

Hoja de trabajo # 6

Modelos de regresión logística

Presentado por:

Andrés Say (carné No. 19705)
Ayleen Rubio (carné No.19003)
Andree Toledo (carné No. 18439)

Bases de datos



Universidad del Valle de Guatemala
Facultad de Ingeniería
Departamento de Ciencias de la Computación
Guatemala, Abril 2021

Análisis

En esta hoja de trabajo se busca poder realizar una clasificación entre casas caras, intermedias y económicas, por lo que es necesario convertir estas a variables a dummy. Luego se escogen las columnas de datos con las que se trabajará, las cuales son las columnas con un tipo de valor numérico y se colocan como 0 los valores que tengan NA para evitar problemas.

Se realizó los pasos necesarios

MoSold	YrSold	SalePrice	datosCasasCaras	datosCasasEconomicas	datosCasasIntermedias
2	2008	208500	0	1	0
5	2007	181500	0	1	0
9	2008	223500	0	1	0
2	2006	140000	0	1	0
12	2008	250000	0	1	0
10	2009	143000	0	1	0
8	2007	307000	0	0	1
11	2009	200000	0	1	0
4	2008	129900	0	1	0
1	2008	118000	0	1	0
2	2008	129500	0	1	0
7	2006	345000	0	0	1
9	2008	144000	0	1	0
8	2007	279500	0	0	1
5	2008	157000	0	1	0
7	2007	132000	0	1	0
3	2010	149000	0	1	0
10	2006	90000	0	1	0
6	2008	159000	0	1	0

Se prosiguió con la realización de los modelos de variables antes de analizar la correlación entre las variables: casas económicas, casas caras y casas intermedias.

Casas económicas:

```
Call: glm(formula = datosCasasEconomicas ~ ., family = binomial(),
  data = train[, c(1:32, 35)], maxit = 100)
```

Coefficients:

(Intercept)	MSSubClass	OverallQual	OverallCond	YearBuilt
1.5508844	0.0067519	-1.5473615	-0.3116421	-0.0149108
YearRemodAdd	BsmtFinSF1	BsmtFinSF2	BsmtUnfSF	TotalBsmtSF
-0.0494980	0.0022016	0.0023132	0.0022702	NA
X1stFlrSF	X2ndFlrSF	LowQualFinSF	GrLivArea	BsmtFullBath
-0.0043926	-0.0028664	0.0089842	NA	-1.4989081
BsmtHalfBath	FullBath	HalfBath	BedroomAbvGr	KitchenAbvGr
-0.1639248	0.7334401	0.9943913	0.2964834	14.9935583
TotRmsAbvGrd	Fireplaces	GarageCars	GarageArea	WoodDeckSF
-0.1299553	-1.0584648	-1.0674851	-0.0006706	-0.0040269
OpenPorchSF	EnclosedPorch	X3SsnPorch	ScreenPorch	PoolArea
-0.0029387	-0.0017157	-0.0002069	-0.0047785	0.0042197
MiscVal	MoSold	YrSold		
-0.0002177	0.0244163	0.0665410		

Degrees of Freedom: 1020 Total (i.e. Null); 990 Residual
 Null Deviance: 862.7
 Residual Deviance: 274.5 AIC: 336.5

En el modelo generado para la predicción de qué casas serán económicas, se cuenta con un valor de intercepto de 1.55 aproximadamente. Otros datos que consideramos importantes es que tanto en TotalBsmtSF como en GrLivArea se cuenta con un valor de NA, por lo que es importante estudiar sus valores y correlaciones.

Casas caras:

```
Call: glm(formula = datosCasasCaras ~ ., family = binomial(), data = train[,
c(1:32, 34)], maxit = 100)
```

Coefficients:

(Intercept)	MSSubClass	OverallQual	OverallCond	YearBuilt
8.068e+03	-6.002e-02	1.099e+01	5.033e-01	4.036e-02
YearRemodAdd	BsmtFinSF1	BsmtFinSF2	BsmtUnfSF	TotalBsmtSF
-4.716e-01	1.706e-02	3.419e-02	8.785e-03	NA
X1stFlrSF	X2ndFlrSF	LowQualFinSF	GrLivArea	BsmtFullBath
3.310e-03	-2.501e-02	-2.052e-01	NA	-1.649e+01
BsmtHalfBath	FullBath	HalfBath	BedroomAbvGr	KitchenAbvGr
3.176e+01	4.318e+01	2.500e+01	-2.020e+01	-3.439e+01
TotRmsAbvGrd	Fireplaces	GarageCars	GarageArea	WoodDeckSF
1.621e+01	-1.323e+01	1.153e+01	-1.162e-01	-3.437e-02
OpenPorchSF	EnclosedPorch	X3SsnPorch	ScreenPorch	PoolArea
-1.736e-01	8.228e-02	7.859e-02	-2.328e-02	9.217e-02
MiscVal	MoSold	YrSold		
-2.728e-02	-8.413e-01	-3.691e+00		

