



# Tecnológico de Monterrey

**Objetivos del negocio y evaluación de la situación actual**

---

**Escuela:** Instituto Tecnológico de Estudios Superiores de Monterrey

**Curso:** Analítica de datos y herramientas de inteligencia artificial II (Gpo 501)

**Nivel académico:** Profesional

**Ciudad:** Puebla

**Autores:**

Maximiliano Espinoza Velez	A01733613
Evelyn Martínez Juárez	A01733906
David Alberto Alvarado Cabrero	A01736390
Roberto Brito Mirando	A01424631

**Fecha de entrega:**

30 de Septiembre de 2024

## Actividad 3.5: Regresión Lineal Simple y Múltiple

Para esta actividad, se realizó el análisis de los datos de las 3 bases de datos de nuestras ciudades utilizando la Regresión Lineal Simple y Múltiple para establecer correlaciones entre las variables de la base de datos. A continuación, se muestran los resultados obtenidos y la comparación entre cada una de las ciudades.

### Regresión Lineal Simple

Para la regresión lineal simple, se dividió la base de datos en 4 sub-bases, cada una perteneciente a un tipo de habitación en particular: Entire Room, Private Room, Shared Room y Hotel Room.

Para posteriormente realizar una regresión lineal simple en base a las siguientes combinaciones: “host\_acceptance\_rate vs host\_response\_rate”, “review\_scores\_location vs review\_scores\_cleanliness”, “host\_acceptance\_rate vs price”, “availability\_365 vs number\_of\_reviews”, “host\_acceptance\_rate vs number\_of\_reviews”, “reviews\_per\_month vs review\_scores\_communication”

A continuación, se muestran entre las 3 ciudades, cuál tenía la mejor correlación con cada una de las 6 combinaciones en los 4 tipos de habitación.

### Cantidad de Registros por Tipo de Habitación

Entire Room: México - 17432, Geneva - 1901, Boston - 2950

Private Room: México - 8581, Geneva - 769, Boston - 1313

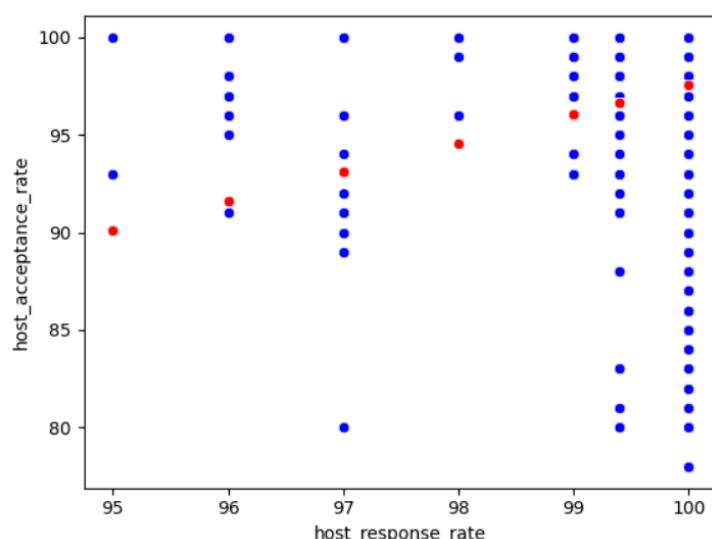
Shared Room: México - 401, Geneva - 13, Boston - 19

