

eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

# SISTEMAS DE APOYO A LA DECISIÓN

GRADO EN INGENIERÍA INFORMÁTICA DE GESTIÓN Y SISTEMAS DE  
INFORMACIÓN

---

2024-2025

---

**Fecha:**

Bilbao, a 25 de mayo de 2025

**Autores:**

Bogajo López, María  
Corral Martín, Ander  
López Sánchez, Peio  
Cotano Romero, Aitor

# Índice general

<b>1. Tableau: Análisis de los Datos iniciales</b>	<b>5</b>
1.1. Contexto . . . . .	5
1.2. Tipo de datos . . . . .	5
1.3. Preprocesado de los datos . . . . .	7
1.4. Gráficos y análisis . . . . .	7
1.4.1. Análisis de la relación entre precio y satisfacción en España y Portugal . . . . .	7
1.4.2. Distribución geográfica de la satisfacción en España y Portugal . . . . .	9
1.4.3. Análisis del cumplimiento del ratio calidad/precio por tipología de alojamiento . . . . .	10
1.4.4. Impacto de las tasas adicionales en la satisfacción del usuario . . . . .	13
1.4.5. Evolución de la satisfacción de los usuarios en España y Portugal a lo largo del tiempo . . . . .	14
1.4.6. Evolución de las valoraciones positivas y negativas en España y Portugal . . . . .	15
1.4.7. Análisis semántico de las reviews: términos frecuentes por sentimiento . . . . .	16
<b>2. Clasificación de sentimientos: Análisis, Preproceso y Experimentación</b>	<b>19</b>
2.1. Graphical Abstract de la solución . . . . .	19
2.2. Datos . . . . .	22
2.2.1. División entre Train, Dev y Test de los datos para entrenar el modelo de predicción del <i>rating</i> . . . . .	22
2.2.2. Descripción del preproceso . . . . .	22
2.2.3. Resultados de la tarea de clasificación . . . . .	25
2.2.4. Resultados de la aplicación de los modelos generativos . . . . .	27
2.3. Algoritmos, link a la documentación y nombre de los hiperparámetros empleados . . . . .	29
2.3.1. Experimentación: Algoritmos empleados y Breve Descripción . . . . .	29
2.3.2. Resultados sobre el Development . . . . .	30
2.3.3. Discusión sobre el proceso de aprendizaje . . . . .	31
2.3.4. Conclusión sobre la tarea de clasificación . . . . .	31
2.4. Algoritmos, link a la documentación y nombre de los hiperparámetros empleados . . . . .	32
2.4.1. Experimentación: Algoritmos empleados y Breve Descripción . . . . .	32
2.4.2. Resultados . . . . .	33
2.4.3. Discusión sobre los descubrimientos realizados . . . . .	35
<b>3. Despliegue y código entregado</b>	<b>37</b>
3.1. Estructura del código entregado . . . . .	37
3.2. Ejecución . . . . .	38
3.3. Requisitos técnicos . . . . .	38
3.4. Observaciones finales . . . . .	39
<b>4. Conclusiones</b>	<b>40</b>
<b>Bibliografía</b>	<b>41</b>

# Índice de figuras

1.1. Relación Precio/Satisfacción de España. . . . .	8
1.2. Relación Precio/Satisfacción de Portugal. . . . .	8
1.3. Distribución de satisfacción en España. . . . .	9
1.4. Distribución de satisfacción en Portugal. . . . .	10
1.13. Palabras positivas más frecuentes en España. . . . .	16
1.14. Palabras positivas más frecuentes en Portugal. . . . .	17
1.15. Palabras negativas más frecuentes en España. . . . .	17
1.16. Palabras negativas más frecuentes en Portugal. . . . .	17
2.1. Flujo esquemático del sistema de clasificación de sentimientos. . . . .	21
2.2. Comparativa de palabras frecuentes en reseñas positivas. . . . .	33
2.3. Comparativa de palabras frecuentes en reseñas negativas. . . . .	34

# Índice de cuadros

1.1. Descripción de los atributos seleccionados del conjunto de datos. . . . .	6
2.1. División Train, Dev y Test de los datos de <code>AirBnBReviews.csv</code> (binario) . . . . .	22
2.2. División Train, Dev y Test de los datos de <code>TripadvisorHotelReviews.csv</code> (multiclase) .	22
2.3. Comentarios totales de los datasets de España y Portugal. . . . .	22
2.4. Resultados sobre el conjunto Dev de <code>AirBnBReviews.csv</code> (clasificación binaria) . . . . .	25
2.5. Resultados sobre el conjunto Dev de <code>trip_advisor_hotel_reviews.csv</code> (multiclase) . . .	25
2.6. Ejemplo de clasificación KNN + BERT . . . . .	26

# Acrónimos

- **LR**: Logistic Regression
- **XGB**: XGBoost
- **MNB**: Multinomial Naive Bayes
- **BoW**: Bag of Words
- **Tf-Idf**: Term frequency – Inverse document frequency

# 1. Tableau: Análisis de los Datos iniciales

## 1.1. Contexto

El presente proyecto se enmarca en una simulación empresarial en la que formamos parte del equipo de análisis de datos de una región gestionada dentro del ecosistema de AirBnB. A raíz de una pérdida de competitividad percibida en los últimos meses, surge la necesidad de realizar un diagnóstico riguroso que permita entender el estado actual del mercado local. Para ello, se dispone de un conjunto de datos compuesto por información detallada sobre propiedades turísticas, evaluaciones de usuarios y reseñas escritas, provenientes tanto de nuestra zona como de un área competidora directa.

Estos datos, recogidos a lo largo de los últimos años, reflejan una gran variedad de aspectos relevantes: desde características físicas de los alojamientos y precios, hasta valoraciones globales y comentarios subjetivos de los huéspedes. Su análisis nos permitirá detectar tanto fortalezas como debilidades de nuestra oferta.

En este caso concreto, el análisis se centra en comparar la zona de España, a la que representamos, con la región de Portugal, que actúa como nuestro principal rival dentro del entorno de AirBnB. El objetivo es obtener una visión clara de las diferencias clave entre ambas zonas, desde el punto de vista de los clientes, y detectar qué estrategias podrían ayudarnos a mejorar nuestro posicionamiento relativo.

El objetivo principal de este estudio es convertir los datos disponibles en conocimiento accionable que permita a la dirección tomar decisiones estratégicas. Se busca entender el posicionamiento relativo de nuestra zona respecto a Portugal, analizar los factores que influyen en la satisfacción de los clientes, y detectar patrones que expliquen los comentarios positivos y negativos. A través del uso de técnicas de análisis de sentimientos, clasificación automatizada y visualizaciones interactivas con Tableau, se pretende construir una narrativa que explique el estado actual del mercado y ofrezca pistas claras sobre cómo mejorar nuestra posición en términos de reputación, ocupación y rentabilidad.

## 1.2. Tipo de datos

Para el desarrollo de este proyecto se nos ha proporcionado un conjunto amplio de datos que incluye información muy diversa sobre alojamientos turísticos listados en AirBnB. No obstante, dado el enfoque comparativo de este estudio, no todos los atributos disponibles resultan relevantes o útiles para el análisis planteado.

En esta sección se describen únicamente aquellas variables que han sido seleccionadas por su valor informativo y su potencial para establecer comparaciones significativas entre la zona española y la portuguesa. Se ha priorizado la inclusión de campos relacionados con la experiencia del usuario, las características de las propiedades, las valoraciones recibidas y la localización, ya que son los que permiten extraer conclusiones más directamente vinculadas a los objetivos del estudio.

A continuación, se presenta una tabla con los atributos seleccionados, acompañados de una breve descripción y su clasificación según criterios estadísticos y de tipo de dato.

Atributo	Descripción	Tipo de dato	Tipo estadístico	Nivel de medición
_id	Identificador único del alojamiento.	int	Cuantitativo	Discreto
name	Título del anuncio publicado.	string	Cualitativo	Nominal
property_type	Tipo de propiedad (casa, apartamento, etc.).	string	Cualitativo	Nominal
room_type	Tipo de habitación ofertada (entera, compartida, etc.).	string	Cualitativo	Nominal
last_scraped	Fecha de la última recopilación del anuncio.	date	Cuantitativo	Ordinal
calendar_last_scraped	Fecha de última actualización del calendario.	date	Cuantitativo	Ordinal
first_review	Fecha de la primera reseña.	date	Cuantitativo	Ordinal
last_review	Fecha de la última reseña.	date	Cuantitativo	Ordinal
number_of_reviews	Total de reseñas recibidas.	int	Cuantitativo	Discreto
price	Precio por noche.	int	Cuantitativo	Discreto
security_deposit	Monto del depósito de seguridad.	float	Cuantitativo	Discreto
cleaning_fee	Tarifa de limpieza adicional.	float	Cuantitativo	Discreto
address	Dirección del alojamiento. Incluye ciudad, región y país.	json	Cualitativo	Nominal
review_scores	Valoraciones sobre distintos aspectos del alojamiento.	json	Cuantitativo	Ordinal
reviews	Comentarios de huéspedes con fecha y texto.	json	Cualitativo	Nominal

1.1. Cuadro: Descripción de los atributos seleccionados del conjunto de datos.

## 1.3. Preprocesado de los datos

Antes de poder realizar los análisis y visualizaciones en Tableau, ha sido necesario llevar a cabo una serie de tareas de limpieza, transformación y enriquecimiento de los datos originales. Este preprocesado ha sido clave para asegurar la calidad del análisis y extraer información realmente útil. A continuación, se describen los pasos más relevantes que se han seguido:

- **Eliminación de columnas irrelevantes o redundantes:** El dataset original contenía una gran cantidad de variables, muchas de ellas sin utilidad directa para los objetivos del proyecto. Estas columnas fueron eliminadas para simplificar el análisis y evitar ruido.
- **Unificación de origen:** Como los datos procedían de diferentes fuentes (España y Portugal), se creó una nueva columna llamada **País** para poder diferenciarlos en Tableau y permitir comparativas directas.
- **Extracción y estructuración de la localización:** En los datos de dirección, se separaron los campos en dos elementos clave:
  - **Ciudad**, extraída de la dirección completa, para facilitar análisis geográficos por zona.
  - **Coordenadas (latitud y longitud)**, para representar cada alojamiento en mapas interactivos y analizar patrones espaciales.
- **Reestructuración de las reseñas:** En las reviews, los comentarios venían agregados. Se realizó una separación individual por reseña, asociándola correctamente con su fecha y el alojamiento correspondiente, lo cual permitió análisis temporales como evolución por año o mes.
- **Análisis de sentimiento:** Una parte fundamental del proyecto ha sido clasificar automáticamente las reseñas según su polaridad. Para ello:
  - Se procesaron los comentarios con un modelo de análisis de sentimiento, que asignó a cada uno una puntuación entre 1 y 9 (cuanto mayor, más positiva la reseña).
  - Además, se generó una variable categórica **Sentimiento**, con valores **positivo** o **negativo**, que ha permitido realizar nubes de palabras y comparativas.
- **Preparación para Tableau:** Finalmente, todos los campos relevantes fueron homogeneizados en formato, renombrados para claridad y exportados en CSV limpios. Esto permitió integrarlos fácilmente en Tableau y facilitar filtros, cálculos y visualizaciones personalizadas.

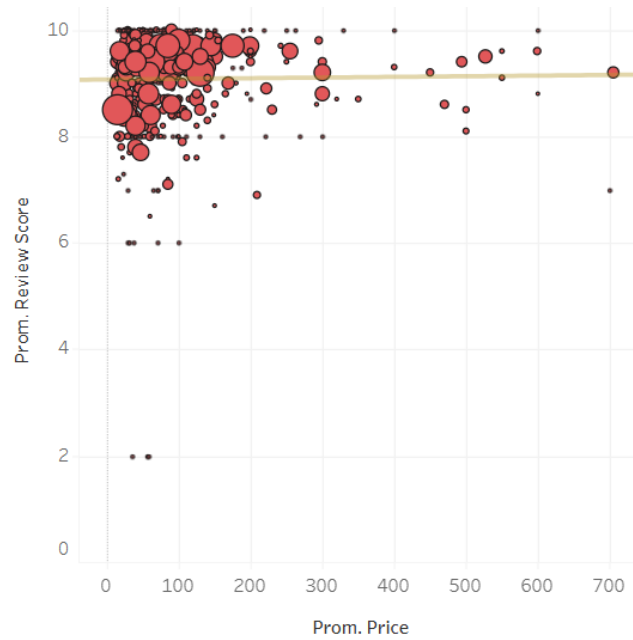
Gracias a este proceso, los datos están en un estado óptimo para su análisis visual y comparativo. Se ha garantizado una estructura coherente, una segmentación clara por país y sentimiento, y la posibilidad de explorar tanto aspectos cuantitativos como cualitativos de la experiencia de los usuarios.

