



Webcare Case

Akanksha Tandon
Madeleine Beck
Ceren Urgan
Andrew Yohanan

OUTLINE

1

Regression with original variables

4

Lasso and Ridge regressions

2

Regression with combined variables

5

Comparison of coefficients

3

Stepwise regression

6

Webcare strategy implications

Research Question

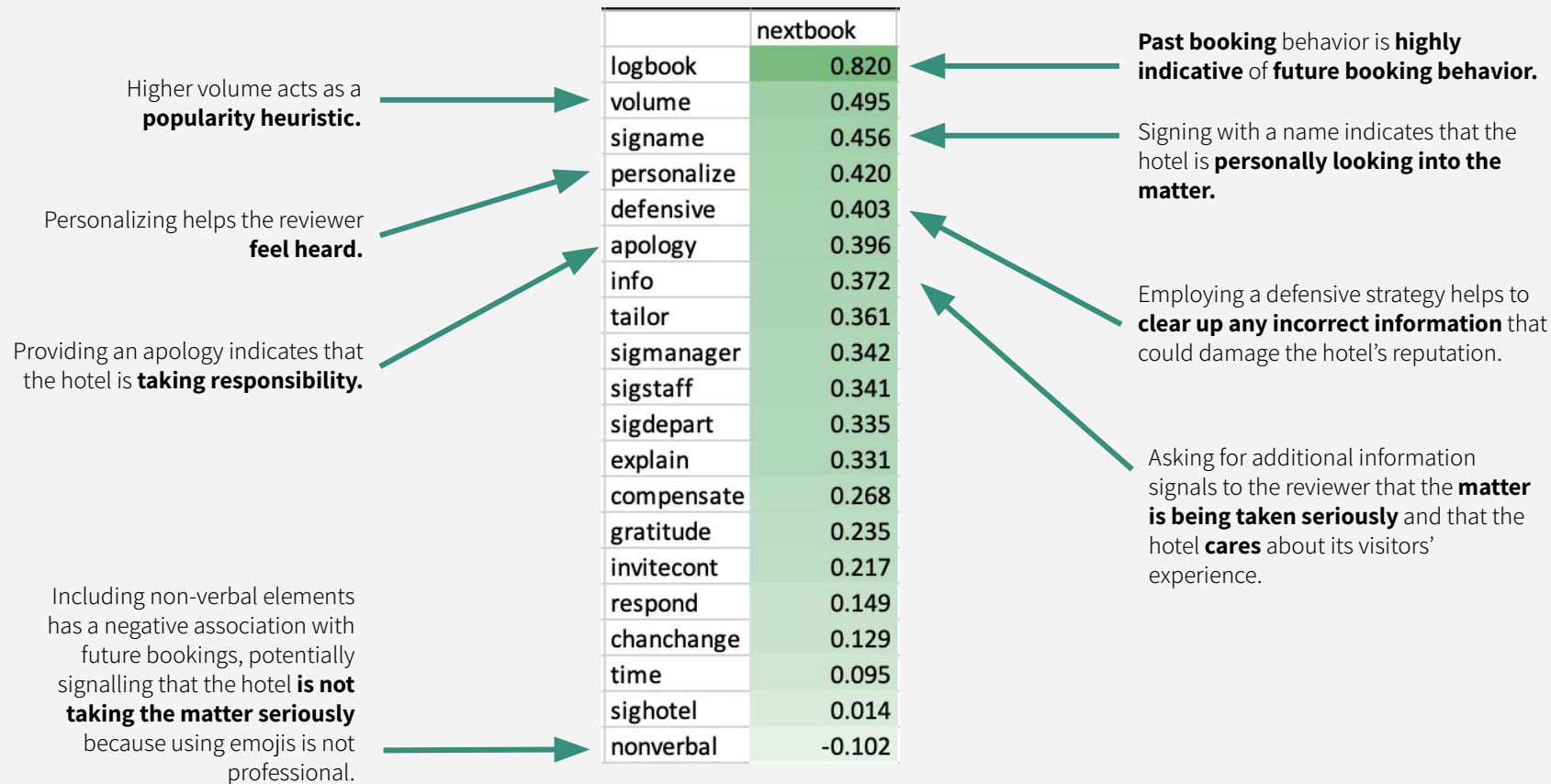
In this analysis, our aim is to understand **how a hotel should respond to customer reviews to maximize future bookings from other consumers reading the dialogue.**

Analyzing booking data for the past 4 years, our main areas of analyses were:

- Which variables are the most significant in determining number of future bookings?
- Which variables are not significant in determining the number of future bookings?
- Based on this understanding, what webcare strategies can a hotel employ to increase the number of future bookings?



