

**Instituto Tecnológico y de Estudios  
Superiores de Monterrey**



**Tecnológico  
de Monterrey**

**Analítica de Datos y Herramientas de Inteligencia Artificial I**

**Actividad 2.1**

**ALUMNO**

**André Calmus González**

**A017333529**

**Grupo 101**

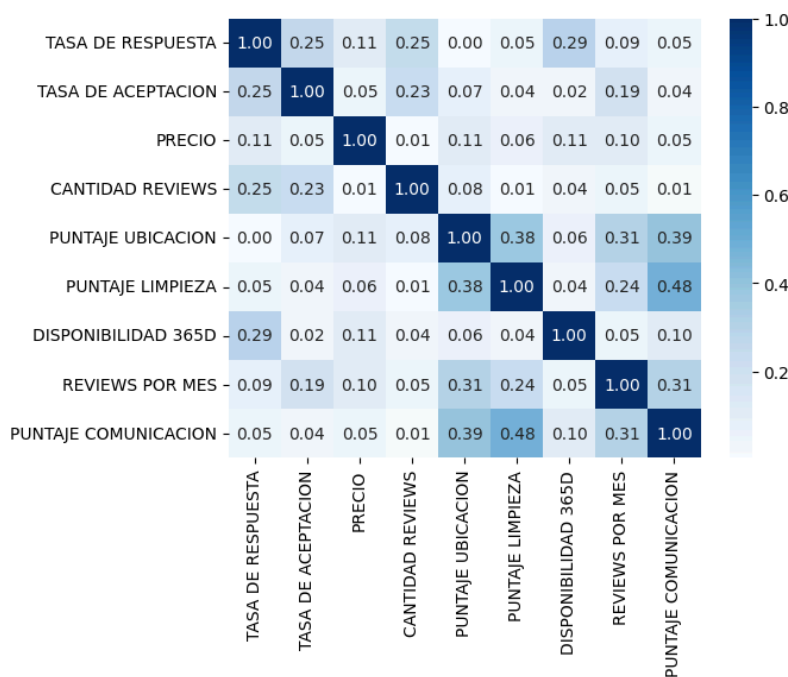
**7, Abril 2025**

## Introducción

En este reporte se analizarán las correlaciones entre cada tipo de habitación con variables seleccionadas que se verán a continuación. Se está trabajando con el dataset de Ginebra, con los datos sacados de Airbnb. Veremos los coeficientes de determinación y correlación más altos por cada tipo de cuarto. De igual manera, haremos modelos de regresión lineal múltiple para las variables cuantitativas seleccionadas.

## Correlaciones

### Entire Home

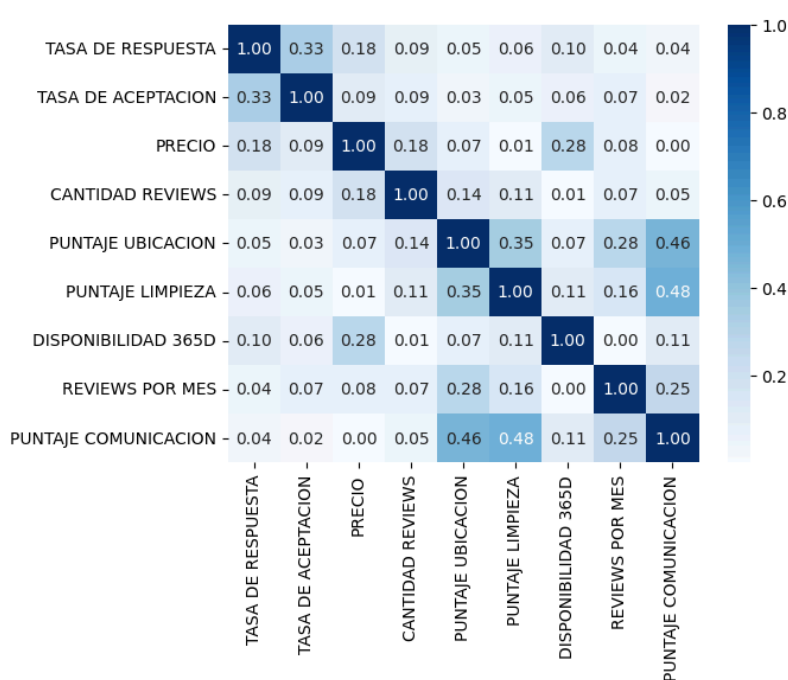


Aquí tenemos nuestro mapa de calor para el tipo de cuarto Entire Home / apt. Podemos observar las 9 variables asignadas para realizar este análisis.

