhood**: Contributes 6.9%, indicating location plays a significant role in predicting occupancy.
- **Maximum Nights** and **Number of Beds**: Also significant, suggesting that booking constraints and property size are relevant factors

Other Moderate Predictors:

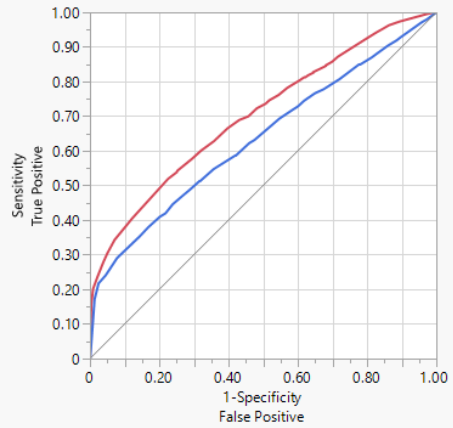
- Reviews per Month
- Review Scores
- Number of Reviews

MODEL COMPARISON: RANDOM FOREST VS NEURAL NETWORK

Why is Random Forest Preferred?

- Better predictive performance (lower misclassification rate and higher AUC).

ROC Curve for occupancy_binary=1



Predictor	Creator	AUC
Prob(occupancy_binary==1)	Bootstrap Forest	0.7025
Probability(occupancy_binary=1)	Neural Model NTanH(3) - Save Profile Formulas	0.6353

Measures of Fit for occupancy_binary

Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RASE	Mean Abs Dev	Misclassification Rate	N	AUC
Bootstrap Forest		0.1362	0.2215	0.5524	0.4317	0.3747	0.2720	6228	0.7025
Neural Model NTanH(3) - Save Profile Formulas		0.0752	0.1271	0.5914	0.4481	0.3954	0.2861	6228	0.6353

TEXT MINING

WORD COUNT / WORD CLOUD

Text Explorer for comments

Number of Terms	Number of Cases	Total Tokens	Tokens per Case	Number of Non-Empty Cases	Portion of Non-Empty Cases
4209	2468	132098	53.5243	2468	1.0000

Term and Phrase Lists

Term	Count
apart	1485
great	1469
stay	1465
locat	1375
place	1280
br	1278
rome	1038
us	889
host	859
clean	796
recommend	783
walk	744
nice	605
help	588
everything	567
perfect	538
love	519
restaur	471
good	460
close	457
need	421
s	421
comfort	411
well	407
also	383

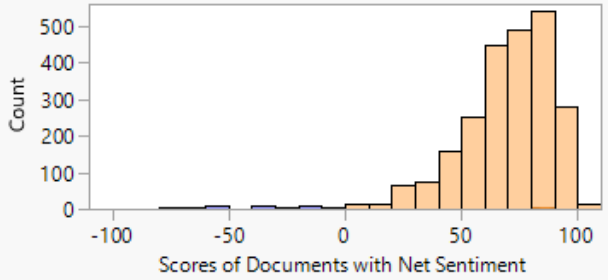
Phrase	Count	N
great location	268	2
walking distance	195	2
highly recommend	191	2
place to stay	127	3
stay in rome	124	3
great host	121	2
great place	104	2
definitely recommend	101	2
gave us	95	2
s place	94	2
come back	93	2
train station	91	2
great stay	88	2
minute walk	88	2
perfect location	83	2
definitely stay	75	2
recommend this place	72	3
everything we needed	69	3
within walking distance	68	3
within walking	68	2
metro station	64	2
trevi fountain	63	2
everything you need	60	3
br the apartment	57	3
enjoyed our stay	57	3

Word Cloud

SENTIMENT ANALYSIS

Sentiment Summary

	N	Mean Score
All Scored Documents	2328	67.2
Net Positive Documents	2302	69.1
Net Negative Documents	22	-36.4
No Sentiment Documents	140	0.0



Document	Positive Sum	Positive Score Mean	Negative Sum	Negative Score Mean	Overall Score
13	156	52	0	0	52
14	90	90	0	0	90
15	254	64	0	0	64
16	427	71	0	0	71
17	240	80	0	0	80
18	0	0	0	0	0
19	60	60	0	0	60
20	176	88	0	0	88
21	115	58	0	0	58
22	60	60	0	0	60

Sentiment	Score	Count
great	80	1393
perfect	90	456
nice	25	335
good	60	306
amazing	90	300
easy	60	259
comfortable	40	253
wonderful	90	234
beautiful	80	212
helpful	35	200
excellent	90	187

[17] Awesome spot right next to the Roman Forum/Colosseum! Andrew's place was our first stop in Italy. Beautifully structured studio apartment and even better location but also had privacy. Very clean, and was perfect for our stay.

