

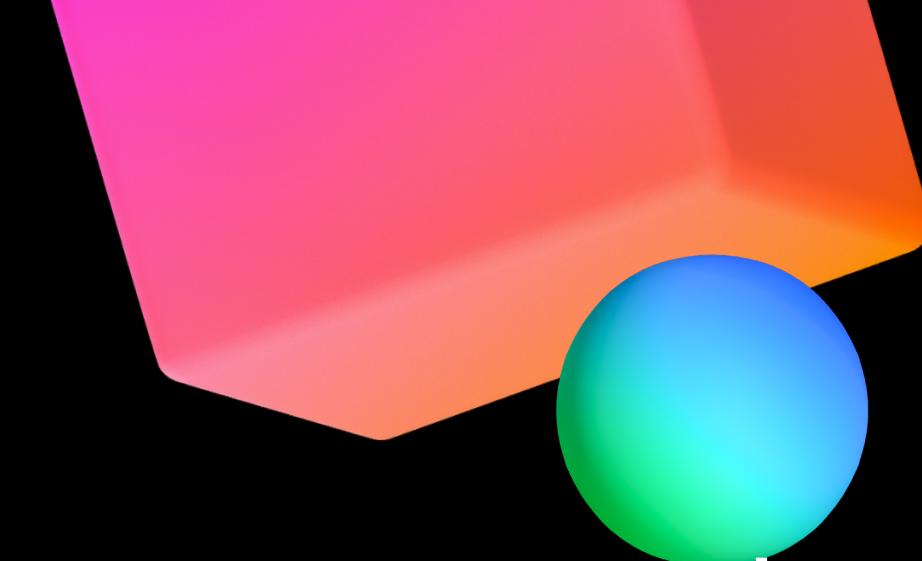
# PROYECTO 2

Daniel Molina  
Benjamín Izquierdo

Por



	X	Posted.On	BHK	Rent	Size	Floor	Area.Locality	Bathroom	City.Bangalore	City.Chennai	Cil
1	1	2022-05-18	2	10000	1100	Ground out of 2	Bandel	2	0	0	
2	2	2022-05-13	2	20000	800	1 out of 3	Phool Bagan, Kankurgachi	1	0	0	
3	3	2022-05-16	2	17000	1000	1 out of 3	Salt Lake City Sector 2	1	0	0	
4	4	2022-07-04	2	10000	800	1 out of 2	Dum dum Park	1	0	0	
5	5	2022-05-09	2	7500	850	1 out of 2	South Dum Dum	1	0	0	
6	6	2022-04-29	2	7000	600	Ground out of 1	Thakurpukur	2	0	0	
7	7	2022-06-21	2	10000	700	Ground out of 4	Malancha	2	0	0	
8	8	2022-06-21	1	5000	250	1 out of 2	Malancha	1	0	0	
9	9	2022-06-07	2	26000	800	1 out of 2	Palm Avenue Kolkata, Ballygunge	2	0	0	
10	10	2022-06-20	2	10000	1000	1 out of 3	Natunhat	2	0	0	
11	11	2022-05-23	3	25000	1200	1 out of 4	Action Area 1, Rajarhat Newtown	2	0	0	
12	12	2022-06-07	1	5000	400	1 out of 1	Keshtopur	1	0	0	
13	13	2022-05-14	1	6500	250	1 out of 4	Tarulia, Keshtopur	1	0	0	
14	14	2022-05-09	1	5500	375	1 out of 2	Dum Dum Metro	1	0	0	
15	15	2022-05-05	3	8500	900	Ground out of 2	Paschim Barisha	2	0	0	
16	16	2022-06-01	3	40000	1286	1 out of 1	New Town Action Area 1	2	0	0	



En este caso no habían valores nulos pero para comprender mejor los datos y aplicando one hot encoding y feature engineering para verificar si un dato nulo pasaba desapercibido..



