



Tecnológico de Monterrey

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE
MONTERREY
CAMPUS PUEBLA

**Analítica de datos y herramientas de inteligencia
artificial I**

**Actividad 1
Módulo 3**

Reporte Comparativo
Regresión Lineal y Regresión Lineal Múltiple

Nombre Alumn(a):
Abril Lizeth Martínez Salas

Matrícula:
A01734613

Resultados

Para la construcción de los modelos de regresión lineal simple y los modelos de regresión lineal múltiple se toma como punto de comparación los modelos de predicción así como los resultados en cuanto al coeficiente de correlación r y el coeficiente de determinación (R^2) de los modelos para 3 tipos de habitaciones distintas (PR=habitación privada, FP= Habitación o departamento completo y SR=habitación compartida) en 3 ciudades de 3 países distintos. En este caso se consideró la Ciudad de México, Montreal y Barcelona.

En general, se considera que la variable de respuesta, *number_of_reviews* (cantidad de comentarios/evaluaciones) estaba débilmente correlacionada con las variables seleccionadas como predictoras dentro de los modelos de regresión lineal. Aunque se lograron identificar otro tipo de correlaciones relativamente altas entre algunas de las variables predictoras, al ser puestas como variables independientes para explicar el modelo de forma matemática del comportamiento de *number_of_reviews* (variable de respuesta), los resultados obtenidos para estas 3 ciudades con sus respectivas tres tipos de habitación fueron similares. De igual forma, tanto la ciudad de México, Montreal y Barcelona tenían un comportamiento similar para las habitaciones compartidas, ya que en los 3 casos tenían los coeficientes de correlación más alto y en algunos casos, los modelos de regresión lineal consideraron las mismas variables predictoras para construir dicho modelo. Para el caso de México, se recurrió a una regresión lineal simple, la cual tuvo los siguientes resultados:

MÉXICO

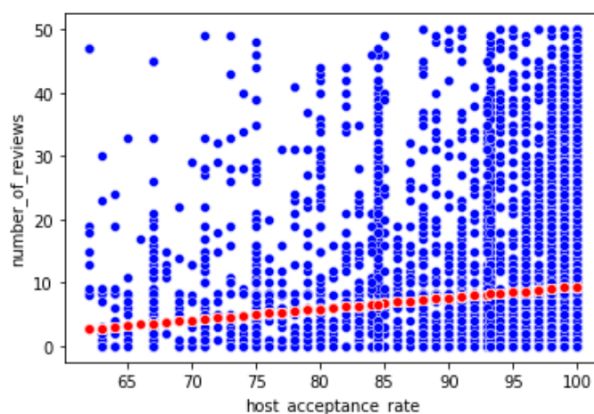
Tipo de Habitación 1. México Habitación Privada

```
1 corr=PR.corrwith(PR['number_of_reviews'])
2 corr
```

```
host_acceptance_rate    0.133364
review_scores_rating     0.112659
price                   -0.046266
review_scores_cleanliness 0.067370
availability_365        -0.090449
review_scores_communication 0.114794
number_of_reviews       1.000000
dtype: float64
```

```
1 #COEFICIENTE DE DETERMINACIÓN R^2
2 coef_deter=pow(corr,2)
3 coef_deter
```

```
host_acceptance_rate    0.017786
review_scores_rating     0.012692
price                   0.002140
review_scores_cleanliness 0.004539
availability_365        0.008181
review_scores_communication 0.013178
number_of_reviews       1.000000
dtype: float64
```



