

# Análisis de Sentimiento de Evaluaciones Airbnb

Luis Felipe Rangel

Enero 2025

## 1 Introducción

El presente reporte analiza las evaluaciones de los huéspedes de Airbnb para el alojamiento **"Wanderer's Delight & Stays"**, un espacio que promete combinar encanto, comodidad y experiencias memorables. Desde sus pintorescas ubicaciones hasta sus extraordinarias comodidades, cada reseña refleja las vivencias sinceras de los visitantes.

El objetivo principal de este análisis es entender la percepción de los huéspedes mediante técnicas de análisis de sentimiento, identificando tendencias positivas y negativas que contribuyen a la experiencia general. Este enfoque permite no solo evaluar la satisfacción de los clientes, sino también proporcionar recomendaciones basadas en los datos para optimizar aún más la oferta del alojamiento.

A lo largo del reporte, se detallarán los pasos metodológicos utilizados, los hallazgos principales y las implicaciones prácticas derivadas del análisis, con el fin de contribuir al desarrollo estratégico y al crecimiento continuo de *Wanderer's Delight & Stays*.

Se toman los datos de Kaggle: <https://www.kaggle.com/datasets/omarsobhy14/airbnbreviews>

## 2 Metodología

El análisis de sentimiento realizado en este estudio se llevó a cabo siguiendo una serie de pasos que se describen a continuación:

- **Tokenización:** Se dividieron las reseñas en unidades más pequeñas, llamadas tokens, para facilitar su procesamiento y análisis.
- **Eliminación de palabras irrelevantes:** Se eliminaron las palabras vacías (*stopwords*) en inglés, que no aportan significado relevante para el análisis, tales como "the", "and", "is", entre otras.
- **Asignación de valores numéricos:** Se utilizó un diccionario léxico basado en el recurso AFINN, obtenido del repositorio de GitHub: <https://raw.githubusercontent.com/fnielsen/afinn-165.txt>. Este diccionario asigna un valor numérico positivo o negativo a cada token, representando su carga emocional.

Con los valores obtenidos para cada token, se calculó un puntaje de sentimiento agregado para cada reseña. Posteriormente, estas reseñas fueron clasificadas como positivas o negativas, según el signo del puntaje final.

Finalmente, se realizó una comparación entre las clasificaciones obtenidas mediante el análisis y los valores reales de sentimiento presentes en la base de datos original. Esto permitió evaluar la precisión del modelo y determinar en qué proporción las reseñas fueron clasificadas correctamente como positivas o negativas.

### 3 Resultados

A continuación, se presenta una muestra de los resultados obtenidos para las reseñas analizadas. La tabla incluye información sobre el género de la reseña (*Genre*), su clasificación como positiva o negativa (*Positive or Negative*), los tokens extraídos, el número de reseña asociado y los valores de sentimiento asignados a cada término.

Genre	Pos or Neg	Token	Rvw Nmbr	Término	Sentimiento
Location	1.0	perfect	0	perfect	3
Location	1.0	popular	0	popular	3
Location	1.0	attractions	0	attractions	2
Cleanliness	1.0	outstanding	1	outstanding	5
Cleanliness	1.0	spotless	1	spotless	2
Neighborhood	1.0	vibrant	2	vibrant	3
Security	1.0	safe	3	safe	1
Security	1.0	secure	3	secure	2
Pet Friendliness	1.0	like	4	like	2
Host	1.0	helpful	5	helpful	2
Cleanliness	0.0	better	7	better	2
Neighborhood	0.0	noisy	8	noisy	-1
Pet Friendliness	0.0	allow	10	allow	1

Table 1: Muestra de los resultados para las primeras reseñas analizadas.

El análisis de sentimiento realizado sobre la totalidad de las reseñas arrojó una proporción de coincidencia del **84%** entre el sentimiento real de las reseñas y el sentimiento predicho. Esto indica un desempeño favorable del modelo para identificar correctamente la polaridad emocional de las reseñas de los huéspedes.

Estos resultados destacan la utilidad de las técnicas de análisis de sentimiento para obtener información valiosa sobre la percepción de los huéspedes, lo que puede servir para mejorar continuamente la experiencia ofrecida por el alojamiento.