## 1.4. Gráficos y análisis

### 1.4.1. Análisis de la relación entre precio y satisfacción en España y Portugal

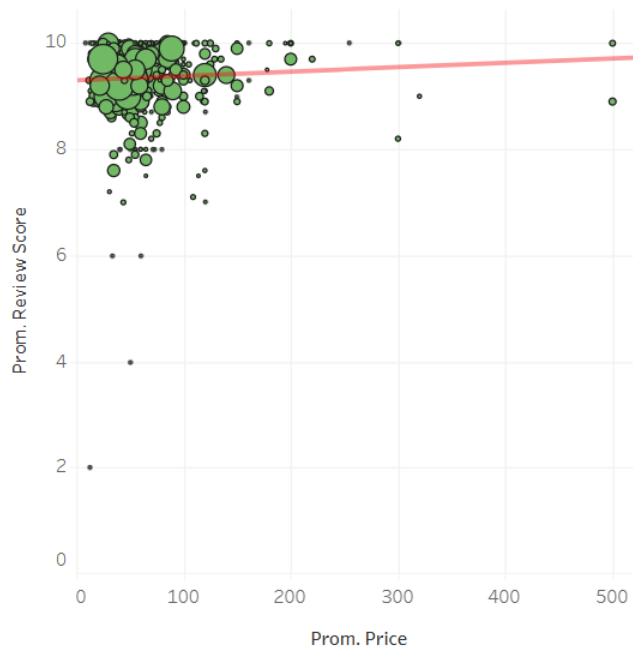
La siguiente visualización representa la relación entre el precio medio de los alojamientos y la puntuación media de satisfacción dada por los usuarios. Cada punto del gráfico corresponde a una propiedad individual. El eje horizontal indica el precio medio (*Prom. Price*) y el eje vertical, la puntuación media de satisfacción (*Prom. Review Score*). El tamaño del punto está asociado al número de reviews que tiene cada propiedad: cuanto más grande el círculo, más comentarios tiene, lo que aporta mayor fiabilidad a los datos. Además, se incluye una *línea de tendencia lineal*, que permite observar si existe correlación entre ambas variables.





1.1. Figura: Relación Precio/Satisfacción de España.

En el caso de España, el gráfico muestra una nube densa de puntos concentrada entre los 0,€ y los 150,€ de precio medio, y puntuaciones de satisfacción mayoritariamente superiores a 8. El tamaño de muchos puntos indica que varias propiedades cuentan con un número considerable de valoraciones, lo que aporta fiabilidad a los promedios calculados. Al aplicar una línea de tendencia lineal, se obtiene una pendiente de aproximadamente 0,00012, con un coeficiente de determinación  $R^2 \approx 0,00016$ . Esto indica que no existe una relación lineal significativa entre el precio y la satisfacción del usuario. En otras palabras, pagar más no garantiza una mejor experiencia según las valoraciones de los clientes. Además, se observa una gran concentración de alojamientos con precios bajos y puntuaciones muy altas (entre 8 y 10), lo que sugiere que muchas propiedades económicas están cumpliendo e incluso superando las expectativas. También se detectan algunos outliers, propiedades con precios significativamente más altos pero con puntuaciones similares o incluso inferiores, lo que indica insatisfacción en la relación calidad/precio.



1.2. Figura: Relación Precio/Satisfacción de Portugal.

En el caso de Portugal, el análisis muestra una ligera diferencia respecto a España. La nube de puntos también se concentra principalmente en rangos de precio entre 0,€ y 150,€, con puntuaciones de satisfacción mayoritariamente superiores a 8. Sin embargo, la línea de tendencia lineal tiene una pendiente algo más pronunciada, con un valor aproximado de 0,00083, y un coeficiente de determinación de  $R^2 \approx 0,00305$ . Aunque la relación sigue siendo débil y no concluyente desde el punto de vista estadístico, estos valores sugieren una mayor coherencia entre lo que los usuarios pagan y lo que perciben en términos de calidad. También se observa una menor presencia de outliers, lo que podría indicar una oferta más homogénea y equilibrada.

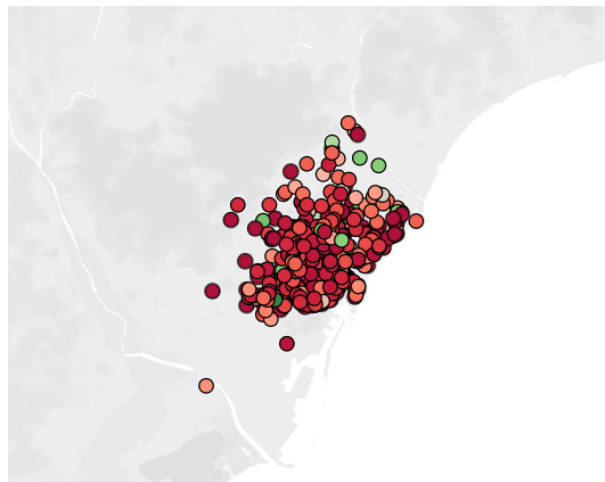
Todo ello sugiere que, aunque ambas zonas ofrecen buenas experiencias a los usuarios, Portugal presenta una relación ligeramente más coherente entre el precio y la satisfacción, lo que podría interpretarse como una oferta más estructurada o una gestión más profesionalizada de los alojamientos.

### Recomendaciones para España.

- Analizar las propiedades outlier con precios elevados y puntuaciones bajas para identificar fallos concretos (limpieza, ubicación, atención, etc.).
- Reforzar el valor diferencial de los alojamientos económicos que ya ofrecen alta satisfacción, comunicando claramente sus puntos fuertes en los anuncios.
- No basar las estrategias de precio únicamente en subir tarifas: al no existir una relación significativa entre precio y satisfacción, subir precios sin mejorar servicios puede perjudicar la percepción de valor.
- Tomar como referencia el modelo portugués, donde se percibe una mayor alineación entre lo que se paga y lo que se recibe. Esto implica trabajar en la coherencia entre precio y calidad, destacando atributos que realmente impactan en la satisfacción.
- Profundizar en atributos específicos de satisfacción (p.ej. limpieza o ubicación) mediante análisis adicionales que permitan identificar qué variables concretas están afectando a la puntuación general.

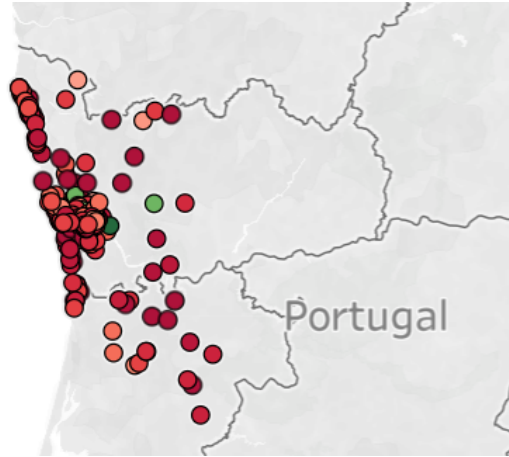
### 1.4.2. Distribución geográfica de la satisfacción en España y Portugal

La siguiente visualización representa un mapa geográfico de las propiedades analizadas en España y Portugal, donde cada punto corresponde a un alojamiento ubicado según sus coordenadas de latitud y longitud. El color de cada punto representa una medida de correlación entre el precio y la satisfacción del usuario: los tonos verdes indican una relación más equilibrada (alta satisfacción con un precio ajustado), mientras que los tonos rojizos representan situaciones donde el precio podría no estar justificado por la satisfacción obtenida.



1.3. Figura: Distribución de satisfacción en España.

En el caso español, el mapa muestra una fuerte concentración de alojamientos en zonas urbanas de Barcelona. Predominan los tonos rojizos, lo que indica que muchas propiedades tienen una relación calidad/precio poco equilibrada. No obstante, también se observan algunos puntos verdes dispersos, especialmente en zonas periféricas o alejadas del centro, lo que revela la existencia de alojamientos con buena relación precio/satisfacción fuera de los focos turísticos más caros. Esta distribución sugiere que las zonas de mayor densidad pueden estar saturadas o tener expectativas no cumplidas por parte de los usuarios.



1.4. Figura: Distribución de satisfacción en Portugal.

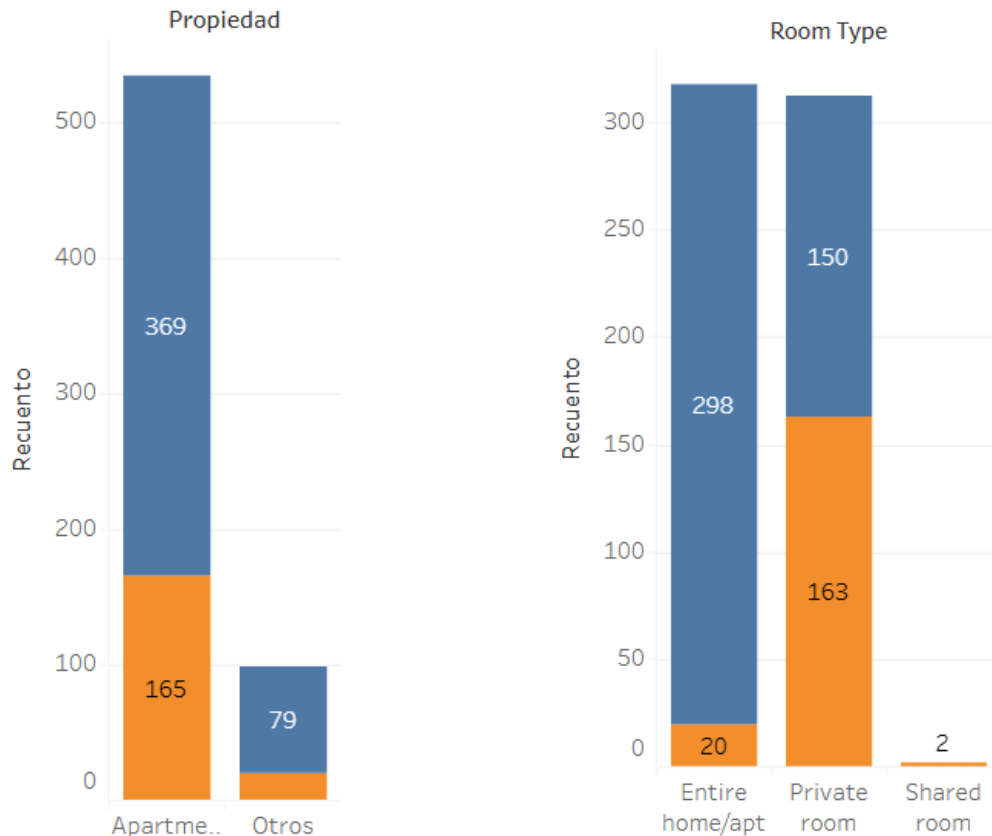
En Portugal, el mapa muestra una concentración de alojamientos en el área de Porto. Aunque también predominan los tonos rojizos, se aprecia una mayor variedad cromática con presencia de verdes en mayor proporción que en España. Esto refuerza la conclusión obtenida en el gráfico anterior: existe una mayor coherencia entre el precio pagado y la satisfacción recibida. La distribución geográfica más dispersa de puntos positivos sugiere una oferta más homogénea en términos de calidad y expectativas.

Cabe destacar que este análisis geográfico no contradice las conclusiones anteriores basadas en la regresión lineal. Ambos enfoques apuntan a que en España, pagar más no implica mejor satisfacción, mientras que en Portugal hay una leve alineación entre precio y calidad.

A partir de este mapa, se recomienda a los responsables de alojamientos en las zonas españolas donde predominan los tonos rojizos realizar un análisis localizado. Es fundamental identificar qué factores concretos podrían estar provocando una baja valoración por parte de los usuarios (por ejemplo, mala ubicación, deficiente atención al cliente o un mal servicio poco). Con estos datos, sería aconsejable ajustar tanto los precios como las condiciones ofertadas, para conseguir una relación más coherente entre el coste y la calidad. Esta medida contribuiría a reducir la diferencia entre las expectativas y la experiencia real del usuario, y a mejorar la posición competitiva frente a zonas con mayor equilibrio como las observadas en Portugal.

#### 1.4.3. Análisis del cumplimiento del ratio calidad/precio por tipología de alojamiento

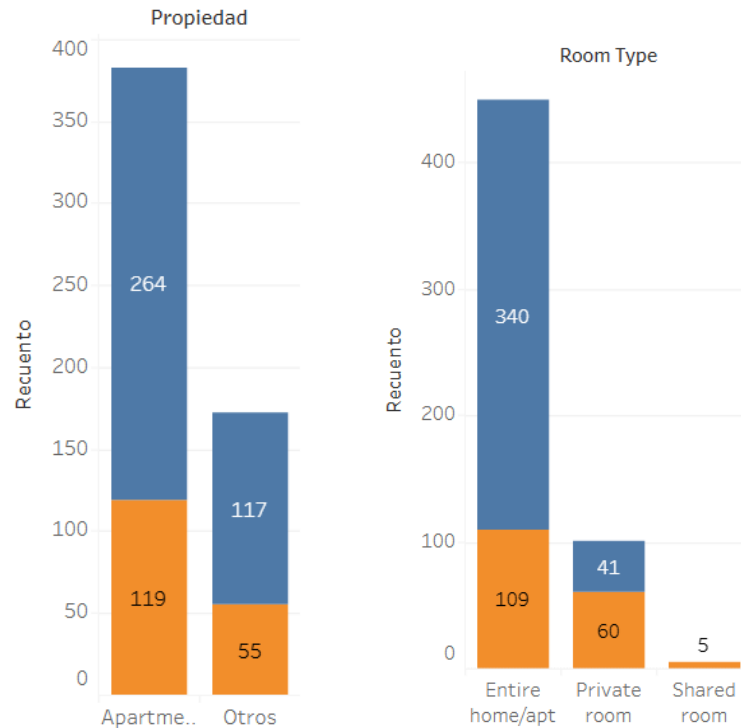
La siguiente visualización se centra en identificar qué tipos de alojamientos presentan un mejor equilibrio entre el precio que se cobra y la satisfacción que generan en los usuarios. Para ello, se ha calculado un *ratio calidad/precio* para cada propiedad, estableciendo un umbral objetivo para considerar que un alojamiento “cumple” con una buena relación. A partir de ese criterio binario (sí/no), se han creado representaciones agrupadas por tipo de alojamiento y tipo de habitación para analizar tendencias generales en España y Portugal.



(a) Cumplimiento ratio calidad/precio por tipo de propiedad en España. (b) Cumplimiento ratio calidad/precio por tipo de habitación en España.