Initial Understanding of the Nature of the Relationship Between Variables and Nextbook



Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.086269	0.046515	44.851	< 2e-16	***
logbook	0.597768	0.007982	74.887	< 2e-16	***
time	-0.058091	0.007649	-7.595	3.08e-14	***
volume	0.034387	0.005569	6.174	6.65e-10	***
respond	0.068268	0.011973	5.702	1.18e-08	***
factor(hotel_id)2	-0.081814	0.011415	-7.167	7.66e-13	***
factor(hotel_id)3	-1.143516	0.034297	-33.342	< 2e-16	***
factor(hotel_id)4	-0.794018	0.025928	-30.624	< 2e-16	***
factor(hotel_id)5	-0.696290	0.019222	-36.224	< 2e-16	***
factor(hotel_id)6	-0.021715	0.012767	-1.701	0.08898	.
factor(hotel_id)7	-0.623308	0.024762	-25.171	< 2e-16	***
tailor	0.021127	0.013104	1.612	0.10691	
defensive	0.037858	0.009485	3.991	6.57e-05	***
invitecont	-0.014976	0.008050	-1.860	0.06284	.
explain	-0.020609	0.010667	-1.932	0.05336	.
nonverbal	-0.046286	0.014386	-3.217	0.00129	**
apology	-0.042778	0.007374	-5.801	6.58e-09	***
compensate	0.014352	0.007936	1.808	0.07055	.
chanchange	0.038511	0.009260	4.159	3.20e-05	***
gratitude	0.002108	0.006662	0.316	0.75172	
info	-0.045118	0.010077	-4.477	7.56e-06	***
personalize	0.001673	0.009656	0.173	0.86244	
sigdepart	-0.032734	0.012870	-2.543	0.01098	*
sigstaff	0.034936	0.013388	2.610	0.00906	**
sigmanager	0.006919	0.013745	0.503	0.61468	
signame	-0.019330	0.014937	-1.294	0.19562	
sighotel	-0.049634	0.010129	-4.900	9.57e-07	***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 120407 on 1143 degrees of freedom
 Residual deviance: 9925 on 1117 degrees of freedom
 AIC: 16543



Key Takeaways

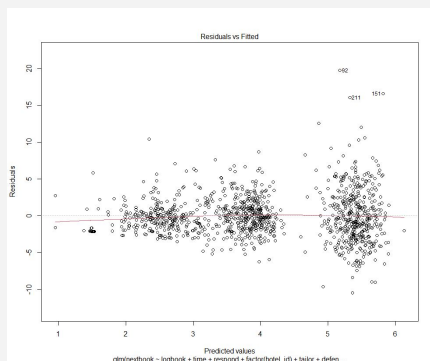
- The overall model is significant, however there are a few variables that are not significant
- The AIC is 16,543
- The MSE is 21,828.09

Likelihood ratio test

Model 1: nextbook ~ logbook + time + volume + respond + factor(hotel_id) +
 tailor + defensive + invitecont + explain + nonverbal + apology +
 compensate + chanchange + gratitude + info + personalize +
 sigdepart + sigstaff + sigmanager + signame + sighotel

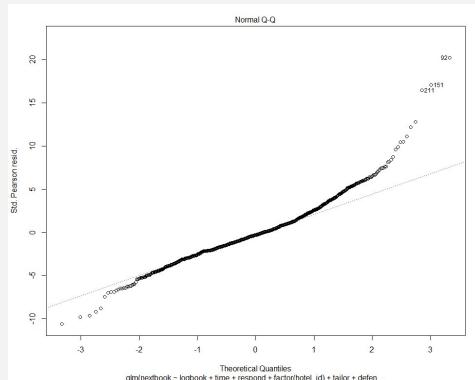
Model 2: nextbook ~ 1
 #Df LogLik Df Chisq Pr(>Chisq)
 1 27 -8245
 2 1 -63486 -26 110482 < 2.2e-16 ***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