Hotel Room: México - 122, Geneva - 2, Boston - 43

#### Entire Room

##### *host\_acceptance\_rate vs host\_response\_rate*

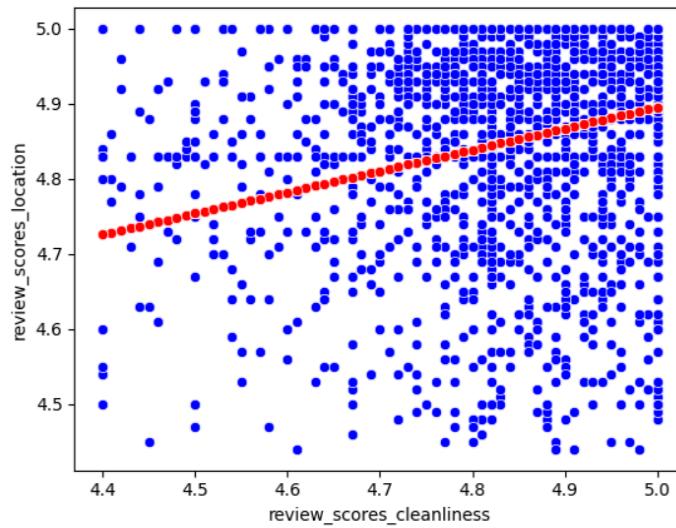
Boston



Coeficiente de Correlación: 0.396353

***review\_scores\_location vs review\_scores\_cleanliness***

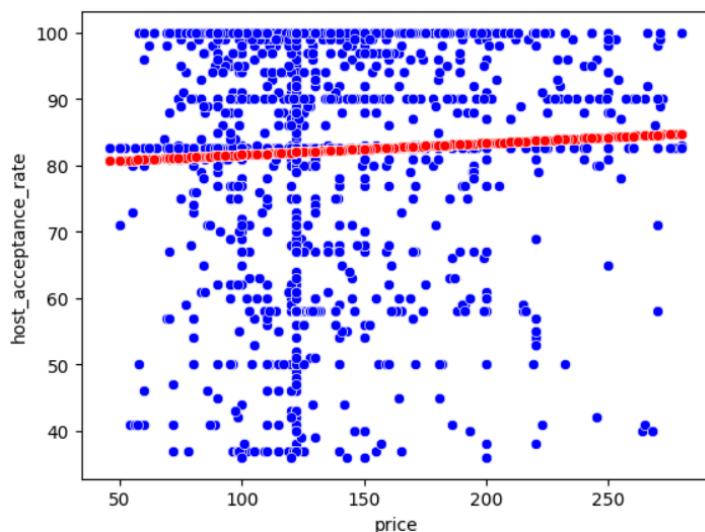
Boston



Coeficiente de Correlación: 0.294000

***host\_acceptance\_rate vs price***

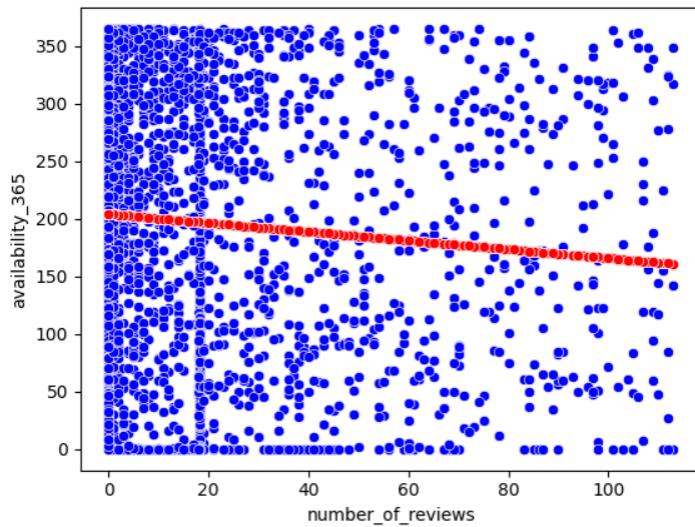
Geneva



Coeficiente de Correlación: 0.042379

***availability\_365 vs number\_of\_reviews***

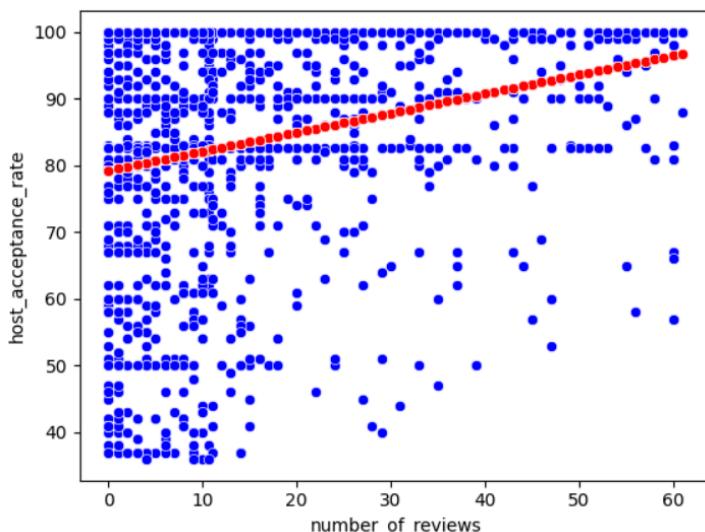
Boston



Coeficiente de Correlación: 0.077874

***host\_acceptance\_rate vs number\_of\_reviews***

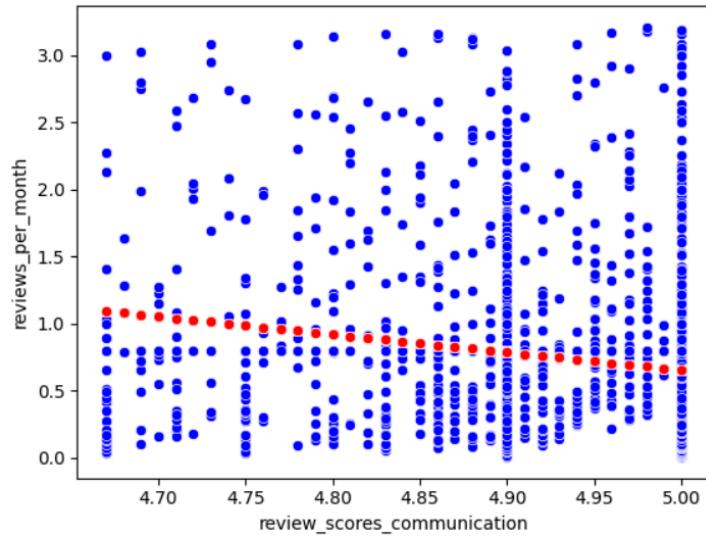
Geneva



Coeficiente de Correlación: 0.228364

***reviews\_per\_month vs review\_scores\_communication***

Geneva

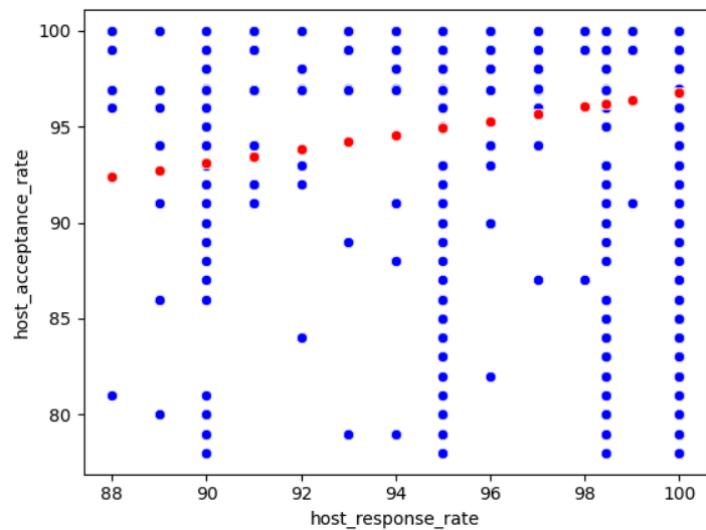


Coeficiente de Correlación: 0.164970

### ***Private Room***

#### ***host\_acceptance\_rate vs host\_response\_rate***

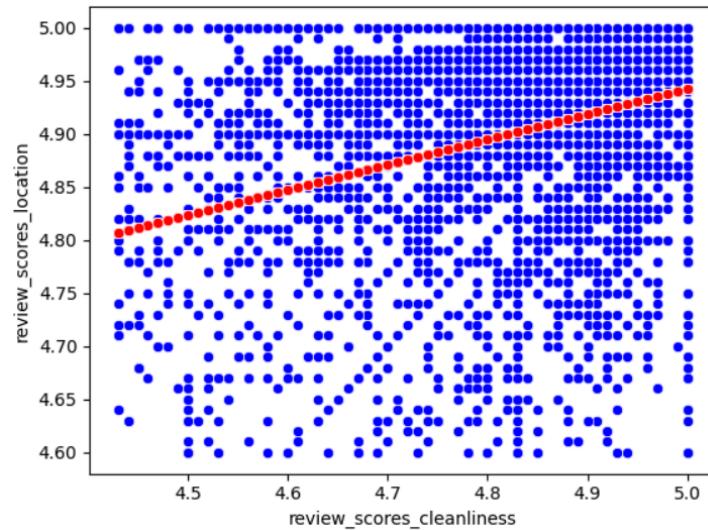