En el caso de España, el análisis revela que la gran mayoría de alojamientos son apartamentos, con más de 500 propiedades analizadas. Dentro de esta categoría, 165 alojamientos cumplen con el ratio de calidad/precio esperado, mientras que 369 no lo hacen. Aunque hay una parte importante que sí lo consigue, la proporción es claramente desfavorable, lo que indica un margen importante de mejora. En la categoría “otros”, compuesta por tipos de alojamiento minoritarios como guest suites, casas o estudios, el patrón es similar: solo 20 propiedades aproximadamente cumplen, frente a unas 79 que no lo hacen. Este comportamiento sugiere que el desafío de equilibrar precio y satisfacción está presente de forma transversal en todas las tipologías, aunque es especialmente visible en los apartamentos por su volumen.

Desde la perspectiva del *tipo de habitación*, los resultados en España también muestran diferencias relevantes. Las habitaciones privadas tienen un desempeño notablemente mejor, con 163 alojamientos cumpliendo frente a 150 que no, lo que indica que este formato puede ofrecer un valor competitivo al usuario. En cambio, los alojamientos enteros muestran cifras inversas: 298 no cumplen frente a solo 20 que sí, una señal clara de que, a menudo, el precio de estas propiedades no está justificado por la satisfacción que generan. Las habitaciones compartidas son muy pocas y apenas aportan información.



(a) Cumplimiento ratio calidad/- precio por tipo de propiedad en Portugal. (b) Cumplimiento ratio calidad/- precio por tipo de habitación en Portugal.

En el caso de Portugal, los patrones son similares pero ligeramente más equilibrados. En la categoría de apartamentos, 119 alojamientos cumplen frente a 264 que no. Aunque la proporción sigue siendo desfavorable, se aprecia un ligero mejor desempeño en comparación con España. En el grupo “otros”, la diferencia también es menos significativa, con 55 propiedades cumpliendo y 117 sin cumplir. Esto podría interpretarse como una señal de mayor consistencia en la oferta portuguesa.

Respecto al tipo de habitación, la diferencia es aún más clara: en Portugal, 109 propiedades de alojamiento completo cumplen con el ratio de calidad/precio, frente a 340 que no lo hacen, lo que representa una proporción algo más favorable que en España. Las habitaciones privadas presentan una relación muy equilibrada: 60 cumplen y 41 no, lo cual refuerza la idea de que este tipo de estancia podría estar ofreciendo una excelente percepción de valor. Al igual que en España, las habitaciones compartidas tienen un peso mínimo.

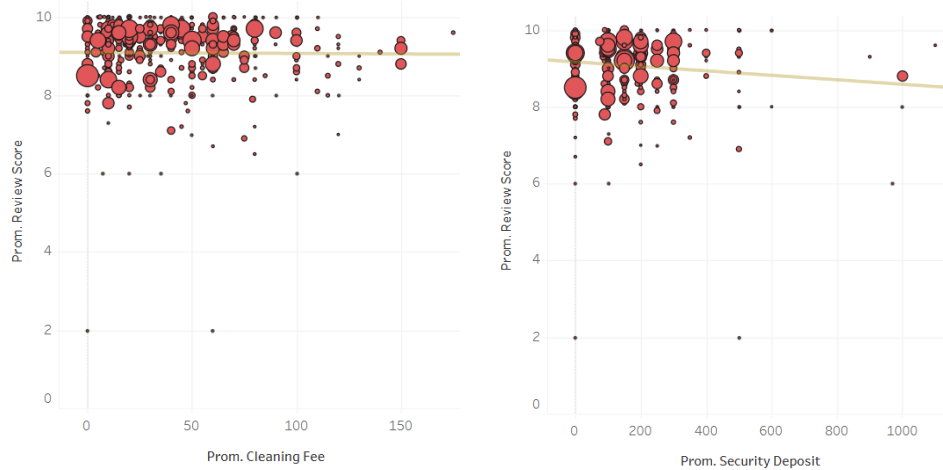
El análisis sugiere que el problema de desequilibrio entre lo que se cobra y lo que se entrega está más acentuado en España que en Portugal, especialmente en el caso de los apartamentos y los alojamientos completos. Esto puede deberse a precios demasiado altos, expectativas no cumplidas o falta de diferenciación en el servicio.

- Se recomienda a los gestores de alojamientos en España revisar su estrategia de precios, especialmente en apartamentos y estancias completas, ajustando el coste a la calidad real que se está ofreciendo.
- Las habitaciones privadas destacan como una opción altamente competitiva en ambos países, por lo que podría incentivarse este modelo de negocio o incorporarlo como línea complementaria.
- Sería útil analizar con más detalle qué factores concretos hacen que algunos alojamientos cumplan y otros no, como limpieza, ubicación, amenities, etc., para replicar buenas prácticas.
- La segmentación de tipos de propiedades debería guiar acciones de marketing y mejora: no es lo mismo comunicar un apartamento turístico que una habitación privada o una casa rural.

- Finalmente, se podría explorar si los alojamientos que no cumplen presentan una peor gestión de expectativas (fotos engañosas, descripciones poco precisas, etc.), ya que esto también impacta en la satisfacción.

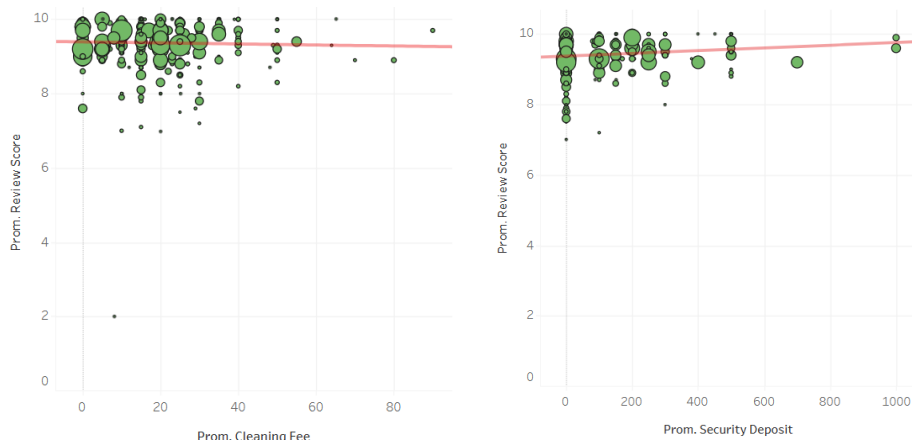
#### 1.4.4. Impacto de las tasas adicionales en la satisfacción del usuario

¿Influyen el depósito de seguridad y las tasas de limpieza en la satisfacción de los clientes? Esta sección analiza si estas tasas extras tienen algún efecto en la percepción de calidad reflejada en las puntuaciones medias de las propiedades.



(a) Satisfacción según tasa de limpieza en España. (b) Satisfacción según depósito de seguridad en España.

En el caso español, los resultados muestran que ni el depósito de seguridad ni la tasa de limpieza presentan una relación estadísticamente significativa con la satisfacción media. En el primer caso, la regresión entre la *Cleaning Fee* y el *Review Score* arroja una pendiente de aproximadamente  $-0,00026$ , un coeficiente de determinación  $R^2$  de  $9,5 \times 10^{-5}$  y una constante de 9,10821. Esto implica que no existe evidencia alguna de que cobrar una tasa de limpieza más alta mejore (ni empeore) la percepción del alojamiento. En cuanto al *Security Deposit*, la pendiente es también negativa ( $-0,00058$ ), con un valor  $R^2$  de 0,0114 y una constante de 9,18019. Aunque este resultado sugiere una posible relación inversa entre el importe del depósito y la satisfacción del usuario, el coeficiente de determinación tan bajo indica que dicha influencia es muy débil.



(a) Satisfacción según tasa de limpieza en Portugal. (b) Satisfacción según depósito de seguridad en Portugal.

La situación en Portugal es similar. En el caso de la *Cleaning Fee*, la pendiente obtenida es  $-0,00138$ , con un coeficiente de determinación  $R^2 = 0,00074$  y una constante de 9,40618, lo cual confirma la falta de relación entre dicha tasa y la satisfacción. Por otro lado, el *Security Deposit* muestra una

pendiente ligeramente positiva (0,00038), con un coeficiente  $R^2 = 0,0100$  y una constante de 9,39156. Estos datos indican que, aunque la relación sigue siendo muy débil, en Portugal el depósito de seguridad no parece perjudicar la experiencia del usuario e incluso podría estar vinculado a estándares más altos de profesionalización del alojamiento.

En conjunto, los resultados muestran que las tasas adicionales no son factores determinantes en la percepción de satisfacción del cliente. En España, un depósito elevado puede generar una ligera insatisfacción, mientras que en Portugal este efecto es más neutro o incluso levemente positivo. Por tanto, se recomienda que los anfitriones españoles revisen cuidadosamente sus políticas de depósitos, asegurándose de que estén justificadas por el valor añadido que ofrecen, y que comuniquen de forma clara y transparente las condiciones asociadas a estas tasas.

Dado que estos costes adicionales no contribuyen significativamente a aumentar la satisfacción, es fundamental no utilizarlos como fuente principal de ingresos adicionales. En su lugar, se sugiere trabajar en aspectos tangibles como la limpieza real del alojamiento, la comunicación con el huésped y la calidad de los servicios ofrecidos, que sí tienen un mayor impacto comprobado en la valoración final del cliente.

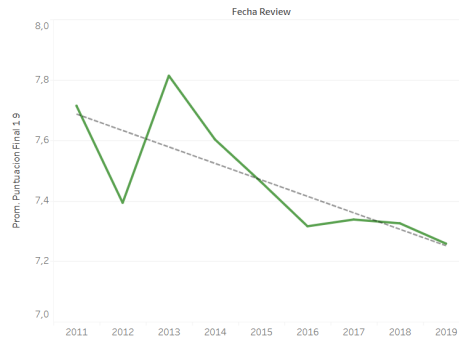
#### 1.4.5. Evolución de la satisfacción de los usuarios en España y Portugal a lo largo del tiempo

La siguiente visualización muestra la evolución temporal de la satisfacción media de los usuarios en España y Portugal, a partir del análisis de sentimiento aplicado a las reseñas. Se ha utilizado una escala de puntuación del 1 al 9, y se han representado los valores promedio por año y por mes. La línea de tendencia lineal permite observar posibles patrones descendentes o ascendentes a lo largo del tiempo, así como identificar estacionalidades que pudieran influir en la experiencia del cliente.

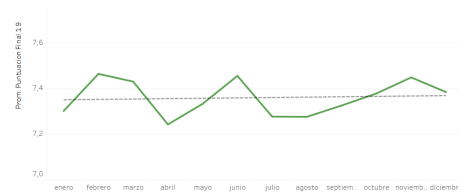


(a) Evolución satisfacción por años en España. (b) Evolución satisfacción por meses en España.

En el caso de España, los resultados reflejan una pendiente moderadamente negativa en las valoraciones medias anuales ( $\approx -0,0296$ ), con un coeficiente de determinación  $R^2 = 0,3035$  y una constante de 67,0955. Aunque no se alcanza una significación estadística clara, los datos sugieren una ligera disminución sostenida en la satisfacción desde 2011. A nivel mensual, la pendiente también es negativa pero mucho más leve ( $\approx -0,0048$ ), con un  $R^2 = 0,0804$ , lo cual indica que no existen patrones estacionales fuertes y que la percepción del usuario se mantiene relativamente estable a lo largo del año.



(a) Evolución satisfacción por años en Portugal.



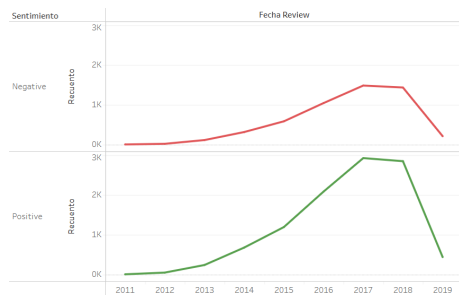
(b) Evolución satisfacción por meses en Portugal.

En el caso de Portugal, se observa una pendiente más acusada en la regresión anual ( $\approx -0,0544$ ), con un coeficiente de determinación significativamente mayor ( $R^2 = 0,5767$ ) y una constante de 117,06, lo que indica una disminución sostenida y más marcada en la satisfacción desde 2011. A nivel mensual, la pendiente es incluso positiva ( $\approx 0,0017$ ), pero el  $R^2 = 0,0059$  sugiere que no hay un patrón estacional relevante.

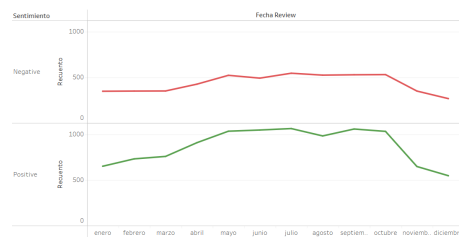
En resumen, España ha mantenido una satisfacción media más estable, con una ligera disminución no significativa, mientras que Portugal muestra una caída más clara y pronunciada a lo largo del tiempo. Esta diferencia representa una oportunidad para reforzar la percepción del valor ofrecido por los alojamientos españoles en un entorno competitivo.

#### 1.4.6. Evolución de las valoraciones positivas y negativas en España y Portugal

La siguiente visualización representa la evolución del número de reseñas clasificadas como positivas y negativas en España y Portugal, tanto a nivel anual como mensual. Estas opiniones han sido previamente etiquetadas mediante análisis de sentimiento, lo que permite detectar patrones en la percepción del usuario, identificar periodos críticos o de mejora, y comparar la estabilidad de ambos mercados.



(a) Evolución valoraciones por años en España.



(b) Evolución valoraciones por meses en España.

