

MINERÍA DE DATOS



*Excelencia que trasciende*

**DEL VALLE**  
GRUPO EDUCATIVO

MIRKA NICOLLE MONZON CABRERA

OSCAR WALDEMAR DE LEON CHAVARRIA

ANDRES EMILIO QUINTO VILLAGRAN

**HOJA DE TRABAJO 6 : REGRESIÓN LOGÍSTICA**

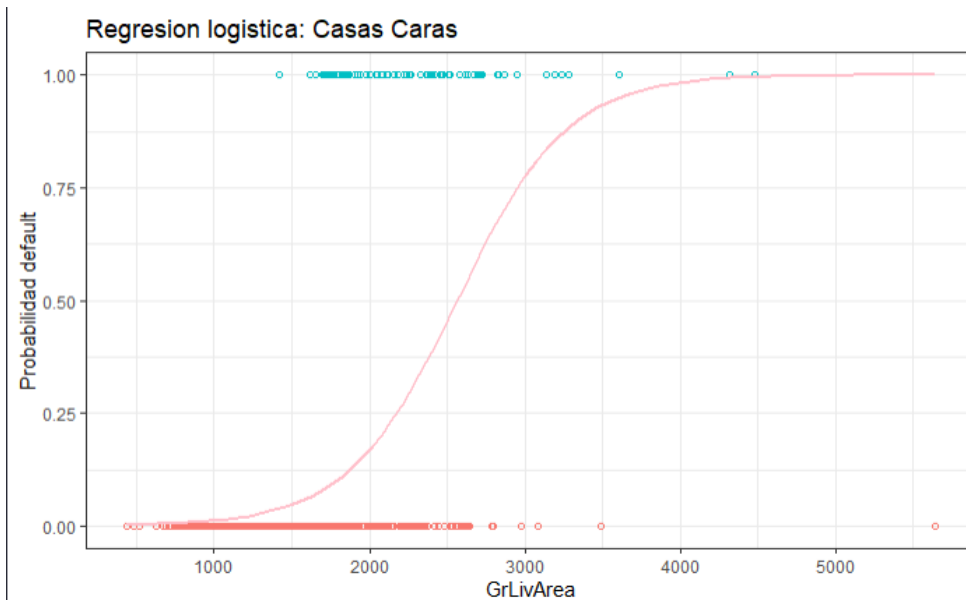
## Regresión logística:

### Caras

```
Call: glm(formula = datos1 ~ ., family = binomial(), data = train[,
c(2:10, 12)], maxit = 100)

Coefficients:
(Intercept)      LotArea      GrLivArea      YearBuilt      BsmtUnfSF      TotalBsmtSF      X1stFlrSF
-3.537e+02      3.307e-05      3.274e-03      3.576e-02     -1.360e-03      4.401e-04      2.386e-03
GarageYrBlt      GarageArea      YearRemodAdd
 6.192e-03      2.482e-03      1.284e-01

Degrees of Freedom: 787 Total (i.e. Null); 778 Residual
Null Deviance: 513.2
Residual Deviance: 201.6      AIC: 221.6
```

