

Diseño de un modelo de análisis de sentimientos basado en machine learning aplicado a
restaurantes de Rehoboth Beach.

Angela Patricia Blanco Gómez

Trabajo de Grado para Optar al Título de Ingeniera Industrial

Director: Henry Lamos Díaz

Ph.D. en Física, Matemática

Universidad Industrial de Santander

Facultad de Ingenierías Físico-mecánicas

Escuela de Estudios Industriales y Empresariales

Bucaramanga

2024

Tabla de contenido

Introducción	13
1. Planteamiento del problema.....	15
2. Objetivo general.....	18
2.1. Objetivos específicos	18
3. Metodología	19
3.1. Fase I: Definición del problema y análisis de la literatura.....	19
3.2. Fase II: Recopilación y procesamiento de datos.....	20
3.3. Fase III: Análisis de sentimientos	20
3.4. Fase IV: Validación del modelo	21
3.5. Fase V: Interpretación de resultados y recomendaciones	21
4. Marco de referencia	22
4.1. Marco de antecedentes.....	22
4.2. Marco teórico	24
4.2.1. Inteligencia artificial	24
4.2.2. Machine learning	26
4.2.3. Deep learning	28
4.2.4. Procesamiento de lenguaje natural.....	29
4.2.5. Word embedding.....	31
4.2.6. Minería de texto	33

4.2.7. Selenium.	35
4.2.8. Análisis de sentimientos.....	36
5. Revisión de la literatura	39
5.1. Análisis bibliométrico	39
5.2. Análisis preliminar de la literatura.....	45
6. Datos	52
7. Análisis de sentimientos.....	57
7.1. Análisis de sentimientos según su polaridad.....	57
7.1.1. Importación de librerías	57
7.1.2. Carga del archivo CSV	58
7.1.3. Estructura para almacenar los resultados	58
7.1.4. Función para asignar etiquetas de sentimiento	59
7.1.5. Leer las reseñas y realizar el análisis	59
7.2. Análisis de sentimientos según su categoría	60
7.2.1. Carga de herramientas y configuración del modelo	61
7.2.2. Carga del modelo y el tokenizer	61
7.2.3. Configuración del dispositivo.....	62
7.2.4. Función para procesar y analizar el texto	62
7.2.5. Lectura y categorización las reseñas.....	63
7.2.6. Almacenamiento de resultados	64

8. Validación del modelo.....	66
8.1. División de Datos (Data Splitting).....	66
8.2. Validación con Datos Reales.....	67
9. Resultados.....	68
9.1. Tipo de cocina.....	70
9.2. Ubicación.....	72
9.3. Rango de precios.....	74
9.4. Tipo de ambiente.....	75
9.5. Precio vs tipo de ambiente.....	76
9.6. Precio vs ubicación.....	78
9.7. Tipo de cocina vs tipo de ambiente.....	80
9.8. Tipo de cocina vs precio.....	81
9.9. Análisis de polaridad.....	82
9.9.1. Aspectos clave.....	84
9.9.2. Rango de precios y tipo de ambiente.....	86
9.9.3. Ubicación.....	89
9.9.4. Tipo de comida.....	90
9.10. Restaurantes mejor calificados.....	91
9.11. Análisis de los restaurantes peor calificados.....	93
10. Estrategias de marketing.....	95

10.1. Marketing de eventos.....	95
10.2. Marketing cultural.....	96
10.3. Marketing de Proximidad	97
10.4. Marketing digital.....	98
10.5. Marketing experiencial	99
11. Conclusiones	102
12. Recomendaciones	104
Referencias Bibliográficas	105