Degrees of Freedom: 1020 Total (i.e. Null); 990 Residual

Null Deviance: 63.17

Residual Deviance: 3.802e-10 AIC: 62

En el modelo generado para la predicción de qué casas serán caras, se cuenta con un valor de intercepto de 8.07e+03 aproximadamente. Otros datos que consideramos importantes es que tanto en TotalBsmtSF como en GrLivArea se cuenta con un valor de NA, por lo que es importante estudiar sus valores y correlaciones de la misma manera.

Casas intermedias:

```
Call: glm(formula = datosCasasIntermedias ~ ., family = binomial(),
  data = train[, c(1:32, 36)], maxit = 100)
```

Coefficients:

(Intercept)	MSSubClass	OverallQual	OverallCond	YearBuilt
2.001e+01	-4.293e-03	1.337e+00	2.368e-01	1.341e-02
YearRemodAdd	BsmtFinSF1	BsmtFinSF2	BsmtUnfSF	TotalBsmtSF
4.452e-02	-3.132e-03	-3.134e-03	-2.729e-03	NA
X1stFlrSF	X2ndFlrSF	LowQualFinSF	GrLivArea	BsmtFullBath
3.794e-03	1.792e-03	-7.460e-03	NA	1.731e+00
BsmtHalfBath	FullBath	HalfBath	BedroomAbvGr	KitchenAbvGr
-4.280e-01	-6.696e-01	-7.249e-01	6.391e-02	-1.430e+01
TotRmsAbvGrd	Fireplaces	GarageCars	GarageArea	WoodDeckSF
1.230e-02	9.721e-01	7.163e-01	1.911e-03	4.093e-03
OpenPorchSF	EnclosedPorch	X3SsnPorch	ScreenPorch	PoolArea
3.972e-03	1.811e-03	1.066e-03	5.103e-03	-1.284e-02
MiscVal	MoSold	YrSold		
2.796e-04	1.013e-02	-6.963e-02		

Degrees of Freedom: 1020 Total (i.e. Null); 990 Residual

Null Deviance: 845.1

Residual Deviance: 322 AIC: 384

En el modelo generado para la predicción de qué casas serán de un valor intermedio, se encuentra con un valor de intercepto de 2.00e+01 aproximadamente. Otros datos que consideramos importantes es que tanto en TotalBsmtSF como en GrLivArea se cuenta con un valor de NA, por lo que es importante estudiar sus valores y correlaciones de la misma manera.

Correlación entre variables

Se decidió hacer un análisis de multicolinealidad entre las variables con el objetivo de observar si hay algún componente "arrastrando" los valores causando sobreajuste y para comprender si las variables que en los modelos anteriores mostraban un valor NA tienen alguna relación con esto y si pueden ser eliminados.

Matriz de correlación:

	MSSubClass	OverallQual	OverallCond	YearBuilt	YearRemodAdd
MSSubClass	1.000000000	0.03262771	-0.059315817	0.027850137	0.040581045
OverallQual	0.032627708	1.00000000	-0.091932343	0.572322769	0.550683924
OverallCond	-0.059315817	-0.09193234	1.000000000	-0.375983196	0.073741498
YearBuilt	0.027850137	0.57232277	-0.375983196	1.000000000	0.592854976
YearRemodAdd	0.040581045	0.55068392	0.073741498	0.592854976	1.000000000
BsmtFinSF1	-0.069835749	0.23966597	-0.046230856	0.249503197	0.128450547
BsmtFinSF2	-0.065648579	-0.05911869	0.040229170	-0.049106831	-0.067758514
BsmtUnfSF	-0.140759481	0.30815893	-0.136840570	0.149040392	0.181133087
TotalBsmtSF	-0.238518409	0.53780850	-0.171097515	0.391452002	0.291065583
X1stFlrSF	-0.251758352	0.476223829	-0.144202784	0.281985859	0.240379268
X2ndFlrSF	0.307885721	0.29549288	0.028942116	0.010307660	0.140023779
LowQualFinSF	0.046473756	-0.03042928	0.025494320	-0.183784344	-0.062419100
GrLivArea	0.074853180	0.59300743	-0.079685865	0.199009714	0.287388520
BsmtFullBath	0.003491026	0.11109779	-0.054941515	0.187598550	0.119469879
BsmtHalfBath	-0.002332535	-0.04015016	0.117820915	-0.038161806	-0.012337032
FullBath	0.131608222	0.55059971	-0.194149489	0.468270787	0.439046484
HalfBath	0.177354389	0.27345810	-0.060769327	0.242655910	0.183330612
BedroomAbvGr	-0.023438028	0.10167636	0.012980060	-0.070651217	-0.040580928
KitchenAbvGr	0.281721040	-0.18388223	-0.087000855	-0.174800246	-0.149597521
TotRmsAbvGrd	0.040380065	0.42745234	-0.057583166	0.095589128	0.191739816
Fireplaces	-0.045569340	0.39676504	-0.023819978	0.147716399	0.112581318
GarageCars	-0.040109793	0.60067072	-0.185757511	0.537850092	0.420622155
GarageArea	-0.098671543	0.56202176	-0.151521371	0.478953820	0.371599809
WoodDeckSF	-0.012579358	0.23892339	-0.003333699	0.224880142	0.205725920
OpenPorchSF	-0.006100121	0.30881882	-0.032588814	0.188685840	0.226297633
EnclosedPorch	-0.012036622	-0.11393686	0.070356184	-0.387267783	-0.193919147
X3SsnPorch	-0.043824549	0.03037057	0.025503660	0.031354513	0.045285810