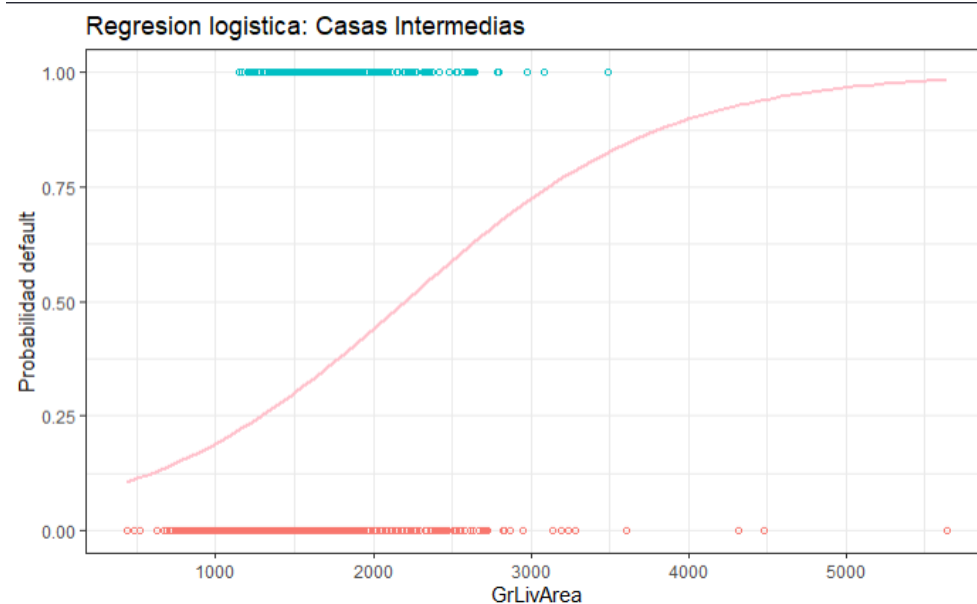


### Intermedias

```
Call: glm(formula = datos2 ~ ., family = binomial(), data = train[,
c(2:10, 13)], maxit = 100)

Coefficients:
(Intercept)      LotArea      GrLivArea      YearBuilt      BsmtUnfSF      TotalBsmtSF      X1stFlrSF
-1.078e+02      8.968e-06      9.128e-04      2.755e-02      8.386e-04     -5.263e-04      9.131e-05
GarageYrBlt      GarageArea      YearRemodAdd
-1.167e-02     -5.879e-04      3.737e-02

Degrees of Freedom: 787 Total (i.e. Null); 778 Residual
Null Deviance: 984.7
Residual Deviance: 758.7      AIC: 778.7
```



## Económicas

```
Call: glm(formula = datos3 ~ ., family = binomial(), data = train[,  
c(2:10, 14)], maxit = 100)
```

### Coefficients:

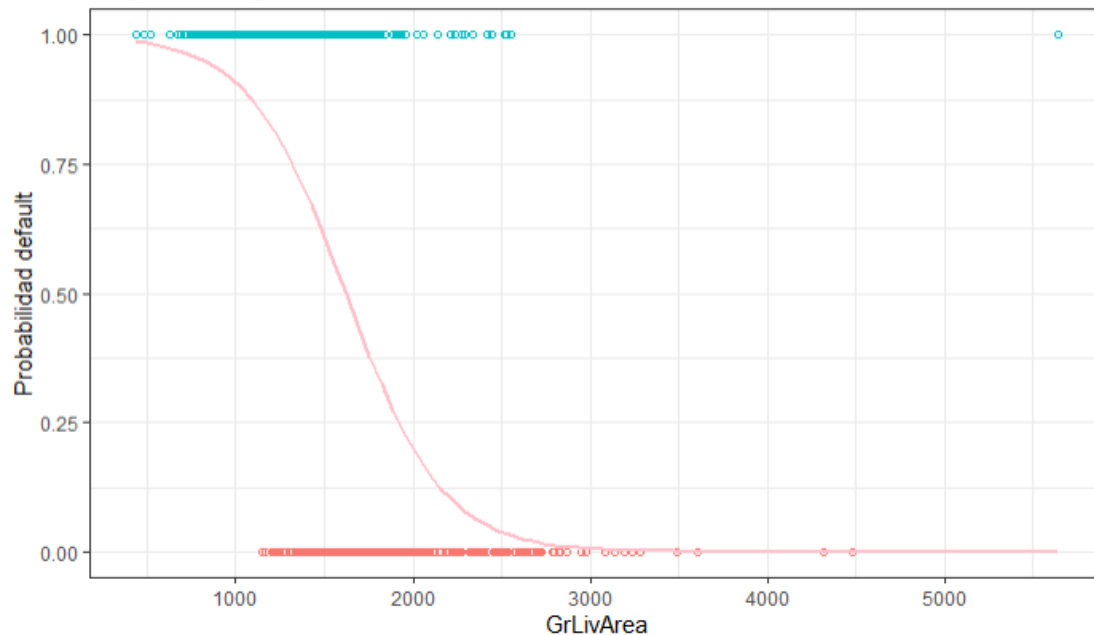
(Intercept)	LotArea	GrLivArea	YearBuilt	BsmtUnfsF	TotalBsmSF	X1stFlrSF
1.877e+02	-5.634e-05	-4.056e-03	-5.325e-02	-1.615e-05	-1.774e-03	5.601e-04
GarageYrBlt	GarageArea	YearRemodAdd				
3.111e-02	-2.810e-03	-6.743e-02				

