



Benemérita Universidad Autónoma de Puebla

*Facultad de Ciencias de la Computación*

## INGENIERIA DE SOFTWARE

### Regresión Logística



Aguilar Gasca Erasmo

Corona Valerio Gabriel

García Zarate Ángel Alfonso



## Introducción:

La regresión logística es un procedimiento básico de la estadística que permite explorar la relación entre una variable dependiente categórica y una o más variables independientes. Su principal función es predecir la probabilidad de que un evento ocurra o no, por ejemplo, si una condición está presente o no. Por consiguiente, se utiliza en distintos campos como medicina, economía, ciencias sociales, entre otros. El modelo de regresión logístico generalmente se aplica cuando suponemos una relación entre variables y se requiere clasificar datos en solo dos grupos. La capacidad de interpretar las relaciones y la medida esencial medida de impacto acciones de prevención adecuadas.

## Desarrollo del análisis:

Lugares a trabajar	Cantidad de registros
<i>Ciudad de Mexico</i>	10523
<i>Euskadi</i>	6891
<i>Los Angeles</i>	45533
<i>Dallas</i>	5307

## Variables dicotómicas a utilizar:

Variable	Valores
<i>host_is_superuser</i>	['f' 't']
<i>instant_bookable</i>	['f' 't']
<i>host_identity_verified</i>	['t' 'f']

## Conversión de variables a dicotómicas

<i>Variable</i>	<i>Valores Originales</i>	<i>Valores Nuevos</i>
<i>room_type</i>	'Entire home/apt', 'Hotel room', 'Private room', 'Shared room'	['Entire home/apt' 'No Entire home/apt']
<i>bathrooms</i>	['1' '2' '3' '4' '5' '6' '7'] *La cantidad varia según el pais	['≤1 baño' '>1 baño']

## Análisis Realizados:

### 1er Analisis:

**Variable dependiente**  
host\_is\_superhost

**Variable(s) Independiente(s)**  
price

## Resultados obtenidos:

<b>Pais</b>	<b>Presicion del modelo</b>	<b>Exactitud del modelo</b>	<b>Sensibilidad del modelo</b>
<b>CDMX</b>	0.6220779220779221	0.6195856873822976	0.0
<b>Euskadi</b>	0.6257686676427525	0.6257686676427525	0.0
<b>Los Angeles</b>	0.6329787234042553	0.6329787234042553	0.0
<b>Dallas</b>	0.660437123851758	0.660437123851758	0.04746317512274959

## 2do Analisis:

### Variable dependiente

instant\_bookable

### Variable(s) Independiente(s)

'review\_scores\_communication',  
'review\_scores\_checkin'

## Resultados obtenidos:

Pais	Presicion del modelo	Exactitud del modelo	Sensibilidad del modelo
<b>CDMX</b>	0.27154046997389036	0.6046879949318974	0.8661228406909789
<b>Euskadi</b>	0.49633699633699635	0.6102514506769826	0.782780410742496
<b>Los Angeles</b>	0.2550335570469799	0.6728404099560761	0.8576923076923076
<b>Dallas</b>	0.5648994515539305	0.6534839924670434	0.7546391752577319

## 3er Analisis:

### Variable dependiente

instant\_bookable

### Variable(s) Independiente(s)

'review\_scores\_rating', 'price',  
'host\_acceptance\_rate'

## Resultados obtenidos:

Pais	Presicion del modelo	Exactitud del modelo	Sensibilidad del modelo
<b>CDMX</b>	0.7047434656340755	0.6997149192271144	0.026595744680851064
<b>Euskadi</b>	0.5534114403859407	0.5425531914893617	0.3298862461220269
<b>Los Angeles</b>	0.5909090909090909	0.7431918008784774	0.9973372781065089
<b>Dallas</b>	0.5223880597014925	0.6064030131826742	0.9333333333333333

#### 4to Analisis:

##### Variable dependiente

instant\_bookable

##### Variable(s) Independiente(s)

'number\_of\_reviews', 'price'

#### Resultados obtenidos:

Pais	Presicion del modelo	Exactitud del modelo	Sensibilidad del modelo
CDMX	0.7047434656340755	0.6046879949318974	0.8661228406909789
Euskadi	0.5534114403859407	0.6102514506769826	0.782780410742496
Los Angeles	0.7463396778916545	0.6728404099560761	0.8576923076923076
Dallas	0.6001255492780917	0.6534839924670434	0.7546391752577319