México



Coeficiente de Correlación: 0.219742

#### ***review\_scores\_location vs review\_scores\_cleanliness***

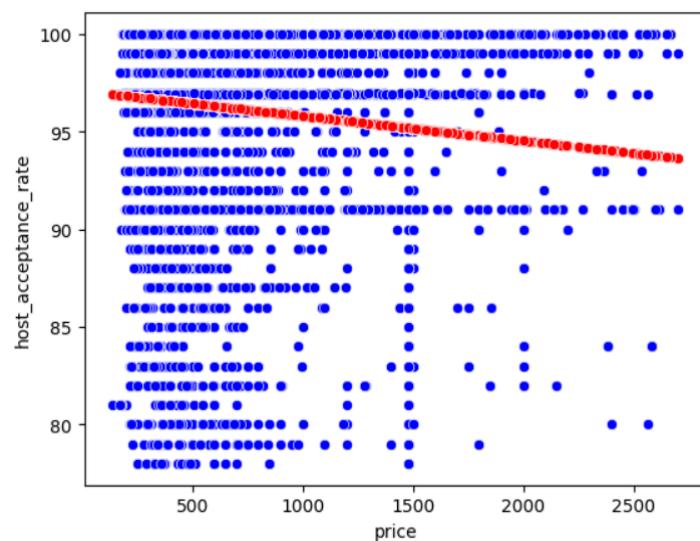
México



Coeficiente de Correlación: 0.353596

#### *host\_acceptance\_rate vs price*

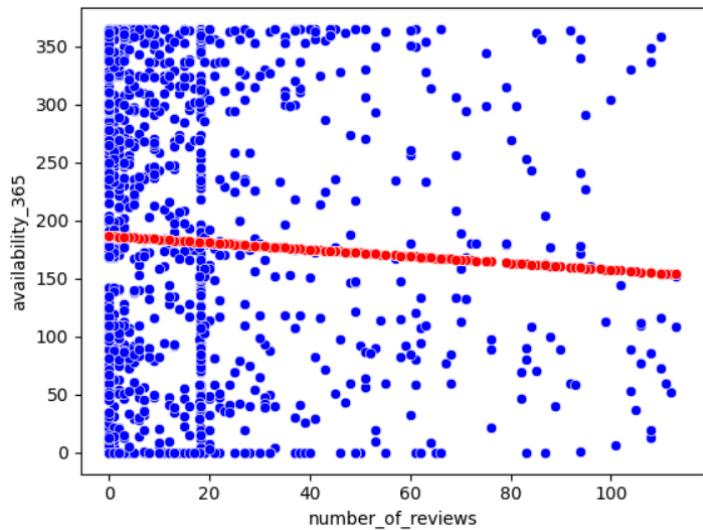
México



Coeficiente de Correlación: 0.130934

#### *availability\_365 vs number\_of\_reviews*

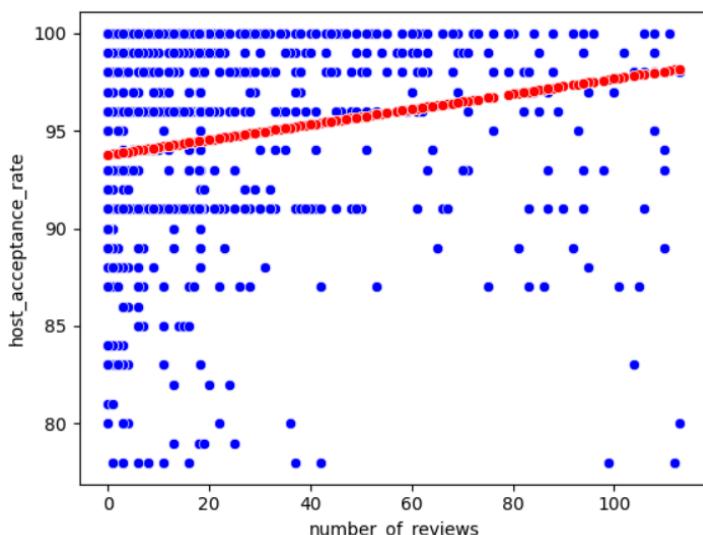
Boston



Coeficiente de Correlación: 0.052596

#### *host\_acceptance\_rate vs number\_of\_reviews*

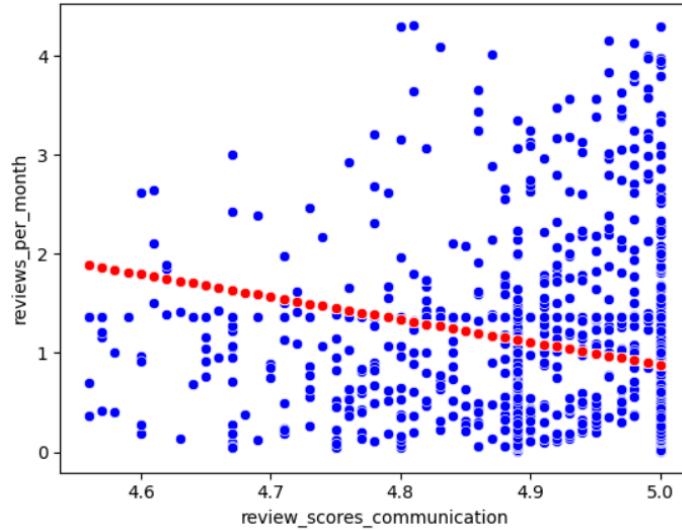
Boston



Coeficiente de Correlación: 0.184835

#### *reviews\_per\_month vs review\_scores\_communication*

Boston

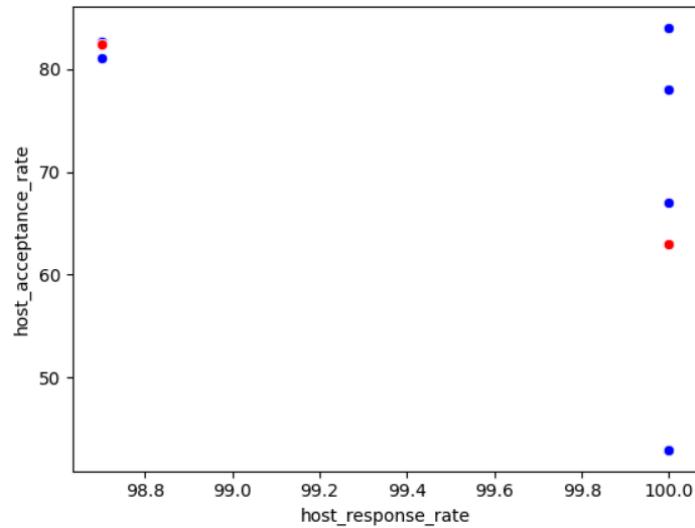


Coeficiente de Correlación: 0.251122

### ***Shared Room***

#### ***host\_acceptance\_rate vs host\_response\_rate***

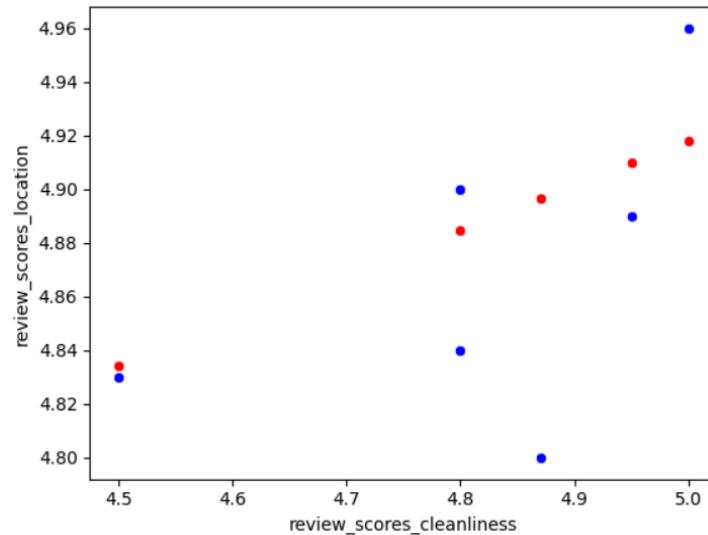
Geneva



Coeficiente de Correlación: 0.662020

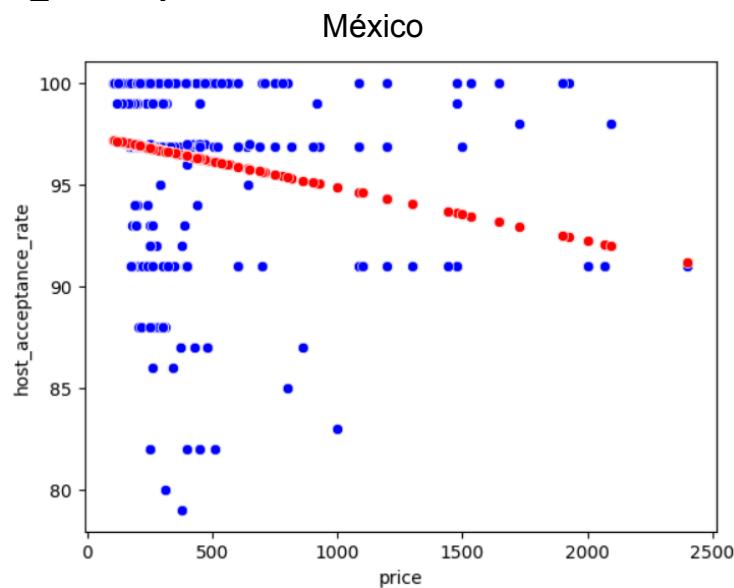
#### ***review\_scores\_location vs review\_scores\_cleanliness***

