



Tecnológico de Monterrey

Actividad 1: Regresión Lineal Simple

María Fernanda Robles Soto - A01736552

04 de Octubre del 2025

Gestión de proyectos de plataformas tecnológicas

Reporte

A continuación, presento el reporte correspondiente al código desarrollado para la Actividad 1, en el cual describo los pasos que seguí, las decisiones que tomé durante el proceso y los principales hallazgos obtenidos en el análisis de regresión lineal y múltiple.

Carga y limpieza de los datos (Valores Nulos y NA)

Como primer paso, cargué las librerías necesarias y la base de datos correspondiente a Brasil, tal como se indicaba en la actividad.

Para no modificar ni evitar la base original, creé una copia de trabajo sobre la cual realicé todo el preprocesamiento. En esta parte, estandaricé los nombres de las columnas, limpié los datos y convertí todos los valores de las columnas correspondientes de porcentuales a tipo float. Asimismo, transformé las variables booleanas y, en los casos donde no se pudo hacerlo, me aseguré de que quedaran como tipo object. Posteriormente, separé las columnas numéricas de las categóricas. Dado que preferí no eliminar valores nulos o NA, opté por imputar los valores utilizando la media para las variables numéricas y la moda para las categóricas. En cuanto al tratamiento de outliers, apliqué el método de los cuartiles (IQR) para identificarlos y analizarlos. Finalmente, realicé una visualización general de la base de datos ya limpia para verificar que las transformaciones se hubieran aplicado correctamente.

Regresión lineal Simple

En esta parte, se hizo un análisis de la correlación en cada tipo de habitación respecto a las variables dependientes e independientes.

Como parte de la exploración de la misma, hice un recuento de los valores por tipo de habitación.

```
# Recuento de valores por tipo de habitación
df["room_type"].value_counts()
```

	count
Entire home/apt	33990
Private room	8152
Shared room	407
Hotel room	23

dtype: int64

Comenzando con el análisis de correlación entre las variables, primero analicé la relación entre el 'host_acceptance_rate' y el 'host_response_rate'. En este caso observé que ambas variables tenían una correlación positiva, lo que significa que los anfitriones que suelen responder más rápido a los huéspedes también tienden a aceptar más solicitudes. Aunque la

relación no fue perfecta, sí mostró una tendencia clara a que un mejor tiempo de respuesta se asocia con una tasa de aceptación más alta.

```
[ ]: # Hacemos el análisis de correlación
# HOST_ACCEPTANCE_RATE y HOST_RESPONSE_RATE

correlaciones1 = {
    "Entire_home_apt": Entire_home_apt["host_acceptance_rate"].corr(Entire_home_apt["host_response_rate"]),
    "Private_room": Private_room["host_acceptance_rate"].corr(Private_room["host_response_rate"]),
    "Shared_room": Shared_room["host_acceptance_rate"].corr(Shared_room["host_response_rate"]),
    "Hotel_room": Hotel_room["host_acceptance_rate"].corr(Hotel_room["host_response_rate"])
}

tabla_corr = pd.DataFrame.from_dict(correlaciones1, orient="index", columns=["Correlación"])
print(tabla_corr)
```

	Correlación
Entire_home_apt	0.460844
Private_room	0.420151
Shared_room	0.296713
Hotel_room	-0.015398

Luego comparé 'review_scores_rating' con 'calculated_host_listings_count'. En esta correlación noté que la relación era débil o casi nula, lo que sugiere que tener más propiedades no necesariamente garantiza una mejor calificación. Esto puede deberse a que los anfitriones con muchas propiedades no siempre logran mantener el mismo nivel de atención o calidad en todas.

```
In [ ]: #REVIEW_SCORES_RATING Y CALCULATED_HOST_LISTINGS_COUNT

correlaciones4 = {
    "Entire_home_apt": Entire_home_apt["review_scores_rating"].corr(Entire_home_apt["calculated_host_listings_count"]),
    "Private_room": Private_room["review_scores_rating"].corr(Private_room["calculated_host_listings_count"]),
    "Shared_room": Shared_room["review_scores_rating"].corr(Shared_room["calculated_host_listings_count"]),
    "Hotel_room": Hotel_room["review_scores_rating"].corr(Hotel_room["calculated_host_listings_count"])
}

tabla_corr = pd.DataFrame.from_dict(correlaciones4, orient="index", columns=["Correlación"])
print(tabla_corr)
```