Degrees of Freedom: 787 Total (i.e. Null); 778 Residual

Null Deviance: 1071

Residual Deviance: 416.3 AIC: 436.3

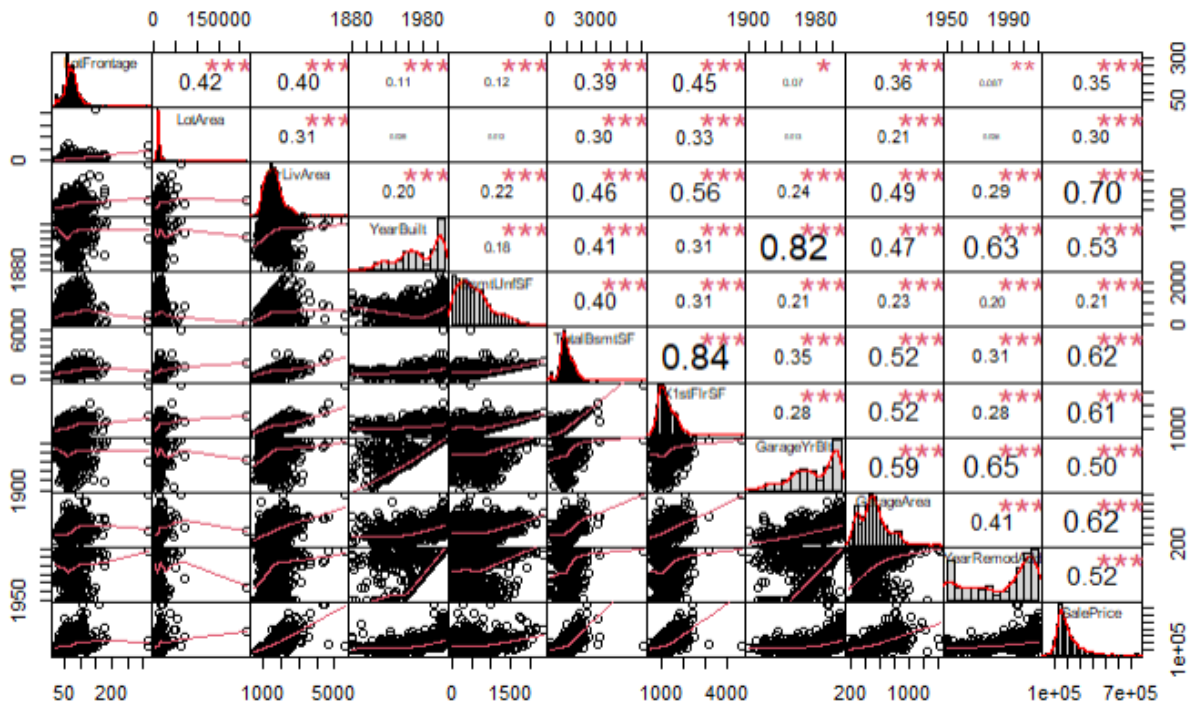
### Regresion logistica: Casas Economicas



## Análisis de las variables del modelo

Como se puede observar en la tabla de correlación, tenemos variables con altos índices, estas son las variables que aportan al modelo de regresión lineal. Esto también nos dice que el modelo encaja con los datos de las mismas.