X	Posted.On	BHK	Size	Floor	Area.Locality	Bathroom	rownum	City.Bangalore	City
---	-----------	-----	------	-------	---------------	----------	--------	----------------	------

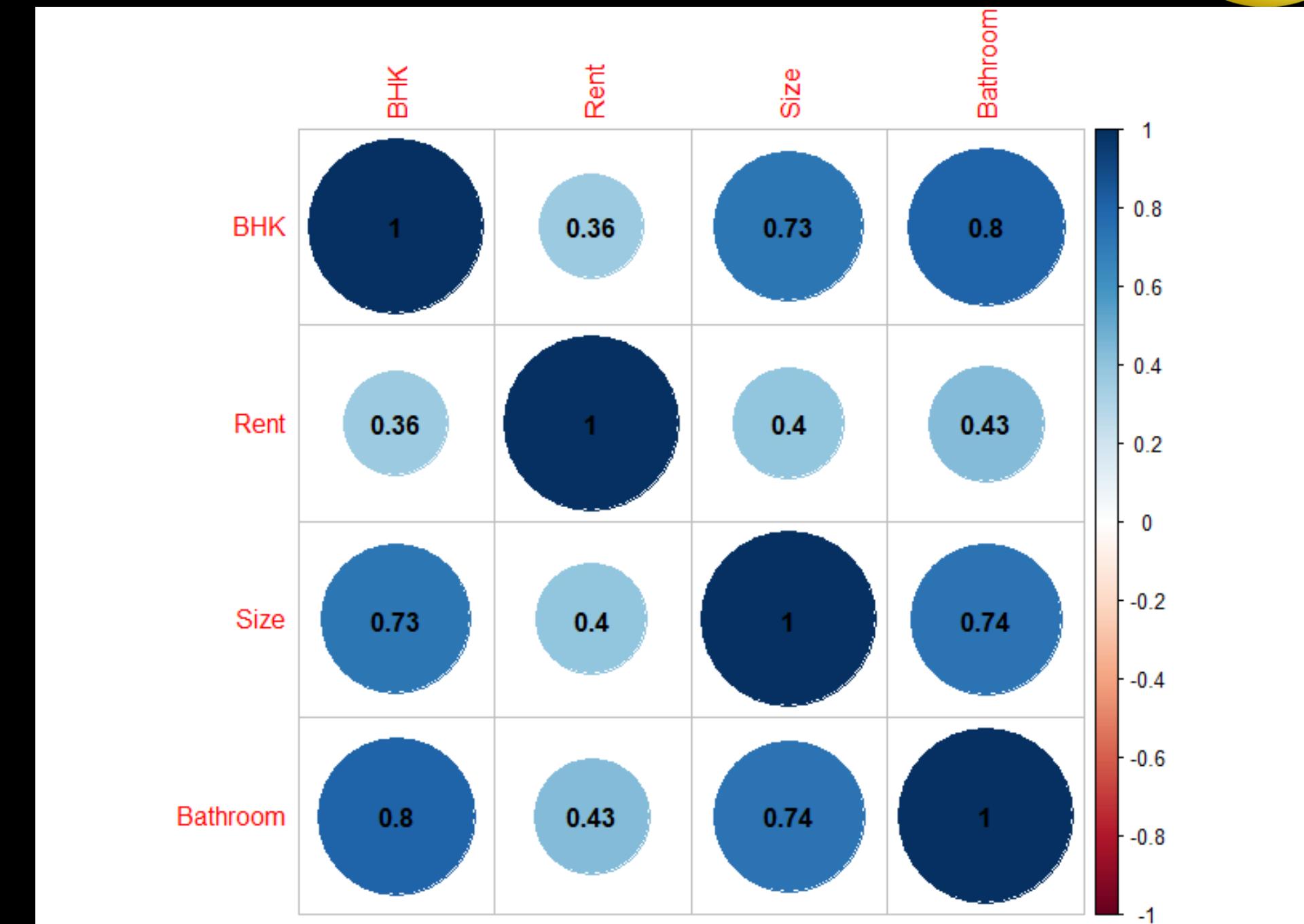
el df\_all1 es el dataframe de rentresults10  
y no trae de por si la columna renta y se  
realizaron pruebas para determinar una  
predicción.

	BHK	Rent	Size	Bathroom
BHK	1.0000000	0.3614742	0.7299383	0.8010655
Rent	0.3614742	1.0000000	0.3980805	0.4346354
Size	0.7299383	0.3980805	1.0000000	0.7366884
Bathroom	0.8010655	0.4346354	0.7366884	1.0000000

Tabla que se utiliza para verificar cuanta correlación hay entre columnas y también puede llegar a ser negativo -1 ninguna correlación.