	Correlación
Entire_home_apt	-0.171162
Private_room	-0.181270
Shared_room	-0.011632
Hotel_room	0.136616

En el caso de 'host_acceptance_rate' frente a 'price', la correlación fue baja y algo dispersa. Esto indica que el precio no influye directamente en la tasa de aceptación de los anfitriones, ya que algunos pueden aceptar muchas reservas a precios altos o bajos dependiendo de su estrategia y del tipo de alojamiento.

```
[ ]: #HOST_ACCEPTANCE_RATE Y PRICE

correlaciones2 = {
    "Entire_home_apt": Entire_home_apt["host_acceptance_rate"].corr(Entire_home_apt["price"]),
    "Private_room": Private_room["host_acceptance_rate"].corr(Private_room["price"]),
    "Shared_room": Shared_room["host_acceptance_rate"].corr(Shared_room["price"]),
    "Hotel_room": Hotel_room["host_acceptance_rate"].corr(Hotel_room["price"])
}

tabla_corr = pd.DataFrame.from_dict(correlaciones2, orient="index", columns=["Correlación"])
print(tabla_corr)
```

	Correlación
Entire_home_apt	-0.152765
Private_room	0.019210
Shared_room	-0.116358
Hotel_room	-0.442452

También analicé 'availability_365' contra 'number_of_reviews'. Aquí encontré que la relación era negativa, lo cual tiene sentido, ya que los alojamientos con más disponibilidad suelen tener menos reservas y, por lo tanto, menos reseñas. Por el contrario, los que se reservan más seguido tienen menos días disponibles y acumulan más comentarios.

```
[ ]: #AVAILABILITY_365 Y NUMBER_OF_REVIEWS

correlaciones5 = {
    "Entire_home_apartment": Entire_home_apartment["availability_365"].corr(Entire_home_apartment["number_of_reviews"]),
    "Private_room": Private_room["availability_365"].corr(Private_room["number_of_reviews"]),
    "Shared_room": Shared_room["availability_365"].corr(Shared_room["number_of_reviews"]),
    "Hotel_room": Hotel_room["availability_365"].corr(Hotel_room["number_of_reviews"])
}

tabla_corr = pd.DataFrame.from_dict(correlaciones5, orient="index", columns=["Correlación"])
print(tabla_corr)
```

	Correlación
Entire_home_apartment	-0.153448
Private_room	-0.139779
Shared_room	-0.168357
Hotel_room	-0.155292

Para 'host_acceptance_rate' frente a 'number_of_reviews' la relación fue muy débil, lo que indica que la cantidad de reseñas no está directamente relacionada con la disposición del anfitrión a aceptar huéspedes. Puede que otros factores, como la ubicación o la política del alojamiento, tengan más peso en esta relación.

```
[ ]: #HOST_ACCEPTANCE_RATE Y NUMBER_OF_REVIEWS

correlaciones3 = {
    "Entire_home_apartment": Entire_home_apartment["number_of_reviews"].corr(Entire_home_apartment["host_response_rate"]),
    "Private_room": Private_room["number_of_reviews"].corr(Private_room["host_response_rate"]),
    "Shared_room": Shared_room["number_of_reviews"].corr(Shared_room["host_response_rate"]),
    "Hotel_room": Hotel_room["number_of_reviews"].corr(Hotel_room["host_response_rate"])
}

tabla_corr = pd.DataFrame.from_dict(correlaciones3, orient="index", columns=["Correlación"])
print(tabla_corr)
```

	Correlación
Entire_home_apartment	0.178520
Private_room	0.113685
Shared_room	0.082269
Hotel_room	-0.618158