Lista de Figuras

Figura 1. Estructura metodológica de la investigación.....	19
Figura 2. Técnicas de aprendizaje automático.....	27
Figura 3. AI vs. machine learning vs. deep learning.....	29
Figura 4. Flujo general del proceso de minería de datos.....	33
Figura 5. Mapa de correlación de palabras clave.....	40
Figura 6. Mapa de correlación de palabras clave.....	41
Figura 7. Distribución de documentos por año.....	42
Figura 8. Distribución de publicaciones según su país o territorio.....	43
Figura 9. Proporción de documentos según su tipo.....	44
Figura 10. Proporción de documentos según su área de investigación.....	45
Figura 11. Extracción de datos: importación de librerías.....	53
Figura 12. Apertura y desplazamiento dentro del navegador.....	54
Figura 13. Extracción de links.....	55
Figura 14. Almacenamiento de base de datos.....	55
Figura 15. Definición de conjunto de etiquetas.....	58
Figura 16. Definición de función assign_label ().....	59
Figura 17. Ciclo de categorización.....	60
Figura 18. Interpretación del lenguaje natural: Tokenizer.....	63
Figura 19. Ciclo de identificación y asignación de categoría según su probabilidad.....	64
Figura 20. Almacenamiento de resultados.....	65
Figura 21. Análisis de los restaurantes de Rehoboth Beach.....	69

Figura 22. Clasificación de restaurantes según el tipo de cocina	70
Figura 23. Clasificación de restaurantes internacionales.....	71
Figura 24. Clasificación de restaurantes según la ubicación	72
Figura 25. Clasificación de restaurantes según el rango de precios.	74
Figura 26. Clasificación de restaurantes según el tipo de ambiente	75
Figura 27. Análisis de la relación precio vs tipo de ambiente	76
Figura 28. Análisis de la relación precio vs ubicación.....	78
Figura 29. Análisis de la relación tipo de ambiente vs tipo de cocina	80
Figura 30. Análisis de la relación tipo de cocina vs precio.....	81
Figura 31. Análisis general de polaridad de las reseñas.....	83
Figura 32. Clasificación de reseñas según de los aspectos clave.....	84
Figura 33. Análisis de polaridad según aspectos clave	85
Figura 34. Análisis de polaridad según rango de precios.....	87
Figura 35. Análisis de polaridad según rango de precios y tipo de ambiente	88
Figura 36. Análisis de polaridad según la ubicación.....	89
Figura 37. Análisis de polaridad según el tipo de comida	90
Figura 38. Distribución de los mejores restaurantes según tipo de comida y ambiente	91
Figura 39. Análisis de polaridad de los mejores restaurantes según aspectos clave.....	91
Figura 40. Distribución de los peores restaurantes según tipo de comida y ambiente	93
Figura 41. Análisis de polaridad de los peores restaurantes según aspectos clave	93

Lista de Tablas

Tabla 1. Cumplimiento de objetivos del proyecto. 14

Lista de Apéndices

(Los apéndices se encuentran en la carpeta adjunta)

Apéndice A. Base de Datos

Apéndice B. Código de Extracción de Datos

Apéndice C. Código de Polaridad

Apéndice D. Código de Clasificación por categorías.

Apéndice E. Artículo Diseño de un modelo de análisis de sentimientos basado en machine learning aplicado a restaurantes de Rehoboth Beach.

Agradecimientos

A Dios, por brindarme sabiduría, resiliencia y salud, que me han permitido culminar esta importante etapa de mi vida.

A mis padres, por su amor incondicional, su dedicación y la fe inquebrantable que han depositado en mí.

A mis amigos, por su apoyo constante y su compañía a lo largo de este camino.

A mi director, Henry Lamos, por compartir generosamente su experiencia y conocimientos durante el proyecto, y por su invaluable orientación.

A la Universidad Industrial de Santander, por formarme como profesional y contribuir significativamente a mi desarrollo personal e interpersonal.