Geneva



Coeficiente de Correlación: 0.471386

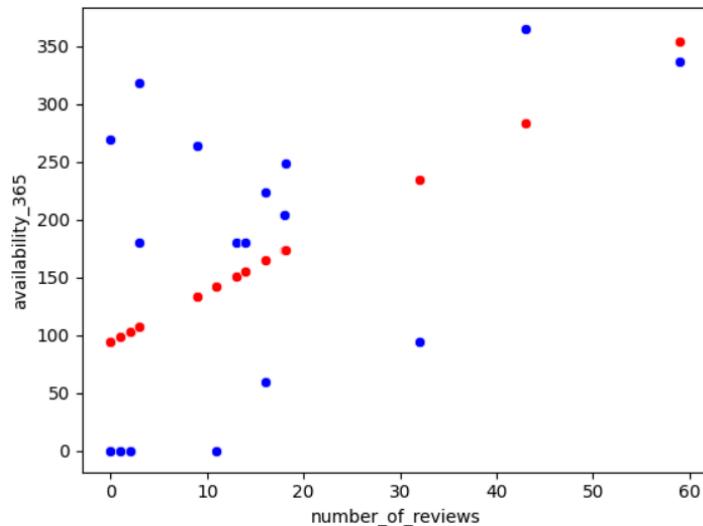
#### *host\_acceptance\_rate vs price*



Coeficiente de Correlación: 0.267020

#### *availability\_365 vs number\_of\_reviews*

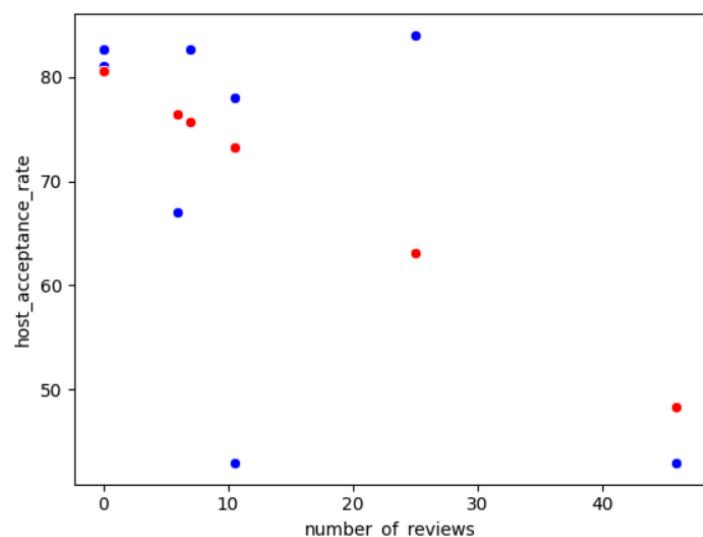
Boston



Coeficiente de Correlación: 0.542943

#### ***host\_acceptance\_rate vs number\_of\_reviews***

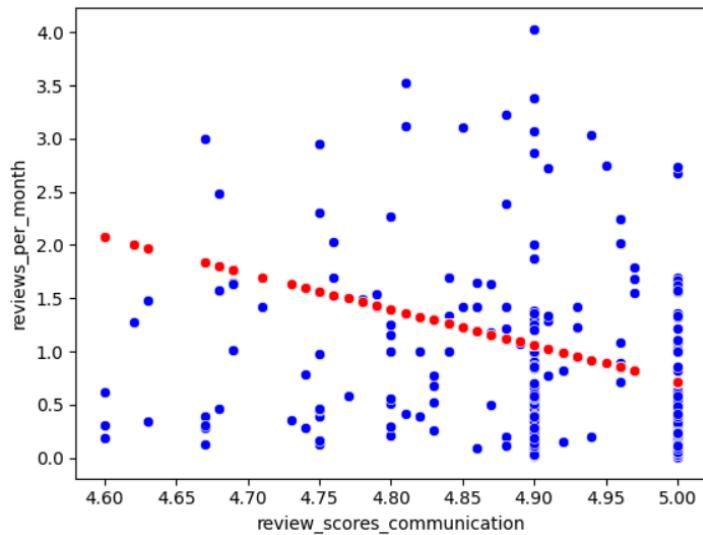
Geneva



Coeficiente de Correlación: 0.639952

#### ***reviews\_per\_month vs review\_scores\_communication***

México

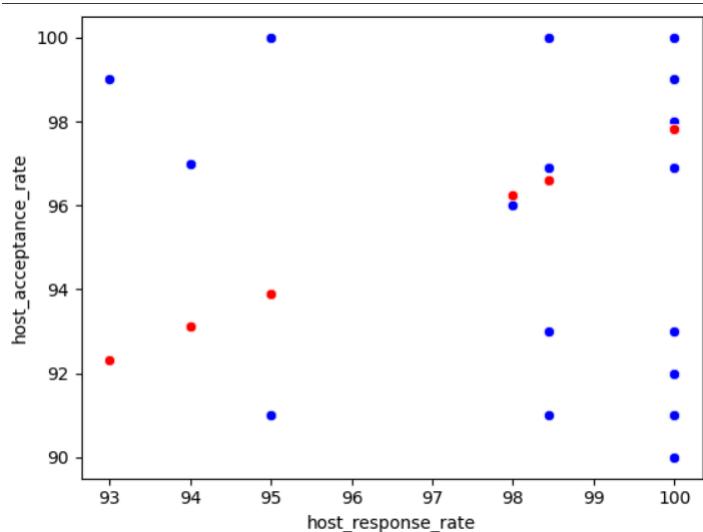


Coeficiente de Correlación: 0.384541

### ***Hotel Room***

#### ***host\_acceptance\_rate vs host\_response\_rate***

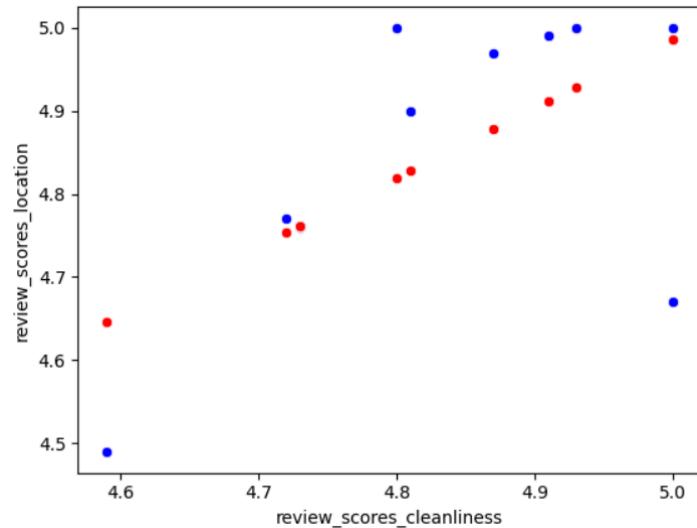
México



Coeficiente de Correlación: 0.414967

#### ***review\_scores\_location vs review\_scores\_cleanliness***

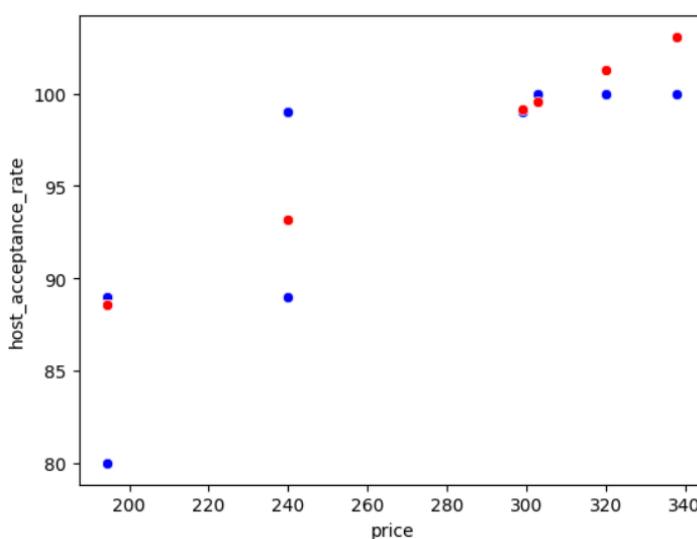
Boston



Coeficiente de Correlación: 0.692168

#### ***host\_acceptance\_rate vs price***

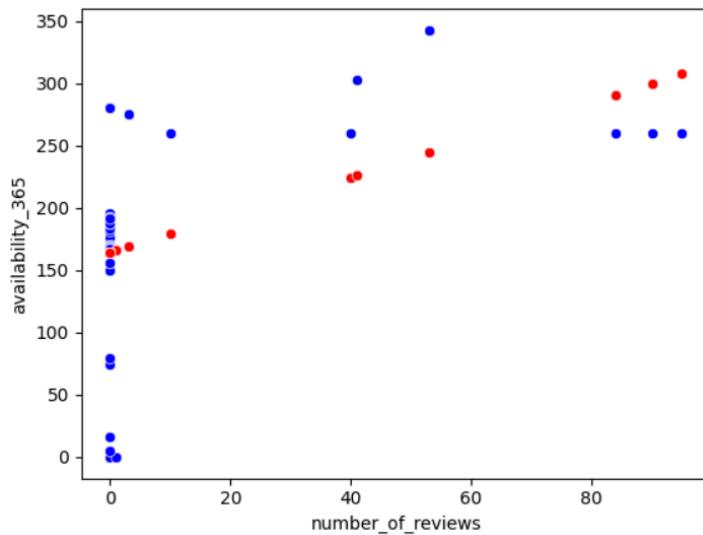
Boston



Coeficiente de Correlación: 0.820757

#### ***availability\_365 vs number\_of\_reviews***

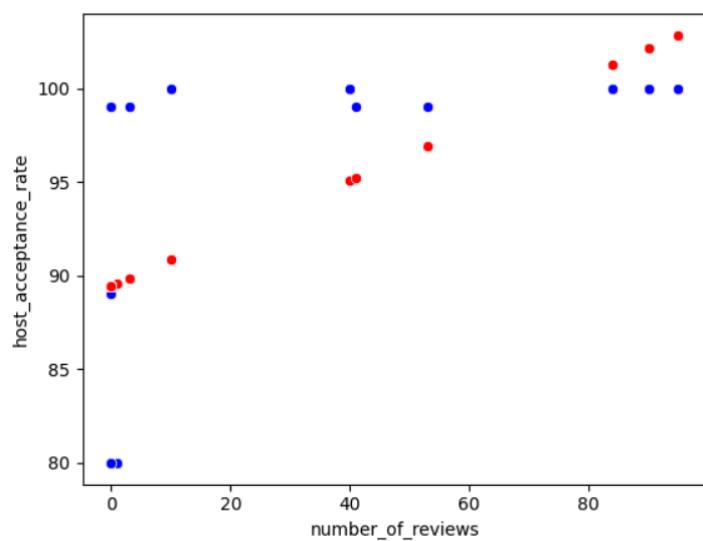
Boston



Coeficiente de Correlación: 0.498864

***host\_acceptance\_rate vs number\_of\_reviews***

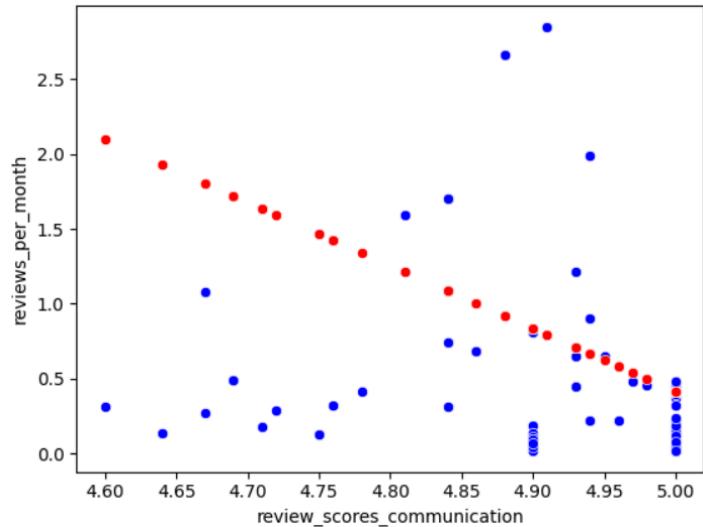
Boston



Coeficiente de Correlación: 0.658785

***reviews\_per\_month vs review\_scores\_communication***

México



Coeficiente de Correlación: 0.489270

