



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Puebla

Actividad 2.1 Regresión lineal simple y múltiple

José Manuel Morales Escalante

Materia:

Analítica de datos y herramientas de inteligencia artificial I

Fecha:

06 de abril de 2025

El documento ipynb realiza un análisis exploratorio y de regresión lineal sobre datos de Airbnb en Río de Janeiro. El análisis se centra en buscar los mejores modelos de regresión de las propiedades según sus características, segmentando los datos por tipo de habitación. El reporte incluye visualizaciones de correlación y preparación de datos para modelos predictivos.

1. host_acceptance_rate VS. host_response_rate

Interpretación general:

- Relación entre la tasa de aceptación de reservas y la rapidez de respuesta del anfitrión.
- Coeficiente positivo indica que anfitriones que responden rápido tienden a aceptar más reservas.

Resultados por tipo de habitación:

| Tipo de habitación | Coeficiente | R ² | Interpretación |
|--------------------|-------------|----------------|---|
| Entire home/apt | 0.263 | 0.069 | Relación débil pero significativa. |
| Private room | 0.285 | 0.081 | Similar a Entire home, pero ligeramente más fuerte. |
| Shared room | 0.102 | 0.010 | Casi no hay relación. |
| Hotel room | -0.150 | 0.022 | Relación inversa (inusual, posiblemente por muestra pequeña). |

Conclusión:

La correlación es más fuerte en **Private rooms**, sugiriendo que anfitriones profesionales (que gestionan varias propiedades) mantienen altas tasas de respuesta y aceptación. En **Hotel rooms**, el signo negativo podría deberse a gestión automatizada.

2. review_scores_location VS. review_scores_cleanliness

Interpretación general:

- Examina si propiedades en mejores ubicaciones tienden a estar más limpias.
- Coeficiente positivo esperado (mejor ubicación → más cuidado en limpieza).

Resultados:

| Tipo de habitación | Coeficiente | R ² | Interpretación |
|--------------------|-------------|----------------|--|
| Entire home/apt | 0.311 | 0.097 | Relación moderada. |
| Private room | 0.288 | 0.083 | Similar a Entire home. |
| Shared room | 0.210 | 0.044 | Relación más débil. |
| Hotel room | 0.450 | 0.202 | La más fuerte (hoteles en áreas turísticas invierten más en limpieza). |

Conclusión:

La relación es más fuerte en **Hotel rooms**, donde la ubicación premium está asociada a altos estándares de limpieza. En **Shared rooms**, la débil correlación sugiere que la limpieza depende más del huésped anterior que del anfitrión.

3. host_acceptance_rate VS. price

Interpretación general:

- Evalúa si anfitriones con alta tasa de aceptación cobran precios distintos.
- Coeficiente negativo sugeriría que anfitriones "flexibles" bajan precios para atraer reservas.

Resultados:

| Tipo de habitación | Coeficiente | R ² | Interpretación |
|--------------------|-------------|----------------|----------------------------------|
| Entire home/apt | -0.024 | 0.001 | Casi nula (no afecta el precio). |
| Private room | -0.042 | 0.002 | Relación muy débil. |

| | | | |
|--------------------|--------|-------|---|
| Shared room | 0.105 | 0.011 | Ligero aumento de precio con mayor aceptación (contrario a lo esperado). |
| Hotel room | -0.300 | 0.090 | Relación inversa significativa (hoteles económicos aceptan más reservas). |

Conclusión:

Solo en **Hotel rooms** hay una relación clara: menor precio → mayor tasa de aceptación. En otros tipos, la decisión de precio parece independiente de la flexibilidad del anfitrión.

4. availability_365 vs. number_of_reviews

Interpretación general:

- Propiedades con más disponibilidad deberían tener más reseñas (más tiempo en el mercado).
- Coeficiente positivo esperado.

Resultados:

| Tipo de habitación | Coeficiente | R ² | Interpretación |
|------------------------|-------------|----------------|---|
| Entire home/apt | 0.180 | 0.032 | Relación débil. |
| Private room | 0.250 | 0.062 | Más fuerte (mayor rotación de huéspedes). |
| Shared room | 0.310 | 0.096 | La más fuerte (alta disponibilidad = más huéspedes = más reseñas). |
| Hotel room | -0.100 | 0.010 | Relación inversa (hoteles con alta demanda limitan disponibilidad). |

Conclusión:

En **Shared rooms**, la alta disponibilidad genera más reseñas. En **Hotel rooms**, la relación inversa sugiere que los más populares están frecuentemente ocupados.

5. host_acceptance_rate vs. number_of_reviews

Interpretación general:

- Anfitriones que aceptan más reservas deberían acumular más reseñas.
- Coeficiente positivo esperado.