Resumen

Título: Diseño de un modelo de análisis de sentimientos basado en machine learning aplicado a restaurantes de Rehoboth Beach.¹

Autor: Angela Patricia Blanco Gómez.²

Palabras clave: Machine learning, Procesamiento de lenguaje natural (PLN), minería de datos, análisis de sentimientos, restaurantes, turismo, marketing

Descripción: El presente trabajo se centra en el desarrollo de un modelo de análisis de sentimientos para evaluar reseñas de restaurantes en Rehoboth Beach. Los restaurantes enfrentan retos como la disminución de visitas y competencia con otras opciones de entretenimiento, por eso se realiza este proyecto. El objetivo es apoyar la toma de decisiones estratégicas en marketing, ayudando a los restaurantes a comprender mejor la percepción y satisfacción de sus clientes. Se emplearon técnicas de web scraping para recopilar reseñas en línea de la plataforma OpenTable, seguidas de métodos de minería de datos y procesamiento de lenguaje natural (PLN) para analizar y clasificar las opiniones según su polaridad y temas clave. Los resultados identifican patrones en la experiencia del cliente y áreas de mejora, lo que permite la formulación de estrategias de marketing dirigidas a mejorar la satisfacción del cliente y revitalizar el turismo. El modelo propuesto ofrece una visión detallada de la percepción del cliente y se presenta como una herramienta valiosa para la toma de decisiones informadas en el sector de la hospitalidad y el turismo en Rehoboth Beach.

¹ Trabajo de grado

² Facultad de Ingenierías Físico-Mecánicas. Escuela de Estudios Industriales y Empresariales. Director: Henry Lamos Díaz. PhD. Física-Matemática.

Abstract

Title: Design of a machine learning-based sentiment analysis model applied to restaurants in Rehoboth Beach. ³

Author: Angela Patricia Blanco Gómez. ⁴

Keywords: Machine learning, natural language processing (NLP), data mining, sentiment analysis, restaurants, tourism, marketing.

Description: This work focuses on the development of a sentiment analysis model to evaluate restaurant reviews in Rehoboth Beach. Restaurants face challenges such as decreasing visits and competition with other entertainment options, which is why this project is undertaken. The goal is to support strategic marketing decision-making by helping restaurants better understand customer perceptions and satisfaction. Web scraping techniques were used to collect online reviews from the OpenTable platform, followed by data mining and natural language processing (NLP) methods to analyze and classify opinions based on their polarity and key themes. The results identify patterns in customer experience and areas for improvement, enabling the formulation of marketing strategies aimed at enhancing customer satisfaction and revitalizing tourism. The proposed model offers a detailed view of customer perceptions and serves as a valuable tool for informed decision-making in the hospitality and tourism sector in Rehoboth Beach

³ Bachelor thesis

⁴ Physical-Mechanical Engineering Faculty. School of Industrial and Business Studies. Supervisor: Henry Lamos Díaz. PhD. Mathematical Physics.

Introducción

La experiencia y la opinión de los clientes son aspectos fundamentales para el desarrollo y la mejora continua de la industria del turismo. Con la llegada de las tecnologías de la información y la comunicación, así como el auge del *big data*, esta industria se ha adaptado significativamente para aprovechar las nuevas oportunidades que ofrecen. Hoy en día, numerosos servicios turísticos, tanto para hoteles como para restaurantes, están disponibles en línea a través de plataformas de reservas, sistemas de pedidos en línea y sitios de reseñas y calificaciones. En este contexto, las opiniones y reseñas de los clientes se han convertido en un factor determinante a la hora de tomar decisiones sobre qué establecimientos visitar. Con el aumento exponencial de estas opiniones en línea, ha surgido una creciente necesidad de procesar de manera automatizada estos vastos conjuntos de datos, ya que resulta imposible para las personas analizar todas estas opiniones de manera individual. En este sentido, el presente proyecto se centra en el diseño y desarrollo de un modelo de análisis de sentimientos basado en *machine learning* aplicado específicamente a los restaurantes ubicados en Rehoboth Beach. El objetivo principal es proporcionar a los propietarios y gerentes de estos establecimientos una herramienta eficaz para monitorear y mejorar la experiencia del cliente. Este modelo permitirá identificar tendencias y patrones en las opiniones de los clientes, así como responder de manera proactiva a las críticas y sugerencias recibidas.

Finalmente, este trabajo pretende servir como referencia para investigaciones futuras en el campo del análisis de sentimientos en la industria restaurantera proporcionando una base sólida para la exploración y el avance continuo en este ámbito.