	BsmtFinSF1	BsmtFinSF2	BsmtUnfSF	TotalBsmtSF
X1stFlrSF	-0.069835749	-0.065648579	-0.140759481	-0.2385184093
MSSubClass	0.239665966	-0.059118693	0.308158927	0.5378084986
-0.251758352	-0.046230856	0.040229170	-0.136840570	-0.1710975146
OverallQual	0.249503197	-0.049106831	0.149040392	0.3914520021
0.476223829	0.128450547	-0.067758514	0.181133087	0.2910655826
OverallCond	1.000000000	-0.050117400	-0.495251469	0.5223960520
-0.144202784	-0.050117400	1.000000000	-0.209294492	0.1048095376
YearBuilt	-0.495251469	-0.209294492	1.000000000	0.4153596052
0.281985859	0.522396052	0.104809538	0.415359605	1.0000000000
YearRemodAdd	0.445862656	0.097117448	0.317987438	0.8195299750
0.240379268	-0.137078986	-0.099260316	0.004469092	-0.1745119501
BsmtFinSF1	-0.064502597	0.014806998	0.028166688	-0.0332453873
0.445862656	0.208171130	-0.009639892	0.240257268	0.4548682025
BsmtFinSF2				
0.097117448				
BsmtUnfSF				
0.317987438				
TotalBsmtSF				
0.819529975				
X1stFlrSF				
1.000000000				
X2ndFlrSF				
-0.202646181				
LowQualFinSF				
-0.014240673				
GrLivArea				
0.566023969				

BsmtFullBath	0.649211754	0.158678061	-0.422900477	0.3073505537
0.244671104				
BsmtHalfBath	0.067418478	0.070948134	-0.095804288	-0.0003145818
0.001955654				
FullBath	0.058543137	-0.076443862	0.288886055	0.3237224136
0.380637495				
HalfBath	0.004262424	-0.032147837	-0.041117530	-0.0488037386
-0.119915909				
BedroomAbvGr	-0.107354677	-0.015728114	0.166643317	0.0504499555
0.127400749				
KitchenAbvGr	-0.081006851	-0.040751236	0.030085868	-0.0689006426
0.068100588				
TotRmsAbvGrd	0.044315624	-0.035226548	0.250647061	0.2855725637
0.409515979				
Fireplaces	0.260010920	0.046920709	0.051574882	0.3395193239
0.410531085				
GarageCars	0.224053522	-0.038263513	0.214175190	0.4345848343
0.439316808				
GarageArea	0.296970385	-0.018226592	0.183302698	0.4866654638
0.489781654				
WoodDeckSF	0.204306145	0.067898326	-0.005316424	0.2320186091
0.235458623				
OpenPorchSF	0.111760613	0.003092562	0.129005415	0.2472637463
0.211671225				
EnclosedPorch	-0.102303306	0.036543339	-0.002537855	-0.0954777367
-0.065291701				
X3SsnPorch	0.026450506	-0.029993398	0.020764006	0.0373837273