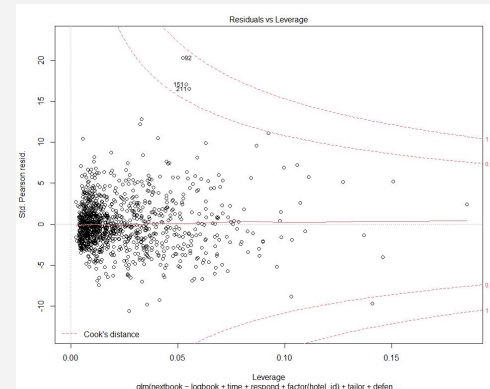
Residual Plot

Clear clusters appearing



QQ Plot

Indicating non-normal errors



Residuals vs. Leverage Plot

Presence of Influential Observations

Analysis of VIFs Uncovers Multicollinearity Issues in the Original Model

	GVIF	Df	GVIF ^{1/(2*Df)}
logbook	5.104663	1	2.259350
time	5.318040	1	2.306088
volume	6.305283	1	2.511032
respond	3.286125	1	1.812767
factor(hotel_id)	102.653862	6	1.471007
tailor	23.601377	1	4.858125
defensive	6.934414	1	2.633328
invitecont	9.437967	1	3.072128
explain	11.067876	1	3.326842
nonverbal	1.184887	1	1.088525
apology	3.501519	1	1.871235
compensate	2.144777	1	1.464506
chancechange	1.353872	1	1.163560
gratitude	4.461332	1	2.112187
info	4.463136	1	2.112613
personalize	10.672287	1	3.266847
sigdepart	17.484526	1	4.181450
sigstaff	18.846875	1	4.341299
sigmanager	28.079019	1	5.298964
signame	37.938932	1	6.159459
sighotel	3.976611	1	1.994144

Key Takeaways

- Delving deeper into the regression diagnostics, we can see high **multicollinearity** among the variables
- This can explain why although the overall model is significant, quite a few of the individual variables are not significant
- In order to refine the model, we can:
 - Remove influential observations
 - Check the correlations between variables using correlation matrices and factor analysis
 - Combine or exclude variables accordingly

1 Diving Deeper into Multicollinearity Issues

Signature Variables

	signame	sigdepart	sighotel	sigmanager	sigstaff
signame	1				
sigdepart	0.474	1			
sighotel	-0.246	-0.101	1		
sigmanager	0.728	-0.142	-0.196	1	
sigstaff	0.477	0.944	-0.086	-0.148	1

Factor Analysis

Uniquenesses:					
	signame	sigdepart	sighotel	sigmanager	sigstaff
	0.005	0.07	0.938	0.129	0.042

Loadings:		
	Factor1	Factor2
signame	0.394	0.916
sigdepart	0.959	0.105
sighotel	-0.238	
sigmanager	-0.246	0.900
sigstaff	0.973	0.102

	Factor1	Factor2
SS loadings	2.087	1.729
Proportion Var	0.417	0.346
Cumulative Var	0.417	0.763

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 2.38 on 1 degree of freedom.
The p-value is 0.123

Multicollinearity

Sigdepart and Sigstaff:

High correlations between posts signed by staff and department, likely because posts are usually signed by both.

Sigmanagers and Signame:

High correlation between posts signed by manager and name, likely due to often signing both together.

Next: Combine Variables

Message Variables

	tailor	defensive	invitecont	explain	nonverbal	apology	compensate	chanchange	gratitude	info	personalize
tailor	1										
defensive	0.830	1									
invitecont	0.879	0.691	1								
explain	0.889	0.874	0.808	1							
nonverbal	0.158	0.076	0.216	0.137	1						
apology	0.740	0.743	0.613	0.731	-0.003	1					
compensate	0.631	0.631	0.600	0.662	0.097	0.622	1				
chanchange	0.431	0.403	0.416	0.425	0.026	0.349	0.305	1			
gratitude	0.784	0.643	0.809	0.725	0.217	0.620	0.548	0.332	1		
info	0.335	0.379	0.005	0.259	0.063	0.308	0.157	0.120	0.106	1	
personalize	0.881	0.750	0.791	0.780	0.075	0.706	0.553	0.373	0.710	0.340	1

Factor Analysis

Loadings:		
	Factor1	Factor2
tailor	0.969	
defensive	0.863	0.269
invitecont	0.921	-0.384
explain	0.920	
nonverbal	0.161	-0.176
apology	0.767	0.241
compensate	0.675	
chanchange	0.450	
gratitude	0.811	-0.163
info	0.275	0.643
personalize	0.885	

	Factor1	Factor2
SS loadings	6.17	0.77
Proportion Var	0.56	0.07
Cumulative Var	0.56	0.63

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 683.77 on 34 degrees of freedom.
The p-value is 5.8E-122

Multicollinearity

Tailoring & Explanations & Personalization:

High correlation between these three variables are likely due to providing "explanations" or "personalization" when a hotel tailors a message.