Finalmente, al analizar 'reviews_per_month' frente a 'review_scores_communication', encontré una relación positiva. Esto sugiere que los anfitriones que se comunican mejor tienden a recibir más reseñas mensuales, lo que refleja una experiencia más positiva por parte de los huéspedes.

```
[ ]: #REVIEWS_PER_MONTH Y REVIEW_SCORES_COMMUNICATION

correlaciones6 = {
    "Entire_home_apartment": Entire_home_apartment["reviews_per_month"].corr(Entire_home_apartment["review_scores_communication"]),
    "Private_room": Private_room["reviews_per_month"].corr(Private_room["review_scores_communication"]),
    "Shared_room": Shared_room["reviews_per_month"].corr(Shared_room["review_scores_communication"]),
    "Hotel_room": Hotel_room["reviews_per_month"].corr(Hotel_room["review_scores_communication"])
}

tabla_corr = pd.DataFrame.from_dict(correlaciones6, orient="index", columns=["Correlación"])
print(tabla_corr)
```

	Correlación
Entire_home_apartment	-0.100496
Private_room	-0.064008
Shared_room	-0.093410
Hotel_room	-0.400261

Además, se utilizó un mapa de calor para identificar las 10 variables con mayor correlación. De ese mismo modo, en el código para cada tipo de alojamiento filtré el dataset y calculé la matriz de correlación entre variables numéricas, me asegure de usar columnas numéricas únicamente, eliminando los duplicados y la diagonal para evitar que saliera una correlación entre la misma variable. Luego ordené las correlaciones absolutas para obtener las 10 variables con mayor correlación con la variable de interés y las presenté en una tabla.

Tabla de las 10 variables con mayor correlación

```
#Calculamos la correlación entre las variables en general de manera automatica
subsets = {
    "Entire_home_apartment": Entire_home_apartment,
    "Private_room": Private_room,
    "Shared_room": Shared_room,
    "Hotel_room": Hotel_room
}

def top10_corr(df):
    # Columnas numéricas únicamente
    num = df.select_dtypes(include=[np.number])
    if num.shape[1] < 2:
        return pd.DataFrame(columns=["Variable_1", "Variable_2", "Abs_Correlación"])

    corr = num.corr(method="pearson")
    corr_abs = corr.abs()

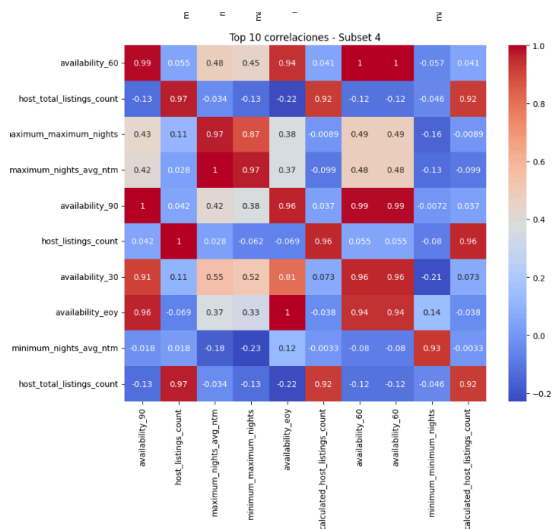
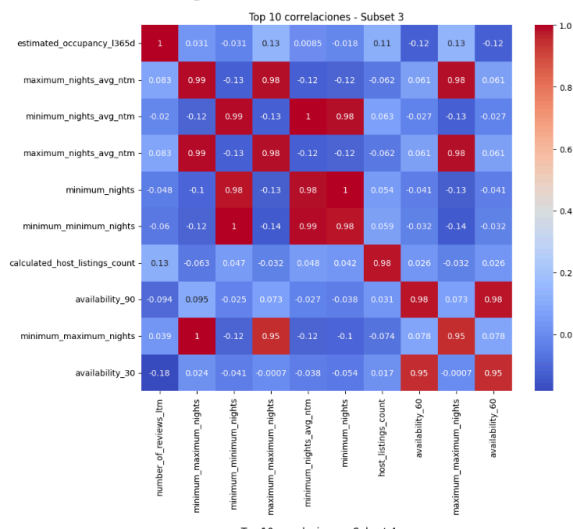
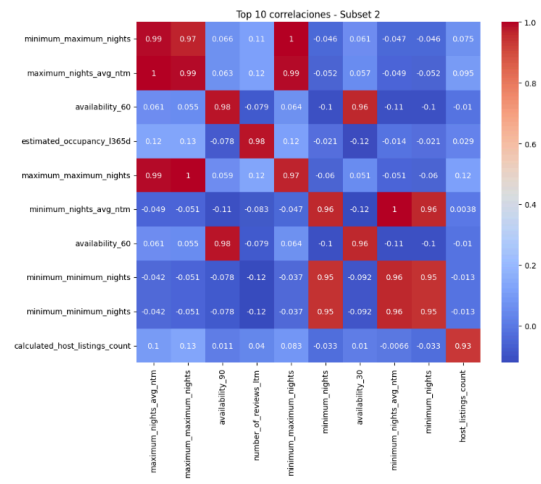
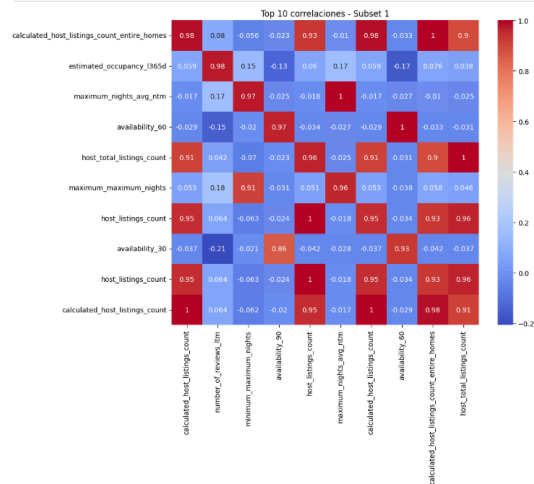
    # Se eliminan los duplicados y la diagonal para evitar correlaciones entre la misma variable
    mask = np.triu(np.ones_like(corr_abs, dtype=bool))
    corr_pairs = corr_abs.mask(mask).stack().sort_values(ascending=False)

    # Toma las 10 más altas
    top10 = corr_pairs.head(20).reset_index()
    top10.columns = ["Variable_1", "Variable_2", "Abs_Correlación"]
    return top10

# Mostrar las tablas para cada subset
top10_dict = {}
for name, sub in subsets.items():
    print("\n" + "="*80)
    print(f"Top 10 correlaciones absolutas - {name}")
    print("="*80)
    top10 = top10_corr(sub)
    top10_dict[name] = top10
    display(top10)
```