A continuación podremos ver los factores de correlación de los conjuntos de variables dependientes e independientes necesarios para el tipo de cuarto en cuestión.

- “TASA DE ACEPTACIÓN vs TASA DE RESPUESTA” = 0.25
- “PUNTAJE UBICACIÓN vs PUNTAJE LIMPIEZA” = 0.38
- “TASA DE ACEPTACIÓN vs PRECIO” = 0.05
- “DISPONIBILIDAD 365D vs CANTIDAD REVIEWS” = 0.04
- “TASA DE ACEPTACIÓN vs CANTIDAD REVIEWS” = 0.23
- “REVIEWS POR MES vs PUNTAJE COMUNICACIÓN” = 0.31

## Private Room

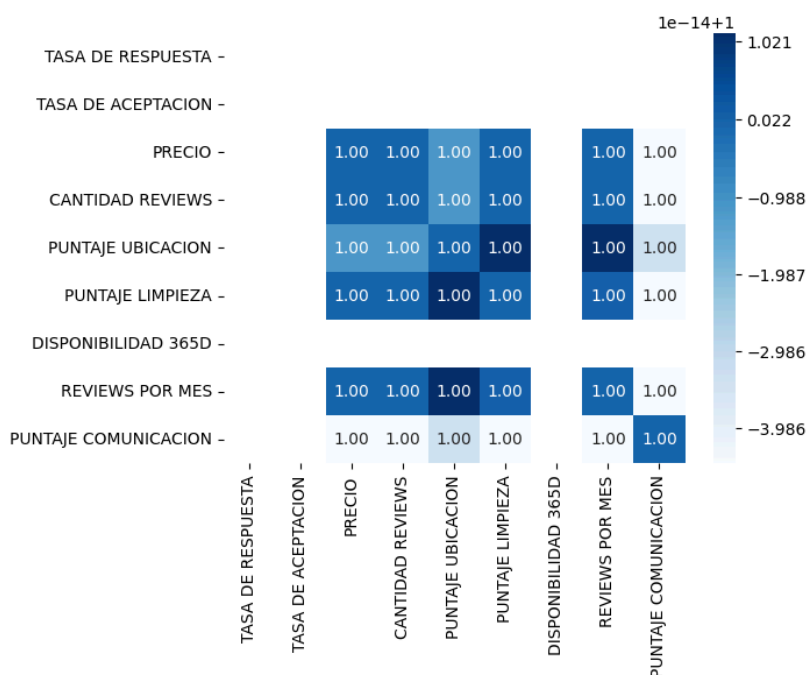


Aquí tenemos nuestro mapa de calor para el tipo de cuarto Private Room. Podemos observar las 9 variables asignadas para realizar este análisis.

A continuación podremos ver los factores de correlación de los conjuntos de variables dependientes e independientes necesarios para el tipo de cuarto en cuestión.

- “TASA DE ACEPTACIÓN vs TASA DE RESPUESTA” = 0.33
- “PUNTAJE UBICACIÓN vs PUNTAJE LIMPIEZA” = 0.35
- “TASA DE ACEPTACIÓN vs PRECIO” = 0.09
- “DISPONIBILIDAD 365D vs CANTIDAD REVIEWS” = 0.01
- “TASA DE ACEPTACIÓN vs CANTIDAD REVIEWS” = 0.09
- “REVIEWS POR MES vs PUNTAJE COMUNICACIÓN” = 0.25