## Regresión Lineal Múltiple

Para la realización de la regresión lineal múltiple a las bases de datos, se aplicó un ajuste estableciendo las variables pertinentes que puedan crear el mejor modelo con la mejor correlación para las siguientes variables: instant\_bookable, host\_acceptance\_rate, host\_is\_superhost, host\_total\_listings\_count, room\_type, accommodates, bedrooms, Price, review\_scores\_value y reviews\_per\_month.

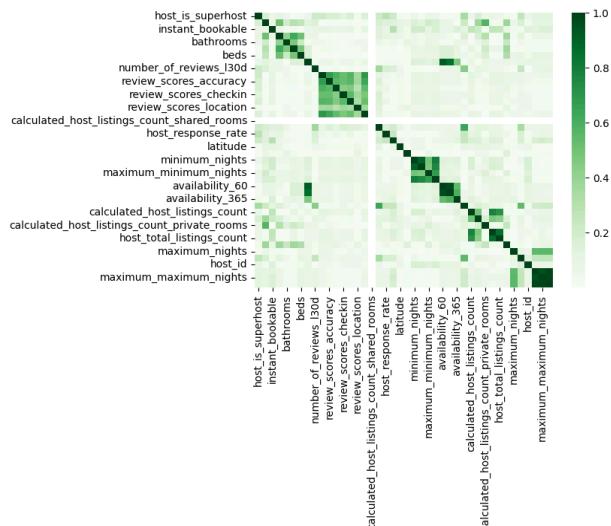
De esta manera, obtenemos 10 modelos en cada ciudad, para de esta manera comparar las regresiones realizadas y establecer la ciudad con el mejor modelo.

## Mapas de Calor

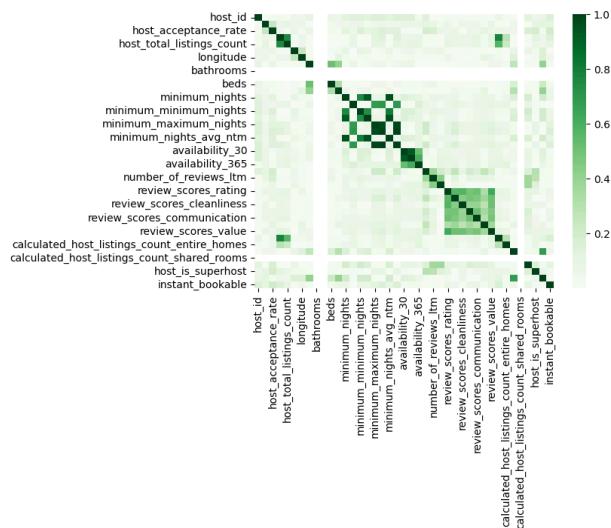
Para poder medir correctamente las mejores correlaciones posibles para cada variable objetivo, es oportuno utilizar el Mapa de Calor, mediante el cuál podemos visualizar fácilmente aquellas variables que poseen mejor correlación con otras.

Al final de esta sección deseamos observar las diferencias entre cada modelo creado para cada ciudad, tanto en las variables independientes utilizadas como también en su correlación resultante.

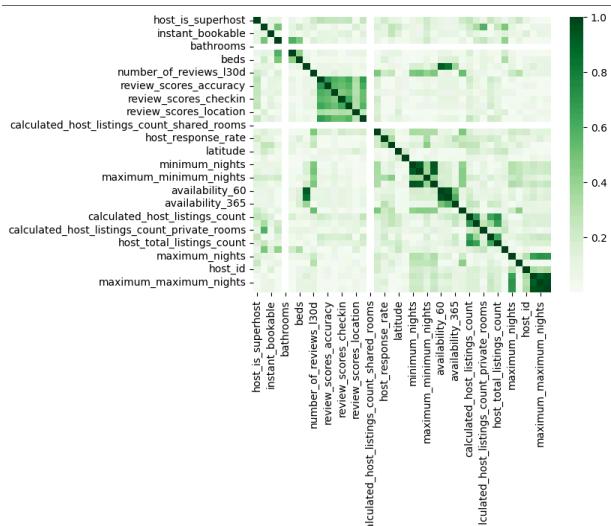
## México



## Geneva



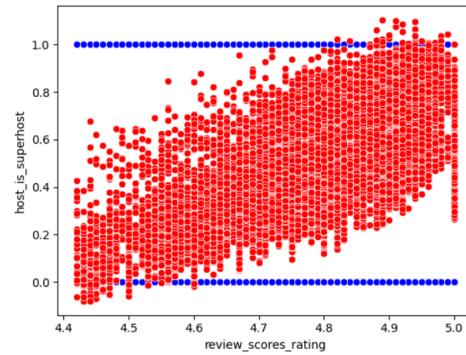
## Boston



## **Host is Superhost**

### **México**

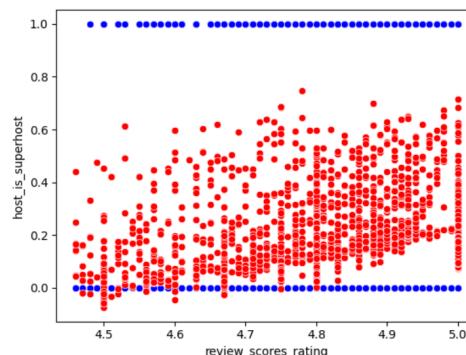
Variables Independientes: 'review\_scores\_rating', 'review\_scores\_communication', 'review\_scores\_value', 'calculated\_host\_listings\_count', 'host\_listings\_count', 'number\_of\_reviews'



Coeficiente de Correlación: 0.3721571777106302

### **Geneva**

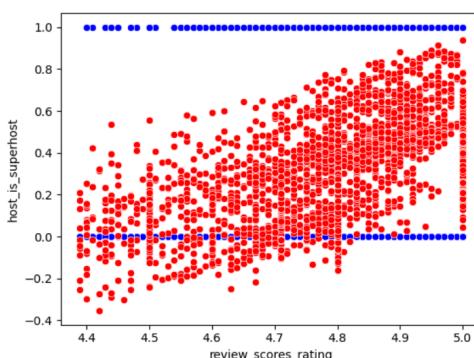
Variables Independientes: 'review\_scores\_rating', 'review\_scores\_communication', 'review\_scores\_value', 'calculated\_host\_listings\_count', 'host\_listings\_count', 'number\_of\_reviews'