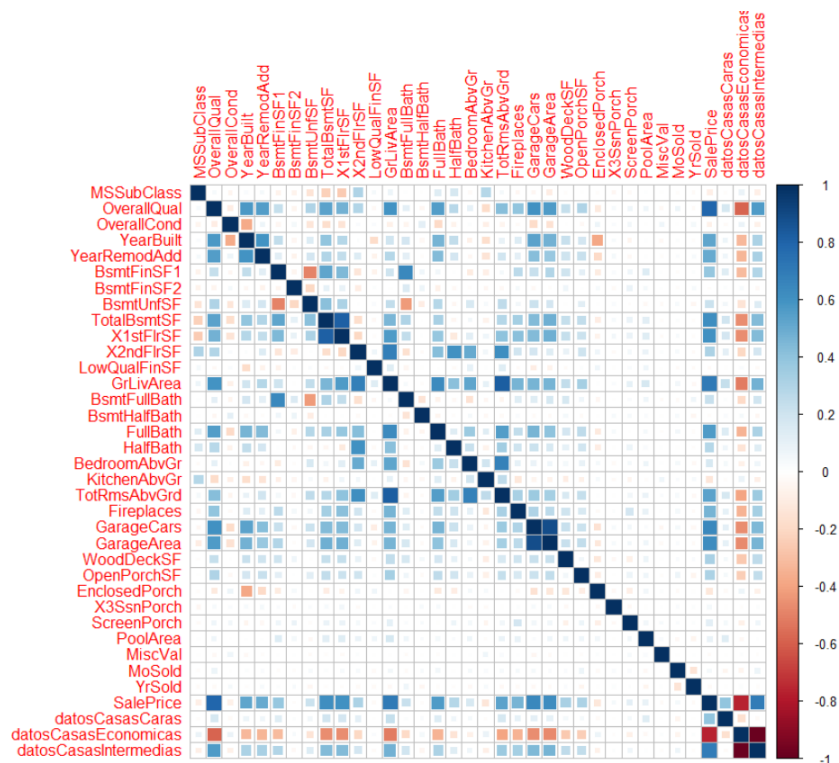
	X2ndFlrSF	LowQualFinSF	GrLivArea	BsmtFullBath
MSSubClass	0.307885721	0.0464737559	0.074853180	0.0034910258
OverallQual	0.295492879	-0.0304292840	0.593007430	0.1110977861
OverallCond	0.028942116	0.0254943199	-0.079685865	-0.0549415154
YearBuilt	0.010307660	-0.1837843444	0.199009714	0.1875985500
YearRemodAdd	0.140023779	-0.0624191001	0.287388520	0.1194698791
BsmtFinSF1	-0.137078986	-0.0645025969	0.208171130	0.6492117536
BsmtFinSF2	-0.099260316	0.0148069979	-0.009639892	0.1586780608
BsmtUnfSF	0.004469092	0.0281666881	0.240257268	-0.4229004774
TotalBsmtSF	-0.174511950	-0.0332453873	0.454868203	0.3073505537
X1stFlrSF	-0.202646181	-0.0142406727	0.566023969	0.2446711042
X2ndFlrSF	1.000000000	0.0633529501	0.687501064	-0.1694939517
LowQualFinSF	0.063352950	1.0000000000	0.134682813	-0.0471434219
GrLivArea	0.687501064	0.1346828130	1.000000000	0.0348360495
BsmtFullBath	-0.169493952	-0.0471434219	0.034836050	1.0000000000
BsmtHalfBath	-0.023854784	-0.0058415048	-0.018918483	-0.1478709605
FullBath	0.421377983	-0.0007095096	0.630011646	-0.0645120486
HalfBath	0.609707300	-0.0270800493	0.415771636	-0.0309049591
BedroomAbvGr	0.502900613	0.1056065685	0.521269511	-0.1506728092
KitchenAbvGr	0.059305753	0.0075217443	0.100063165	-0.0415025464
TotRmsAbvGrd	0.616422635	0.1311847760	0.825489374	-0.0532752361
Fireplaces	0.194560892	-0.0212721434	0.461679134	0.1379277084
GarageCars	0.183925583	-0.0944795202	0.467247419	0.1318812244
GarageArea	0.138346959	-0.0676014132	0.468997477	0.1791894804
WoodDeckSF	0.092165418	-0.0254436480	0.247432821	0.1753151901
OpenPorchSF	0.208026063	0.0182510391	0.330223962	0.0673414614
EnclosedPorch	0.061988691	0.0610812378	0.009113210	-0.0499106491
X3SsnPorch	-0.024357648	-0.0042956104	0.020643190	-0.0001060915