Gratitude and Invite Customer

Expect that comments that express gratitude are also likely to invite the customer for another visit.

Next: Combine Variables

2 Regression Model - Combined Variables

Likelihood ratio test

```
Model 1: nextbook ~ logbook + time + volume + respond + factor(hotel_id) +
  tailor + defensive + invitecont + explain + nonverbal + apology +
  compensate + chanchange + gratitude + info + personalize +
  sigdept + sigstaff + sigmanager + signame + sighotel
```

```
Model 2: nextbook ~ 1
#Df LogLik Df Chisq Pr(>Chisq)
```

```
1 27 -8245
2 1 -63486 -26 110482 < 2.2e-16 ***
```

```
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.158781	0.057366	37.632	< 2e-16 ***
logbook	0.592157	0.009341	63.396	< 2e-16 ***
time	-0.027776	0.009306	-2.985	0.00284 **
volume	0.012308	0.007004	1.757	0.07888 .
respond	0.139434	0.014278	9.766	< 2e-16 ***
factor(hotel_id)2	-0.038618	0.014491	-2.665	0.00770 **
factor(hotel_id)3	-1.183334	0.039204	-30.184	< 2e-16 ***
factor(hotel_id)4	-0.830616	0.030551	-27.188	< 2e-16 ***
factor(hotel_id)5	-0.652557	0.021306	-30.628	< 2e-16 ***
factor(hotel_id)6	0.051030	0.016677	3.060	0.00221 **
factor(hotel_id)7	-0.592134	0.028987	-20.428	< 2e-16 ***
tep	-0.009381	0.021758	-0.431	0.66635
defensive	0.016519	0.011864	1.392	0.16380
invitecont	-0.028436	0.016350	-1.739	0.08199 .
nonverbal	-0.069166	0.019220	-3.599	0.00032 ***
apology	-0.011325	0.009972	-1.136	0.25608
compensate	-0.006409	0.010349	-0.619	0.53570
chanchange	0.053566	0.011937	4.488	7.20e-06 ***
ig	-0.012779	0.018346	-0.697	0.48607
info	-0.005549	0.015996	-0.347	0.72867
sigdept_stf	0.006549	0.008970	0.730	0.46537
sigman_name	-0.020667	0.014612	-1.414	0.15723
sighotel	-0.071008	0.013909	-5.105	3.31e-07 ***

```
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for poisson family taken to be 1)
```