En el caso del overfitting puede que con algunas variables se esté dando el caso, ya sea en las variables que tienen poco índice de correlación entre ellas.



## Eficiencia del algoritmo para clasificar o predecir

2	3	9	22	26	27	29	34	35	52	53	54
-8.24470545	-2.88931095	-13.50814450	-15.65852386	-1.43934844	-7.73499406	-4.35682942	-8.73369368	-1.48557076	-15.23806937	-12.69241886	-0.31937020
55	56	58	61	63	64	70	75	80	81	97	98
-11.02847296	-9.87362996	-3.56534858	-3.60206646	-3.96368549	-13.48231411	-8.54391006	-15.29568265	-12.10888394	-7.06722218	-2.47856660	-12.19921173
103	106	107	110	111	116	119	122	124	130	131	138
-6.56688398	-2.30288342	-11.59510907	-5.40246262	-8.66665119	-5.48571552	2.31659216	-15.95875152	-6.45171751	-7.59687669	-7.32253410	-3.23611552
139	140	144	146	155	159	172	173	175	177	180	198
-2.58647580	-3.50833957	-3.27486785	-3.94409347	-15.74142767	-1.88923184	-8.13264323	-5.40432027	-3.32110501	-3.62462173	-9.25074706	-0.99481662
202	210	213	216	218	227	229	232	237	240	248	256
-7.92806242	-11.29038123	-0.93142732	-7.30288754	-15.90350286	-0.85072660	-11.92223317	1.15611471	-3.99400810	-14.57147820	-12.74633841	-2.74111209
261	264	271	283	284	287	291	295	298	302	304	304
-7.96816247	-7.18928567	-0.98228984	-3.29496297	-2.42029881	-1.93032646	-7.30216027	-1.82306464	-9.82981604	-2.93828577	-0.55539863	-10.53129157
313	316	321	323	332	335	341	345	353	360	366	372
-15.11934105	-4.47027780	1.04898655	-0.09946956	-8.52518411	-3.92516387	-3.32006719	-11.17851708	-15.64819783	-2.20900803	-8.87472468	-9.58885205
373	375	391	392	397	400	403	409	410	414	415	419
-10.17651590	-3.38428884	-14.99348127	-2.29620076	-10.59536714	-2.58599617	-11.71086521	-0.39693208	0.26369689	-16.33005339	-2.82579097	-15.75446238
421	426	433	436	437	439	444	445	450	451	461	461
-2.97911977	-9.87009271	-12.69091123	-10.48091306	-4.22220249	-16.21795653	-10.80797000	-4.06888832	-3.80279967	-8.33751340	-10.49707147	-1.46313113
462	464	468	473	475	476	478	483	486	487	490	495
-7.49347851	-8.06807220	-7.55735390	-4.46153993	-1.90933413	-11.74285044	1.88007251	-7.62688246	-7.05131064	-11.04868021	-9.48133644	-10.33256200
500	503	504	511	512	513	516	524	525	531	533	544
-10.77553155	-10.60849805	-2.80597855	-5.86990292	-4.06100777	-13.15547619	2.78541982	14.89837434	-0.15951197	-4.75464768	-7.94925048	-7.56287786
552	553	554	562	564	566	569	574	575	578	580	582
-14.52245032	-1.13639209	-6.39867572	-7.64575419	-13.67053946	-15.46274400	0.57620402	-5.42864888	-7.30112748	-9.33468888	-12.07261241	2.11986465
584	585	586	589	592	603	618	626	629	631	632	632
-2.61994783	-8.80094208	2.36188310	-3.11619443	2.84786769	-3.98205626	-15.22185647	-8.21435660	-12.68885849	-7.35484630	-10.12757574	-2.79788150