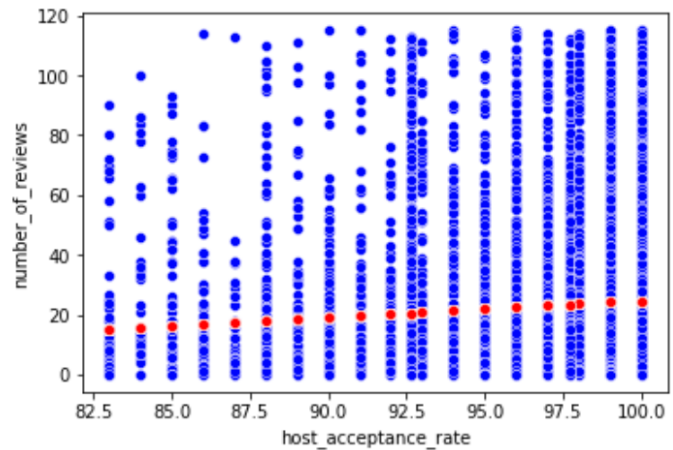
Tipo de Habitación 2. México Depto/Hab. completo

```
1 corr=FP.corrwith(FP['number_of_reviews'])
2 corr
```

```
host_acceptance_rate      0.073297
review_scores_rating      -0.077706
price                    -0.058850
review_scores_cleanliness -0.088539
availability_365          -0.072587
review_scores_communication -0.081994
number_of_reviews         1.000000
dtype: float64
```

```
1 #COEFICIENTE DE DETERMINACIÓN R^2
2 coef_deter=pow(corr,2)
3 coef_deter
```

```
host_acceptance_rate      0.005372
review_scores_rating      0.006038
price                    0.003463
review_scores_cleanliness 0.007839
availability_365          0.005269
review_scores_communication 0.006723
number_of_reviews         1.000000
dtype: float64
```



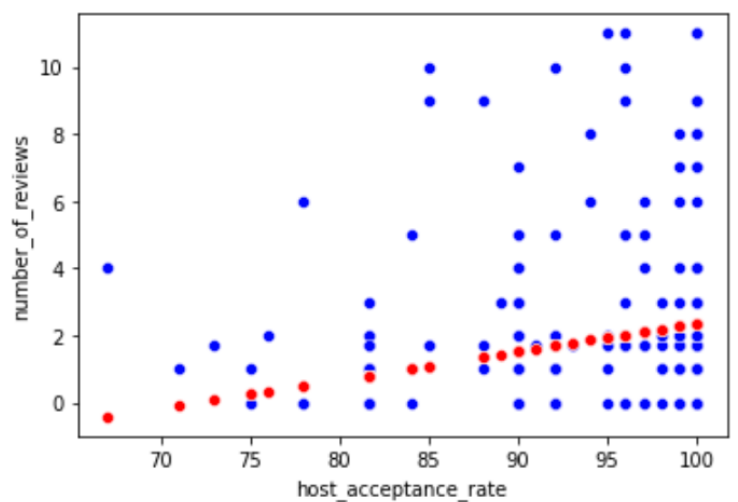
Tipo de Habitación 3. México Habitación Compartida

```
1 corr=SR.corrwith(SR['number_of_reviews'])
2 corr
```

```
host_acceptance_rate      0.264243
review_scores_rating      0.284745
price                    -0.174675
review_scores_cleanliness 0.301884
availability_365          -0.022350
review_scores_communication 0.299565
number_of_reviews         1.000000
dtype: float64
```

```
1 #COEFICIENTE DE DETERMINACIÓN R^2
2 coef_deter=pow(corr,2)
3 coef_deter
```

```
host_acceptance_rate      0.069824
review_scores_rating      0.081080
price                    0.030511
review_scores_cleanliness 0.091134
availability_365          0.000500
review_scores_communication 0.089739
number_of_reviews         1.000000
dtype: float64
```



Análisis

Se puede observar que el 3er tipo de habitación tenía la correlación más alta con respecto a la mayoría de las variables predictoras, para los 3 casos se consideró como comparativa la variable de *Host_acceptance_rate*. De igual forma, es observable que aunque se realizó la regresión lineal simple en este caso, el comportamiento de los valores reales es sumamente distinto, lo cual se puede ver al consultar el valor de R^2 del modelo.

MONTREAL

Para la base de datos de la lista de propiedades de Airbnb de Montreal se hizo un Análisis de Regresión Múltiple para poder encontrar una combinación de variables que permitiera incrementar el R^2 del modelo. Los valores numéricos obtenidos en el análisis de los 3 tipos de habitación indican que las correlaciones eran ligeramente más altas que aquellas obtenidas en el análisis de la Ciudad de México. De igual forma, el tercer tipo de habitación obtuvo el R^2 y los coeficientes de correlación más altos. Dado que en este caso se estaba tratando de un análisis de regresión lineal múltiple, se hicieron distintas corridas hasta obtener el número de residuales menor considerando las variables con la correlación más alta.