```
Null deviance: 94285.2 on 1007 degrees of freedom
Residual deviance: 7254.5 on 985 degrees of freedom
AIC: 12878
```

Combined Variables

$Te_p = (Tailor + Explanation + Personalize)/3$

$IG = (Invite + Gratitude)/2$

$Sigdept_staff = (Sigdept + Sigstaff)/2$

$Sigman_name = (Sigman + Signame)/2$

Key Takeaways

- The AIC is **12,878** which indicates an improvement from the original Poisson regression (compare Model 1 AIC 16,580)
- **Model 2 MSE: 15,875.45**
- Model 1 MSE: 21,828.12
- VIFs are better but a few still high (tep, sigman_name)

	GVIF	Df	GVIF^(1/(2*Df))
logbook	5.855413	1	2.419796
time	5.814358	1	2.411298
volume	7.776637	1	2.788662
respond	3.994997	1	1.998749
factor(hotel_id)	189.514537	6	1.548116
tep	39.293120	1	6.268422
defensive	7.984113	1	2.825617
invitecont	28.354297	1	5.324875
nonverbal	1.191878	1	1.091732
apology	4.838079	1	2.199563
compensate	2.528104	1	1.590001
chanchange	1.365083	1	1.168368
ig	28.278231	1	5.317728
info	6.300572	1	2.510094
sigdept_stf	6.255405	1	2.501081
sigman_name	19.528509	1	4.419107
sighotel	2.599865	1	1.612410

Likelihood ratio test

Model 1: nextbook ~ logbook + time + volume + respond + factor(hotel_id) +
 tailor + defensive + invitecont + explain + nonverbal + apology +
 compensate + chanchange + gratitude + info + personalize +
 sigdepart + sigstaff + sigmanager + signame + sighotel

Model 2: nextbook ~ 1

#Df LogLik Df Chisq Pr(>Chisq)

1 27 -8245
 2 1 -63486 -26 110482 < 2.2e-16 ***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.164640	0.056974	37.994	< 2e-16 ***
logbook	0.592483	0.009270	63.912	< 2e-16 ***
time	-0.027321	0.008774	-3.114	0.001845 **
volume	0.011317	0.006814	1.661	0.096757 .
respond	0.138426	0.013890	9.966	< 2e-16 ***
factor(hotel_id)2	-0.038829	0.013838	-2.806	0.005018 **
factor(hotel_id)3	-1.200643	0.039503	-30.394	< 2e-16 ***
factor(hotel_id)4	-0.828320	0.030613	-27.058	< 2e-16 ***
factor(hotel_id)5	-0.647551	0.020824	-31.096	< 2e-16 ***
factor(hotel_id)6	0.053163	0.014424	3.686	0.000228 ***
factor(hotel_id)7	-0.621292	0.026828	-23.158	< 2e-16 ***
defensive	0.033584	0.013085	2.567	0.010272 *
invitecont	-0.031299	0.009404	-3.328	0.000874 ***
explain	-0.046299	0.014413	-3.212	0.001317 **
nonverbal	-0.039251	0.021281	-1.844	0.065125 .
chanchange	0.053818	0.011918	4.516	6.31e-06 ***
personalize	0.028465	0.012691	2.243	0.024899 *
sigdepart	-0.077441	0.025767	-3.005	0.002652 **
sigstaff	0.048693	0.025426	1.915	0.055483 .
sigmanager	-0.046839	0.012221	-3.833	0.000127 ***
sighotel	-0.088484	0.012394	-7.140	9.37e-13 ***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 94285.2 on 1007 degrees of freedom
 Residual deviance: 7237.6 on 987 degrees of freedom
 AIC: 12857

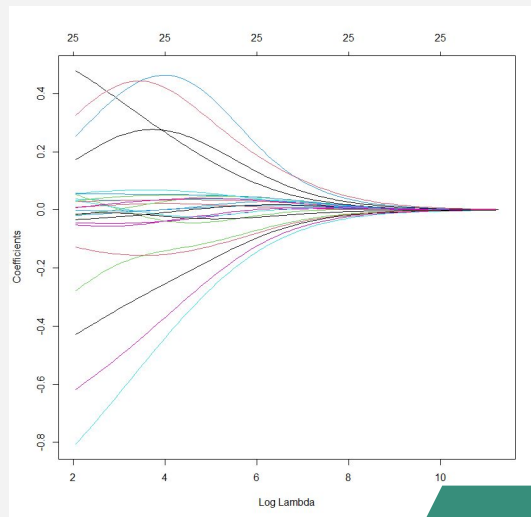
	GVI	Df	GVI ^A (1/(2*Df))
logbook	5.767292	1	2.401519
time	5.168554	1	2.273445
volume	7.356904	1	2.712361
respond	3.780838	1	1.944438
factor(hotel_id)	105.174966	6	1.473984
defensive	9.716673	1	3.117158
invitecont	9.391117	1	3.064493
explain	14.976150	1	3.869903
nonverbal	1.468157	1	1.211675
chanchange	1.361105	1	1.166664
personalize	13.752863	1	3.708485
sigdepart	51.893250	1	7.203697
sigstaff	50.774582	1	7.125629
sigmanager	14.628002	1	3.824657
sighotel	2.063120	1	1.436356

Key Takeaways

- The AIC is 12,857, which marks an improvement
- Based on this model, webcare that includes non-verbal elements does not significantly impact the logs of expected next week's bookings.