En el caso español, se observa un crecimiento estable de las reseñas positivas hasta 2017, seguido de una caída apreciable a partir de 2018. En contraste, las reseñas negativas se mantienen estables o incluso aumentan levemente en los últimos años. Este cambio sugiere una posible pérdida de calidad percibida o un aumento en las expectativas no satisfechas. A nivel mensual, las valoraciones positivas siguen siendo predominantes en todos los meses del año, aunque tienden a decrecer entre agosto y octubre. Las negativas, por otro lado, no presentan grandes picos, lo que descarta una fuerte estacionalidad como causa principal.





(a) Evolución valoraciones por años en Portugal. (b) Evolución valoraciones por meses en Portugal.

En el caso de Portugal, las reseñas positivas han crecido hasta 2017 y se han mantenido estables desde entonces. Las reseñas negativas se han mantenido bajas y sin aumentos notables. A nivel mensual, se repite un patrón muy parecido al de España: las opiniones positivas dominan el año entero, y no se detectan concentraciones estacionales de opiniones negativas. Esta estabilidad refuerza la percepción de una experiencia de usuario más consistente.

En conjunto, los datos anuales y mensuales confirman que la evolución del sentimiento es más favorable en Portugal, donde las valoraciones positivas superan de forma sostenida a las negativas sin grandes cambios. En cambio, en España se aprecia un deterioro progresivo en el volumen de reseñas positivas y una creciente convergencia con las negativas, especialmente en los últimos años.

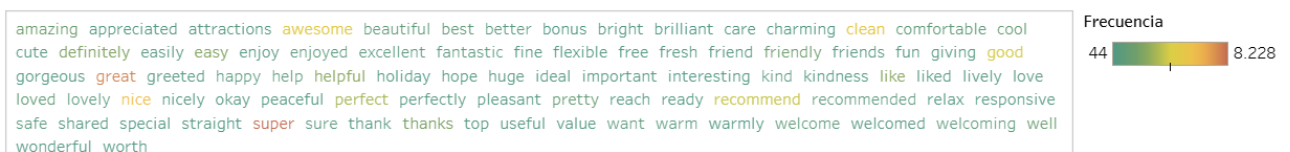
### Recomendaciones para España.

- Profundizar en las causas del descenso de reseñas positivas desde 2018 mediante un análisis de contenido de comentarios recientes.
- Reforzar las estrategias que inviten a los usuarios a dejar valoraciones si han tenido una experiencia positiva.
- Identificar puntos críticos de insatisfacción, como fallos en la gestión, limpieza o calidad del alojamiento.
- Tomar como referencia la consistencia observada en Portugal, implementando buenas prácticas que contribuyan a estabilizar y mejorar la percepción del usuario a lo largo del año.
- Garantizar que la comunicación y las expectativas del usuario estén alineadas con lo que realmente ofrece cada propiedad.

### 1.4.7. Análisis semántico de las reviews: términos frecuentes por sentimiento

Esta sección explora los patrones de lenguaje utilizados en las reseñas de usuarios, diferenciando entre comentarios positivos y negativos. Para ello, se han generado nubes de palabras con las 100 palabras más frecuentes clasificadas por país (España y Portugal) y por sentimiento (positivo o negativo). El tamaño de cada palabra representa su frecuencia relativa, y el color refleja la intensidad del sentimiento expresado. Este análisis ayuda a entender qué aspectos destacan los clientes cuando evalúan su experiencia.

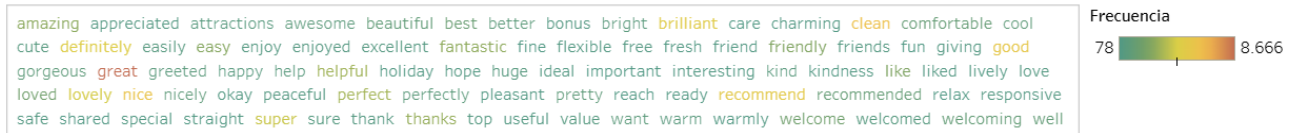
#### Reseñas positivas en España.



1.13. Figura: Palabras positivas más frecuentes en España.

Las palabras más frecuentes en los comentarios positivos de los usuarios españoles reflejan una clara orientación hacia el servicio y el confort. Términos como “*super*”, “*clean*”, “*helpful*”, “*great*”, “*recommend*” o “*nice*” muestran que los huéspedes valoran especialmente la limpieza, la atención del anfitrión y la calidad general de la estancia. Sin embargo, el vocabulario positivo es algo más limitado que en Portugal, con menos referencias emocionales o expresiones de entusiasmo, lo que sugiere una experiencia más funcional que emocional.

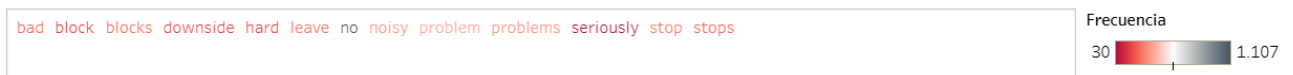
### Reseñas positivas en Portugal.



1.14. Figura: Palabras positivas más frecuentes en Portugal.

En comparación, las reseñas positivas en Portugal presentan una mayor riqueza léxica y emocional. Palabras como “*fantastic*”, “*brilliant*”, “*beautiful*”, “*ideal*”, “*excellent*” o “*enjoyed*” destacan por su frecuencia y connotaciones más intensas. Esto sugiere que los alojamientos portugueses están generando experiencias satisfactorias en un plano más emocional, lo cual representa una clara ventaja competitiva respecto a España.

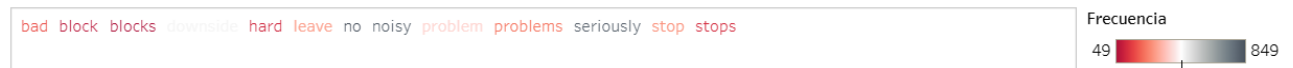
### Reseñas negativas en España.



1.15. Figura: Palabras negativas más frecuentes en España.

Las reseñas negativas españolas están marcadas por términos como “*problem*”, “*noisy*”, “*stop*”, “*hard*” o “*leave*”. La presencia de estas palabras indica que las críticas se centran principalmente en molestias concretas (ruido, dificultad, incomodidad), y que son suficientemente frecuentes como para ser tenidas en cuenta. Aunque el número total de palabras negativas no es alto, su peso sugiere problemas específicos que pueden resolverse con ajustes operativos relativamente simples.

### Reseñas negativas en Portugal.



1.16. Figura: Palabras negativas más frecuentes en Portugal.

En el caso portugués, los comentarios negativos contienen muchas de las mismas palabras, como “*problem*”, “*blocks*”, “*stop*” o “*hard*”, pero su frecuencia general es algo menor. Esto indica que, aunque existen puntos de mejora también en Portugal, los usuarios parecen vivir menos experiencias negativas, o al menos las expresan con menor frecuencia.

### Conclusiones y recomendaciones para España.

- Reforzar el vínculo emocional con los huéspedes: fomentar experiencias que generen memorabilidad, como en Portugal, puede enriquecer el vocabulario positivo usado por los clientes y mejorar su percepción.
- Ampliar el enfoque de los servicios ofrecidos: muchas palabras positivas portuguesas aluden a sensaciones y emociones, no solo a servicios funcionales. España debería buscar que sus alojamientos transmitan más que eficiencia: también hospitalidad y calidez.

- Atender los focos específicos de insatisfacción: problemas con el ruido o accesibilidad deben ser abordados prioritariamente, ya que son recurrentes en los comentarios negativos.
- Tomar inspiración en el lenguaje emocional del competidor: el uso frecuente en Portugal de términos como “*brilliant*”, “*beautiful*” o “*enjoyed*” refleja un tipo de experiencia distinta. Imitar este patrón puede ayudar a mejorar la reputación y competitividad de la zona.

## 2. Clasificación de sentimientos: Análisis, Pre-proceso y Experimentación

### 2.1. Graphical Abstract de la solución

La siguiente figura representa de manera esquemática el flujo completo del sistema de clasificación de sentimientos diseñado en este proyecto. El objetivo general ha sido procesar reseñas de alojamientos turísticos, predecir una puntuación asociada al sentimiento expresado, y permitir un análisis posterior por zonas.

El flujo de trabajo se compone de los siguientes pasos:

- **Fuentes de datos:** Se trabajó con dos conjuntos etiquetados para el entrenamiento y validación de modelos:
  - `AirBnBReviews.csv`, diseñado para una tarea de clasificación binaria (**positivo** / **negativo**).
  - `TripadvisorHotelReviews.csv`, con puntuaciones de 1 a 5, para clasificación multiclase.
- El conjunto central del estudio, `airbnb.csv`, contiene reseñas sin etiqueta que serán procesadas tras entrenar los modelos.
- **Traducción automática:** Las reseñas del fichero `airbnb.csv` se encontraban en varios idiomas. Para homogeneizarlas, se empleó un sistema de traducción automática utilizando el modelo generativo `gemma3:4b`, ejecutado localmente con `Ollama` e integrado mediante `LangChain`. Las reseñas fueron traducidas al inglés para facilitar la vectorización y el análisis por modelos preentrenados.
- **Preprocesado:** Todos los textos fueron sometidos a un pipeline común implementado en `limpiador.py`, `clasificador.py` y `sentimentAnalisis_v3.py`, que incluye:
  - Limpieza básica del texto (minúsculas, símbolos, espacios).
  - Vectorización mediante TF-IDF (`TfidfVectorizer`) y transformación a matrices dispersas.
  - Cálculo de características adicionales como longitud del texto, número de exclamaciones, proporción de mayúsculas o polaridad, normalizadas con `StandardScaler`.
  - Unión de todas las variables numéricas con la representación textual.
- **Modelos tradicionales:** Se entrenaron varios clasificadores utilizando los conjuntos etiquetados:
  - K-Nearest Neighbors (KNN)
  - Random Forest
  - `StackingClassifier`, combinando varios modelos base.

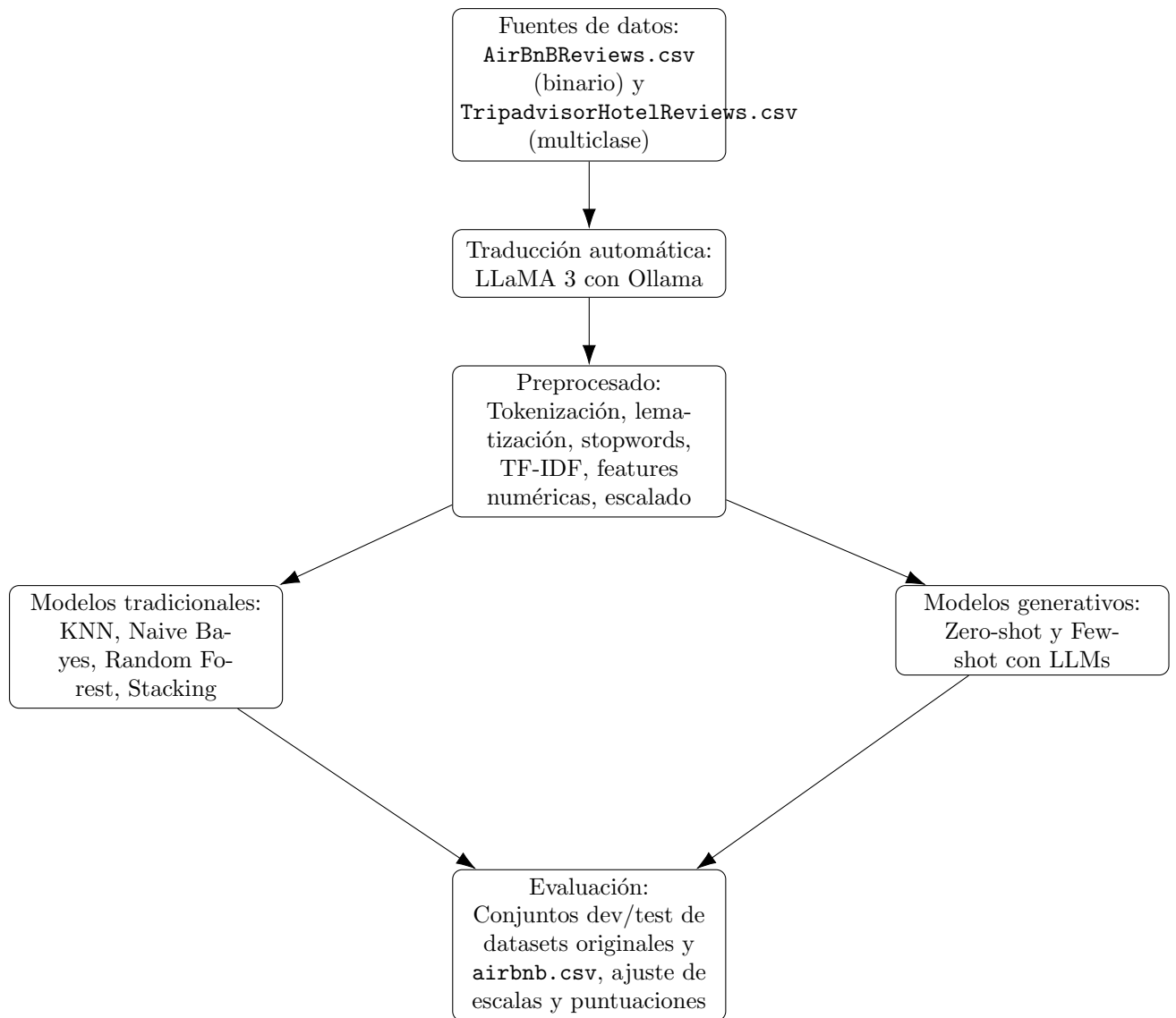
Los modelos fueron validados sobre conjuntos de desarrollo y prueba usando métricas como precisión, recall y F1-score (binaria), y desviación cuadrática (multiclase).