	BsmtHalfBath	FullBath	HalfBath	BedroomAbvGr
KitchenAbvGr				
MSSubClass	-0.0023325346	0.1316082224	0.177354389	-0.023438028
0.281721040				
OverallQual	-0.0401501577	0.5505997094	0.273458099	0.101676356
-0.183882235				
OverallCond	0.1178209151	-0.1941494887	-0.060769327	0.012980060
-0.087000855				
YearBuilt	-0.0381618057	0.4682707872	0.242655910	-0.070651217
-0.174800246				
YearRemodAdd	-0.0123370321	0.4390464839	0.183330612	-0.040580928
-0.149597521				
BsmtFinSF1	0.0674184779	0.0585431369	0.004262424	-0.107354677
-0.081006851				
BsmtFinSF2	0.0709481337	-0.0764438620	-0.032147837	-0.015728114
-0.040751236				
BsmtUnfSF	-0.0958042882	0.2888860555	-0.041117530	0.166643317
0.030085868				
TotalBsmtSF	-0.0003145818	0.3237224136	-0.048803739	0.050449956
-0.068900643				
X1stFlrSF	0.0019556536	0.3806374950	-0.119915909	0.127400749
0.068100588				
X2ndFlrSF	-0.0238547839	0.4213779829	0.609707300	0.502900613
0.059305753				
LowQualFinSF	-0.0058415048	-0.0007095096	-0.027080049	0.105606569
0.007521744				
GrLivArea	-0.0189184832	0.6300116463	0.415771636	0.521269511
0.100063165				
BsmtFullBath	-0.1478709605	-0.0645120486	-0.030904959	-0.150672809
-0.041502546				
BsmtHalfBath	1.0000000000	-0.0545358120	-0.012339900	0.046518848
-0.037944350				
FullBath	-0.0545358120	1.0000000000	0.136380589	0.363251983
0.133115214				
HalfBath	-0.0123399001	0.1363805887	1.000000000	0.226651484
-0.068262549				
BedroomAbvGr	0.0465188484	0.3632519830	0.226651484	1.000000000
0.198596758				
KitchenAbvGr	-0.0379443502	0.1331152142	-0.068262549	0.198596758
1.000000000				
TotRmsAbvGrd	-0.0238363413	0.5547842535	0.343414858	0.676619936
0.256045409				
Fireplaces	0.0289755866	0.2436705031	0.203648508	0.107569681
-0.123936235				
GarageCars	-0.0208910590	0.4696720433	0.219178152	0.086106438
-0.050633892				
GarageArea	-0.0245355796	0.4056562085	0.163549364	0.065252530
-0.064433047				
WoodDeckSF	0.0401612233	0.1877032138	0.108080303	0.046853773
-0.090130273				
OpenPorchSF	-0.0253237579	0.2599774255	0.199740148	0.093809572
-0.070090610				
EnclosedPorch	-0.0085553339	-0.1150929635	-0.095316526	0.041570435
0.037312385				

X3SsnPorch	0.0351136309	0.0353530166	-0.004972488	-0.024477796	
-0.024600359					
	TotRmsAbvGrd	Fireplaces	GarageCars	GarageArea	WoodDeckSF
MSSubClass	0.040380065	-0.045569340	-0.04010979	-0.09867154	-0.012579358
OverallQual	0.427452343	0.396765038	0.60067072	0.56202176	0.238923392
OverallCond	-0.057583166	-0.023819978	-0.18575751	-0.15152137	-0.003333699
YearBuilt	0.095589128	0.147716399	0.53785009	0.47895382	0.224880142
YearRemodAdd	0.191739816	0.112581318	0.42062215	0.37159981	0.205725920
BsmtFinSF1	0.044315624	0.260010920	0.22405352	0.29697039	0.204306145