En cada tipo de alojamiento elegido, observé que las variables más relacionadas estaban principalmente vinculadas con la calidad del servicio y la comunicación del anfitrión. Esto refuerza la idea de que la interacción entre anfitrión y huésped tiene un papel clave en las evaluaciones y en la actividad de los alojamientos

```
In [ ]:
for i, subset in enumerate([entire_home_aprt, Private_room, shared_room, Hotel_room], start=1):
    plt.figure(figsize=(10,8))
    corr = subset.corr(numeric_only=True)
    top_vars = corr.abs().unstack().sort_values(ascending=False).drop_duplicates()[1:11].index
    sns.heatmap(corr.loc[top_vars.get_level_values(0), top_vars.get_level_values(1)], annot=True, cmap="coolwarm")
    plt.title(f"Top 10 correlaciones - Subset (i)")
    plt.show()
```



Regresión Lineal Múltiple

En la parte de regresión lineal múltiple busque crear el mejor modelo para cada variable cuantitativa en la base de datos, que son: review_scores_rating, host_acceptance_rate, host_is_superhost, host_total_listings_count, accommodates, bedrooms, price, review_scores_value, bathrooms y reviews_per_month.

Primero me asegure de que se estuviera seleccionando una variable numérica, para después sacar la correlación de la variable, en este caso, review_scores_rating con las demás variables, de modo que se pudieran seleccionar las 3 variables independientes con mayor correlación y crear el modelo a partir de eso.

Regresión Multiple

Modelo de regression lineal multiple para cada variable cuantitativa

```
1 [ ]: # Review_scores_rating

# Seleccionar solo columnas numéricas
num_cols = df.select_dtypes(include=np.number).columns

# Correlación de review_scores_rating con todas las numéricas
corr_review = df[num_cols].corrwith(df["review_scores_rating"]).sort_values(ascending=False)

print(corr_review)
```

review_scores_rating	1.000000
review_scores_accuracy	0.773660
review_scores_value	0.724238
review_scores_cleanliness	0.711896
review_scores_communication	0.642535
review_scores_checkin	0.601993
review_scores_location	0.400439
id	0.150113
price	0.107588