- **Modelos generativos:** Para la inferencia directa sobre reseñas sin etiqueta, se utilizaron:
    - El modelo LLaMA 3 formulando un *prompt* que pide una puntuación del 1 al 9.
- Prompts utilizados:**

- **0-shot:**  
*Give a rating from 1 to 9 for the following hotel review.*  
*Only answer with the number:*  
*Review: {comment}*  
*Rating:*
- **few-shots:**  
*Give a rating from 1 to 9 for the following hotel review. Only answer with the number.*  
*Example 1:*  
*Review: "We had a splendid time in the old centre of Porto. The apartment is very well situated next to the old Ribeira square. It's perfect to have such an apartment to your disposal, you feel home, and have a place to relax between the exploration of this very nice city. We thank Ana & Gonçalo, and we hope the apartment is free when we go back next year. Porto is charming original"*  
*Rating: 9*  
*Example 2:*  
*Review: "It's a nice flat. The location is good even though you have to walk a bit to the metro and quite a bit to the historic center. I liked that there was a restaurant in the same building."*  
*Rating: 6*  
*Example 3:*  
*Review: "Lot of Traffic noise due to old windows toward the street. And the guys where delayed more than 1 hour we where waiting"*  
*Rating: 2*  
*Example 4:*  
*Review: "The apartment was as we expected and I recommend it to other travelers."*  
*Rating: 5*  
*Example 5:*  
*Review: "It was our first time using Airbnb' and it was really good. Thanks for everything!"*  
*Rating: 8*  
*Now, rate this review:*  
*Review: {comment}*  
*Rating:*
- **1-shots:**  
*El mismo prompt que few-shots pero solo con el ejemplo 1.*
- **2-shots:**  
*El mismo prompt que few-shots pero con los ejemplos 2 y 3.*
- Modelos de Hugging Face mediante la API de **transformers**, usados sin entrenamiento adicional para análisis comparativo.

Estos enfoques no requieren vectorización previa y se evalúan por su capacidad de predecir directamente una puntuación.

- **Evaluación y análisis posterior:** Las predicciones de los modelos fueron comparadas con los valores reales en el conjunto `airbnb.csv` mediante cálculo de desviación cuadrática media. Posteriormente, las puntuaciones fueron agrupadas por alojamiento, barrio o zona, y se calcularon promedios por grupo utilizando los scripts `review_fusion.py` y `break_review_scores.py`.



2.1. Figura: Flujo esquemático del sistema de clasificación de sentimientos.

## 2.2. Datos

### 2.2.1. División entre Train, Dev y Test de los datos para entrenar el modelo de predicción del *rating*

Se han utilizado dos *datasets* distintos para la tarea de clasificación:

- **AirBnBReviews.csv**: Clasificación binaria (**positivo** / **negativo**).
- **TripadvisorHotelReviews.csv**: Clasificación multiclase (valoraciones del 1 al 5).

Para ambos conjuntos se ha realizado una división en tres particiones: **Train**, **Dev** y **Test Final**, siguiendo una proporción de 60 %, 20 % y 20 % respectivamente.

Conjunto de Datos	% de instancias	Num. de instancias
Train	60 %	212
Dev	20 %	70
Test	20 %	72

2.1. Cuadro: División Train, Dev y Test de los datos de **AirBnBReviews.csv** (binario)

Conjunto de Datos	% de instancias	Num. de instancias
Train	60 %	12 294
Dev	20 %	4 098
Test	20 %	4 099

2.2. Cuadro: División Train, Dev y Test de los datos de **TripadvisorHotelReviews.csv** (multiclase)

Conjunto de Datos	Num. de instancias
airbnb_spain_id_comments.csv	15737
airbnb_portugal_id_comments.csv	20561

2.3. Cuadro: Comentarios totales de los datasets de España y Portugal.

### 2.2.2. Descripción del preproceso

El preprocesado de los datos ha sido una etapa esencial en la preparación de los modelos de clasificación binaria y multiclase. Este proceso fue diseñado para transformar y normalizar los textos de las reseñas, adaptar las entradas a los algoritmos clásicos de aprendizaje automático y facilitar su uso por modelos generativos. A continuación, se detalla el pipeline seguido.

#### A. Separación de los datasets

El fichero **airbnb.csv** contiene comentarios de distintos países, para conseguir los comentarios de España y Portugal utilizamos el script **airbnbCreateDatasetsForPojectAlumno.py**, que dado el input **airbnb.csv** generó **airbnb\_spain.csv** y **airbnb\_portugal.csv**.

## B. Limpieza de los datasets

El script `limpiador.py` recibe como input un dataset y genera otros dos, por ejemplo: dado el dataset `airbnb-portugal.csv` genera `airbnb-portugal_limpio.csv` y `airbnb-portugal_id_reviews.csv`.

El primer dataset elimina las siguientes columnas:

```
"listing_url", "summary", "space", "description", "neighborhood_overview",
"notes", "transit", "access", "interaction", "house_rules", "bed_type",
"minimum_nights", "maximum_nights", "cancellation_policy", "accommodates",
"bedrooms", "beds", "bathrooms", "amenities", "extra_people", "guests_included",
"images", "host", "availability", "weekly_price", "monthly_price",
"text_embeddings" y "image_embeddings".
```

(Esto da información suficiente para el análisis realizado con Tableau.)

El segundo dataset es un paso previo para conseguir el dataset objetivo sobre el que hacer la clasificación. Este contiene solo el id y las reviews.

## C. Traducción de reseñas al inglés

Los ficheros `airbnb-spain.csv` y `airbnb-portugal.csv` contenían comentarios escritos en varios idiomas. Para unificar el análisis en inglés, se tradujeron automáticamente las reseñas no inglesas utilizando el modelo `gemma3:4b` ejecutado en local con `Ollama`, mediante la librería `LangChain`. Esto se implementa en el script `translate_reviews.py`, donde se define un *prompt* para automatizar la traducción:

```
You are a professional translator. Respond with only the translation.
Translate to informal English if not already English.
Use straight apostrophes (') instead of curly ones (').
Text: {text}
Translation:
```

Cada comentario traducido se almacena donde estaba el comentario sin traducir, para tener una historia del avance de los datos fue creado un nuevo dataset con los comentarios ya traducidos. Como el proceso de traducción es muy lento y hay muchos comentarios, se implementó una forma de traducir conjuntos de reviews especificando el rango de las mismas a incluir, por ejemplo:

```
python3 ScriptsDeApoyo/translate_reviews.py --row_range 0-10 --model gemma3:4b
--input_csv Datos/airbnb-portugal.csv --output_csv Datos/airbnb-portugal_trans.csv
--review_columns reviews --subcolumn comments
```

De esta forma se ha podido traducir sin la necesidad de tener el CPU trabajando durante días sin descanso. (las iteraciones siguientes tenían el mismo archivo de input y output)

## D. Preparación de los datasets

Una vez traducido todo terminamos de tratar los datos con los siguientes scripts:

- **procesarAdressyScore.py** Añaden columnas con información sobre la ciudad de origen, su longitud y su latitud.
- **unir.py** Concatena los datasets de España y Portugal.
- **separeHotelReviews.py** Genera un nuevo dataset que contiene solo el país del destino, el id del hotel/apartamento, la fecha en la que se escribió la review y el propio comentario.

El dataset restante ya está preparado para la clasificación.



## E. Limpieza, normalización y vectorización del texto

La limpieza y la normalización del texto se ha llevado a cabo en la función `preprocesar_texto` de cada clasificador, ejemplo del script `entrenar_TripAdvisor.py`:

```
def preprocesar_texto(texto):
    texto = re.sub(r'[^a-zA-Z\s]', '', texto).lower()
    lemmatizer = WordNetLemmatizer()
    stop_words = set(stopwords.words('english'))
    palabras = texto.split()
    palabras = [lemmatizer.lemmatize(palabra)
                 for palabra in palabras if palabra not in stop_words]
    return ' '.join(palabras)
```

Esta función elimina caracteres especiales y saltos de línea mediante expresiones regulares y escribe todo en minúsculas. Después lematiza las palabras que no sean stopwords (lo comprueba con la librería `stopwords` de `nlTK.corpus`) con la librería `WordNetLemmatizer` de `nlTK.stem`.

La representación numérica del texto se realizó con `TfidfVectorizer`, como se define en `clasificador.py` y `entrenar_pos_neg.py`. Los parámetros configurados fueron:

```
vectorizer = TfidfVectorizer(
    max_features=20000,
    ngram_range=(1, 3),
    stop_words="english"
)
```

Una vez ajustado sobre el conjunto de entrenamiento, el vectorizador se guardó para reutilización posterior:

```
joblib.dump(vectorizer, vectorizer_path)
```

## F. División en conjuntos de entrenamiento, desarrollo y prueba

La división se realiza mediante la función `train_test_split` con estratificación, ejemplo del script `entrenar_pos_neg.py`:

```
X_train, X_temp, y_train, y_temp = train_test_split(
    X_tfidf, y, test_size=0.4, stratify=y, random_state=42
)
X_dev, X_test, y_dev, y_test = train_test_split(
    X_temp, y_temp, test_size=0.5, stratify=y_temp, random_state=42
)
```

Esta división se aplica tanto en el binario como en el multiclase.

## G. Almacenamiento y reutilización de modelos

Los modelos entrenados se serializaron con `joblib`, como por ejemplo:

```
joblib.dump(knn, "modelo_knn.pkl")
```

Posteriormente, se cargan para predicción en scripts como `predecirComentariosBERT.py`:

```
knn_model = joblib.load(args.knn_model)
knn_vectorizer = joblib.load(args.knn_vectorizer)
```

Esto permite reproducir exactamente el entorno de inferencia sin necesidad de volver a entrenar.

### 2.2.3. Resultados de la tarea de clasificación

En esta sección se presentan los resultados obtenidos con los distintos algoritmos de clasificación aplicados a las tareas de análisis de sentimiento binario y multiclase. Asimismo, se incluye la evaluación sobre el conjunto objetivo.

#### A. Clasificación binaria: AirBnBReviews.csv

Se entrenó un modelo sobre el dataset `AirBnBReviews.csv`, donde cada reseña estaba etiquetada como **positiva** o **negativa**:

- **K-Nearest Neighbors (KNN)**, que clasifica cada reseña según la clase mayoritaria entre sus  $k$  vecinos más cercanos en el espacio TF-IDF.

Este modelo se entrenó sobre el conjunto `train`, y se evaluó sobre el subconjunto `test` del mismo dataset después de hacer un barrido de hiperparámetros sobre `dev`, barrido en el cual se concluyó que el valor óptimo para el hiperparámetro  $k$  es 1. La métrica principal fue la **precisión**, aunque también se registraron *recall* y *F1-score* en evaluaciones adicionales. La clase a evaluar fue Positive or Negative.

Los resultados han sido obtenidos con los scripts `entrenar_pos_neg.py` (genera un modelo knn después de un barrido de hiperparámetros) y `clasificador.py` (plantilla genérica).

Algoritmo	Hiperparámetros	Accuracy	Recall	F1-score
KNN	k=1	0.9706	0.98	0.96

2.4. Cuadro: Resultados sobre el conjunto Dev de `AirBnBReviews.csv` (clasificación binaria)

#### B. Clasificación multiclase: TripadvisorHotelReviews.csv

En el conjunto `TripadvisorHotelReviews.csv`, cada reseña presenta un **rating** de 1 a 5. El objetivo aquí fue predecir la clase correcta (rating) usando clasificadores multiclase. Para conseguir esto se ha recurrido a un modelo de Stacking que combina un modelo de Random Forest y otro de Logistic Regression, y a otro modelo de BERT. (Hay scripts diferentes para clasificar tanto con BERT como con stacking)

Algoritmo	Hiperparámetros	Accuracy
RandomForest	MaxDepth=20, MinSamplesSplit=2, NumberOfEstimators=200	-
LogisticRegression	C=1, Solver = lbfgs	-
Stacking	-	0.6328
BERT	BatchSize=16, LearningRate=2e-5, Epochs=3	0.67

2.5. Cuadro: Resultados sobre el conjunto Dev de `trip_advisor_hotel_reviews.csv` (multiclase)

Los modelos fueron entrenados con los conjuntos `train` y evaluados en `dev` y `test`. La métrica principal fue la **accuracy**.

### C. Evaluación sobre el dataset objetivo

Finalmente, los modelos que obtuvieron mejores resultados sobre `dev` y `test` fueron aplicados al conjunto central del estudio que contiene el país de destino, el id del alojamiento, la fecha en la que se realizó la review y el propio comentario.

Los datasets de entrenamiento emplean métricas y métodos de evaluación distintos, para abordar este problema en primer lugar se realizó una evaluación del comentario utilizando el modelo KNN entrenado con el conjunto de datos `airbnb_reviews.csv` y se añadió una columna denominada *Sentimiento* al dataset. Posteriormente, se utilizó el clasificador BERT entrenado con el conjunto de datos `trip_advisor_hotel_reviews.csv`, el cual añade una columna denominada *Score\_BERT\_1\_5* con una puntuación en una escala del 1 al 5. Por último, a través de las predicciones obtenidas con ambos modelos, se infirió una puntuación final en una escala del 1 al 9. Por ejemplo, si el sentimiento es negativo y la puntuación asignada por BERT es alta, la puntuación final se ubicará en el rango superior de los valores menores que 5. En cambio, si el sentimiento es positivo y la puntuación de BERT es baja, la puntuación final se ubicará dentro en el rango inferior de los valores de los valores mayores o iguales que 5.

Sentimiento	PuntuaciónBERT(1-5)	PuntuaciónFinal(1-9)
Positivo	5	9
Positivo	1	5
Negativo	5	4
Negativo	1	1

2.6. Cuadro: Ejemplo de clasificación KNN + BERT

Esta evaluación se encuentra implementada en los scripts `predecirComentarios.py` (modelo Stacking de RandomForest + LogisticRegression) y `predecirComentarioBERT(mod CSV).py` (modelo BERT). En ambos casos, el vectorizador y el modelo entrenado fueron cargados desde disco y aplicados al conjunto completo de reseñas preprocesadas.