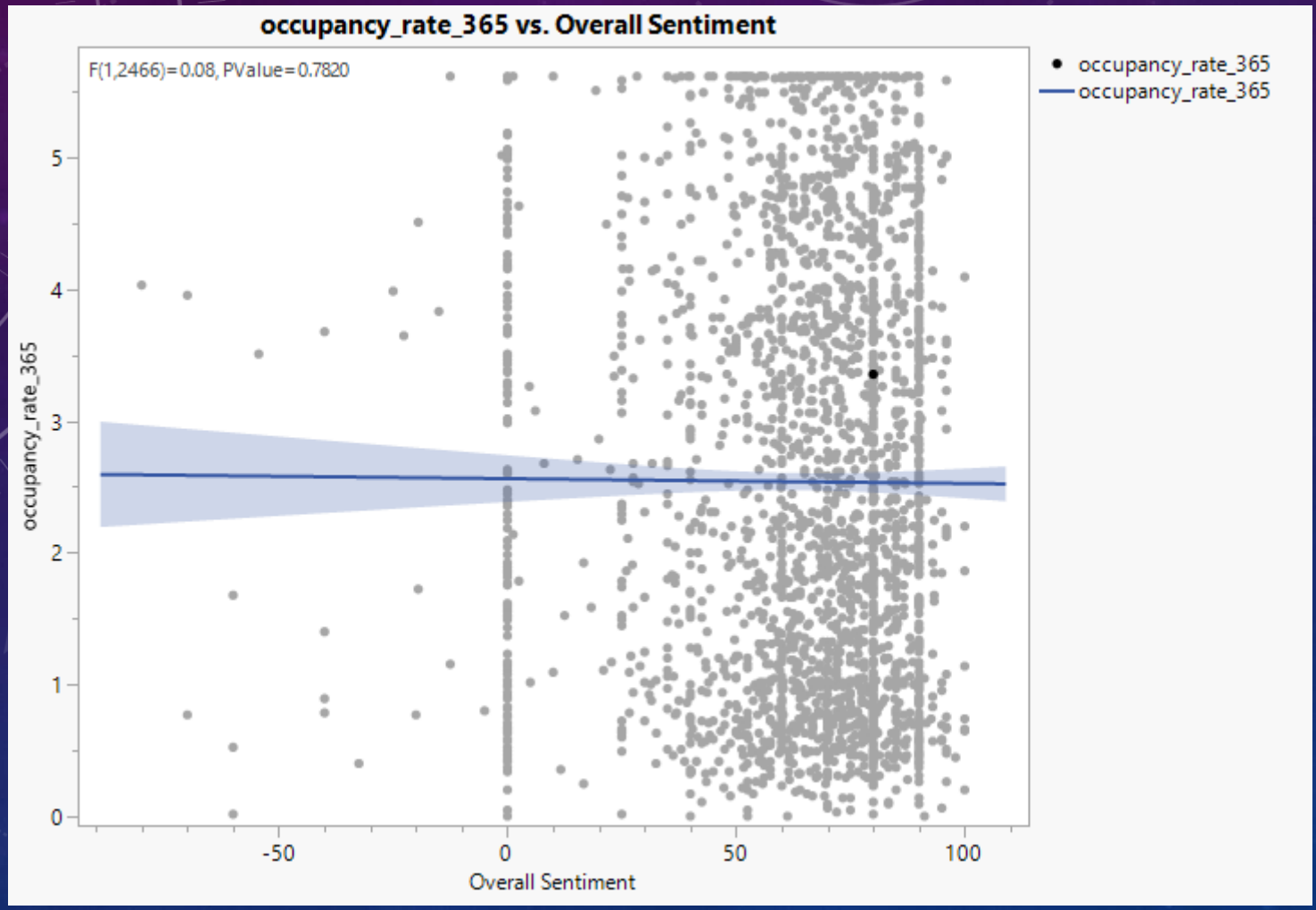
Document Sentiments

awesome
perfect
better

+Sentiment xIntensifier

+90	x1.6
+60	x1.2
+30	x0.8
-30	x0.3
-60	
-90	

OVERALL SENTIMENT AND OCCUPANCY RATE



TOPIC MODELLING

SVD Centered and Scaled TF IDF

Principal Components of 893 terms by 2468 documents using TF IDF weighting Centered and Scaled making 100 vectors