- Tailor, gratitude, compensate, info, sigmanager were dropped from the model indicating that they did not improve the model.
- Past bookings were a good index to measure future the next week of bookings, and serves as a baseline.
- Stepfit MSE: 21,829.41
- Combined Stepwise MSE: 21,828.1

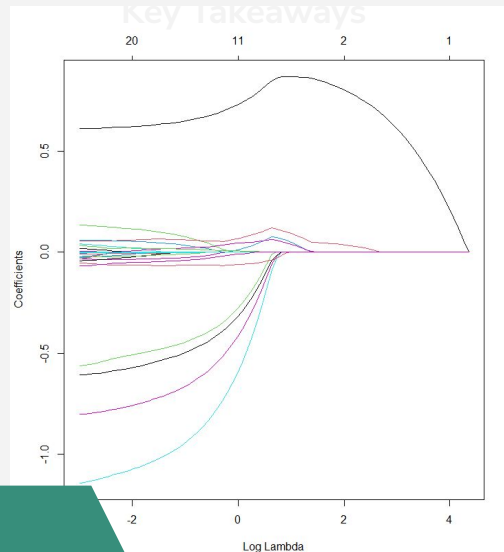
	Ridge Coefficients	Lasso Coefficients
(Intercept)	21.3753411	1.948
logbook	22.1984903	0.633
time	-8.7456895	-0.005
volume	6.0313179	0.0093
respond	0.364269	0.101
factor(hotel_id)2	29.0170624	0.000
factor(hotel_id)3	-63.5990503	-1.024
factor(hotel_id)4	-67.5030202	-0.721
factor(hotel_id)5	-72.1665208	-0.540
factor(hotel_id)6	58.6821509	0.065
factor(hotel_id)7	-50.8589369	-0.482
tailor	-1.2615917	0.000
defensive	1.5064316	0.004
invitecont	-8.6422521	-0.032
explain	-2.4335561	-0.009
nonverbal	-9.1217277	-0.066
apology	1.1284316	-0.014
compensate	-2.8414266	0.000
chancechange	-0.2345125	0.047
gratitude	-1.6796395	-0.003
info	29.2728563	0.016
personalize	5.2715994	0.000
sigdepart	6.0492648	0.000
sigstaff	8.1552612	0.018
sigmanager	8.4980557	0.000
signame	4.4815459	0.000
sighotel	0.228796	-0.047
tep	0.9305629	0.000
ig	-6.045843	-0.008

Ridge:



Ridge MSE: 2072.4

Lasso:



Lasso MSE: 9.26

- Looking at **both** models, we see that the highlighted variables remain consistent throughout.
- To have **robust** recommendations, the focus will be on these variables.

Comparison of Coefficients

+: Positive coefficient -: Negative coefficient 0 : Coefficient of 0 or insignificant

	Combined Poisson	Stepwise	Ridge	Lasso
(Intercept)	+	+	+	+
logbook	+	+	+	+
volume				
personalize				
defensive	+	+	+	+
apology	-	-	-	-
info	0		+	+
tailor				
explain		-	-	-
compensate	0		-	0
gratitude			-	-
invitecont	-	-	-	-
respond	+	+	+	+

	Combined Poisson	Stepwise	Ridge	Lasso
chancechange	+	+	+	+
time	0	-	+	-
nonverbal	-		-	-
signame		-	+	0
sigmanager			+	0
sigstaff		+	+	+
sigdepart		-	+	0
sighotel	-	-	-	-
hotel_id	_*	_*	_*	_*
tep	0		+	0
ig	0		+	-
sigman_name	0			
sigdept_staff	0			

* All hotel_id are negative except for hotel_id6

Webcare Strategy Implications



Do: Encourage reviewers to email or call hotel to further discuss their experience.



Do: Respond to all communications from customers regarding their experience.



Do: Provide defensive clarifications to present the reviewer and readers with the facts.



Don't: Sign with only hotel name. Sign your responses with hotel staff.



Don't: Apologize or explain in the public view. Perhaps do so in a different channel (i.e., privately).



Don't: Include non-verbal elements (e.g., emojis). Stay professional with guests.



Don't: Invite reviewers back to your hotel in public view. Perhaps send a more personal message.

Sample Review and Response

Good Practice

Website Review:

Dear Marriott Antwerp,

I just had the worst hotel experience in my life. The bed was small and the room smelled of fish.

- **T.J. Hastie, 1 STAR**

Yet the message is solution oriented by suggesting channel change.

Response:

Mr. Hastie,

All of our beds are at least a queen and our customers find them large and extremely comfortable. We would like to reach out to you via email, please feel free to message us and we can resolve this issue.

- **Robert Tibshirani, Front Desk**

The response is signed by staff, not the hotel name.

Has a defensive tone without explaining or apologizing.