## 2.2.4. Resultados de la aplicación de los modelos generativos

Además de los clasificadores tradicionales, se evaluaron modelos generativos de lenguaje para la predicción de puntuaciones de sentimiento directamente sobre el conjunto objetivo `airbnb.csv`. En este caso, no se entrenó un clasificador supervisado, sino que se emplearon técnicas de *zero-shot*, *one-shot* y *few-shot learning*, formulando las predicciones mediante *prompts* adaptados.

### A. Uso de modelos generativos locales (Ollama + LLaMA 3)

El script `predecirComentariosOllama.py` permite evaluar reseñas traducidas al inglés utilizando un modelo LLM ejecutado localmente a través de la herramienta `Ollama`, concretamente con el modelo `llama3`. Se eligió este modelo tras realizar varias pruebas con otras opciones disponibles como `gemma` o `mistral`, concluyendo que `llama3` ofrece un equilibrio adecuado entre rendimiento y coste computacional.

El prompt definido solicita explícitamente una puntuación del 1 al 10, presentando previamente dos ejemplos de referencia:

```
prompt = f"""Give a rating from 1 to 10 for the following hotel review. Only answer
with the number.

Example 1:
Review: "Lot of Traffic noise due to old windows toward the street. And the guys where
delayed more than 1 hour we where waiting
in front of the building and couldnt get in contact with the guys. Finally we came
inside then the cleaning team has not been there so it was dirty used beeds beer
cans used towels bed linen etc. so i had to wait another 2 hours before i could use
the apartment . that was not nice after a long day of traveling"
Rating: 2

Example 2:
Review: "It's a nice flat. The location is good even though you have to walk a bit to
the metro and quite a bit to the historic center. I liked that there was a
restaurant in the same building."
Rating: 5

Now, rate this review:
Review: "{comment}"
Rating: ""
```

Analizando los resultados obtenidos con *zero-shot*, *one-shot*, *two-shots* y *few-shots*, se observa una alta dependencia de la calidad de los ejemplos utilizados en el prompt. A modo ilustrativo, se muestra la puntuación obtenida para una reseña concreta:

It was our first time using Airbnb and it was really good. Thanks for everything!

- Zero-shot: 8
- One-shot: 4
- Two-shots: 9
- Few-shots: 3

Este comportamiento demuestra que el modelo puede producir resultados incoherentes cuando los ejemplos incluidos no son variados o relevantes. Así, aunque *zero-shot* pueda parecer más fiable en ciertos casos, un *prompt* bien diseñado con ejemplos representativos ofrece un mejor rendimiento en *few-shot*.

Las recomendaciones extraídas del proceso son:

- Incluir ejemplos variados y similares a las reseñas reales.
- Cubrir casos extremos (muy positivos y muy negativos).
- No introducir demasiados ejemplos para evitar lentitud y confusión.
- Asegurar coherencia temática entre los ejemplos.

## B. Traducción de comentarios utilizando Ollama

Para traducir las reseñas de los datasets, se empleó el script `translate_reviews.py`, el cual sustituye el texto original por su equivalente en inglés. Para esta tarea se utilizó el modelo `gemma3:4b` proporcionado por Ollama, dado que los modelos de la familia `gemma` son significativamente más rápidos y ligeros, lo cual supone una ventaja crucial al trabajar con grandes volúmenes de texto y permite el uso de la gpu.

Se descartó el uso de modelos más potentes como `mistral` por su mayor consumo de recursos. Dentro de la familia `gemma`, se eligió `gemma3:4b` frente a `gemma2:2b` por ser el modelo más actualizado y preciso disponible.

Durante la implementación del prompt se identificaron varios desafíos técnicos:

1. **Detección de idioma:** Algunos modelos no distinguían correctamente si el texto ya estaba en inglés, generando traducciones innecesarias.
2. **Formato de salida:** En ocasiones se añadían explicaciones o reformulaciones pese a las instrucciones del prompt.
3. **Caracteres especiales:** Comillas tipográficas (‘’), emoticonos y caracteres HTML causaban errores o truncamientos.
4. **Estructuras complejas:** Algunas entradas contenían listas o diccionarios, lo que requería procesar únicamente los campos `comments`.
5. **Longitud del prompt:** Para textos muy extensos, el modelo truncaba la respuesta o fallaba en la inferencia.

Finalmente, se definió el siguiente prompt:

```
template = (
    "You are a professional translator. The response needs to be just 1 option, and\n"
    "only that, nothing else. "\n"
    "Translate the following text into informal English, only if the original text is\n"
    "not in English. "\n"
    "If the original text is already in English, return it as is, just change\n"
    "and never use\n"
    "use ' instead.\n"
    "Text: {text}\nTranslation:"
)
```

## C. Sentiment Analysis de palabras con VADER

Para analizar la polaridad de las palabras presentes en los comentarios, se utilizó el script `sentimentAnalysis.py`, el cual asigna una puntuación de sentimiento a cada término según su frecuencia, polaridad y país de origen de la reseña.

Se evaluaron distintas herramientas, decidiéndose finalmente por VADER (Valence Aware Dictionary and sEntiment Reasoner), disponible en `nlTK`. Aunque `TextBlob` y algunos modelos en `Hugging Face` ofrecen funcionalidades similares, se optó por VADER por las siguientes razones:

### ¿Por qué VADER y no TextBlob?

- **Lexicón limitado:** `TextBlob` no cubre muchas palabras informales o expresiones típicas de reseñas.
- **Sensibilidad al contexto:** Tiene menor capacidad para interpretar lenguaje informal, errores ortográficos o abreviaturas.
- **Neutralidad excesiva:** Clasifica muchas palabras relevantes como neutras (por ejemplo, “clean”, “friendly”), reduciendo la utilidad práctica.

### Ventajas de VADER:

- Diseñado específicamente para textos sociales (opiniones, redes, reviews).
- Reconoce patrones del lenguaje informal, uso de mayúsculas, exclamaciones, emoticonos.
- Ofrece una puntuación compuesta continua que distingue positividad, negatividad y neutralidad.

## ¿Por qué no usar modelos de Hugging Face?

Aunque modelos como BERT o RoBERTa pueden captar mejor los matices del lenguaje, presentan limitaciones importantes:

- Requieren hardware potente (idealmente GPU) para procesar texto a gran escala.
- Inferencia más lenta y costosa para un análisis palabra a palabra.
- Menor interpretabilidad frente a métodos basados en diccionarios.

## 2.3. Algoritmos, link a la documentación y nombre de los hiperparámetros empleados

### 2.3.1. Experimentación: Algoritmos empleados y Breve Descripción

Durante el desarrollo de este proyecto se emplearon varios algoritmos de clasificación, tanto tradicionales como basados en modelos de lenguaje, para abordar tareas de clasificación binaria y multiclase. A continuación se enumeran los algoritmos utilizados, acompañados de una breve descripción, los hiperparámetros más relevantes aplicados en los scripts, y un enlace a su documentación oficial.

#### ■ K-Nearest Neighbors (KNN)

Clasificador basado en la distancia entre vectores. Para una instancia desconocida, calcula la clase mayoritaria entre los  $k$  vecinos más cercanos en el espacio vectorial generado por TF-IDF. Este método es no paramétrico y se encuentra implementado en `clasificador.py` y `entrenar_pos_neg.py`.

##### Hiperparámetros utilizados:

- `n_neighbors = 1`
- `weights = 'distance'`
- `metric = 'euclidean'`

##### Link:

<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

#### ■ Random Forest (RF)

Conjunto de clasificadores basados en árboles de decisión. Utiliza muestreo aleatorio para crear múltiples árboles, y luego hace votación. Se utilizó en las tareas binaria y multiclase. Implementado en `entrenar_TripaAdvisor.py` y `clasificador.py`.

##### Hiperparámetros utilizados:

- `n_estimators = 200`
- `max_depth = 20`
- `min_samples_split = 2`
- `random_state = 42`

##### Link:

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

### ■ Logistic Regression

Un modelo de clasificación lineal que estima la probabilidad de que una instancia pertenezca a una clase mediante una función sigmoide (o softmax para multiclase), y toma decisiones en función de un umbral de probabilidad.

#### Hiperparámetros utilizados:

- `C = 1`
- `solver = lbfgs`

#### Link:

[https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

### ■ BERT

Modelo basado en transformadores entrenado mediante *fine-tuning* sobre datos clasificados. Utiliza la arquitectura `bert-base-uncased` desde la librería `transformers`. Su entrenamiento se realiza en el script `predecirComentarioBert.py`, directamente sobre reseñas.

#### Hiperparámetros utilizados:

- `batch_size = 16`
- `learning_rate = 2e-5`
- `epochs = 3`

#### Link:

[https://huggingface.co/transformers/model\\_doc/bert.html](https://huggingface.co/transformers/model_doc/bert.html)

### ■ LLaMA 3 (modelo generativo local con LangChain + Ollama)

Este modelo se utilizó para predicción de puntuaciones directamente desde reseñas textuales, sin entrenamiento, mediante prompts diseñados específicamente para tareas de *zero-shot* o *few-shot learning*. La integración se realiza en `predecirComentariosOllama.py`, usando `LangChain` y `ChatOllama`.

**Hiperparámetros utilizados:** No aplican como tales; la configuración se controla mediante el prompt y la arquitectura cargada en Ollama.

**Link:** <https://ollama.com/library/llama3>

## 2.3.2. Resultados sobre el Development

Los modelos entrenados fueron evaluados sobre el conjunto de desarrollo (`dev`) de cada dataset. La evaluación se centró en medir la capacidad de los clasificadores para detectar correctamente las distintas clases en el caso binario (`AirBnBReviews.csv`). Se calcularon métricas estándar de clasificación: precisión, *recall* y *F1-score*, tanto de forma general como optimizando el rendimiento sobre una clase específica.

En los tres enfoques siguientes, los resultados fueron extraídos de los scripts `entrenar_pos_neg.py` y `clasificador.py`, aplicando las funciones de evaluación de `sklearn.metrics` como `classification_report()` y `precision_score()`, entre otras.

### 2.3.3. Discusión sobre el proceso de aprendizaje

Durante el desarrollo del proyecto nos enfrentamos a diversos desafíos que requirieron un esfuerzo mayor al inicialmente previsto. Entre los más relevantes se encuentran la correcta traducción de los comentarios, el tratamiento adecuado de los datos para su posterior clasificación, y el diseño de un método que permitiera entrenar un modelo considerando dos escalas de evaluación distintas.

La traducción de los comentarios supuso numerosos contratiempos, siendo el principal la extensa cantidad de tiempo necesaria para traducir la totalidad de los textos, cuya magnitud no es despreciable. Con el fin de optimizar este proceso, inicialmente desarrollamos un script que traducía reseñas individuales, lo que nos permitió experimentar con distintos prompts y modelos de lenguaje. Tras identificar una combinación que ofrecía resultados satisfactorios, descubrimos que la herramienta *ollama* estaba utilizando por defecto la CPU, lo que ralentizaba considerablemente las traducciones. Luego de varios intentos por forzar el uso de la GPU, concluimos que la causa del problema era el tamaño del modelo utilizado, por lo que optamos por emplear una versión con menos parámetros (*gemma3:4b*).

Adicionalmente, surgieron complicaciones con el formato del archivo CSV, ya que la columna *reviews* (que contiene los comentarios) almacenaba los datos en forma de arreglo. Las primeras versiones del script *translatereviews.py* traducían el arreglo completo, lo que resultaba en la pérdida de información. Tras identificar este problema, implementamos una función que permite seleccionar un rango específico de reseñas, lo que facilitó el procesamiento en bloques más manejables y redujo la carga continua sobre el equipo. A pesar de estas mejoras, fue necesario dividir el trabajo entre dos personas (una traduciendo las reseñas de España y la otra las de Portugal), empleando un total aproximado de 32 horas para completar la traducción de ambos conjuntos de datos.

En lo relativo al tratamiento de los datos, surgió un debate sobre qué variables eran realmente relevantes y cuáles podían descartarse para optimizar el procesamiento. Tras varias pruebas y evaluaciones, se definieron los pasos descritos en el apartado 2.2.2, tomando en cuenta principalmente las necesidades del análisis posterior en Tableau.

En cuanto al trabajo realizado con Tableau, el proceso de aprendizaje fue especialmente significativo, dado que al inicio del proyecto ningún integrante del equipo tenía experiencia previa con la herramienta. La curva de aprendizaje inicial fue pronunciada, especialmente en lo relativo a la importación y limpieza de los datos y la configuración de relaciones entre tablas. A medida que avanzaba el análisis, fue necesario familiarizarse con conceptos específicos del entorno de Tableau como medidas calculadas, jerarquías, filtros contextuales y mapas de dispersión, entre otros. También se presentaron dificultades técnicas, como la imposibilidad de trabajar con datos anidados o formatos de texto complejos directamente, lo cual obligó a realizar una limpieza previa de los archivos. El uso de filtros, leyendas personalizadas y gráficos requirió un trabajo adicional de documentación. Por a ello, el proceso permitió adquirir una comprensión sólida sobre cómo estructurar la información para generar visualizaciones claras, comparativas y orientadas a la toma de decisiones.

Además del aprendizaje técnico, fue necesario definir cuidadosamente una narrativa visual. Se priorizaron visualizaciones que permitieran comparar España y Portugal de forma directa, mantener una estructura reconocible (por ejemplo, gráficos por país, por año y por mes) y garantizar una lectura intuitiva. Este enfoque narrativo facilitó la unión del contenido visual con los objetivos del proyecto.