SVD Plots

Singular Values

Topic Analysis for 8 topics

Top Loadings by Topic

Topic 1		Topic 2		Topic 3		Topic 4		Topic 5	
Term	Loading	Term	Loading	Term	Loading	Term	Loading	Term	Loading
talk	0.42052	us	0.49984	dri	0.67445	pasta	0.65194	qualiti	0.61182
hear	0.41703	rome	0.38392	tram	0.53355	couch	0.47983	holiday	0.60197
not	0.40758	arriv	0.35036	rout	0.52478	son	0.47833	iron	0.50689
mention	0.39038	recommend	0.32748	kid	0.49367	driver	0.43232	luxuri	0.44854
night	0.38556	apart	0.31658	slept	0.46873	squar	0.43091	generous	0.42868
no	0.37798	stay	0.31083	problems	0.41244	euro	0.41307	interior	0.35208
loud	0.37500	airport	0.28793	bus	0.39902	bread	0.38506	claudio	0.34943
bedroom	0.35759	first	0.27649	line	0.38634	fresh	0.37120	san	0.34922
concern	0.35582	book	0.27575	start	0.35111	prepar	0.36532	dine	0.34877
nois	0.35532	even	0.27392	note	0.35040	milk	0.36482	prefer	0.30802
toilet	0.33199	emanuela	0.26989	point	0.34277	meal	0.35663	enough	0.30431
br	0.33081	make	0.25570	wash	0.33994	breakfast	0.35223	alreadi	0.29364
street	0.31273	also	0.25562	week	0.33393	market	0.35211	departur	0.29119
quite	0.31222	home	0.25045	pastri	0.30844	food	0.34937	effici	0.28612
alessandro	0.30206	went	0.24487			open	0.34720		
c	0.30013	easy	0.24265			wine	0.34707		