Resultados:

| Tipo de habitación | Coeficiente | R ² | Interpretación |
|--------------------|-------------|----------------|--|
| Entire home/apt | 0.150 | 0.022 | Relación débil. |
| Private room | 0.210 | 0.044 | Moderada (más aceptación = más huéspedes). |
| Shared room | 0.050 | 0.003 | Casi nula. |
| Hotel room | -0.200 | 0.040 | Relación inversa (hoteles con baja aceptación tienen reseñas por gestión profesional). |

Conclusión:

En **Private rooms**, la relación es clara: aceptar más reservas genera más reseñas. En **Hotel rooms**, las reseñas pueden venir de canales externos (ej. agencias de viajes).

6. reviews_per_month VS. review_scores_communication

Interpretación general:

- ¿Propiedades con más reseñas mensuales tienen mejor comunicación?
- Coeficiente positivo sugeriría que buena comunicación impulsa más reseñas.

Resultados:

| Tipo de habitación | Coeficiente | R ² | Interpretación |
|--------------------|-------------|----------------|------------------|
| Entire home/apt | 0.085 | 0.007 | Relación mínima. |
| Private room | 0.120 | 0.014 | Ligera mejora. |

| | | | |
|--------------------|--------|-----------|---|
| Shared room | 0.250 | 0.06 2 | Fuerte relación (comunicación clave en habitaciones compartidas). |
| Hotel room | -0.050 | 0.00 3 | Sin relación clara. |

Conclusión:

En **Shared rooms**, la comunicación es crucial para generar reseñas. En **Hotel rooms**, las reseñas dependen más de otros factores (ej. limpieza, ubicación).

Reporte de Modelos de Regresión

Múltiple con R² Ajustado

1. Modelos por Tipo de Habitación

1.1. Entire Home/Apt

| Variables Incluidas | R ² Ajustado | Explicación |
|---|----------------------------|--|
| accommodates + bedrooms + host_is_superhost | 0.48 | El modelo explica el 48% de la variación en precios |
| + review_scores_value | 0.51 | Añadir valor percibido mejora ligeramente el modelo |
| + host_total_listings_count | 0.52 | Incluir el número de propiedades del anfitrión aporta poco |
| Modelo completo | 0.53 | Máximo poder explicativo alcanzado |

Interpretación:

Las características físicas (bedrooms, accommodates) son los principales predictores. El estatus Superhost y las valoraciones añaden capacidad explicativa limitada.

1.2. Private Room

| Variables Incluidas | R ² Ajustado | Explicación |
|----------------------------------|-------------------------|---|
| accommodates + host_is_superhost | 0.32 | Modelo base con capacidad moderada |
| + review_scores_value | 0.38 | Las valoraciones mejoran significativamente el modelo |
| + reviews_per_month | 0.39 | Frecuencia de reseñas aporta poco |
| Modelo completo | 0.40 | Techo explicativo para este tipo |

Patrón:

La reputación (review_scores_value) tiene mayor peso relativo que en Entire Homes, mientras que el tamaño pierde importancia.

1.3. Shared Room

| Variables Incluidas | R ² Ajustado | Explicación |
|-----------------------|-------------------------|--|
| host_is_superhost | 0.35 | El estatus del anfitrión es clave |
| + review_scores_value | 0.45 | Las valoraciones duplican el poder explicativo |
| + reviews_per_month | 0.47 | Las reseñas frecuentes añaden poco |
| Modelo completo | 0.48 | Buen ajuste considerando la simplicidad |

Dinámica única:

Las variables sociales (reputación, estatus) dominan completamente sobre los atributos físicos.

1.4. Hotel Room

| Variables Incluidas | R ² Ajustado | Explicación |
|---------------------------|-------------------------|--|
| host_total_listings_count | 0.45 | El tamaño de la cadena es el mejor predictor |
| + review_scores_value | 0.52 | Las valoraciones añaden poder explicativo |
| + host_is_superhost | 0.53 | El estatus aporta mínimo valor |
| Modelo completo | 0.55 | Máxima capacidad explicativa |

Observación clave:

Los hoteles muestran patrones distintos, donde la escala operativa (host_total_listings_count) es más importante que factores individuales.

2. Comparativa General de Poder Explicativo

| Tipo de Habitación | R ² Ajustado Máximo | Variables Clave |
|--------------------|--------------------------------|---|
| Entire Home/Apt | 0.53 | Tamaño (bedrooms, accommodates) + Superhost |
| Private Room | 0.40 | Reputación (review_scores_value) + Capacidad |
| Shared Room | 0.48 | Reputación + Estatus Superhost |
| Hotel Room | 0.55 | Escala operativa (listings_count) + Valor percibido |

○