Tipo de Habitación 1. Montreal Habitación Privada

```
1 corr=PR.corrwith(PR['number_of_reviews'])
2 corr
```

```
host_acceptance_rate    0.169878
review_scores_rating     0.071857
price                   0.068198
review_scores_cleanliness 0.039665
availability_365        0.101026
review_scores_communication 0.030691
number_of_reviews       1.000000
dtype: float64
```

```
1 #COEFICIENTE DE DETERMINACIÓN R^2
2 coef_deter=pow(corr,2)
3 coef_deter
```

```
host_acceptance_rate    0.028858
review_scores_rating     0.005163
price                   0.004651
review_scores_cleanliness 0.001573
availability_365        0.010206
review_scores_communication 0.000942
number_of_reviews       1.000000
dtype: float64
```

```
1 y_pred= model.predict(X=PR[['host_acceptance_rate','review_scores_rating','availability_365']])
2 y_pred
array([6.15250162, 5.14831768, 5.63494879, ..., 6.88220061, 7.06158276,
       4.31115031])
```

```
1 #Coeficiente de Determinación del modelo
2 coef_Deter=model.score(X=Vars_Indep, y=Var_Dep)
3 coef_Deter
0.03862605471819569
```

```
1 coef_Correl=np.sqrt(coef_Deter)
2 coef_Correl
0.19653512337034235
```

Tipo de Habitación 2. Montreal Depto/Hab. completo

```
] 1 corr=FP.corrwith(FP['number_of_reviews'])
   2 corr
```

```
host_acceptance_rate      0.201476
review_scores_rating      -0.035192
price                     0.067977
review_scores_cleanliness -0.034166
availability_365          0.056843
review_scores_communication -0.076977
number_of_reviews         1.000000
dtype: float64
```

```
➤ 1 #COEFICIENTE DE DETERMINACIÓN R^2
   2 coef_deter=pow(corr,2)
   3 coef_deter
```

```
➤ host_acceptance_rate      0.040593
review_scores_rating      0.001238
price                     0.004621
review_scores_cleanliness 0.001167
availability_365          0.003231
review_scores_communication 0.005925
number_of_reviews         1.000000
dtype: float64
```

```
1 y_pred= model.predict(X=FP[['host_acceptance_rate','price','availability_365','review_scores_communication']])
2 y_pred
```

```
array([[14.52183341,  8.98010943, 15.67473492, ..., 13.066321 ,
        12.64174119,  8.13159415])
```

```
1 #Coeficiente de Determinación del modelo
2 coef_Deter=model.score(X=Vars_Indep, y=Var_Dep)
3 coef_Deter
```

```
0.047502834897851565
```

```
1 coef_Correl=np.sqrt(coef_Deter)
2 coef_Correl
```

```
0.21795145078170863
```

Tipo de Habitación 3. Montreal Habitación Compartida

```
1 corr=SR.corrwith(SR['number_of_reviews'])
2 corr
```

```
host_acceptance_rate      0.471754
review_scores_rating      -0.012890
price                     0.050458
review_scores_cleanliness -0.126576
availability_365           0.123311
review_scores_communication -0.037588
number_of_reviews         1.000000
dtype: float64
```

```
1 #COEFICIENTE DE DETERMINACIÓN R^2
2 coef_deter=pow(corr,2)
3 coef_deter
```

```
host_acceptance_rate      0.222552
review_scores_rating      0.000166
price                     0.002546
review_scores_cleanliness 0.016022
availability_365           0.015206
review_scores_communication 0.001413
number_of_reviews         1.000000
dtype: float64
```

```
1 y_pred= model.predict(X=SR[['host_acceptance_rate','review_scores_cleanliness','availability_365']])
2 y_pred
```

```
] 1 #Coeficiente de Determinación del modelo
2 coef_Deter=model.score(X=Vars_Indep, y=Var_Dep)
3 coef_Deter
```

```
0.24241574729261495
```

```
] 1 coef_Correl=np.sqrt(coef_Deter)
2 coef_Correl
```

```
0.49235733699480394
```

En términos generales, este último modelo obtuvo el coeficiente de correlación más alto de las 3 ciudades y de los 3 tipos de habitación. Se puede ver que el coeficiente de correlación está en .49 que es un equivalente a un 49%. Aunque el coeficiente de determinación es significativamente menor. Al igual que en México, en la mayoría de los modelos, el *host_acceptance_rate* fue una de las variables predictoras o independientes que se utilizaron para modelar casi todas las regresiones.