Con eso, después de declarar las variables independientes y dependiente, definí el modelo ajustando a su vez las variables para sacar los coeficientes de determinación y verificar los coeficientes de correlación. Por último, vemos el modelo matemático.

```
[ ]: from sklearn.linear_model import LinearRegression
import numpy as np

# Declaramos las variables independientes y dependiente
Vars_Indep = df[['review_scores_accuracy', 'review_scores_value', 'review_scores_cleanliness']]
Var_Dep = df['review_scores_rating']

# Definimos el modelo
model = LinearRegression()
# Ajustamos el modelo con las variables
model.fit(X=Vars_Indep, y=Var_Dep)

# Definimos los coeficientes e intercepto
print("Coeficientes obtenidos:", model.coef_)
print("Intercepto:", model.intercept_)

# Verificamos los coeficientes de determinación
coef_Deter = model.score(X=Vars_Indep, y=Var_Dep)
print("Coeficiente de determinación:", coef_Deter)

# Verificamos los coeficientes de correlación
coef_Correl = np.sqrt(coef_Deter)
print("Coeficiente de correlación:", coef_Correl)

# Vemos el modelo matemático
modelo_str = f"{Var_Dep.name} = {model.intercept_:.4f}"
for c, v in zip(model.coef_, Vars_Indep.columns):
    signo = "+" if c >= 0 else "-"
    modelo_str += f" {signo} {abs(c):.4f}*{v}"
print("Modelo matemático:")
print(modelo_str)
```

Podemos ver que las variables explican en conjunto el 71% de la variabilidad en la calificación general, con un coeficiente de correlación de 0.84 que muestra una relación fuerte y positiva. Entre ellas, la precisión es la que más influye en la puntuación total con un coeficiente 0.43, seguida de la limpieza 0.23 y el valor percibido 0.22, esto nos sugiere que los alojamientos con mejores evaluaciones en precisión, limpieza y valor tienden a recibir calificaciones generales significativamente más altas.

```
Coefficientes obtenidos: [0.42777344 0.21794589 0.22574528]
Intercepto: 0.6455669176948255
Coeficiente de determinación: 0.7091273884500802
Coeficiente de correlación: 0.842097018430822
Modelo matemático:
review_scores_rating = 0.6456 + 0.4278*review_scores_accuracy + 0.2179*review_scores_value + 0.2257*review_scores_cleanliness
```

Para las siguientes variables utilice el mismo método.

- host_acceptance_rate

```
Coefficientes obtenidos: [4.86696760e-01 5.66022349e-03 2.05235889e-04]
Intercepto: 0.29112957584221744
Coeficiente de determinación: 0.2710738894241317
Coeficiente de correlación: 0.5206475673852051
Modelo matemático:
host_acceptance_rate = 0.2911 + 0.4867*host_response_rate + 0.0057*number_of_reviews_ltm + 0.0002*estimated_occupancy_l365d
```

Podemos ver que la variable depende principalmente de host_response_rate, con un coeficiente de 0.4867 que muestra que, a mayor rapidez o disposición del anfitrión para responder, mayor es su tasa de aceptación. Las variables number_of_reviews_ltm y estimated_occupancy_l365d tienen un efecto positivo pero muy pequeño. Con un coeficiente de determinación de 0.2711 y un coeficiente de correlación de 0.5206, el modelo explica alrededor del 27% de la variabilidad en host_acceptance_rate, lo que indica una relación moderada y sugiere que existen otros factores no incluidos que también influyen en esta variable.

- host_is_superhost

```
Coefficientes obtenidos: [1.52049649e-03 7.24322718e-03 3.53901248e-06]
Intercepto: 0.1390872113126133
Coeficiente de determinación: 0.23461362874567226
Coeficiente de correlación: 0.48436931028469615
Modelo matemático:
host_is_superhost = 0.1391 + 0.0015*estimated_occupancy_l365d + 0.0072*number_of_reviews_ltm + 0.0000*estimated_revenue_l365d
```

El modelo se ve influido principalmente por number_of_reviews_ltm, con un coeficiente de 0.0072, lo que indica que un mayor número de reseñas se asocia ligeramente con una mayor probabilidad de que el anfitrión sea superhost. Las variables estimated_occupancy_l365d y estimated_revenue_l365d también tienen un efecto positivo pero no tan relevante. Con un coeficiente de determinación de 0.2346 y un coeficiente de correlación de 0.4844, el modelo explica cerca del 23% de la variabilidad en host_is_superhost, reflejando una relación débil a

moderada y sugiriendo que existen otros factores más relevantes para determinar si un anfitrión alcanza el estatus de superhost.