BsmtFinSF2	-0.035226548	0.046920709	-0.03826351	-0.01822659	0.067898326
BsmtUnfSF	0.250647061	0.051574882	0.21417519	0.18330270	-0.005316424
TotalBsmtSF	0.285572564	0.339519324	0.43458483	0.48666546	0.232018609
X1stFlrSF	0.409515979	0.410531085	0.43931681	0.48978165	0.235458623
X2ndFlrSF	0.616422635	0.194560892	0.18392558	0.13834696	0.092165418
LowQualFinSF	0.131184776	-0.021272143	-0.09447952	-0.06760141	-0.025443648
GrLivArea	0.825489374	0.461679134	0.46724742	0.46899748	0.247432821
BsmtFullBath	-0.053275236	0.137927708	0.13188122	0.17918948	0.175315190
BsmtHalfBath	-0.023836341	0.028975587	-0.02089106	-0.02453558	0.040161223
FullBath	0.554784254	0.243670503	0.46967204	0.40565621	0.187703214
HalfBath	0.343414858	0.203648508	0.21917815	0.16354936	0.108080303
BedroomAbvGr	0.676619936	0.107569681	0.08610644	0.06525253	0.046853773
KitchenAbvGr	0.256045409	-0.123936235	-0.05063389	-0.06443305	-0.090130273
TotRmsAbvGrd	1.000000000	0.326114480	0.36228857	0.33782212	0.165983884
Fireplaces	0.326114480	1.000000000	0.30078877	0.26914124	0.200018796
GarageCars	0.362288571	0.300788766	1.000000000	0.88247541	0.226342138
GarageArea	0.337822121	0.269141238	0.88247541	1.000000000	0.224666307
	datosCasasCaras	datosCasasEconomicas	datosCasasIntermedias		
MSSubClass	-0.013666157	0.08172304	-0.080153403		
OverallQual	0.174316225	-0.58948541	0.563195060		
OverallCond	-0.026984819	0.08113947	-0.076929612		
YearBuilt	0.070978031	-0.32657833	0.317071436		
YearRemodAdd	0.060507088	-0.33385991	0.326523162		
BsmtFinSF1	0.131083340	-0.31567171	0.294135375		
BsmtFinSF2	0.013106170	0.03195476	-0.034987202		
BsmtUnfSF	0.051120953	-0.14814267	0.140090269		
TotalBsmtSF	0.192588952	-0.46564672	0.434031053		
X1stFlrSF	0.180620760	-0.46709764	0.437867301		
X2ndFlrSF	0.119702861	-0.20867903	0.187910855		
LowQualFinSF	-0.008345866	0.02722209	-0.025949511		
GrLivArea	0.231547505	-0.51447197	0.475832826		
BsmtFullBath	0.019547056	-0.19953676	0.198435610		
BsmtHalfBath	0.024813178	0.04348704	-0.048992744		
FullBath	0.144837791	-0.34222365	0.318336579		
HalfBath	0.065481423	-0.14872338	0.137840990		
BedroomAbvGr	0.047844511	-0.07754071	0.069158590		
FullBath	0.144837791	-0.34222365	0.318336579		
HalfBath	0.065481423	-0.14872338	0.137840990		
BedroomAbvGr	0.047844511	-0.07754071	0.069158590		
KitchenAbvGr	-0.014676796	0.08787182	-0.086187552		
TotRmsAbvGrd	0.173160808	-0.39097421	0.362164499		
Fireplaces	0.103224672	-0.34851675	0.332940691		
GarageCars	0.114546267	-0.46071065	0.444450065		
GarageArea	0.109910699	-0.47711835	0.462000990		
WoodDeckSF	0.041174236	-0.26155807	0.257041258		
OpenPorchSF	0.044070404	-0.24974330	0.244490593		
EnclosedPorch	-0.024940406	0.07523221	-0.071344622		
X3SsnPorch	-0.008075001	-0.01199899	0.013760920		