## Hotel Room



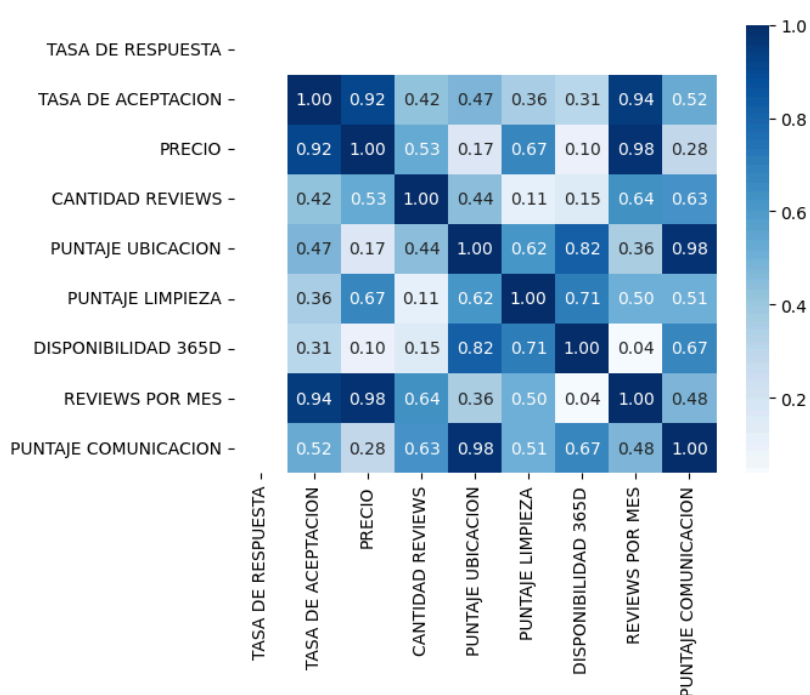
Aquí tenemos nuestro mapa de calor para el tipo de cuarto Hotel Room. Podemos observar las 9 variables asignadas para realizar este análisis.

Para este tipo de cuarto sólo tenemos 2 entradas en el DataFrame por lo que los datos no tienen la suficiente información para dar una correlación ni para realizar un modelo. Por esta razón

tenemos zonas del mapa del calor sin correlación alguna.

- “TASA DE ACEPTACIÓN vs TASA DE RESPUESTA” = NaN
- “PUNTAJE UBICACIÓN vs PUNTAJE LIMPIEZA” = 1.00
- “TASA DE ACEPTACIÓN vs PRECIO” = NaN
- “DISPONIBILIDAD 365D vs CANTIDAD REVIEWS” = NaN
- “TASA DE ACEPTACIÓN vs CANTIDAD REVIEWS” = NaN
- “REVIEWS POR MES vs PUNTAJE COMUNICACIÓN” = 1.00

## Shared Room



Aquí tenemos nuestro mapa de calor para el tipo de cuarto Shared Room. Podemos observar las 9 variables asignadas para realizar este análisis.

Para este tipo de cuarto sólo tenemos 4 entradas en el DataFrame por lo que hay variables en las que no se puede obtener una correlación y en muchas de las demás variables

podremos encontrar correlaciones más altas. La falta de información sesga nuestra análisis.

- “TASA DE ACEPTACIÓN vs TASA DE RESPUESTA” = NaN
- “PUNTAJE UBICACIÓN vs PUNTAJE LIMPIEZA” = 0.62
- “TASA DE ACEPTACIÓN vs PRECIO” = 0.92
- “DISPONIBILIDAD 365D vs CANTIDAD REVIEWS” = 0.15
- “TASA DE ACEPTACIÓN vs CANTIDAD REVIEWS” = 0.42
- “REVIEWS POR MES vs PUNTAJE COMUNICACIÓN” = 0.48

## Comparativa Correlaciones Variables Seleccionadas por Tipo de Cuarto

	Variables Comparadas	Entire Home	Private Room	Hotel Room	Shared Room
0	TASA DE ACEPTACIÓN vs TASA DE RESPUESTA	0.25	0.33	NaN	NaN
1	PUNTAJE UBICACIÓN vs PUNTAJE LIMPIEZA	0.38	0.35	1.0	0.62
2	TASA DE ACEPTACIÓN vs PRECIO	0.05	0.09	NaN	0.92
3	DISPONIBILIDAD 365D vs CANTIDAD REVIEWS	0.04	0.01	NaN	0.15
4	TASA DE ACEPTACIÓN vs CANTIDAD REVIEWS	0.23	0.09	NaN	0.42
5	REVIEWS POR MES vs PUNTAJE COMUNICACIÓN	0.31	0.25	1.0	0.48

Para analizar las correlaciones más altas sería óptimo no tener en cuenta los tipo de cuarto Hotel Room y Shared Room, esto debido al número de entradas que tiene cada tipo de cuarto, dejándonos sin la suficiente información para realizar un buen análisis. La correlación más alta la tiene el tipo de cuarto Entire Home, en la comparación de PUNTAJE UBICACIÓN vs PUNTAJE LIMPIEZA con un 0.38.