Coeficiente de Correlación: 0.3624944115795811

### **Boston**

Variables Independientes: 'review\_scores\_rating', 'review\_scores\_communication', 'review\_scores\_value', 'calculated\_host\_listings\_count', 'host\_listings\_count', 'number\_of\_reviews'



Coeficiente de Correlación: 0.4440036194079203

## **Room Type**

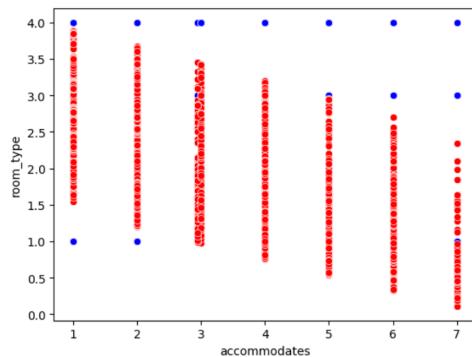
### **México**

Variables

Independientes:

'accommodates',

'calculated\_host\_listings\_count\_private\_rooms', 'price'



Coeficiente de Correlación: 0.6846673216026932

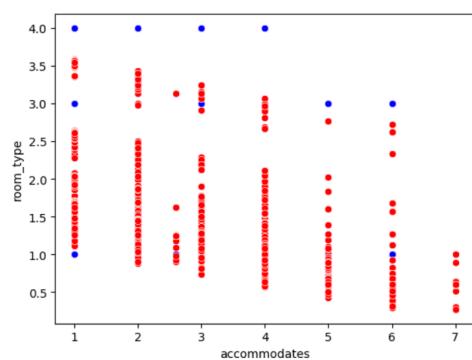
### **Geneva**

Variables

Independientes:

'accommodates',

'calculated\_host\_listings\_count\_private\_rooms', 'price'



Coeficiente de Correlación: 0.7331526817139237

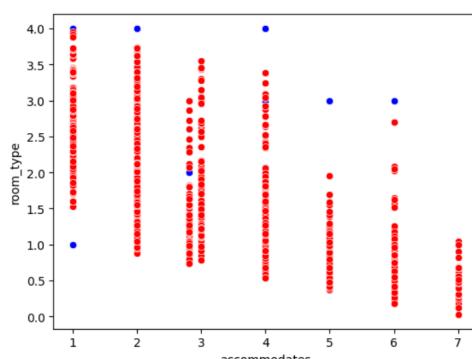
### **Boston**

Variables

Independientes:

'accommodates',

'calculated\_host\_listings\_count\_private\_rooms', 'price'

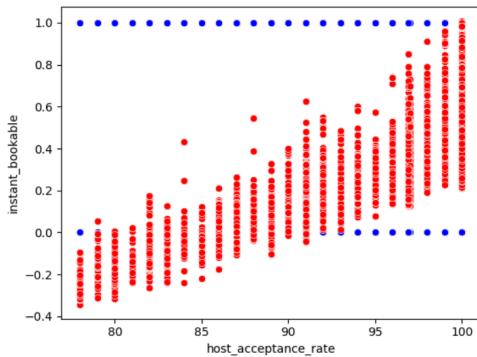


Coeficiente de Correlación: 0.7224072278730591

## ***Instant Bookable***

### **México**

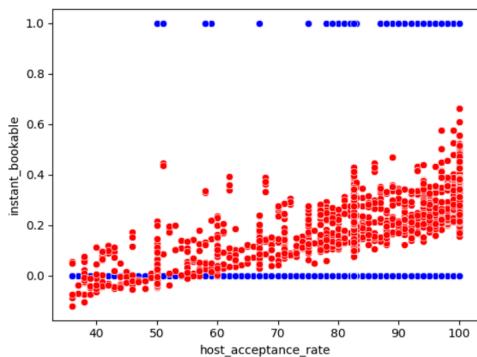
Variables Independientes: 'host\_acceptance\_rate', 'maximum\_minimum\_nights', 'calculated\_host\_listings\_count', 'host\_id'



Coeficiente de Correlación: 0.40035330831425586

### **Geneva**

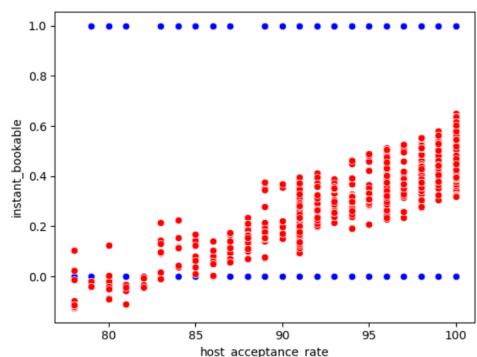
Variables Independientes: 'host\_acceptance\_rate', 'maximum\_minimum\_nights', 'calculated\_host\_listings\_count', 'host\_id'



Coeficiente de Correlación: 0.2943111755174787

### **Boston**

Variables Independientes: 'host\_acceptance\_rate', 'maximum\_minimum\_nights', 'calculated\_host\_listings\_count', 'host\_id'

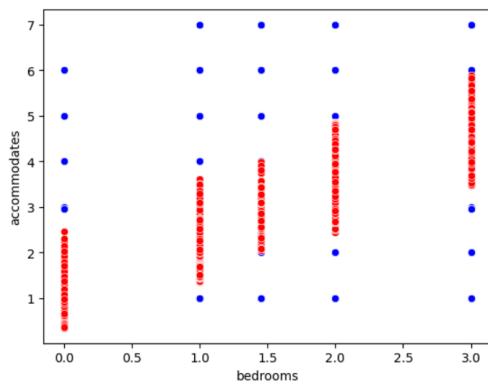


Coeficiente de Correlación: 0.3066987425410336

## **Accommodates**

### **México**

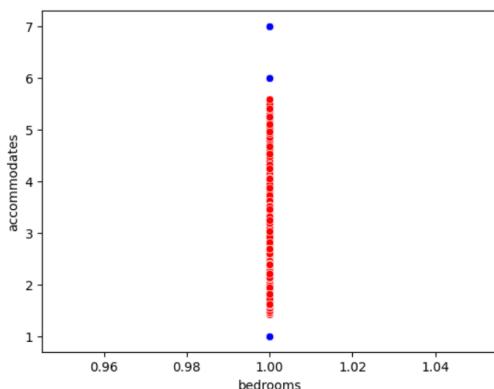
Variables Independientes: 'instant\_bookable', 'bedrooms', 'beds','price'



Coeficiente de Correlación: 0.684707197593561

### **Geneva**

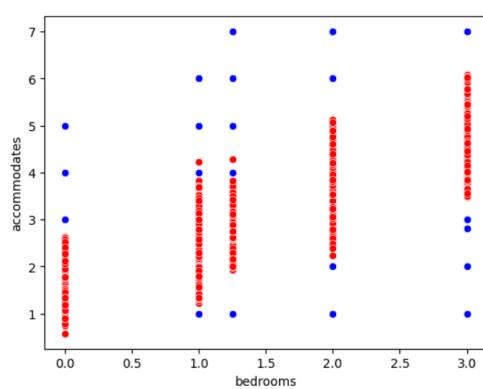
Variables Independientes: 'instant\_bookable', 'bedrooms', 'beds','price'



Coeficiente de Correlación: 0.5796946116294518

### **Boston**

Variables Independientes: 'instant\_bookable', 'bedrooms', 'beds','price'



