

# Actividad 2:COMPLEMENTO

Victoria Eugenia Téllez Castillo A01732258

Alberto Rodriguez Porras A01721497

7 de octubre del 2022

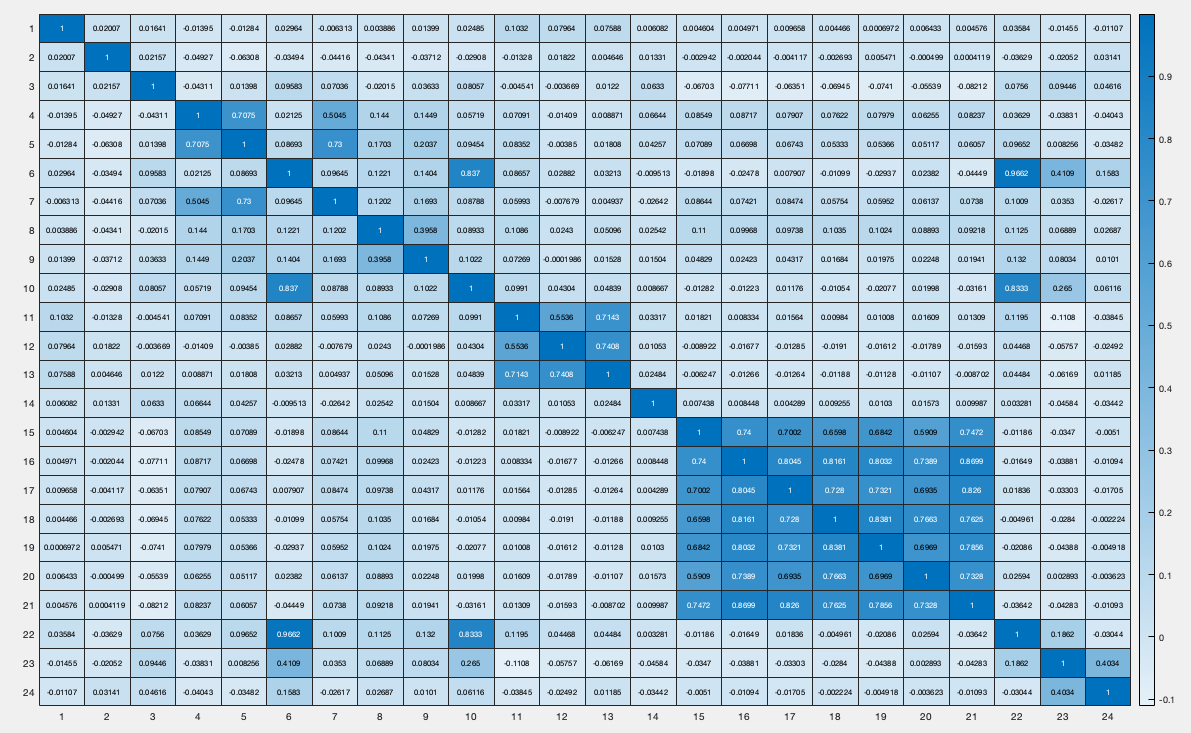
Nombre de la materia: Gestión de proyectos de plataformas tecnológicas.

**Análisis de correlación**:

Dentro de este reporte se comentarán los resultados obtenidos a partir del análisis de correlación realizado en la herramienta de “matlab”; la base de datos utilizada para la realización de esta actividad lleva el nombre de “méxico largo”, esta base de datos fue obtenida de la plataforma de Airbnb, lleva un aproximado de 70 variables dentro de las cuales fueron seleccionadas únicamente 27 para realizar este análisis de correlación. Las variables seleccionadas fueron las siguientes:

* price
* minimun\_nights
* availability\_365
* number\_of\_reviews
* number\_of\_reviews\_ltm
* calculated\_host\_listing\_count
* reviews\_per\_month
* host\_response\_rate
* host\_acceptance\_rate
* host\_total\_listings\_count
* accommodates
* bedrooms
* beds
* maximum\_nights
* reviwe\_scores\_rating
* review\_scores\_accuracy
* review\_scores\_cleanliness
* review\_scores\_checkin
* review\_scores\_communication
* review\_score\_location
* review\_score\_value
* calculated\_host\_listings\_count\_entire\_homes
* calculated\_host\_listings\_count\_entire\_private\_rooms
* calculated\_host\_listings\_count\_entire\_shared\_rooms
* room\_type
* property\_type
* neighbourhood\_cleansed

Posteriormente, limpiamos la base de datos removiendo los valores nulos de toda la base, y reemplazando los valores atípicos de las variables cuantitativas que vamos a usar dentro de nuestro análisis; a continuación realizamos un mapa de calor para poder determinar la correlación entre las variables de manera más visual y se obtiene lo siguiente.

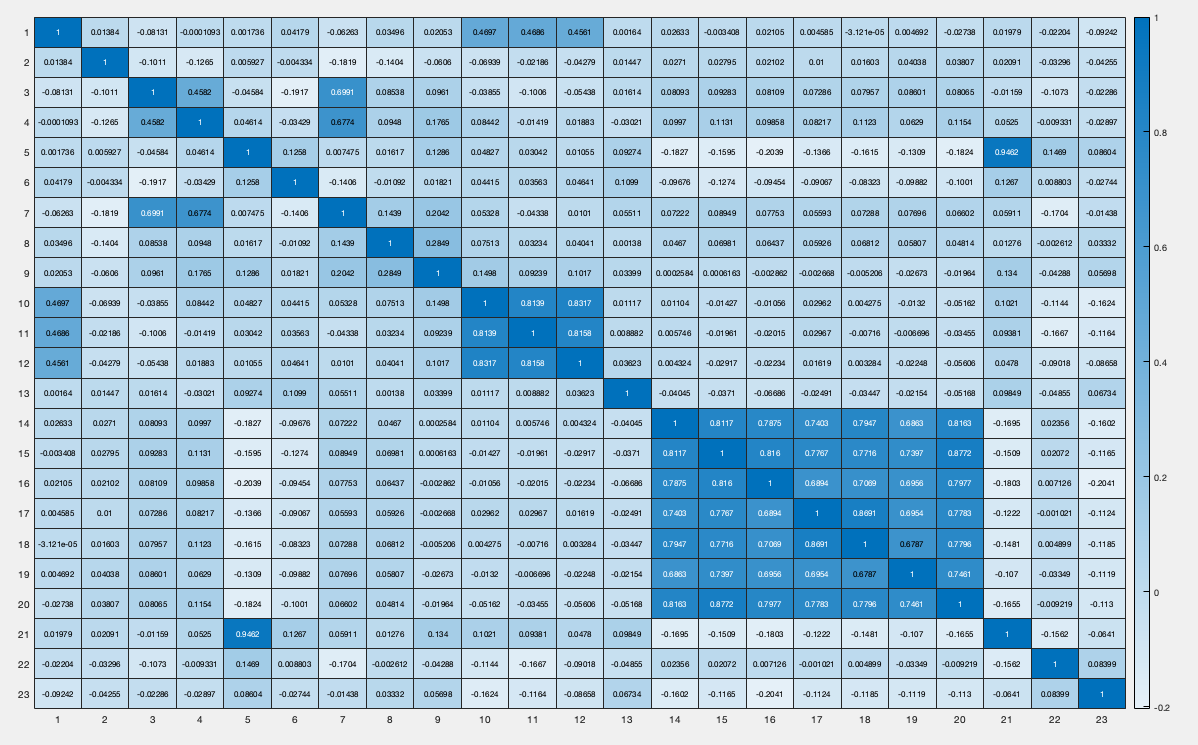


Los colores nos ayudan a determinar el mayor nivel de correlación, entre más oscuras se encuentran las variables, mayor correlación tienen; las variables que mejor correlación presentan son las siguientes:

|  |  |  |
| --- | --- | --- |
| **CORRELACIÓN** | **VARIABLE UNO**  **X** | **VARIABLE DOS**  **Y** |
| **.97** | **6= calculated\_host\_listings\_count** | **22= calculated\_host\_listings\_count\_entire\_homes** |
| **.83** | **10= host\_total\_listings\_count** | **22= calculated\_host\_listings\_count\_entire\_homes** |
| **.84** | **6= calculated\_host\_listings\_count** | **10= host\_total\_listings\_count** |
| .74 | 12= bedrooms | 13= beds |
| .71 | 11= accommodates | 13= beds |
| .71 | 4= number\_of\_reviews | 5=number\_of\_reviews\_ltm |
| .74 | 15=review\_scores\_rating | 16= review\_scores\_accuracy |
| .80 | 16=review\_scores\_accuracy | 17=review\_scores\_cleanliness |
| .81 | 16=review\_scores\_accuracy | 18=review\_scores\_checkin |
| .76 | 18=review\_scores\_checkin | 21= review\_scores\_value |
| -.04 | 3=availability\_365 | 4=price |
| .10 | 11= accommodates | 1= price |
| -0.06 | 15= review\_scores\_rating | 3=availability\_365 |

Posterior al análisis realizado en México y al listado de las variables que mayor correlación presentan, vamos a analizar otro país con características similares a México con el fin de poder realizar comparaciones que nos ayuden a encontrar hallazgos los cuales utilicemos a la hora de crear y proponer nuestras estrategías que tienen la finalidad de ayudar a la empresa de Airbnb a expandirse e incrementar su presencia en el territorio mexicano, en la tabla superior se encuentran las variables que mayor correlación presentan, y de igual manera decidimos incluir tres variables extras las cuales nosotros habíamos pensado podrían tener correlación, sin embargo, vemos que en el territorio mexicano, estas variables cuentan con muy baja relación entre ellas e incluso negativa; el país que fue seleccionado para realizar el análisis comparativo fue Estados Unidos, la ciudad de Texas consideramos que sería un perfecto ejemplo gracias a la similitud que tiene con México, posterior a la manipulación de la base de datos, realizamos otro mapa de calor y encontramos las siguientes correlaciones:

Texas.



|  |  |  |
| --- | --- | --- |
| **CORRELACIÓN** | **VARIABLE UNO**  **X** | **VARIABLE DOS**  **Y** |
| .81 | 14= review\_scores\_rating | 15=review\_scores\_accuracy |
| .81 | 10=accommodates | 11=bedrooms |
| .83 | 10=accommodates | 12=beds |
| .82 | 11=bedrooms | 12=beds |
| .80 | 16=review\_scores\_cleanliness | 20= review\_scores\_value |
| .82 | 15=review\_scores\_accuracy | 16=review\_scores\_cleanliness |
| **.88** | **15=review\_scores\_accuracy** | **20=review\_scores\_value** |
| .75 | 19= review\_scores\_location | 20=review\_scores\_value |
| **.87** | **17=review\_scores\_checkin** | **18=review\_scores\_communication** |
| **.95** | **5= calculated\_host\_listings\_count** | **21=calculated\_host\_listings\_count\_entire\_homes** |
| .46 | 3=availability\_365 | 4=price |
| .47 | 11= accommodates | 1= price |
| .09 | 15= review\_scores\_rating | 3=availability\_365 |

Al comparar las variables que mayor correlación presentan en Texas encontramos los siguientes hallazgos:

* las últimas tres variables comparadas en ambos países, Texas presenta una mayor relación entre variables, sin embargo, esta relación continúa siendo muy baja.
* Las variables que mayor correlación presentan en México son:
  + calculated\_host\_listings\_count con calculated\_host\_listings\_count\_entire\_homes
  + host\_total\_listings\_count con calculated\_host\_listings\_count\_entire\_homes
  + calculated\_host\_listings\_count con host\_total\_listings\_count.
* Las variables que mayor correlación presentan en Texas son:
  + calculated\_host\_listings\_count con calculated\_host\_listings\_count\_entire\_homes
  + review\_scores\_accuracy con review\_scores\_value
  + review\_scores\_checkin con review\_scores\_communication
* Las variables que mayor correlación presentaron fueron las de la base de datos de México.