[reached getOption("max.print") -- omitted 9 rows]

Adicionalmente se realizó un gráfico de correlación para identificar qué variables podrían estar causando multicolinealidad.

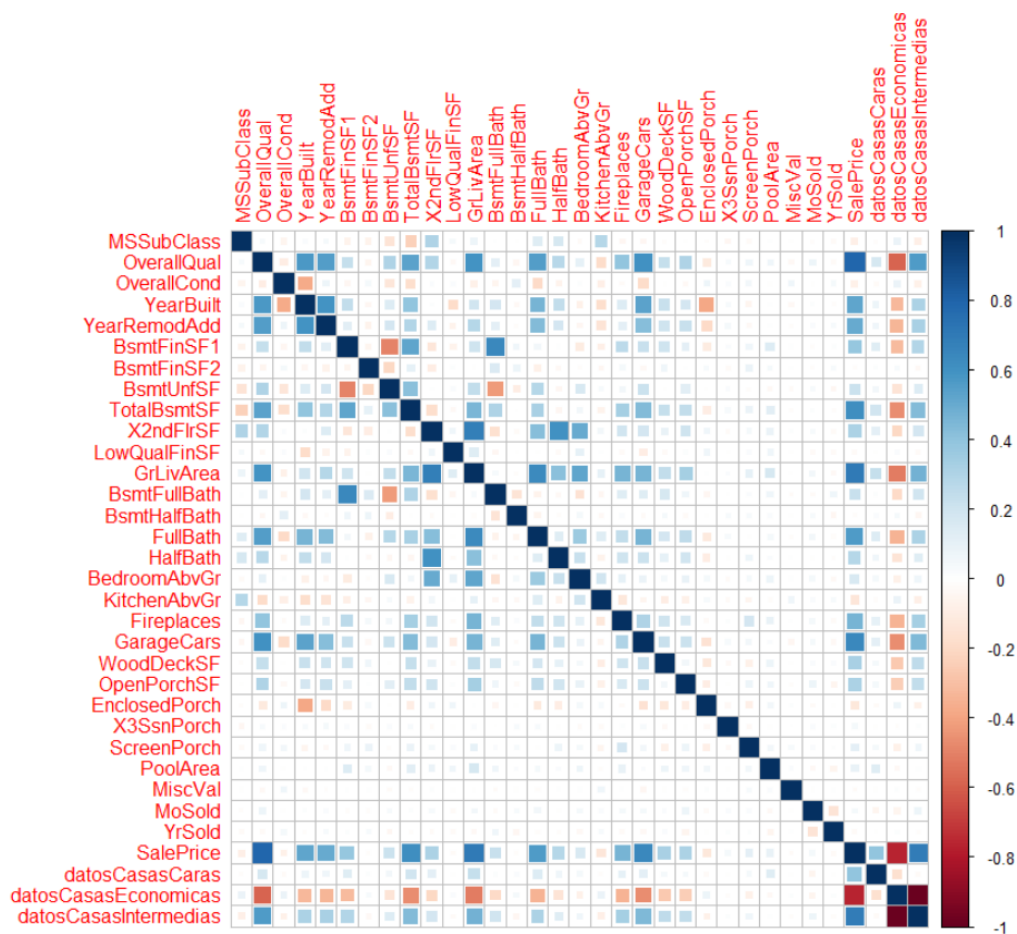


Se hizo un gráfico de correlación para evaluar la multicolinealidad, al igual que una matriz de correlación. Utilizando esta información se determinó que habían varias variables que se relacionaban entre sí y mantenerlas podía aumentar el riesgo de overfitting, en base a eso se utilizó la gráfica para evaluar cuales eran las variables correlacionadas y en base a la matriz se eliminó las variables que tenían menos relación con la variable respuesta.

X1stFlrSF-totalBsmtSF 0.8195 es su nivel de correlación
 totRmsAbvGrd-GrLivArea 0.825 es su nivel de correlación
 GarageArea-GarageCars 0.8824 es su nivel de correlación
 evaluando su correlación con y
 X1stFlrSF 0.6
 totalBsmtSF 0.61
 totRmsAbvGrd 0.53
 GrLivArea 0.7
 GarageArea 0.62
 GarageCars 0.64

Evaluando esto, para evitar multicolinealidad y por ende overfitting, se eliminará X1stFlrSF, totRmsAbvGrd y GarageArea.

Luego de eliminar la multicolinealidad se realizó de nuevo el gráfico para determinar si no habían otras variables que podían estar afectando.



Como se puede ver, ninguno de los datos muestra ser factor de multicolinealidad.

Modelo con el conjunto de prueba para predecir y matriz de confusión

Casas económicas

```

Coefficients:
(Intercept)      MSSubClass      overallQual      overallCond      YearBuilt      YearRemodAdd      BsmtFinSF1
BsmtFinSF2      BsmtUnfSF      -3.456e+00      -1.827e+00      -7.903e-02      -2.047e-01      -1.411e-02
2.118e+02      2.735e-02
-1.435e-02      -1.230e-02
X2ndFlrSF      LowQualFinSF      GrLivArea      BsmtFullBath      BsmtHalfBath      FullBath      HalfBath
BedroomAbvGr      KitchenAbvGr      -3.399e-03      -3.664e+00      -6.068e+00      -2.609e+00      3.537e+00
-1.440e-02      1.416e-02
-2.213e-01      1.128e+01
Fireplaces      GarageCars      WoodDeckSF      OpenPorchSF      EnclosedPorch      X3SsnPorch      ScreenPorch
PoolArea      MiscVal      -3.324e+00      -2.490e+00      2.972e-05      2.200e-03      -2.631e-03      -2.216e-02      -3.655e-03
5.739e-02      -9.468e-04
MoSold      YrSold
-3.145e-01      2.125e-01

Degrees of Freedom: 438 Total (i.e. Null); 410 Residual
Null Deviance: 357.5
Residual Deviance: 40.25      AIC: 98.25
Confusion Matrix and Statistics

```

```

      Reference
Prediction 0 1
0      58  4
1       4 373

      Accuracy : 0.9818
      95% CI : (0.9644, 0.9921)
No Information Rate : 0.8588
P-Value [Acc > NIR] : <2e-16

      Kappa : 0.9249

McNemar's Test P-Value : 1

      Sensitivity : 0.9355
      Specificity : 0.9894
      Pos Pred value : 0.9355
      Neg Pred value : 0.9894
      Prevalence : 0.1412
      Detection Rate : 0.1321
      Detection Prevalence : 0.1412
      Balanced Accuracy : 0.9624

'Positive' class : 0

```

Puede observarse que en el modelo generado ya no se encuentran las variables con valores NA.