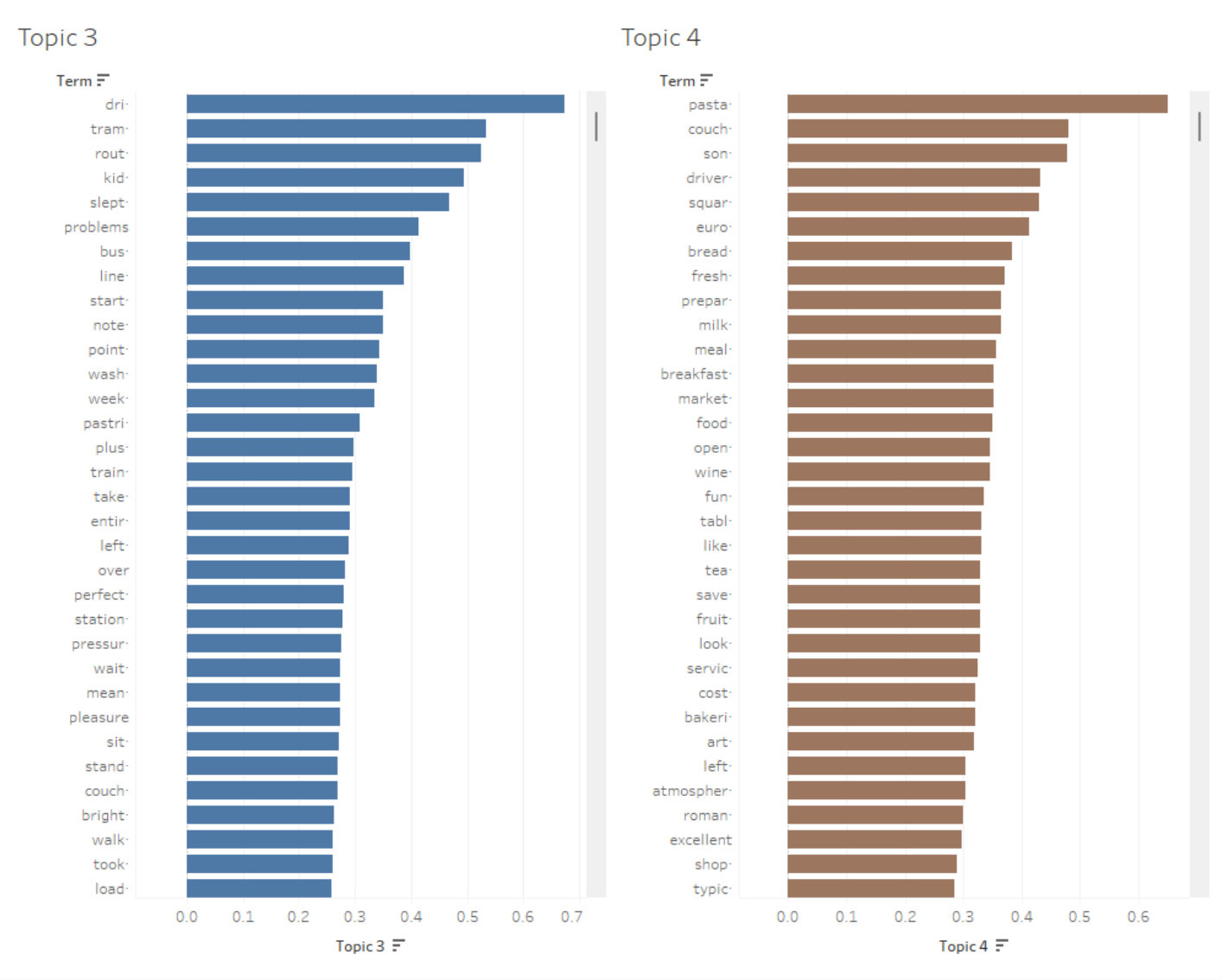
Topic 6		Topic 7		Topic 8	
Term	Loading	Term	Loading	Term	Loading
mattress	0.60042	campo	0.59798	broken	0.34456
sheet	0.52886	fiori	0.52871	no	0.33830
instead	0.45314	di	0.41621	toilet	0.32806
pay	0.44889	lucky	0.32854	paper	0.30625
euro	0.43679	if	0.31515	furnitur	0.29930
seem	0.41618	favorit	0.31310	laundri	0.28168
not	0.37032	piazza	0.30171	br	0.27748
now	0.36805	don	0.28550	monti	0.27713
luxuri	0.36467	gelato	0.28043	washer	0.27345
upon	0.36121	sleeper	0.27950	room	0.26691
later	0.36065	ll	0.26927	line	0.25927
staff	0.34848	de	0.26513	common	0.25579
checkout	0.34421	market	0.26099	wine	0.25535
pictur	0.32474	front	0.25367	design	0.24689
cover	0.31177	navona	0.25101	full	0.24531
				bathroom	0.24294

Theme:

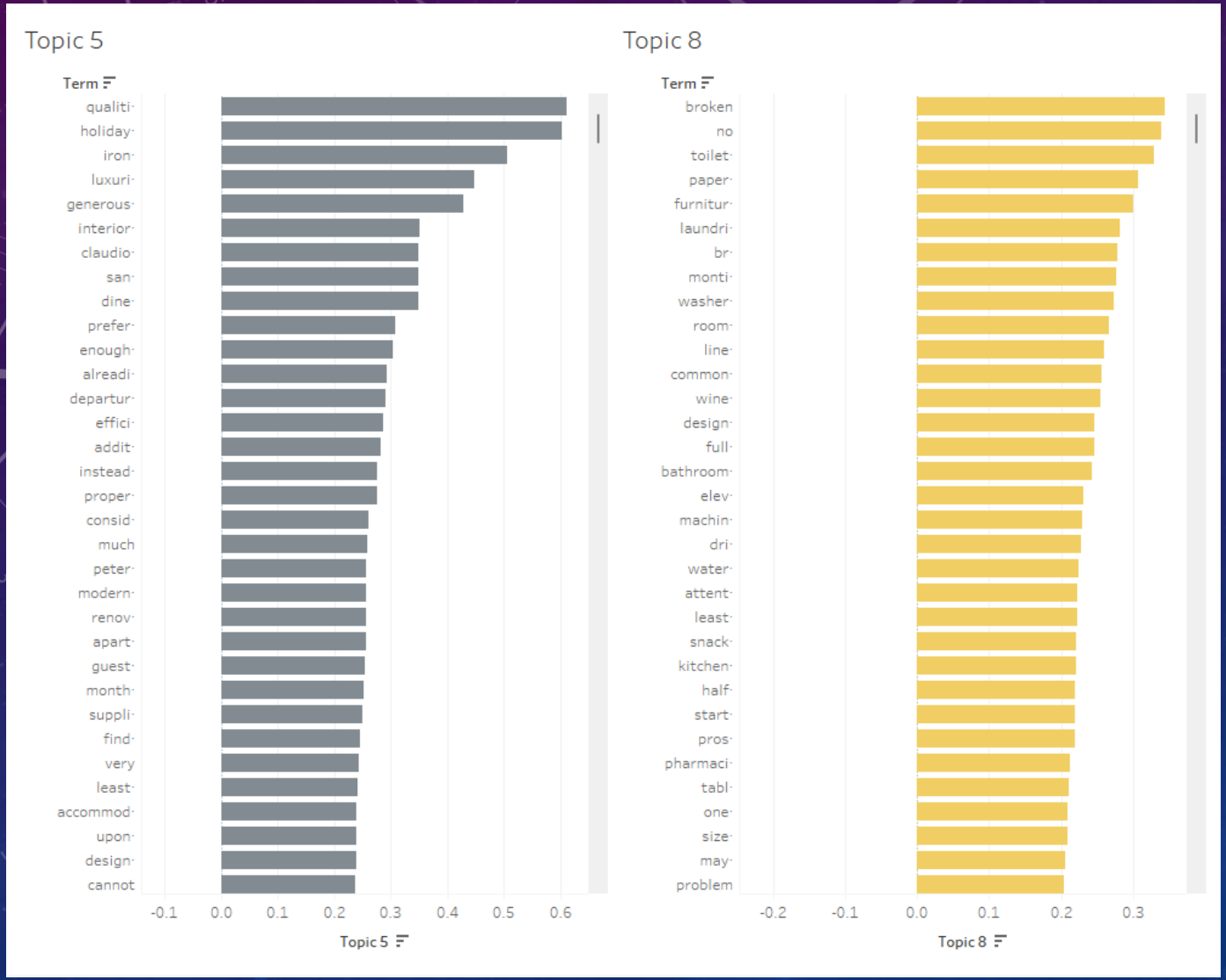
- **Topic 1: Noise and Quietness**
  - Guests discussing noise levels, quietness, or disturbances
- **Topic 2: Arrival and Booking Experience.**
  - Comments related to the booking process, arriving at the location, and initial impressions of the apartment.
- **Topic 3: Transportation**
  - Accessibility and use of public transport, navigating the city.
- **Topic 4: Food and Dining**
  - Food-related experiences, possibly discussing dining in the apartment, local markets, or nearby restaurants.
- **Topic 5: Luxury and Comfort**
  - The overall quality, luxury, and comfort of the stay, with a focus on aesthetics and dining experiences.
- **Topic 8: Maintenance Issue**
  - Comments about maintenance problems, such as broken furniture, toilet issues, or laundry facilities.

NOISE & QUIETNESS AND ARRIVAL & BOOKING EXPERIENCE

TRANSPORTATION AND FOOD & DINING



LUXURY & COMFORT AND MAINTENANCE ISSUE



RECOMMENDATIONS

01

Optimize Pricing:

- Ensure competitive pricing, as price is the strongest predictor of occupancy.

02

Focus on Accommodations:

- Highlight the size and amenities of your property in listings to attract more guests.

03

Improve Booking Flexibility:

- Adopt flexible booking policies to cater to a wider audience.

04

Enhance Guest Experience:

- Focus on cleanliness and quietness, as these are recurring themes in guest feedback.

05

Engage with Guests:

- Active communication and responsiveness can boost reviews and occupancy.

06

Leverage Location:

- Market your property's proximity to attractions or public transport to increase its appeal.


THANK YOU!