Coeficiente de Correlación: 0.6825781272100805

## ***Bedrooms***

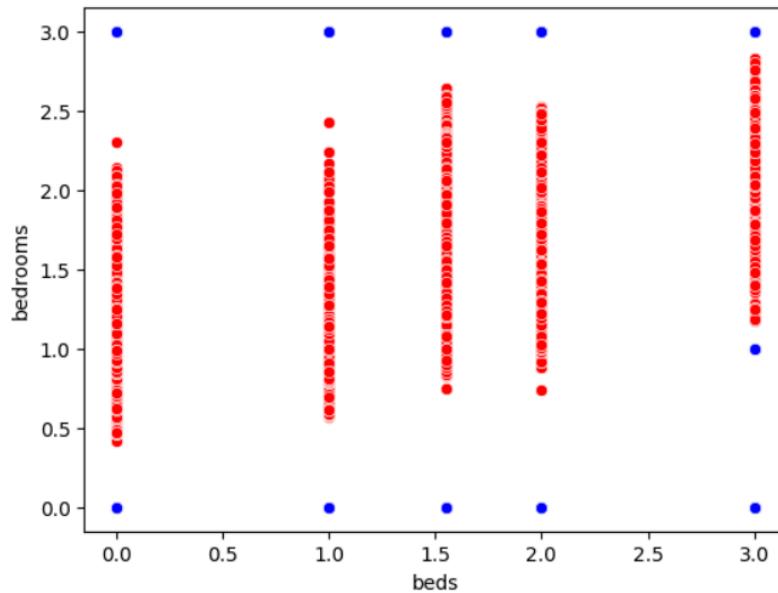
### ***México***

Variables

Independientes:  
'calculated\_host\_listings\_count','price'

'accommodates',

'beds',



Coeficiente de Correlación: 0.6726662322952779

### ***Geneva***

NAN

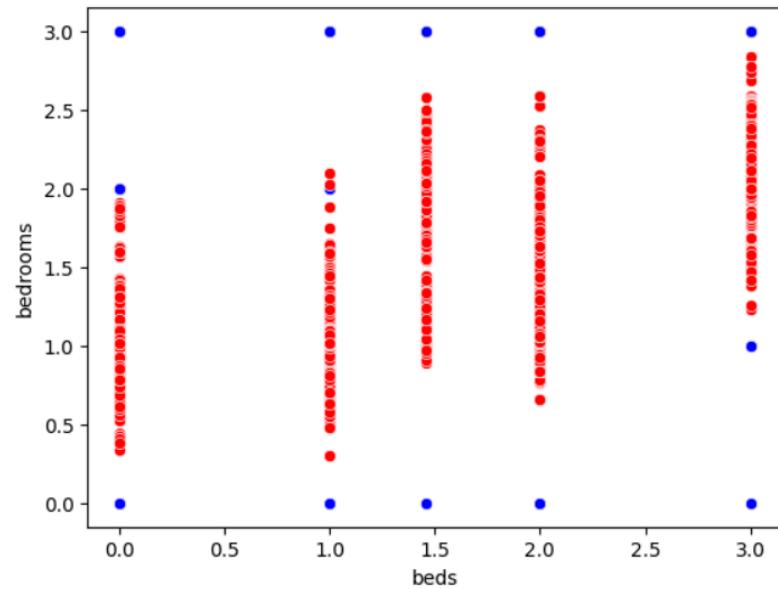
### ***Boston***

Variables

Independientes:  
'calculated\_host\_listings\_count','price'

'accommodates',

'beds',

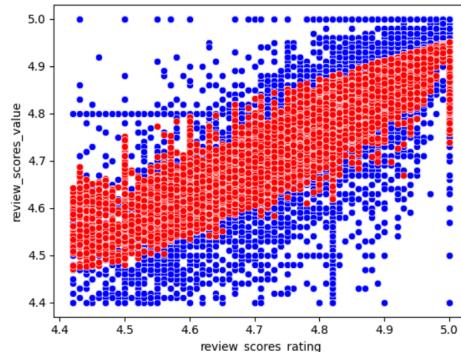


Coeficiente de Correlación: 0.632966467696344

## **Review Scores Value**

### **México**

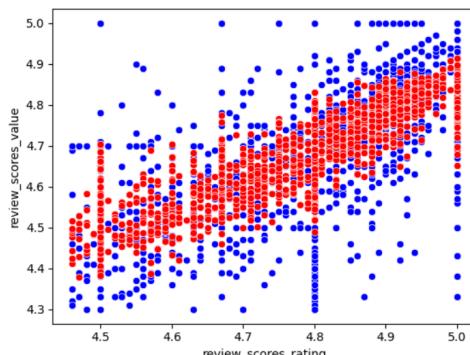
Variables Independientes: 'review\_scores\_rating', 'review\_scores\_accuracy', 'review\_scores\_cleanliness', 'review\_scores\_checkin', 'review\_scores\_communication'



Coeficiente de Correlación: 0.7448017666087589

### **Geneva**

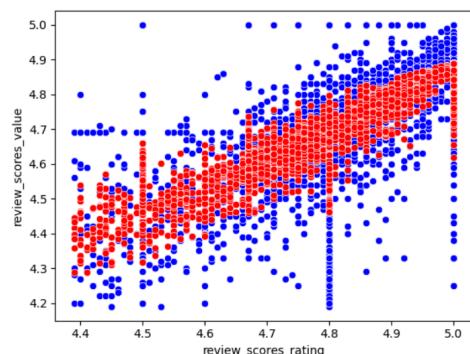
Variables Independientes: 'review\_scores\_rating', 'review\_scores\_accuracy', 'review\_scores\_cleanliness', 'review\_scores\_checkin', 'review\_scores\_communication', 'review\_scores\_location'



Coeficiente de Correlación: 0.6962177116118189

### **Boston**

Variable Independiente: 'review\_scores\_rating', 'review\_scores\_accuracy', 'review\_scores\_cleanliness', 'review\_scores\_checkin', 'review\_scores\_communication'

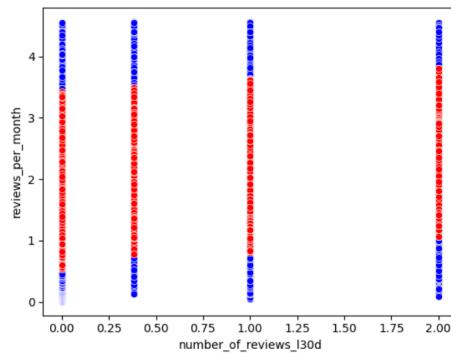


Coeficiente de Correlación: 0.7321024250076902

## **Reviews Per Month**

### **México**

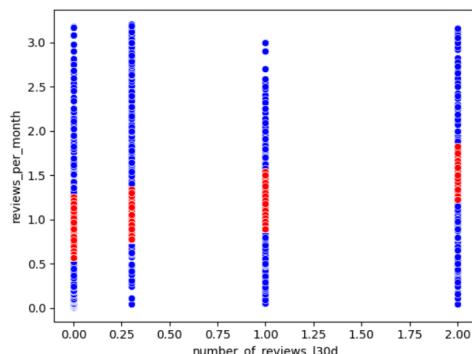
Variables Independientes: 'number\_of\_reviews\_l30d', 'minimum\_nights',  
'minimum\_minimum\_nights','maximum\_minimum\_nights','minimum\_nights\_avg\_ntm'  
'number\_of\_reviews\_ltm'



Coeficiente de Correlación: 0.657130009711513

### **Geneva**

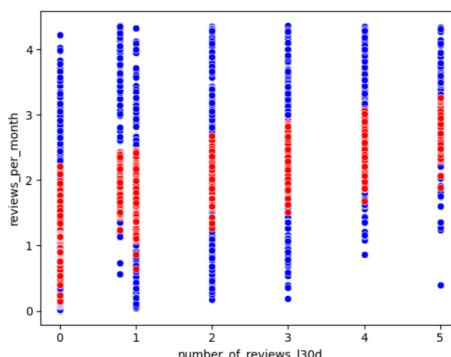
Variables Independientes: 'number\_of\_reviews\_ltm', 'number\_of\_reviews\_l30d'



Coeficiente de Correlación: 0.4304400909858852

### **Boston**

Variables Independientes: 'number\_of\_reviews\_l30d', 'minimum\_nights',  
'minimum\_minimum\_nights','maximum\_minimum\_nights','minimum\_nights\_avg\_ntm'  
'number\_of\_reviews\_ltm'

