

**Benemérita Universidad Autónoma de Puebla**

**FACULTAD DE CIENCIAS DE LA COMPUTACIÓN**

**Herramientas web**

**Juan Pablo García Garibay**

**Luis Enrique Becerril Diaz**

**Jesus Rodríguez Hernández**

**Índice:**

[**Introducción** 3](#_Toc194356278)

[**qué se usa** 4](#_Toc194356279)

[**Proceso** 6](#_Toc194356280)

[**Variables:** 7](#_Toc194356281)

[**Análisis:** 9](#_Toc194356282)

[**Revisión** 10](#_Toc194356283)

[**Evaluación** 11](#_Toc194356284)

[**Conclusión** 12](#_Toc194356285)

# **Introducción**

El objetivo de este análisis es examinar cómo se comportan las variables en las bases de datos de Airbnb para comparar el rendimiento de la plataforma en diferentes países. Este enfoque busca identificar diferencias y similitudes en el desempeño de los hosts, con especial interés en México, ya que se ha observado una posible disminución en el rendimiento de Airbnb en el país tras la pandemia, en comparación con otras ciudades a nivel internacional.

Por ello, estamos evaluando el comportamiento de los datos de diversas ciudades, incluyendo México, con el propósito de predecir el desempeño de los hosts y entender en qué aspectos varían. Los resultados de este análisis permitirán implementar mejoras significativas para optimizar la calidad del servicio ofrecido por los hosts en México.

Las cuidades a analizar son: México, Seattle y Denver

# **qué se usa**

se usan diversas librerías, además de las bases de datos respectivas de las ciudades antes mencionadas (México, Seattle, Denver) así como un compilador el cual cuenta con el lenguaje de programación Python incluyendo varias librerías las cuales nos ayudaron al proceso que llevamos a cabo las cuales son:

1. Pandas : Se utiliza para la manipulación y análisis de datos estructurados (como bases de datos o tablas). Es ideal para cargar, limpiar, y realizar operaciones en conjuntos de datos.
2. Numpy : Proporciona herramientas para cálculos matemáticos y manejo de matrices y arreglos multidimensionales. Es fundamental para realizar operaciones numéricas de alta eficiencia.
3. Matplotlib: Esta librería permite la creación de gráficos y visualizaciones, como histogramas, gráficos de líneas, y otros tipos de diagramas.
4. Scipy,especial: Incluye funciones matemáticas avanzadas como funciones gamma, logarítmicas y exponenciales que son útiles en modelos matemáticos complejos.
5. Scipy.optimize.curve\_fit: Se utiliza para ajustar curvas a los datos mediante métodos de optimización no lineal.
6. Seaborn: Facilita la creación de visualizaciones de datos estadísticos con gráficos más elegantes y fáciles de interpretar (como mapas de calor y gráficos de distribución).
7. Sklearn.metrics\_score: Calcula la métrica R² (coeficiente de determinación), que mide qué tan bien un modelo de regresión ajusta los datos.
8. Sklearn.model\_section.train\_test\_split: Se usa para dividir un conjunto de datos en datos de entrenamiento y prueba, lo que es crucial para entrenar y evaluar modelos predictivos.
9. Sklearn.processing.standarScaler: Escalador que transforma los datos para que tengan una media de 0 y desviación estándar de 1, útil para normalizar datos antes de aplicar algoritmos de aprendizaje.

Texto

El contenido generado por IA puede ser incorrecto.

# **Proceso**

Después de importar las librerías se abren las bases de datos respectivas de las cuidades, se hace una copia en un dataframe, al cual le vamos a quitar columnas innecesarias.

Pars poder manejar los datos de manera adecuada se transforman los tipos de daros de datos, ya sea a valores numéricos o categóricos, ya sea el caso, para poder transformarlos a variables de tipo diptonica, la cual solo poder dos valores, a diferencia de una variable de tipo Bool, este puede tomar más valores aparte de 0 y 1, por ejemplo, ‘’masculino’’ y ‘’femenino’’, ‘’mayor de edad’’, ‘’menor de edad’’, la cual nos va a servir para hacer la regresión logística asignando este tipo de variable a las variables dependientes, y las vamos a manejar como ‘’X’’. Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.

