



# Tecnológico de Monterrey

Analítica de datos y herramientas de inteligencia artificial

## **Actividad 6.1 (Regresión lineal simple)**

José Ignacio Hernández Rodríguez A01703130

Tecnológico de Monterrey, Campus Querétaro

5/05/2023

El presente reporte aterriza en las conclusiones obtenidas de la comparación de los resultados obtenidos de los cálculos realizados de la regresión lineal de las siguientes variables: “host\_acceptance\_rate vs host\_response\_rate”, “host\_acceptance\_rate vs price”, “host\_acceptance\_rate vs number\_of\_reviews”, “review\_scores\_location vs review\_scores\_cleanness”, “availability\_365 vs number\_of reviews”, “reviews\_per\_month vs review\_scores\_communication”. Los resultados a comparar son los coeficientes de determinación y correlación, así como también los modelos matemáticos. El análisis de los resultados es sobre las bases de datos del DF y California.

### Modelos matemáticos

Hotel Room		
Modelo matemático	DF	California
host_acceptance_rate vs host_response_rate	$y = 0.67835143x + 0.26812711148753066$	$y = 1.02523785x + 0.0381930481684023$
host_acceptance_rate vs price	$y = 0.62852689x + 0.3277244893054201$	$y = 0.55689264x - 1.4832788258780762$
host_acceptance_rate vs number_of_reviews	$y = -1.47290562e-05x + 0.9112905846283037$	$y = 1.37986599e-05x + 0.8094552245787601$
review_scores_location vs review_scores_cleanness	$y = -4.66947697e-07x + 0.8743659535692$	$y = -2.29750443e-05x + 0.7918208554587263$
availability_365 vs number_of reviews	$y = 0.00132107x + 0.8466371558698922$	$y = 0.00046365x + 0.7962857128180519$
reviews_per_month vs review_scores_communication	$y = 0.00057995x + 0.8485155168835032$	$y = 0.00064089x + 0.757769852849461$
Entire home / apt		
host_acceptance_rate vs host_response_rate	$y = 0.71790653x + 1.41489079217234$	$y = -0.4289354x + 6.51301164996563$

host_acceptance_rate vs price	$y = 0.43967956x + 2.749463839514829$	$y = 0.3234047x + 3.2520123203186584$
host_acceptance_rate vs number_of_reviews	$y = -0.35915992x + 254.19984142300686$	$y = 0.10665693x + 167.77747368533454$
review_scores_location vs review_scores_cleanliness	$y = -0.05945972x + 243.2786788430396$	$y = -0.13598232x + 171.7788273097436$
availability_365 vs number_of_reviews	$y = 0.37722635x + 0.05123634291469825$	$y = 0.55689264x + -1.4832788258780762$
reviews_per_month vs review_scores_communication	$y = 0.37722635x + 0.05123634291469825$	$y = 20.03799278x + -1.4832788258780762$

Tomando en cuenta los resultados de los modelos matemáticos de regresión de las 6 variables para ambos tipos de habitación, se puede identificar que los modelos de regresión de las variables del conjunto de datos de California cuentan con valores de intercepción en el eje de las y menores a comparación de los modelos del DF. Por otra parte, los resultados de x en su mayoría de los modelos, son mayores en los modelos del conjunto de datos de California.

### **Coeficientes**

Hotel Room		
Modelo matemático	DF	California
host_acceptance_rate vs host_response_rate	D: 0.581784017666669 C: 0.7627476762774626	D: 0.951897521440099 C: 0.9756523568567336
host_acceptance_rate vs price	D: 0.4797559816175936 C: 0.6926441955416891	D: 0.011944222106396452 C: 0.10928962487993292

host_acceptance_rate vs number_of_reviews	D: 0.12329092094701422 C: 0.351128069152858	D: 0.009216635348141544 C: 0.09600330904787367
review_scores_location vs review_scores_cleanness	D: 0.0005551002247745629 C: 0.02356056503512942	D: 0.0014117967321803926 C: 0.03757388364516493
availability_365 vs number_of reviews	D: 0.04618901087177962 C: 0.2149162880560234	D: 0.027736171539034737 C: 0.16654180117626546
reviews_per_month vs review_scores_communication	D: 0.017284601653650422 C: 0.1314709156188182	D: 0.03110234127819611 C: 0.1763585588458811
Entire home / apt		
host_acceptance_rate vs host_response_rate	D: 0.5019476993591685 C: 0.7084826740006903	D: 0.05515867016996512 C: 0.23485883030017227
host_acceptance_rate vs price	D: 0.3114627519253965 C: 0.5580884803733155	D: 0.10139448343468849 C: 0.31842500441185284
host_acceptance_rate vs number_of_reviews	D: 0.014844412000572182 C: 0.014844412000572182	D: 0.01122457035412916 C: 0.10594607285845549
review_scores_location vs review_scores_cleanness	D: 0.0010078697650548474 C: 0.03174696465892208	D: 0.009459283016082787 C: 0.09725884543877121
availability_365 vs number_of reviews	D:0.00543470686489455 6 C: 0.07372046435620544	D:0.01194422210639645 2 C: 0.10928962487993292
reviews_per_month vs review_scores_communication	D: 0.005434706864894556 C: 0.07372046435620544	D: 0.011944222106396452 C: 0.10928962487993292

**Tomando en cuenta los resultados de los coeficientes de determinación y correlación de** las 6 variables para ambos tipos de habitación, se puede identificar que los coeficientes de determinación, que son aquellos que hacen referencia a la predicción de modelos futuros, son mayores en su mayoría en la base de datos de DF. Esto indica que los modelos de DF cuentan con una mejor capacidad de realizar predicciones. Mientras que los coeficientes de correlación son superiores en la base de datos de California. Esto indica que la correlación entre variables en la base de datos del DF es muy baja. Mientras que en la base de datos de California existe una mayor correlación entre variables. Sin embargo, esta correlación no se puede considerar muy significativa ya que los valores no se aproximan a 1.