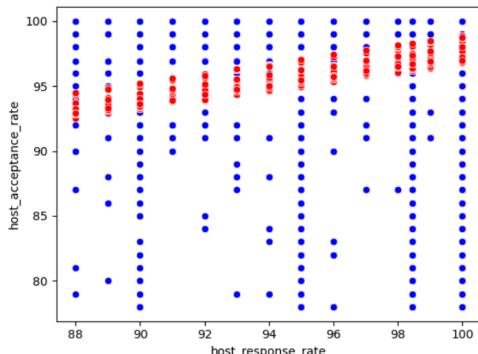


Coeficiente de Correlación: 0.5515720796693809

## **Host Acceptance Rate**

### **México**

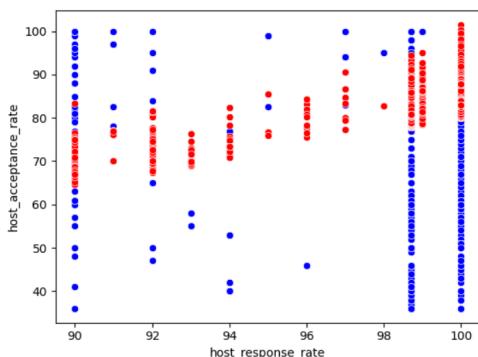
Variables Independientes: 'number\_of\_reviews\_I30d', 'host\_response\_rate', 'maximum\_minimum\_nights'



Coeficiente de Correlación: 0.2624983182237149

### **Geneva**

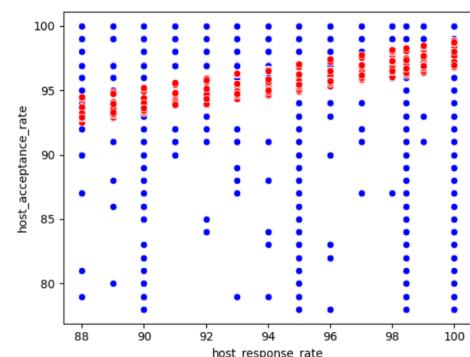
Variables Independientes: 'host\_response\_rate', 'number\_of\_reviews', 'reviews\_per\_month'



Coeficiente de Correlación: 0.3246754070377149

### **Boston**

Variables Independientes: 'number\_of\_reviews\_I30d', 'host\_response\_rate', 'maximum\_minimum\_nights'



Coeficiente de Correlación: 0.4910849651579349

## ***Host Total Listings Count***

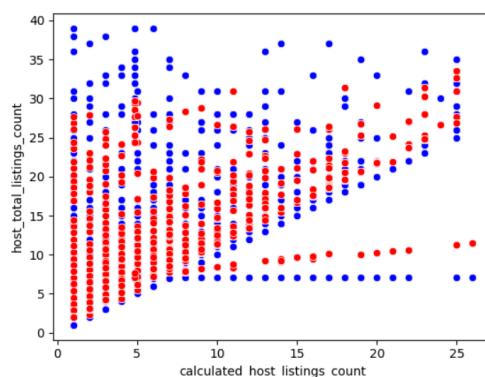
### **México**

Variables

Independientes:

'calculated\_host\_listings\_count',

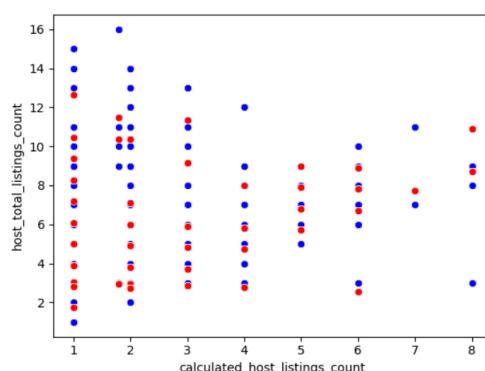
'calculated\_host\_listings\_count\_entire\_homes', 'host\_listings\_count'



Coeficiente de Correlación: 0.8637485629975565

### **Geneva**

Variables Independientes: 'host\_listings\_count', 'calculated\_host\_listings\_count'



Coeficiente de Correlación: 0.7573016195865145

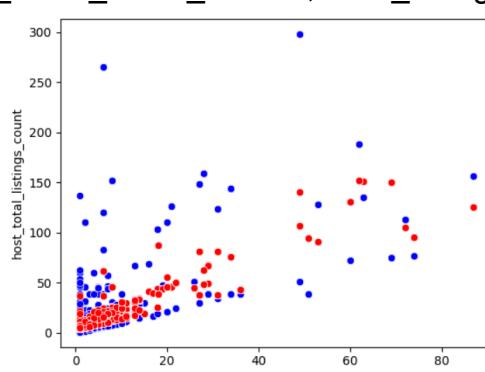
### **Boston**

Variables

Independientes:

'calculated\_host\_listings\_count',

'calculated\_host\_listings\_count\_entire\_homes', 'host\_listings\_count'

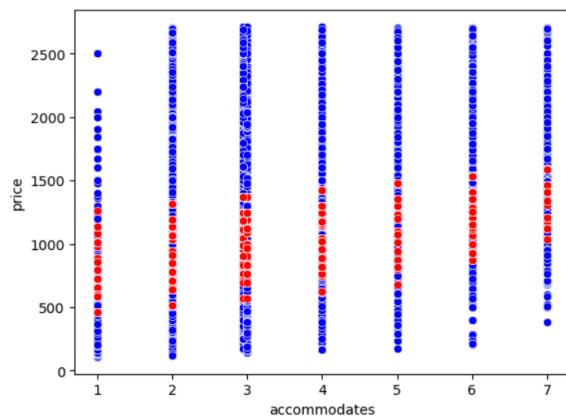


Coeficiente de Correlación: 0.8058108970314181

## **Price**

### **México**

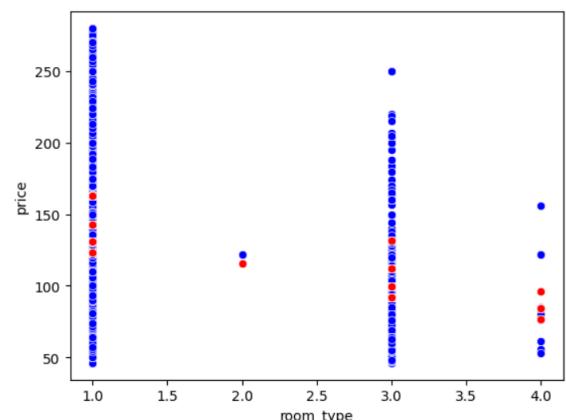
Variables Independientes: 'room\_type', 'accommodates', 'beds'



Coeficiente de Correlación: 0.4581447421266185

### **Geneva**

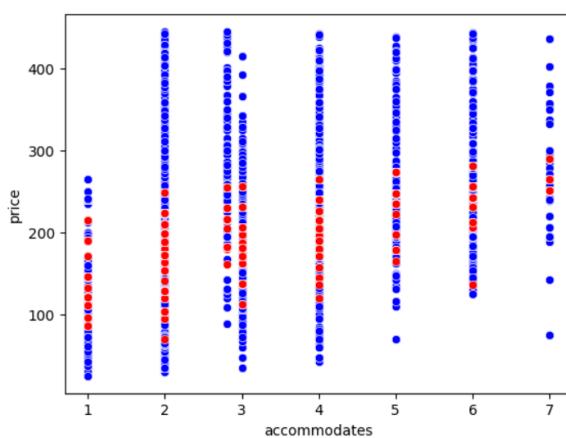
Variables Independientes: 'bathrooms', 'beds', 'room\_type'



Coeficiente de Correlación: 0.46498824961445834

### **Boston**

Variables Independientes: 'room\_type', 'accommodates', 'beds'



Coeficiente de Correlación: 0.5496706784525577

## ***Observaciones***

Entre las tres ciudades, al menos de manera visual gracias a las gráficas, se puede notar que los modelos creados son muy similares entre ellos, sin embargo, la diferencia reside en el coeficiente de correlación. Mismo el cuál nos ayuda a determinar aquellos modelos con la mayor efectividad para predecir los datos entre las tres ciudades.

Estas diferencias, aparentemente pueden proceder de los datos que contiene cada ciudad, pues más que la diferencia en la cantidad de datos, parece que la correlación depende significativamente de cómo cada variable de la base interactúa entre sí con los datos que contiene. Es decir, que cada ciudad contiene no sólo una cantidad diferente de datos, sino que dentro de estos datos existe una diferencia significativa en sus valores.