#### 5to Analisis:

##### Variable dependiente

host\_identity\_verified

##### Variable(s) Independiente(s)

'number\_of\_reviews', 'price',  
'host\_total\_listings\_count'

#### Resultados obtenidos:

Pais	Presicion del modelo	Exactitud del modelo	Sensibilidad del modelo
CDMX	0.9322141273360786	0.9322141273360786	0.0
Euskadi	0.9782398452611218	0.9782398452611218	0.0
Los Angeles	0.8927525622254758	0.8927525622254758	0.0
Dallas	0.9221594475831764	0.9221594475831764	0.0

## 6to Analisis:

### Variable dependiente

host\_identity\_verified

### Variable(s) Independiente(s)

'host\_acceptance\_rate',  
'review\_scores\_rating'

## Resultados obtenidos:

Pais	Presicion del modelo	Exactitud del modelo	Sensibilidad del modelo
<b>CDMX</b>	0.9306303452644916	0.9306303452644916	0.0
<b>Euskadi</b>	0.9758220502901354	0.9758220502901354	0.0
<b>Los Angeles</b>	0.8943631039531479	0.8943631039531479	0.0
<b>Dallas</b>	0.9234149403640929	0.9234149403640929	0.0

## 7mo Analisis:

### Variable dependiente

room\_type

### Variable(s) Independiente(s)

'accommodates',  
'host\_acceptance\_rate', 'price'

## Resultados obtenidos:

Pais	Presicion del modelo	Exactitud del modelo	Sensibilidad del modelo
<b>CDMX</b>	0.7591836734693878	0.7994931897370922	0.8438327157226045
<b>Euskadi</b>	0.7572463768115942	0.8583172147001934	0.9101274312541918
<b>Los Angeles</b>	0.7198492462311558	0.813103953147877	0.9334791687381923
<b>Dallas</b>	0.96	0.9001883239171374	0.9978021978021978

## 8vo Analisis:

### Variable dependiente

room\_type

### Variable(s) Independiente(s)

'number\_of\_reviews', 'bedrooms'

## Resultados obtenidos:

Pais	Presicion del modelo	Exactitud del modelo	Sensibilidad del modelo
<b>CDMX</b>	0.6649789029535865	0.7329743427304403	0.7935517420696828
<b>Euskadi</b>	0.6164383561643836	0.7770793036750484	0.8266850068775791
<b>Los Angeles</b>	0.08560311284046693	0.7241581259150806	0.9767441860465116
<b>Dallas</b>	.7994931897370922	0.8681732580037664	0.9942487419122933

## 9no Analisis:

### Variable dependiente

bathrooms\_text

### Variable(s) Independiente(s)

'number\_of\_reviews', 'price',  
'accommodates'

## Resultados obtenidos:

Pais	Presicion del modelo	Exactitud del modelo	Sensibilidad del modelo
<b>CDMX</b>	0.6799410029498525	0.7272727272727273	0.894249512670565
<b>Euskadi</b>	0.7495741056218058	0.7558027079303675	0.8842519685039371
<b>Los Angeles</b>	0.8349318525996972	0.8233528550512446	0.9238915396252764
<b>Dallas</b>	0.8324873096446701	0.8305084745762712	0.8935483870967742

## 10mo Analisis:

### Variable dependiente

bathrooms\_text

### Variable(s) Independiente(s)

'host\_acceptance\_rate',

'review\_scores\_rating'

## Resultados obtenidos:

Pais	Presicion del modelo	Exactitud del modelo	Sensibilidad del modelo
<b>CDMX</b>	0.6566360468799494	0.6566360468799494	0.0
<b>Euskadi</b>	0.6058994197292069	0.6058994197292069	0.0
<b>Los Angeles</b>	0.6168374816983895	0.6168374816983895	0.0
<b>Dallas</b>	0.583804143126177	0.583804143126177	0.8935483870967742

## Conclusión:

El análisis de regresión logística realizado nos dio la oportunidad de evaluar qué variables afectan ciertas características de alojamiento en diferentes ciudades y en qué medida. Los modelos sondeados varían en precisión, exactitud y sensibilidad, porque los datos son heterogéneos y las relaciones entre las variables estudiadas son más complejas de lo esperado. En particular, encontramos que hay ciertas variables que predicho con más eficiencia que otras. Esto recibe el asunto bajo análisis directo y puede ser tomado en cuenta para análisis y futuros estudios.