- host_total_listing_count

```
Coeficientes obtenidos: [ 1.46488753 -0.05708612  0.10720052]
Intercepto: -0.19326634786755026
Coeficiente de determinación: 0.9219449973765208
Coeficiente de correlación: 0.96017966932055
Modelo matemático:
host_total_listings_count = -0.1933 + 1.4649*host_listings_count - 0.0571*calculated_hos
t_listings_count + 0.1072*calculated_host_listings_count_entire_homes
```

host_listings_count es la que presenta una mayor relación, con un coeficiente de 1.4649, lo que indica que a medida que aumenta el número de propiedades registradas por el anfitrión, también crece significativamente el total de listados. Las variables calculated_host_listings_count y calculated_host_listings_count_entire_homes tienen efectos menores. Con un coeficiente de determinación de 0.9219 y un coeficiente de correlación de 0.9602, el modelo presenta un ajuste excelente, explicando más del 92% de la variabilidad en host_total_listings_count, lo que evidencia una relación muy fuerte entre las variables del modelo.

- accommodates

```
Coeficientes obtenidos: [0.94389345 0.60512636 0.20297077]
Intercepto: 0.8491712781277982
Coeficiente de determinación: 0.5918716881649602
Coeficiente de correlación: 0.7693319752648788
Modelo matemático:
accommodates = 0.8492 + 0.9439*bedrooms + 0.6051*beds + 0.2030*bathrooms
```

El modelo de regresión indica que accommodates está principalmente determinado por bedrooms, seguido de beds y bathrooms, lo que sugiere que el número de habitaciones es el factor más importante para determinar cuántas personas puede alojar una propiedad. Con un intercepto de 0.8492, el modelo tiene un coeficiente de determinación de 0.5919 y un coeficiente de correlación de 0.7693, explicando aproximadamente el 59% de la variabilidad en accommodates.

- bedrooms

```
Coeficientes obtenidos: [0.15232952 0.10274264 0.42697096]
Intercepto: 0.051278209329793434
Coeficiente de determinación: 0.6087950383965693
Coeficiente de correlación: 0.7802531886487676
Modelo matemático:
bedrooms = 0.0513 + 0.1523*accommodates + 0.1027*beds + 0.4270*bathrooms
```

El modelo de regresión muestra que bedrooms está influido principalmente por bathrooms seguido de accommodates y beds, indicando que a mayor número de baños, mayor es la cantidad de habitaciones de la propiedad. Con un intercepto de 0.0513, el modelo presenta un coeficiente de determinación de 0.6088 y un coeficiente de correlación de 0.7803, explicando

alrededor del 61% de la variabilidad en bedrooms y mostrando una relación entre las variables predictoras y el número de habitaciones.

- price

```
Coeficientes obtenidos: [ 75.50376882  19.63416089 119.79533308]
Intercepto: 53.10234598687293
Coeficiente de determinación: 0.3360275947851551
Coeficiente de correlación: 0.5796788721224494
Modelo matemático:
price = 53.1023 + 75.5038*bedrooms + 19.6342*accommodates + 119.7953*bathrooms
```

La variable price está principalmente influido por bathrooms, seguido de bedrooms y accommodates, lo que sugiere que la cantidad de baños tiene el mayor impacto en el precio de la propiedad, mientras que el número de habitaciones y la capacidad de alojamiento también aumentan el precio, pero en menor medida. Con un intercepto de 53.1023, el modelo tiene un coeficiente de determinación (R^2) de 0.3360 y un coeficiente de correlación de 0.5797, explicando aproximadamente el 34% de la variabilidad en price, lo que indica una relación moderada y sugiere que otros factores no incluidos también influyen en el precio.