El modelo para predecir las casas económicas se ha equivocado al predecir que 4 cosas sí eran económicas y al decir que 4 casas no lo eran, por esta razón, tiene un porcentaje de acierto de 98.18%

Casas caras

```

Coefficients:
(Intercept)      MSubClass      overallQual      overallCond      YearBuilt      YearRemodAdd      BsmtFinSF1
BsmtFinSF2      BsmtUnfSF      5.068e+03      1.831e-02      2.143e+00      5.637e+00      1.386e-02      -2.307e-01      2.970e-02
4.107e-02      2.643e-02      x2ndFlrSF      LowQualFinSF      GrLivArea      BsmtFullBath      BsmtHalfBath      FullBath      HalfBath
BedroomAbvGr      KitchenAbvGr      2.865e-02      -4.679e-02      4.748e-03      -1.992e+00      2.552e+00      1.129e+01      -6.820e+00
-2.517e-01      -2.390e+01      Fireplaces      GarageCars      WoodDeckSF      OpenPorchSF      EnclosedPorch      X3SsnPorch      ScreenPorch
PoolArea      MiscVal      -1.812e+01      4.873e+00      -6.031e-03      5.850e-02      -1.213e-01      5.311e-02      6.390e-02
6.134e-02      5.123e-03      MoSold      YrSold      -4.140e+00      -2.371e+00

Degrees of Freedom: 438 Total (i.e. Null); 410 Residual
Null Deviance: 25.56
Residual Deviance: 2.226e-10 AIC: 58
Confusion Matrix and Statistics

```

```

      Reference
Prediction 0 1
0      437  0
1       0  2

      Accuracy : 1
      95% CI : (0.9916, 1)
No Information Rate : 0.9954
P-Value [Acc > NIR] : 0.1347

      Kappa : 1

McNemar's Test P-Value : NA

      Sensitivity : 1.0000
      Specificity : 1.0000
      Pos Pred value : 1.0000
      Neg Pred value : 1.0000
      Prevalence : 0.9954
      Detection Rate : 0.9954
      Detection Prevalence : 0.9954
      Balanced Accuracy : 1.0000

'Positive' class : 0

```

Puede observarse que en el modelo generado ya no se encuentran las variables con valores NA.

En el modelo solo se ha predicho que dos casas están consideradas como caras y las ha predicho correctamente, contando con un porcentaje de acierto de 1, creemos que es posible que esto suceda debido a la poca cantidad de casas caras que se encuentran en el conjunto de datos

Casas Intermedias:

```

Coefficients:
(Intercept)      MSSubClass    overallQual    overallCond      YearBuilt      YearRemodAdd
  4.232e+00     -5.194e-03     1.321e+00     2.700e-01     1.511e-02     4.575e-02
 BsmFinSF1      BsmFinSF2      BsmUnFSF      X2ndFlrSF      LowQualFinSF      GrLivArea
 -2.919e-03     -2.975e-03     -2.489e-03     -1.858e-03     -1.069e-02     3.876e-03
 BsmFullBath    BsmHalfBath      FullBath      HalfBath      BedroomAbvGr      KitchenAbvGr
  1.737e+00     -4.179e-01     -7.704e-01     -7.291e-01     2.496e-02     -1.430e+01
 Fireplaces      GarageCars      WoodDeckSF      OpenPorchSF      EnclosedPorch      X3SsnPorch
  9.345e-01     1.148e+00     4.004e-03     4.245e-03     1.725e-03     1.206e-03
 ScreenPorch     PoolArea      MiscVal      MoSold      YrSold
  5.000e-03     -1.298e-02     2.786e-04     1.301e-02     -6.464e-02

Degrees of Freedom: 1020 Total (i.e. Null);  992 Residual
Null Deviance:      845.1
Residual Deviance: 323.9      AIC: 381.9
Confusion Matrix and Statistics

```

```

      Reference
Prediction 0  1
0    367  12
1     15  45

Accuracy : 0.9385
95% CI : (0.9118, 0.9591)
No Information Rate : 0.8702
P-Value [Acc > NIR] : 2.36e-06

Kappa : 0.7338

McNemar's Test P-Value : 0.7003

Sensitivity : 0.9607
Specificity : 0.7895
Pos Pred Value : 0.9683
Neg Pred Value : 0.7500
Prevalence : 0.8702
Detection Rate : 0.8360
Detection Prevalence : 0.8633
Balanced Accuracy : 0.8751

'Positive' Class : 0

```

Puede observarse que en el modelo generado ya no se encuentran las variables con valores NA.

En cuanto a las casas intermedias, se ha equivocado al clasificar 15 casas como intermedias y 12 como no intermedias, por lo que tiene un porcentaje de acierto de 93.85%.

Se ha obtenido un mayor éxito en los aciertos de las casas caras, sin embargo, debido a la poca cantidad de estas que existen, creemos que eso puede influenciar en el resultado, por esto mismo hemos decidido que el modelo que ha hecho un mejor

trabajo es el modelo para predecir qué casas son económicas con un porcentaje de acierto de 98.18%.