# DIAGRAMA UTILIZADO

Diagrama que muestra los datos de las tablas y la importancia en la que se muestra en cada circulo identificado con cada número dependiendo sea más grande el circulo más grande su valor



# MODELO DE LOS DATOS

```
model<-lm(Rent~ BHK + Size + Bathroom + City.Bangalore + City.Chennai + City.Delhi + City.Hyderabad  
+ City.Kolkata + City.Mumbai + Furnishing.Status.Furnished + Furnishing.Status.SemiFurnished +  
Furnishing.Status.Unfurnished + Area.Type.BuiltArea + Area.Type.CarpetArea +  
Area.Type.SuperArea + Tenant.Preferred.Bachelors + Tenant.Preferred.BachelorsFamily +  
Tenant.Preferred.Family + Point.of.Contact.ContactAgent + Point.of.Contact.ContactBuilder +  
Point.of.Contact.ContactOwner, data=train)  
summary(model)
```

Name	Type	Value
model	list [12] (S3: lm)	List of length 12
coefficients	double [22]	-3983.2 1507.8 36.5 12411.8 -45185.8 -52355.7 ...
residuals	double [4278]	-16367 17562 7264 6992 -1205 -1122 ...
effects	double [4278]	-2255362 1799010 -977346 908153 -162634 -535788 ...
rank	integer [1]	17
fitted.values	double [4278]	26367 2438 9736 3008 8705 8122 ...
assign	integer [22]	0 1 2 3 4 5 ...
qr	list [5] (S3: qr)	List of length 5
df.residual	integer [1]	4261
xlevels	list [0]	List of length 0
call	language	lm(formula = Rent ~ BHK + Size + Bathroom + City.Bangalore + City.Chennai + ...
terms	formula	Rent ~ BHK + Size + Bathroom + City.Bangalore + City.Chennai + City.Delhi + ...
model	list [4278 x 22] (S3: data.frame)	A data.frame with 4278 rows and 22 columns



Para terminar no se utilizo la columna floor y la columna Area.locality porque presentan poca correlación en el sistema y no afecta mucho en la predicción final del proceso.

	rownum,"Rent"
1,	-1911.40029910226
2,	-9380.49653117808
3,	15750.3843489292
4,	-8704.45175948704
5,	-3117.25108047443
6,	-980.945547431493
7,	3157.19375494778
8,	31270.9778223932
9,	6117.46576736163
10,	2712.86622551908
11,	49354.3512776824
12,	31074.8288132748
13,	14236.3671873848
14,	19462.9293375017
15,	-4864.17176526486
16,	5453.2015500958
17,	61697.7973279078
18,	16971.0104215032
19,	2490.11651225542
20,	18209.7162924518
21,	60872.4838720479
22,	5322.41091619177
23,	10067.3503308252
24,	14790.1636914554
25,	46382.9125312769
26,	18817.4831588211
27,	5795.25628051509

Columna de renta en la cual se utiliza para determinar finalmente como queda la predicción final del proyecto.

```
model<-lm(Rent~ BHK + Size + Bathroom, data=train)
summary(model)
#-----
```

```
model<-lm(Rent~ BHK + Size + Bathroom + city.Bangalore + city.chennai + city.Delhi + city.Hyderabad
+ city.Kolkata + city.Mumbai, data=train)
summary(model)
#-----
```

	A	B	C
1	rownum,"Rent"		
2	1,717.251064851764		
3	2,497.272053957757		
4	3,33399.0690704385		
5	4,-848.851889473353		
6	5,-602.623000512274		
7	6,1597.16710842779		
8	7,2450.83327393674		
9	8,43544.3534496298		
10	9,2477.08315200382		
11	10,-1702.51805498231		
12	11,65693.4288103713		
13	12,35845.0880683396		
14	13,3796.95721736785		
15	14,34498.9641249085		
16	15,-2802.41310945234		
17	16,4650.62338287681		
18	17,72292.7991371915		
19	18,33157.0921584551		
20	19,1350.93821946671		
21	20,34498.9641249085		
22	21,42444.4583951598		
23	22,743.500942918832		
24	23,32299.1740159684		
25	24,30099.3839070284		
26	25,38044.8781772797		
27	26,32299.1740159684		
28	27,-4148.53705288345		

