

## Actividad 1 – Regresión

Antonio Jaime López – A01424138

Los data frames a analizar fueron los listings de Aribnb, los cuales nos hablan de los datos recolectados sobre las distintas opciones para la renta de habitaciones de distintas ciudades, en este caso, yo ocupe México, Berlín y Florence. Dentro de las variables a analizar nos centramos solo en 6 siendo utilizadas como mis variables independientes (`['host_acceptance_rate', 'review_scores_rating', 'price', 'review_scores_cleanliness', 'availability_365', 'review_scores_communication']`) y como variable dependiente `'Number_of_reviews'`.

Para poder analizar de manera más profunda y distinguir patrones, los 3 dataframes se terminaron dividiendo para analizar por separado los tipos de habitación que Airbnb ofrece, por la naturaleza y tamaño de los mismos solo utilice 3 (Entire Home/Apt, Private Room y Shared Room), siendo estos 3 los que contaban con más instancias en las 3 ciudades.

Los resultados del análisis fueron algo inconclusos debido a que tanto las correlaciones como los coeficientes de determinación daban resultados demasiado bajos, a continuación los correspondientes a la Cd de México.

Para el caso de la Cd. de México se realizaron regresiones lineales simples, siempre usando como variable dependiente a `"Number_of_reviews"`

Entire Home

R2	Correlacion
0.013	0.12
7.55e-05	0.0087
0.0044	-0.067
0.00012	0.011
0.0068	-0.083
0.0010	-0.032

Private Room

R2	Correlacion
0.0090	0.094
1.91e-06	0.00014
0.010	-0.1
1.82e-07	0.00043
0.0080	0.09

0.0038	-0.062
Shared Room	
R2	Correlacion
0.0088	0.094
0.0032	-0.057
0.0065	-0.081
0.0034	-0.058
0.0021	-0.047
0.015	-0.12

Para el análisis de Berlín y Florence, se hicieron regresiones utilizando todas las variables para analizar su comportamiento:

### Berlín:

Entire Home

R2 0.083



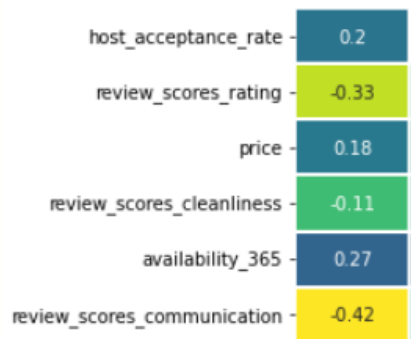
Private Room

R2 0.049



## Shared Room

R2 0.24



## Florence:

### Entire Home

R2 0.057



### Private Room

R2 0.058



Shared Room

R2 0.10

host_acceptance_rate	-0.016
review_scores_rating	-0.21
price	-0.18
review_scores_cleanliness	-0.11
availability_365	-0.21
review_scores_communication	-0.14

Al final no se puede sacar una conclusión precisa o correcta por la naturaleza de las correlaciones y los valores obtenidos de los coeficientes de determinación. Posiblemente haya técnicas al momento de tratar los valores nulos o outliers que hacen ruido dentro de los datos que puedan ayudar a mejorar los resultados.