634	635	640	647	656	659	660	662	665	668	670	676
-6.96821207	-8.23226214	-1.12395537	-14.23373028	-12.81194702	-14.02139253	-5.11067997	3.40419153	3.85808491	-1.98382041	-15.57681994	-9.11288559
678	684	690	695	697	702	705	707	716	733	734	734
-10.67426824	-1.08631776	-4.07103112	-13.30419135	-18.19767869	-11.06346021	-2.32219946	-4.60844851	-9.17297112	-15.25402309	-2.73625580	-9.27214783
736	741	744	754	756	761	772	781	797	797	801	801
-11.62686065	-8.95625095	-4.36755718	-1.15865506	-5.46382294	-13.63584996	-12.72602014	-6.43812740	-9.68990880	-5.40671429	-8.59813359	-4.45955724
802	803	815	831	831	832	840	850	851	865	871	874
-10.29342692	-4.07006460	-16.49327539	-6.10612734	-10.71045036	-6.16041020	-9.31580253	-4.91774911	-6.31054722	-3.49747710	-14.18114336	-14.00352060
876	879	882	884	885	890	893	896	897	906	907	913
0.29726870	-6.48501301	-4.43714240	-12.37026452	-12.78252221	-10.75290981	-7.18772799	-9.74648541	-16.03654351	-13.82048281	-1.45887775	-15.47210417
914	919	925	931	933	934	939	946	950	965	973	974
-10.71018008	-2.90640737	-4.19199914	-2.95433799	-0.53313618	-3.41265610	-0.36626901	-6.74684759	-3.25941589	-2.55284566	-11.23727993	-3.26067543
988	990	991	993	1008	1009	1013	1016	1017	1020	1028	1041
2.02502080	-3.80368924	0.52785281	-5.29494280	-12.13652452	-2.24561430	-9.36229687	-3.65438440	-3.96148010	-3.59040693	0.56829819	-2.77747767
1051	1052	1053	1054	1055	1064	1070	1075	1077	1080	1083	1086
-4.18303265	-4.03459616	-6.59552562	-5.24781686	-1.35219788	-11.26737713	-8.13480613	-3.69384537	-7.64174280	-7.53844558	-4.07302559	-6.79778308
1088	1090	1093	1095	1106	1107	1126	1127	1128	1131	1135	1136
-1.02736055	-3.27602798	-12.43583510	-8.71367630	0.57349588	-4.04678345	-13.68135861	-2.43599216	-1.17926342	-10.88469116	-5.42448552	-17.10420862
1140	1146	1148	1158	1159	1160	1164	1167	1168	1170	1171	1173
-14.22462378	-14.91655706	-13.94678112	-2.42234891	-1.65569911	-7.55018903	-9.14170826	-1.26178127	-3.45016398	5.89049016	-9.60076820	-5.08284431
1175	1176	1197	1203	1205	1215	1221	1222	1226	1229	1230	1233
-10.48203189	2.32033194	-2.41305377	-9.79766611	-9.44335874	-12.44960022	-12.52448749	-10.31936052	-12.49127698	1.49943891	-10.03136792	-11.00131559
1240	1241	1247	1250	1255	1257	1265	1268	1283	1289	1290	1293
-2.81041957	-2.96517123	-3.95807381	-14.87982108	-4.52334555	-0.55453260	-5.09991799	0.43941715	-5.52277825	-1.91679887	-1.54533478	-9.94666027
1295	1307	1314	1320	1323	1330	1336	1337	1346	1348	1360	1373
-10.07412014	-4.05561799	-0.05418656	-14.57113907	-3.49849488	-5.05851256	-7.72139152	-8.11617373	-16.37455999	-0.93523550	1.70112863	-1.68153848
1375	1376	1377	1378	1380	1383	1386	1391	1394	1400	1402	1403

Con la predicción realizada se puede ver el resultado para las variables respuesta.