# **Variables:**

Las variables que ocupamos para el análisis son las siguientes, las cuales vamos a proseguir a describirlas:

* + host response rate: Porcentaje de mensajes respondidos por el anfitrión en un tiempo determinado.
  + host aseptance rate: Porcentaje de solicitudes de reserva aceptadas por el anfitrión.
  + host is superhost: Indica si el anfitrión tiene la distinción de "Superhost", lo que refleja un alto nivel de rendimiento y satisfacción del cliente.
  + host is identify: Confirma si la identidad del anfitrión ha sido verificada por Airbnb.
  + room\_type: Tipo de alojamiento ofrecido, como "entire home/apt", "private room" o "shared room".
  + has\_availability: Indica si el alojamiento está disponible en las fechas seleccionadas.
  + Acomodates: Número de personas que el alojamiento puede hospedar cómodamente.
  + Price: Costo por noche del alojamiento
  + instant bookable: Especifica si el alojamiento se puede reservar instantáneamente sin necesidad de aprobación previa del anfitrión
  + review score rating: Calificación promedio de las reseñas que los huéspedes han dejado para el alojamiento.

# **Análisis:**

Se analizan los datos de las bases de datos, así como los valores únicos de las columnas como lo son room type y bathroom text. Y se verifica la distribución de valores categóricos, para así poder hacer la regresión logística, de las cuales escogemos las variables dependientes, Host\_is\_superhost y como variable independiente Price, el cual es el modelo uno, de las cuales podemos obtener una matriz de confusión, porcentaje de Precisión, exactitud y sensibilidad.

Interfaz de usuario gráfica, Texto

El contenido generado por IA puede ser incorrecto.

# **Revisión**

Los datos obtenidos de las variables indepnedinetes son las características o factores que se obtiene o que influyen en el resultado, es decir, cada vez que se corra el código va a soltar un análisis diferente por lo cual solo tomamos la primera corrida de código, y en este caso son:

* review\_scores\_rating
* review\_scores\_communication
* review\_scores\_checkin

las cuales se ocupan en los análisis que hemos hecho, ademas aquí se está repitiendo el proceso de entrenamiento y evaluación del modelo para validar y ajustar su rendimiento. Esto generalmente implica dividir los datos en conjuntos de entrenamiento y prueba, entrenar el modelo con los datos de entrenamiento y evaluar su precisión o eficacia con los datos de prueba.

# 

# **Evaluación**

Después de todo el proceso se presentan los resultados los cuales se presenta en:

* Matriz de confusión.
* Precisión del modelo.
* Exactitud.
* Sensibilidad.

Captura de pantalla de computadora

El contenido generado por IA puede ser incorrecto.

# **Conclusión**

Estas variables son fundamentales para evaluar y predecir el rendimiento de los hosts de Airbnb en diferentes ciudades y países. Analizarlas permite identificar patrones y aspectos clave que influyen en la calidad del servicio y la satisfacción de los huéspedes.

1. Host Response Rate y Host Acceptance Rate: La rapidez y frecuencia con la que los hosts responden y aceptan reservas son indicadores cruciales de su profesionalismo y disponibilidad. Mejorar estas métricas puede aumentar la confianza de los huéspedes
2. Host is Superhost y Host Identity Verified: Ser un "Superhost" y tener la identidad verificada son factores que generan mayor credibilidad y satisfacción, influyendo directamente en la percepción de los huéspedes
3. Accommodates y Price: La capacidad de alojamiento y los precios competitivos son elementos clave para atraer distintos tipos de usuarios. Analizar la relación entre precio y comodidad puede optimizar la estrategia de cada host
4. Instant Bookable: La posibilidad de reservar instantáneamente mejora la experiencia del usuario, promoviendo decisiones rápidas y prácticas.
5. Review Score Rating: Las calificaciones reflejan la calidad del servicio y son una métrica crítica para evaluar el éxito de los hosts y sus áreas de mejora.

Usar estas variables para comparar el comportamiento de los hosts en México frente a los de otras ciudades permite identificar oportunidades de mejora en el desempeño .Por ejemplo, podría destacarse la importancia de mejorar el tiempo de respuesta y la comunicación con los huéspedes para incrementar las calificaciones y atracción de nuevos clientes.

Si necesitas un modelo detallado o técnicas para implementar el análisis con estas variables, ¡puedo ayudarte!