## Modelos Matemáticos Regresión Lineal y Coeficientes de Determinación y Correlación

### Entire Home

Coeficiente de Determinación: 0.002278300502072894

Coeficiente de Correlación: 0.04773154619403078

Modelo Matemático:  $y = 1.41047493x + 9.099808852938489$

El coeficiente de determinación es 0.0023, lo que indica que solo el 0.23% de la variación en la variable dependiente se explica por el modelo. Además, el coeficiente de correlación es 0.0477, lo que sugiere una relación extremadamente débil y casi nula entre las variables.

### Private Room

Coeficiente de Determinación: 0.005279435756309403

Coeficiente de Correlación: 0.07265972582049428

Modelo Matemático:  $y = 1.96211895x + 9.223581710822662$

El coeficiente de determinación es 0.0053, lo que indica que solo el 0.53% de la variación en la variable dependiente se explica por el modelo. Además, el coeficiente de correlación es 0.0727, lo que sugiere una relación muy débil entre las variables.

### **Hotel Room**

Coefficiente de Determinación: 1.0

Coefficiente de Correlación: 1.0

Modelo Matemático:  $y = 80.0x + 2.0$

Como explicamos anteriormente, si bien tenemos resultados aparentemente perfectos, son cero confiables y viables, esto debido a que este modelo se entrenó con tan sólo 2 entradas, por lo que no tiene peso ni relevancia alguna.

### **Shared Room**

Coefficiente de Determinación: 0.4117320524713872

Coefficiente de Correlación: 0.6416635040824648

Modelo Matemático:  $y = -40.77417948x + 57.79450441609422$

De igual manera que con el tipo de cuarto Hotel Room, para Shared Room tan sólo contamos con 4 entradas, por lo que los coeficientes aunque se vean más coherentes que los anteriores, siguen sin ser confiables debido a la tan poca cantidad de información con la que se entrenó al modelo.

### **Modelos Regresión Lineal Múltiple Variables Cuantitativas**

#### **NOMBRE ANFITRIÓN**

Coefficiente de Determinación: 0.016599779159455763

Coefficiente de Correlación: 0.1288401302368783

#### **TASA DE ACEPTACIÓN**

Coefficiente de Determinación: 0.0006053687675641184

Coefficiente de Correlación: 0.024604242877278675

### **SUPERANFITRIÓN**

Coefficiente de Determinación: 0.00990142752289258

Coefficiente de Correlación: 0.09950591702452966

### **CANTIDAD TOTAL ANUNCIOS**

Coefficiente de Determinación: 0.021406153340873257

Coefficiente de Correlación: 0.14630841855776194

### **TIPO DE CUARTO**

Coefficiente de Determinación: 1.0

Coefficiente de Correlación: 1.0

Esto debido a la poca información en algunos tipos de cuarto.

### **CAPACIDAD**

Coefficiente de Determinación: 0.029313448693067956

Coefficiente de Correlación: 0.1712117072313338

### **DORMITORIOS**

Coefficiente de Determinación: 0.18019714945651877

Coefficiente de Correlación: 0.42449634798961317

### **PRECIO**

Coefficiente de Determinación: 0.21419612746308325

Coefficiente de Correlación: 0.46281327494258767

### **PUNTAJE RATING**

Coefficiente de Determinación: 0.6153231210379457



Coeficiente de Correlación: 0.7844253444643063

## **REVIEWS POR MES**

Coeficiente de Determinación: 0.14247660496590198

Coeficiente de Correlación: 0.37746073301192795

## **Conclusión**

Los modelos de regresión lineal simple y múltiple aplicados no presentan un buen desempeño, ya que muestran coeficientes de determinación y correlación muy bajos. Esto indica que la capacidad de los modelos para explicar la variabilidad de la variable dependiente es prácticamente nula y la relación entre las variables es débil o inexistente. Sólo el modelo de regresión lineal múltiple para puntaje rating dio valores buenos que pueden indicar la veracidad del modelo.

Dado esto, se sugiere explorar en el futuro modelos de regresión no lineal u otras técnicas más avanzadas que puedan capturar relaciones más complejas entre las variables, y así obtener resultados más precisos y representativos de las variables analizadas.