### 2.3.4. Conclusión sobre la tarea de clasificación

A partir de los resultados obtenidos en las fases de desarrollo y prueba, se extrajeron diversas conclusiones relevantes para la selección de los modelos finales, considerando los objetivos específicos de cada enfoque: detección de clases, balance general o regresión multiclase.



El proceso de entrenamiento de los modelos de clasificación presentó ciertos desafíos y generó algunas dudas metodológicas. Además de la estrategia que finalmente se adoptó (descrita en el apartado 2.2.3, sección C), se valoró la posibilidad de clasificar el conjunto de datos `airbnbreviews.csv`, originalmente binario, utilizando una escala del 1 al 9 mediante modelos de lenguaje. Paralelamente, se consideró escalar las puntuaciones del conjunto `tripadvisorhotelreviews.csv` (con valores de 1 a 5) a la misma escala aplicando la función  $f(x) = 2x - 1$ . Sin embargo, esta opción fue descartada debido a la falta de garantías sobre la calidad de las predicciones del modelo de lenguaje, así como por las limitaciones del escalado propuesto, que no permitiría representar valores pares en la nueva escala.

Por otro lado, los modelos generativos demostraron potencial para tareas complementarias o análisis de tipo interpretativo, al permitir obtener una puntuación directamente a partir del texto mediante enfoques de *zero-shot* con modelos como LLaMA 3. No obstante, su mayor complejidad operativa, la sensibilidad a la formulación del *prompt* y el elevado coste computacional impidieron que superaran en rendimiento a los clasificadores tradicionales entrenados específicamente para esta tarea.

## 2.4. Algoritmos, link a la documentación y nombre de los hiperparámetros empleados

Tras aplicar el modelo de predicción de sentimiento sobre el conjunto central, se procedió a un análisis semántico de los resultados obtenidos, con el objetivo de detectar diferencias de percepción entre zonas y evaluar la coherencia global del sistema. Este proceso constituye una extensión analítica del pipeline de clasificación, orientada a la interpretación agregada de los resultados por ubicación o propiedad.

### 2.4.1. Experimentación: Algoritmos empleados y Breve Descripción

Los ratings predichos fueron obtenidos a partir de los textos procesados y traducidos previamente, utilizando los modelos de clasificación seleccionados en la fase de experimentación. Este proceso se encuentra implementado principalmente en el script `sentimentAnálisis.v3.py`, donde se aplican los clasificadores entrenados, se generan predicciones por reseña, y se calcula un valor numérico de salida que puede ser agregado por alojamiento o zona.

Los modelos empleados en este análisis fueron aquellos seleccionados como óptimos en las tareas previas (clasificación binaria o regresión sobre puntuación). La inferencia de sentimiento se realizó cargando los modelos previamente entrenados (por ejemplo, `modelo_knn.pkl`) y aplicándolos sobre los vectores generados a partir del texto (TF-IDF concatenado con variables estructurales).

Este procedimiento produjo una predicción para cada reseña, que fue posteriormente agregada para calcular el `sentiment_score_predicho` medio por alojamiento. Este valor fue comparado con el `rating_media` real disponible en el dataset original `airbnb.csv`, permitiendo calcular métricas de error como la desviación cuadrática media o analizar el sesgo de las predicciones en distintas zonas geográficas.

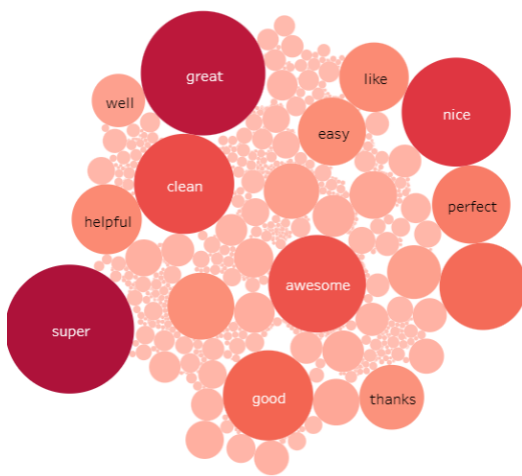
- **Clasificadores tradicionales (TF-IDF + features):** se aplicaron los modelos KNN y Random Forest, cargados desde disco y ejecutados sobre los datos preprocesados. Las predicciones se escalaron o ajustaron si era necesario para mantener coherencia con la escala del rating real (de 1 a 9).
- **Modelos generativos (LLM):** como los implementados en `prededirComentariosOllama.py`, permitieron generar una puntuación directamente desde el texto sin entrenamiento. En estos casos, se definió un *prompt* fijo que solicitaba un valor del 1 al 9, en formato textual, que luego fue transformado a valor numérico y almacenado junto con la reseña original y el rating real.

La agregación y análisis de los resultados predichos permitió observar diferencias semánticas entre propiedades, zonas y patrones lingüísticos, así como validar el funcionamiento global del sistema con datos reales.

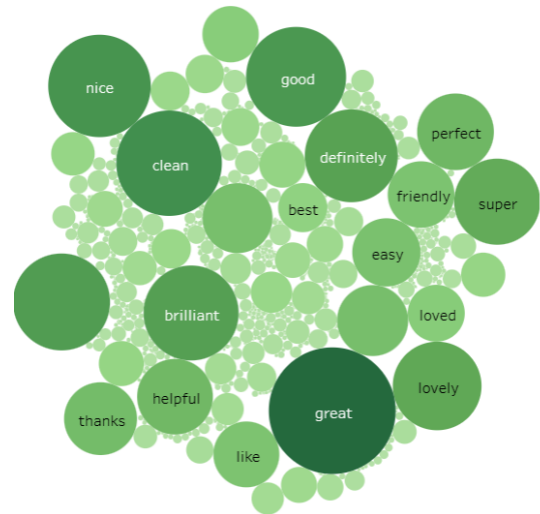
### 2.4.2. Resultados

A continuación se presentan los principales hallazgos obtenidos mediante visualizaciones gráficas elaboradas en Tableau, orientadas al análisis semántico y comparativo del sentimiento en reseñas por país y por polaridad. En este análisis, las reseñas de **España** representan el entorno propio del estudio, mientras que las de **Portugal** actúan como referencia competidora.

Se construyeron **nubes de palabras** para representar las palabras más frecuentes en reseñas positivas y negativas, diferenciadas por país.



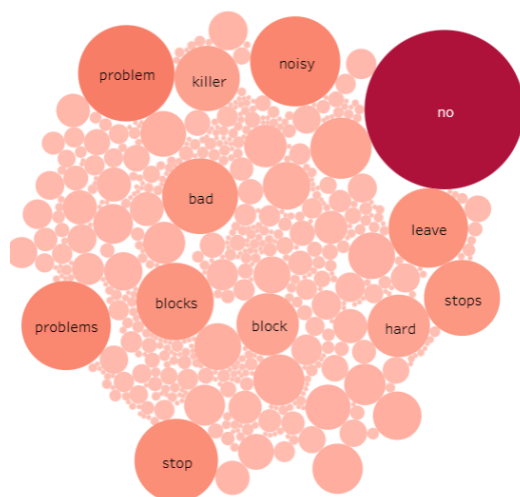
(a) Nube de palabras positivas en España.



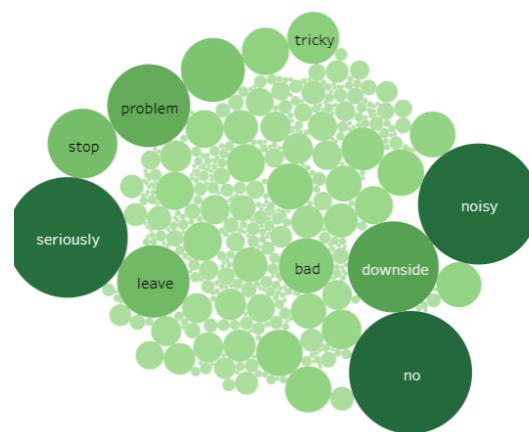
(b) Nube de palabras positivas en Portugal.

2.2. Figura: Comparativa de palabras frecuentes en reseñas positivas.

Estas visualizaciones muestran que, en ambos países, las reseñas positivas están dominadas por términos que reflejan una valoración general del servicio como “great”, “clean”, “perfect”, “recommend”, “super” o “helpful”. Estas palabras se corresponden directamente con atributos clave del alojamiento como limpieza, estado general, atención del anfitrión y calidad de la experiencia percibida.



(a) Nube de palabras negativas en España.

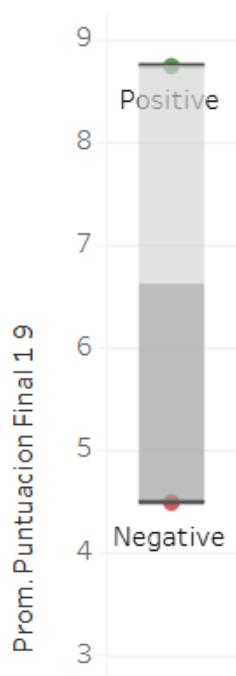


(b) Nube de palabras negativas en Portugal.

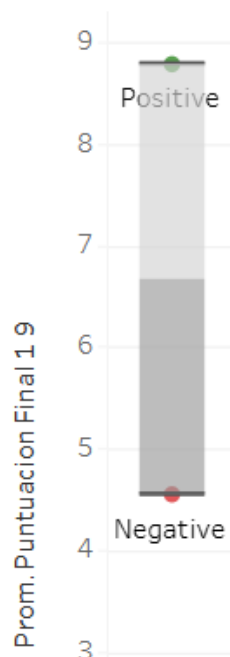
2.3. Figura: Comparativa de palabras frecuentes en reseñas negativas.

En las reseñas negativas, destacan términos como “problem”, “no”, “leave”, “bad”, “stop”, “blocks” o “noisy”. En el caso de España, se observa mayor presencia de términos como “blocks” o “bother”, mientras que en Portugal emergen con más frecuencia “seriously”, “regret” o “shame”. Esto sugiere diferencias en el tono de las quejas, posiblemente asociadas a distintos estándares culturales o a la gestión operativa en cada zona.

A nivel numérico, se elaboró un **boxplot** que muestra la distribución de la puntuación final por sentimiento (positivo o negativo), diferenciada también por país.



(a) Distribución de puntuación final por sentimiento en España.



(b) Distribución de puntuación final por sentimiento en Portugal.

Las reseñas etiquetadas como positivas presentan una media y mediana claramente más altas que las negativas. En ambos países, la mediana de puntuaciones para reseñas positivas se sitúa cerca

de 6.6–6.7, mientras que las negativas se concentran por debajo del 4.5. Este patrón valida la consistencia entre los modelos de clasificación aplicados y los valores reales presentes en el dataset.

En conjunto, estas representaciones gráficas confirman que las palabras más significativas en cada grupo están fuertemente ligadas a los atributos centrales del producto turístico (limpieza, comunicación, ubicación, estado del alojamiento). Sin embargo, también emergen conceptos adicionales que van más allá de los atributos estructurados, como expresiones de decepción emocional (“regret”, “shame”, “hesitate”) o entusiasmo (“super”, “fantastic”, “awesome”), lo cual sugiere que los modelos capturan dimensiones más subjetivas y emocionales en las opiniones de los usuarios.

Por tanto, las gráficas no solo reflejan una buena alineación entre las predicciones de sentimiento y los atributos clave, sino que también permiten detectar diferencias semánticas sutiles entre España y Portugal. Estas diferencias, aunque no radicales, pueden ofrecer ventajas competitivas si se entienden como señales para ajustar la estrategia de comunicación, atención al cliente o condiciones operativas.

Desde una perspectiva comparativa, los resultados sugieren que **Portugal** presenta una mayor consistencia emocional en las reseñas positivas, con términos que evocan entusiasmo y satisfacción (“amazing”, “fantastic”, “highly recommend”), mientras que **España** destaca en aspectos tangibles como limpieza, ubicación o trato (“clean”, “perfect”, “friendly”). Sin embargo, en el plano negativo, las quejas en España tienden a centrarse en aspectos físicos o logísticos como “blocks”, “noisy” o “bother”, mientras que en Portugal predominan expresiones de decepción emocional como “regret” o “shame”, lo que podría indicar una mayor sensibilidad al servicio o a la experiencia global. Para mejorar la percepción general en España, se recomienda reforzar la comunicación previa con el huésped, clarificar instrucciones y reforzar elementos de confort emocional, lo que podría suavizar la aparición de términos negativos relacionados con la frustración o la confusión. Asimismo, se podría potenciar el uso de lenguaje positivo en la gestión post-estancia para fomentar la aparición de términos más emocionales en las valoraciones positivas.

### 2.4.3. Discusión sobre los descubrimientos realizados

El análisis conjunto de las métricas objetivas y del sentimiento expresado en las reseñas permite extraer conclusiones relevantes sobre el posicionamiento comparativo entre España y Portugal dentro del ecosistema de alojamientos turísticos en plataformas como AirBnB.

Desde el punto de vista de la relación calidad/precio, Portugal muestra una mayor coherencia en la percepción del valor por parte de los usuarios. Tanto en los análisis de regresión como en los mapas de satisfacción, se observa una menor proporción de alojamientos con baja puntuación pese a un precio elevado, lo cual sugiere una oferta más equilibrada y homogénea. En contraste, España presenta una mayor cantidad de outliers negativos, especialmente concentrados en zonas de alta densidad como Barcelona, lo que indica posibles problemas de sobreprecio o expectativas no cumplidas.

En términos de evolución temporal, Portugal mantiene una mayor estabilidad en sus valoraciones, mientras que España ha experimentado un descenso progresivo en la cantidad de reseñas positivas desde 2018, así como un crecimiento en las negativas. Esto sugiere una pérdida de percepción favorable por parte de los usuarios que debe ser abordada mediante estrategias de refuerzo de calidad y comunicación transparente.