**MEXICO VARIABLES QUE NOSOTROS SELECCIONAMOS**

|  |  |  |
| --- | --- | --- |
| **CORRELACIÓN** | **PRICE** | **VARIABLE DOS** |
| .10 | 1= Price | 11= accommodates |
| .02 | 2= minimum\_nights | 3= availability\_365 |
| **.096** | **3= availability\_365** | **6= calculated\_host\_listings\_count** |
| .71 | 4= number\_of\_reviews | 5=number\_of\_reviews\_ltm |
| .73 | 5= number\_of\_reviews\_ltm | 7= reviews\_per\_month |
| .17 | 7= reviews\_per\_month | 9= host\_acceptance\_rate |
| .40 | 8= host\_response\_rate | 9= host\_acceptance\_rate |
| .55 | 11= accommodates | 12= bedrooms |
| .70 | 15= review\_scores\_rating | 17= review\_scores\_cleanliness |
| **.87** | **16= review\_scores\_accuracy** | **21= review\_scores\_value** |

Para llevar a cabo un mejor análisis, vamos a seleccionar 10 relaciones entre variables y las vamos a comparar las mismas en ambos países ahora que ya comprobamos que entre países no comparten las mismas variables con alta correlación.

**TEXAS VARIABLES QUE NOSOTROS SELECCIONAMOS**

|  |  |  |
| --- | --- | --- |
| **CORRELACIÓN** | **PRICE** | **VARIABLE DOS** |
| .47 | 1= Price | 10= accommodates |
| -0.0043 | 2= minimum\_nights | 6= availability\_365 |
| 0.125 | 6= availability\_365 | 5= calculated\_host\_listings\_count |
| **.70** | **3= number\_of\_reviews** | **7=number\_of\_reviews\_ltm** |
| .68 | 7= number\_of\_reviews\_ltm | 4= reviews\_per\_month |
| .176 | 4= reviews\_per\_month | 9= host\_acceptance\_rate |
| .28 | 8= host\_response\_rate | 9= host\_acceptance\_rate |
| .81 | 10= accommodates | 11= bedrooms |
| .79 | 14= review\_scores\_rating | 16= review\_scores\_cleanliness |
| **.88** | **15= review\_scores\_accuracy** | **20= review\_scores\_value** |

Ahora, teniendo unas variables seleccionadas podemos realizar un análisis más profundo entre las ciudades seleccionadas, en primera instancia se observa que la mayoría de las variables presentan más correlación en la ciudad de Texas que en el Df, las únicas que son mayores en correlación es en la relación de las siguientes variables:

* minimum\_nights con availability\_365.
* number\_of\_reviews\_ltm con reviews\_per\_month.
* host\_response\_rate con host\_acceptance\_rate

La diferencia entre estas variables no es muy grande excepto la relación entre number\_of\_reviews\_ltm con reviews\_per\_month ya que en Texas esta correlación es presentada negativa, gracias al análisis de estas variables podemos deducir lo siguiente:

* En Estados Unidos, más específico en la ciudad de Texas, la empresa Airbnb cuenta con un segmento de mercado más extenso y una mayor presencia que en México.
* La relación que tiene la variable precio con la capacidad de albergar no es muy significativa, sin embargo tiene más significancia en Texas que en México.
* La relación entre la variable de mínimo de noches con la disponibilidad 365 no tiene relación en ninguno de los países.
* Las variables que más relación presentaron fueron precisión de las puntuaciones de las revisiones con la variable de valor de puntuación de revisión; y las variables de número de reseñas con número de reseñas en los últimos 12 meses.
* La variable de capacidad de albergue y habitaciones presenta una relación fuerte, esta fue una de las hipótesis planteadas, más adelante podremos tener una mayor comprobación de su significancia.
* La variable que pensamos que iba a tener una mayor correlación y estuvimos equivocados fue la correlación entre tasa de respuesta del host con tasa de aceptación de anfitrión.

Estos fueron unos de los principales hallazgos encontrados, más adelante podremos tener más apoyos y herramientas para poder tener una conclusión mucho más certera.