Barcelona

Al igual que en los modelos anteriores se les dió prioridad a aquellas variables que tenían la correlación más alta y se eligió aquella combinación que redujera la magnitud de los residuales. Al igual que en los primeros dos datasets, el comportamiento en cuanto a la parte estadística fue similar. De igual manera, el tercer tipo de habitación SR (compartida) junto con su regresión lineal múltiple, obtuvo el coeficiente de determinación y correlación más altos.

Tipo de Habitación 1. Barcelona Habitación Privada

```
1 corr=PR.corrwith(PR['number_of_reviews'])
2 corr
```

```
host_acceptance_rate      0.120518
review_scores_rating      0.043048
price                     0.047164
review_scores_cleanliness 0.044245
availability_365          -0.024027
review_scores_communication 0.005611
number_of_reviews         1.000000
dtype: float64
```

```
1 #COEFICIENTE DE DETERMINACIÓN R^2
2 coef_deter=pow(corr,2)
3 coef_deter
```

```
host_acceptance_rate      0.014525
review_scores_rating      0.001853
price                     0.002224
review_scores_cleanliness 0.001958
availability_365          0.000577
review_scores_communication 0.000031
number_of_reviews         1.000000
dtype: float64
```

```
1 y_pred= model.predict(X=PR[['host_acceptance_rate','review_scores_cleanliness','review_scores_rating']])
2 y_pred
```

```
array([[11.90386658,  9.6223611 ,  7.90322954, ...,  9.39295154,
        8.35054617,  7.20727248])
```

```
1 #Coeficiente de Determinación del modelo
2 coef_Deter=model.score(X=Vars_Indep, y=Var_Dep)
3 coef_Deter
```

```
0.01670896388457066
```

```
1 coef_Correl=np.sqrt(coef_Deter)
2 coef_Correl
```

```
0.12926315749110673
```

Tipo de Habitación 2. Montreal Depto/Hab. completo

```
1 corr=FP.corrwith(FP['number_of_reviews'])
2 corr
```

```
host_acceptance_rate      0.144971
review_scores_rating      0.011787
price                     0.245415
review_scores_cleanliness 0.055694
availability_365          0.023797
review_scores_communication -0.013044
number_of_reviews         1.000000
dtype: float64
```

```
1 #COEFICIENTE DE DETERMINACIÓN R^2
2 coef_deter=pow(corr,2)
3 coef_deter
```

```
host_acceptance_rate      0.021016
review_scores_rating      0.000139
price                     0.060229
review_scores_cleanliness 0.003102
availability_365          0.000566
review_scores_communication 0.000170
number_of_reviews         1.000000
dtype: float64
```

```
1 y_pred= model.predict(X=FP[['host_acceptance_rate','price','review_scores_cleanliness']])
2 y_pred
```

```
array([24.4972013 , 21.54953699, 25.49594771, ..., 28.84894016,
       18.68770521, 35.03517147])
```

```
1 #Coeficiente de Determinación del modelo
2 coef_Deter=model.score(X=Vars_Indep, y=Var_Dep)
3 coef_Deter
```

```
0.07781072940185796
```

```
1 coef_Correl=np.sqrt(coef_Deter)
2 coef_Correl
```

```
0.2789457463412159
```


Tipo de Habitación 3. Montreal Habitación Compartida

```
1 corr=SR.corrwith(SR['number_of_reviews'])
2 corr
```

```
host_acceptance_rate      -0.288342
review_scores_rating       -0.164162
price                     -0.102100
review_scores_cleanliness  0.087063
availability_365          -0.133140
review_scores_communication -0.098495
number_of_reviews         1.000000
dtype: float64
```

```
1 #COEFICIENTE DE DETERMINACIÓN R^2
2 coef_deter=pow(corr,2)
3 coef_deter
```

```
host_acceptance_rate      0.083141
review_scores_rating       0.026949
price                     0.010424
review_scores_cleanliness  0.007580
availability_365          0.017726
review_scores_communication 0.009701
number_of_reviews         1.000000
dtype: float64
```

```
1 y_pred= model.predict(X=SR[['host_acceptance_rate','review_scores_rating','price','availability_365']])
2 y_pred
```

```
1 #Coeficiente de Determinación del modelo
2 coef_Deter=model.score(X=Vars_Indep, y=Var_Dep)
3 coef_Deter
```

```
0.10082949987331713
```

```
1 coef_Correl=np.sqrt(coef_Deter)
2 coef_Correl
```

```
0.31753661186281673
```