- review_scores_value

```
Coeficientes obtenidos: [0.56306288 0.36868291 0.14945948]
Intercepto: -0.478204728963604
Coeficiente de determinación: 0.5687698375356742
Coeficiente de correlación: 0.7541683084933192
Modelo matemático:
review_scores_value = -0.4782 + 0.5631*review_scores_rating + 0.3687*review_scores_accuracy + 0.1495*review_scores_cleanliness
```

El modelo de regresión sugiere que la calificación general del alojamiento es el factor más determinante para la percepción de valor, mientras que la precisión y la limpieza también contribuyen positivamente pero en menor medida. Con un intercepto de -0.4782, el modelo tiene un coeficiente de determinación de 0.5688 y un coeficiente de correlación de 0.7542, explicando aproximadamente el 57% de la variabilidad en review_scores_value y mostrando una relación moderada a fuerte entre las variables.

- bathrooms

```
Coeficientes obtenidos: [0.39906093 0.04350349 0.0005866 ]
Intercepto: 0.40399297569493575
Coeficiente de determinación: 0.49216886944978755
Coeficiente de correlación: 0.701547481963828
Modelo matemático:
bathrooms = 0.4040 + 0.3991*bedrooms + 0.0435*accommodates + 0.0006*price
```

El modelo de regresión muestra que el número de habitaciones es el factor más relevante para definir la cantidad de baños en la propiedad. Con un intercepto de 0.4040, el modelo presenta un coeficiente de determinación de 0.4922 y un coeficiente de correlación de 0.7015, explicando cerca del 49% de la variabilidad en bathrooms y mostrando una relación moderada entre las variables predictoras y el número de baños.

- reviews_per_month

```

Coeficientes obtenidos: [ 0.06852936 -0.00300428  0.31616376]
Intercepto: 0.4796484507885932
Coeficiente de determinación: 0.7047014125868905
Coeficiente de correlación: 0.8394649561398561
Modelo matemático:
reviews_per_month = 0.4796 + 0.0685*number_of_reviews_ltm - 0.0030*estimated_occupancy_1
365d + 0.3162*number_of_reviews_l30d

```

Estos resultados sugieren que la cantidad de reseñas recientes y acumuladas impulsa la frecuencia mensual de reseñas, mientras que la ocupación estimada tiene un impacto casi despreciable. Con un intercepto de 0.4796, el modelo presenta un coeficiente de determinación de 0.7047 y un coeficiente de correlación de 0.8395, explicando aproximadamente el 70% de la variabilidad en reviews_per_month y mostrando una relación fuerte entre las variables predictoras y la frecuencia de reseñas.

Conclusiones y hallazgos generales

En general, los modelos mostraron que algunas variables, como el número de habitaciones y el número de personas que puede alojar la propiedad (accommodates), tienen una relación fuerte con el precio. Esto significa que, en general, a mayor capacidad del alojamiento, mayor es el precio. También observé que la variable 'review_scores_value' suele estar influenciada por la calificación general del alojamiento, lo que indica que los huéspedes tienden a evaluar el valor con base en su satisfacción global.

Al comparar los coeficientes de la regresión múltiple con las correlaciones simples del heatmap, noté que algunos valores cambiaban ligeramente de signo o intensidad. Esto ocurre porque en la regresión múltiple las variables se analizan de forma conjunta, considerando el efecto de todas al mismo tiempo. Por ejemplo, una variable que parecía importante en la correlación simple puede perder relevancia cuando se incluyen otras que explican mejor la variación.

En general, el modelo múltiple tuvo un buen ajuste, con valores de R^2 aceptables para explicar las variaciones de las variables dependientes. Aunque no todos los predictores fueron significativos, sí se logró identificar qué factores tienen más peso en el comportamiento de los alojamientos. Esto demuestra que el análisis múltiple permite entender de manera más completa la relación entre las características del anfitrión, el alojamiento y las opiniones de los huéspedes.

En conclusión, el análisis de regresión lineal simple me permitió observar relaciones directas entre pares de variables, como la conexión entre comunicación y reseñas, o la disponibilidad y el número de comentarios. Por otro lado, la regresión múltiple me ayudó a ver una visión más completa, considerando cómo varias variables influyen al mismo tiempo. Aprendí que, aunque algunas correlaciones simples parecen fuertes, es necesario usar modelos más complejos para entender realmente el comportamiento de los datos.