rownum,"Rent"
1,1012.60946818617
2,628.994540491025
3,21601.5611008345
4,-8137.12665021091
5,-1289.08009798474
6,2547.06917896678
7,-2382.90273478364
8,45712.2793993425
9,4081.52888974738
10,-3207.15473646049
11,54352.1461271547
12,32285.7569300123
13,6383.21845591829
14,23519.6357393103
15,-5125.22937493626
16,1453.24654216788
17,65860.5939580092
18,21179.5846803699
19,-4300.97737325939
20,23519.6357393103
21,43794.2047608668
22,7477.04109271719
23,19683.4864623588
24,15847.3371854073
25,36121.9062069638
26,19683.4864623588

# EXPERIMENTOS EFECTUADOS

```
model<-lm(Rent~ BHK + Size + Bathroom + City.Bangalore + City.Chennai + City.Delhi + City.Hyderabad  
+ City.Kolkata + City.Mumbai + Furnishing.Status.Furnished + Furnishing.Status.SemiFurnished -  
Furnishing.Status.Unfurnished, data=train)  
summary(model)  
#-----  
test$Rent<-predict(model,test)  
  
head(test)  
df <- data.frame(test)  
df  
write.csv(df,file="C:/Users/benja/Downloads/house1.csv")  
write.csv(test[,c("rownum","Rent")], file="C:/Users/benja/Downloads/housexp1.csv",row.names=FALSE)
```

rownum	Rent
1	1,415.270271113382
2	2,35.0790689395508
3	3,20909.7303784675
4	4,-8785.2253160690
5	5,-2380.6745435521
6	6,1936.03507980873
7	7,4910.92492749498
8	8,44937.6828504296
9	9,3456.79988850407
10	10,4226.4492581576
11	11,53722.122425405
12	12,31630.990774345
13	13,13731.229312503
14	14,22810.686389336
15	15,-5667.788963668
16	16,8712.8369492333
17	17,65127.858490620
18	18,19976.722454453
19	19,5408.110805053

```
model<-lm(Rent~ BHK + Size + Bathroom + City.Bangalore + City.Chennai + City.Delhi + City.Hyderabad  
+ City.Kolkata + City.Mumbai + Furnishing.Status.Furnished + Furnishing.Status.SemiFurnished +  
Furnishing.Status.Unfurnished + Area.Type.BuiltArea + Area.Type.CarpetArea +  
Area.Type.SuperArea + Tenant.Preferred.Bachelors + Tenant.Preferred.BachelorsFamily +  
Tenant.Preferred.Family, data=train)  
summary(model)
```

rownum	Rent
1	-2992.55086048058
2	-6373.98210267532
3	17217.5009275739
4	-3981.20540021723
5	-5492.09288216297
6	-155.780436242772
7	1917.82691898574
8	31316.8151029299
9	8171.90066526937
10	1174.24144498366
11	49595.7022212005
12	34663.7724382224
13	17467.5261444153
14	19114.0732127929
15	-931.646303781459
16	5710.97148942359
17	60975.1359325141
18	16576.5997454061
19	-8541.65197859874
20	19114.0732127929
21	54382.4574824818
22	3470.83834934925
23	12318.8118572041

	rownum,"Rent"
1,	-14704.7838371085
2,	-12252.4721814195
3,	15117.2018491879
4,	-7886.34917299626
5,	-4189.95875744333
6,	-2067.74285870966
7,	-13509.3667048651
8,	33848.4030800506
9,	4373.78606025075
10,	-5517.7466403353
11,	57933.7765184974
12,	26957.8688417083
13,	13862.755504566
14,	18650.628471404
15,	-1982.04234106616
16,	-1241.20876651325
17,	62836.2793770497
18,	18157.6932049395
19,	1224.76945103403
20,	17909.9277368922
21,	63194.6391318286
22,	7584.78239724318
23,	8000.304300000005

# CSV DEFINITIVO

7	G1			71076.48110	7	2h
---	----	--	--	-------------	---	----