A nivel semántico, las diferencias también son notables. En las reseñas positivas, el lenguaje usado por los usuarios en Portugal incluye una mayor carga emocional y expresiones de entusiasmo (“brilliant”, “fantastic”), mientras que en España predominan términos más funcionales y descriptivos (“clean”, “helpful”). Esto sugiere que la experiencia en Portugal no solo es valorada como correcta, sino como memorable. En el caso de las reseñas negativas, las críticas en España tienden

a centrarse en aspectos operativos (“blocks”, “noisy”), mientras que en Portugal aparecen términos más relacionados con la decepción emocional (“regret”, “shame”), lo cual puede estar ligado a expectativas distintas según el perfil del viajero .

En lo que respecta a los modelos de predicción de sentimiento, se ha constatado que los algoritmos tradicionales (como Random Forest y Naive Bayes) ofrecen resultados consistentes y bien alineados con las valoraciones reales, lo que valida su uso para tareas de análisis automatizado. Los modelos generativos, si bien útiles en entornos sin datos etiquetados, presentan una mayor variabilidad y dependencia del diseño del prompt, lo que puede limitar su aplicabilidad operativa.

En resumen, nuestras fortalezas en España residen en una oferta económica con alto nivel de satisfacción, particularmente en habitaciones privadas, y en un volumen elevado de reseñas que aportan una base sólida para el análisis. Sin embargo, los principales puntos débiles están en la coherencia del valor percibido, especialmente en los alojamientos enteros, y en la pérdida de consistencia temporal en las opiniones positivas.

En cambio, Portugal destaca por una percepción más uniforme y emocionalmente positiva de la experiencia turística, con una estructura de precios más alineada con la satisfacción. Para mejorar nuestra posición competitiva, recomendamos:

- Reforzar la calidad percibida en alojamientos completos mediante revisión de precios, atención al cliente y servicios añadidos.
- Introducir mejoras que generen vínculos emocionales (experiencias memorables) con el huésped.
- Aprovechar la fortaleza de las habitaciones privadas como línea diferenciadora en campañas de marketing.
- Analizar los puntos críticos identificados en las reseñas negativas para ejecutar mejoras específicas.

Este enfoque permitiría no solo igualar el rendimiento observado en Portugal, sino superar sus resultados mediante una oferta más diversa, emocionalmente rica y adaptada a las expectativas reales del usuario.

## 3. Despliegue y código entregado

### 3.1. Estructura del código entregado

El proyecto se ha desarrollado íntegramente en Python y se organiza en un conjunto de scripts modulares que permiten una ejecución flexible, segmentada por tareas específicas dentro del flujo de análisis de sentimiento. Esta estructura favorece la reutilización y facilita la depuración o mejora de componentes individuales sin necesidad de modificar el conjunto completo.

A continuación, se describe el propósito de cada uno de los scripts entregados:

- **limpiador.py**: encargado de la limpieza básica de los textos de reseñas. Aplica operaciones como conversión a minúsculas, eliminación de símbolos, espacios en blanco redundantes, saltos de línea y signos de puntuación.
- **translate\_reviews.py**: traduce automáticamente las reseñas no escritas en inglés. Utiliza **LangChain** con un modelo **LLaMA 3** ejecutado en local mediante **Ollama**. Se define un *prompt* fijo para traducción directa y se almacena el resultado en un nuevo archivo para su uso posterior.
- **airbnbCreateDatasetsForPojectAlumno.py**: genera versiones procesadas del dataset original **airbnb.csv**, separando subconjuntos y extrayendo información relevante como identificadores, puntuaciones y texto preprocesado.
- **entrenar\_pos\_neg.py**: entrena modelos de clasificación binaria sobre el dataset **AirBnBReviews.csv**. Incluye preprocesado textual, vectorización TF-IDF, oversampling y entrenamiento de clasificadores como Naive Bayes, KNN y Random Forest.
- **entrenar\_TripAdvisor.py**: realiza una operación equivalente a la anterior pero sobre el dataset multiclase **TripAdvisorHotelReviews.csv**. Utiliza los mismos clasificadores adaptados a clasificación con cinco clases (ratings del 1 al 5).
- **clasificador.py**: contiene funciones auxiliares utilizadas en múltiples scripts: tokenización, vectorización, escalado de variables, creación de conjuntos Train/Dev/Test y carga/guardado de modelos mediante **joblib**.
- **predecirComentarios.py**: aplica modelos ya entrenados (cargados desde disco) sobre nuevas reseñas para predecir una puntuación de sentimiento. Devuelve valores que luego pueden ser evaluados frente a los ratings reales.
- **predecirComentariosOllama.py**: utiliza modelos generativos (**LLaMA 3**) en local para asignar una puntuación a reseñas traducidas, usando enfoques *zero-shot* o *few-shot*. Funciona de forma autónoma, sin necesidad de entrenamiento supervisado.
- **predecirComentarioBert.py**: realiza inferencia sobre reseñas usando un modelo **BERT** previamente finetuneado. El modelo es cargado mediante la librería **transformers**.
- **sentimentAnalysis\_v3.py**: combina la salida de predicción de los modelos con características numéricas adicionales (longitud del texto, uso de mayúsculas, puntuación, polaridad) para calcular una puntuación final agregada por reseña.
- **review\_fusion.py**: une las predicciones generadas por los modelos con los metadatos del alojamiento (por ejemplo, zona, tipo de alojamiento, precio), creando un dataset unificado que sirve de base para los análisis en Tableau.

- `break_review_scores.py`: calcula la media de las puntuaciones predichas y reales por alojamiento, lo cual permite comparar el rendimiento de cada anuncio individual.
- `review_counter.py`: analiza la longitud, frecuencia y distribución de las reseñas por alojamiento, útil para detectar anomalías o diferencias de comportamiento entre zonas.
- `huggingFace.py`: alternativa generativa que utiliza modelos alojados en Hugging Face, permitiendo pruebas con modelos externos como GPT-2, BLOOM o BERT para análisis comparativo.
- `SentimentVader.py`: emplea el analizador de sentimientos VADER como herramienta adicional para exploración de polaridad sin entrenamiento.

Se incluyen también los siguientes objetos serializados:

- `modelo_knn.pkl`: clasificador KNN entrenado sobre el conjunto de entrenamiento binario.
- `vectorizador_tfidf.pkl`: objeto TF-IDF vectorizer entrenado, necesario para convertir texto en matrices numéricas consistentes.

## 3.2. Ejecución

El flujo completo del sistema puede reproducirse de forma modular mediante la ejecución secuencial de los scripts, que han sido diseñados para ejecutarse de manera independiente:

1. `airbnbCreateDatasetsForPojectAlumno.py` (*generación de subconjuntos*)
2. `limpiador.py` (*limpia los datasets*)
3. `translate_reviews.py` (*traducción de reseñas*)
4. `entrenar_pos_neg.py` y `entrenar_TripAdvisor.py` (*entrenamiento de modelos sobre `airbnb_reviews.csv` y `trip_advisor_hotels_reviews.csv` respectivamente*)
5. `procesarAdressyScore.py`
6. `unir.py` (*junta los datasets de España y Portugal*)
7. `predecirComentarios.py`, `predecirComentariosOllama.py`, o `predecirComentarioBert.py` (*evaluación de reseñas sin etiquetar*)
8. `sentimentAnalisis_v3.py` (*cálculo de la puntuación final combinada*)

Para utilizar el clasificador de BERT (`predecirComentarioBert.py`) hay que descargar el modelo que hemos entrenado con el script `BERT.py` desde: [https://drive.google.com/file/d/1NIka\\_3Cyix5RQ40vf3FG2Sy8gYjduB5v/view?usp=drive\\_link](https://drive.google.com/file/d/1NIka_3Cyix5RQ40vf3FG2Sy8gYjduB5v/view?usp=drive_link) y extraerlo en dentro de la carpeta Modelos del repositorio.

## 3.3. Requisitos técnicos

El entorno de desarrollo requiere las siguientes versiones y bibliotecas:

- **Python**: 3.10 o superior.
- **Principales dependencias**:
  - `pandas`, `numpy`, `scikit-learn`
  - `nlTK`, `textblob`, `chardet`
  - `imblearn`, `colorama`
  - `transformers`, `torch` (para el uso de BERT y modelos de Hugging Face)
  - `langchain`, `langchain-ollama`, **Ollama** (aplicación externa utilizada para traducción y predicción generativa)

El sistema ha sido probado en ejecución local sobre CPU. Aunque algunas tareas, como la inferencia con BERT, podrían beneficiarse del uso de GPU, esta no es estrictamente necesaria para el funcionamiento general del flujo de trabajo.

### 3.4. Observaciones finales

La modularidad y claridad de los scripts entregados permite extender fácilmente el sistema a nuevas tareas. Por ejemplo, se podría incorporar detección de ironía, análisis multilingüe o personalización por zona geográfica simplemente añadiendo nuevos módulos.

El pipeline completo, tal y como está diseñado, permite procesar, etiquetar y evaluar conjuntos de reseñas sin intervención manual, lo que lo hace adecuado para análisis periódicos o integración en herramientas de apoyo a la toma de decisiones turísticas.



## 4. Conclusiones

El proyecto ha abordado con éxito la tarea de análisis de sentimiento aplicado a reseñas turísticas, integrando métodos clásicos de aprendizaje supervisado, modelos generativos de lenguaje y técnicas de análisis semántico automatizado. El objetivo principal ha sido diseñar un sistema capaz de transformar texto libre en puntuaciones numéricas que reflejen el grado de satisfacción del usuario, incluso en ausencia de etiquetas explícitas.

Desde la fase inicial de adquisición y preparación de los datos, se diseñó un pipeline completo y modular que abarca:

- Traducción automática de reseñas no escritas en inglés, estandarizando la entrada textual para todos los modelos.
- Limpieza y normalización del texto mediante técnicas de procesamiento del lenguaje natural.
- Vectorización TF-IDF y extracción de características adicionales como longitud, uso de mayúsculas y polaridad léxica.
- División en conjuntos Train/Dev/Test con estratificación de clases.

Se desarrollaron y compararon múltiples modelos de clasificación para distintas tareas:

- **Clasificación binaria:** detección de sentimiento positivo o negativo usando `AirBnBReviews.csv`.
- **Clasificación multiclase:** predicción del rating (1–5) en `TripadvisorHotelReviews.csv`.
- **Predicción directa:** inferencia de puntuación (1–9) a partir del texto mediante modelos generativos.

Los modelos evaluados incluyen:

- KNN.
- Random Forest.
- BERT.
- LLaMA 3 vía Ollama.

Los resultados obtenidos evidencian los siguientes hallazgos clave:

- Los modelos generativos han demostrado utilidad para predecir puntuaciones sin necesidad de entrenamiento, aunque su rendimiento depende fuertemente del diseño del *prompt*.
- Se ha encontrado una correlación razonable entre las puntuaciones predichas y las reales tanto a nivel de reseña individual como a nivel agregado por alojamiento.
- El análisis semántico muestra diferencias sutiles en el lenguaje de las reseñas según país, lo que permite detectar oportunidades estratégicas de mejora.

La arquitectura de código desarrollada se basa en scripts independientes pero interconectables, lo que facilita la reutilización, el mantenimiento y la extensión futura. Gracias a esta modularidad, se puede sustituir un modelo, modificar el vectorizador o actualizar los conjuntos sin alterar la estructura global.

Además, la inclusión de modelos generativos permite replicar el sistema en dominios sin etiquetado, lo que lo hace aplicable a sectores donde los datos son abundantes pero no estructurados.

El análisis agregado por zona geográfica ha permitido observar patrones diferenciados en el lenguaje y la valoración de usuarios entre España (zona propia) y Portugal (zona competidora). Mientras que España destaca por atributos tangibles como limpieza y ubicación, Portugal obtiene un mayor peso en lenguaje emocional y valoración global.

Desde el punto de vista operativo, se han identificado puntos de mejora específicos para los alojamientos en España:

- Optimizar alojamientos enteros que presentan mayor desajuste entre precio y satisfacción.
- Reforzar comunicación pre y post-estancia para mejorar la percepción global.
- Incentivar la generación de reseñas positivas mediante experiencia diferenciadora.
- Vigilar aspectos logísticos que aparecen recurrentemente en reseñas negativas (ruido, accesibilidad, check-in).

A nivel técnico, el trabajo realizado valida el uso combinado de IA simbólica y conexionista en entornos reales de decisión. La integración de modelos entrenables con técnicas de generación directa ofrece una solución híbrida robusta que puede adaptarse a distintos escenarios de datos.

En conclusión, el sistema desarrollado demuestra la viabilidad de aplicar técnicas de procesamiento de lenguaje natural y machine learning para el análisis automatizado de reseñas turísticas, con una capacidad real de traducir texto libre en métricas cuantificables. Este enfoque puede escalarse fácilmente a otros contextos como restaurantes, servicios locales o comercio electrónico, consolidando su valor como herramienta de apoyo a la toma de decisiones basada en experiencia de usuario.

# Bibliografía

Garza Olvera, B. (2014). *Estadística y probabilidad* (G. Govea Anaya T. Betancourt Pérez, Eds.; 1.a ed.). Pearson. <https://drive.google.com/file/d/13FQobNnJU3Ou0tRhosDn9Fn9t1H7Q0HM/view>

Pat, Lucio Menchaca, Aarón Pat, Juan Luis, David. (2013). *Introducción a los Modelos de Regresión*.

Knaflitz, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson.

Kotler, P., Keller, K. L. (2015). *Marketing Management* (15th ed.). Pearson.

Aggarwal, C. C. (2018). *Machine Learning for Text*. Springer.

Bendle, N., Farris, P., Pfeifer, P., Reibstein, D. (2016). *Marketing Metrics: The Manager's Guide to Measuring Marketing Performance* (3rd ed.). Pearson FT Press.