Tabla 1*Cumplimiento de objetivos del proyecto.*

Objetivos	Cumplimiento
Realizar una revisión de la literatura relacionada con el análisis de sentimientos, el procesamiento de lenguaje natural (PLN) y el machine learning aplicado a reseñas de restaurantes y turismo.	Revisión de la literatura
Realizar la recopilación y almacenamiento de reseñas de los principales restaurantes de Rehoboth Beach a través de técnicas de web scraping y minería de datos.	Datos
Seleccionar y adaptar un modelo de análisis de sentimientos pre-entrenado, utilizando técnicas de procesamiento de lenguaje natural (PLN), para clasificar las reseñas recopiladas de los principales restaurantes.	Análisis de sentimientos
Evaluar la precisión y eficacia del modelo de análisis de sentimientos diseñado	Validación del modelo Resultados
Proponer recomendaciones y estrategias de marketing para los restaurantes de Rehoboth Beach, basadas en los resultados del análisis de sentimientos.	Estrategias de marketing

1. Planteamiento del problema

El auge del turismo en Delaware, Estados Unidos, ha sido destacado en los últimos tiempos. A pesar de ser el segundo estado más pequeño del país, con una población de aproximadamente 1 millón de habitantes, el apodado "Primer Estado" registró la llegada de 28.3 millones de turistas en 2021. El sector turístico generó 47,760 empleos, a tiempo completo y parcial, lo que representa el 64 % de los nuevos puestos de trabajo creados en el estado en ese año. El turismo y la hospitalidad se posicionan como el cuarto mayor empleador privado del estado (Welch, 2022) y contribuyeron con \$4 mil millones al producto interno bruto en el año 2022, marcando un récord en la región tras la pandemia de COVID-19.

Kenneth McGill, director gerente en Rockport Analytics afirma que el condado de Sussex es el condado del estado que más rápidamente se recuperó y crece a pasos agigantados, en parte debido a su economía de playa y hospitalidad. (Morris, 2023). Según el reporte de Delaware Sea Grant College, las playas de Delaware generan \$6.9 mil millones anualmente y más de \$711 millones en ingresos fiscales. (Delaware Sea Grant College, 2017)

La ciudad de Rehoboth Beach, que se hace llamar "La Capital de Verano de la Nación" es el resort de playa más popular y visitado del estado. Esta encantadora ciudad costera atrae a personas de todo el mundo debido a sus numerosas actividades recreativas y de ocio, con una destacada oferta de restaurantes. (Schultz, 2024). Dado que los restaurantes son uno de los principales atractivos de la región, su gestión es de vital importancia, ya que no solo impulsan el crecimiento económico de la ciudad al atraer a millones de turistas, sino que también representan un pilar fundamental para los miles de trabajadores de esta industria.

Aunque el condado ha mostrado signos de recuperación tras la pandemia de COVID-19, el sector restaurantero sigue enfrentando desafíos importantes. Con el aumento de los precios, los consumidores son más cautelosos en sus hábitos de gasto, lo que ha añadido una presión adicional a los restaurantes. Estos establecimientos no solo compiten entre sí, sino también con alternativas como servicios de streaming y la compra de comestibles para consumo en casa. Esta situación ha generado una disminución notable en las ventas, evidenciada por una menor afluencia de personas que prefieren salir a comer fuera. (Osborne, 2024) Los clientes, ahora más que nunca, valoran las experiencias culinarias únicas, buscando algo más que lo básico. Al ser una ciudad turística, los restaurantes de la región deben esforzarse por destacar frente a las opciones de otras ciudades, reduciendo al máximo el margen de malas experiencias para fidelizar a su clientela y atraer nuevos visitantes. Ante estos retos, los restaurantes necesitan implementar nuevas herramientas y estrategias innovadoras para adaptarse a las condiciones actuales y superar las dificultades del mercado.

Cada día, el volumen de datos disponibles a través de plataformas digitales aumenta. A diario, miles de personas comparten sus opiniones, sugerencias y experiencias en los distintos restaurantes de la ciudad, evaluando aspectos como el servicio, el ambiente y la calidad de la comida, entre otros factores. Una herramienta que facilita la recopilación y gestión de estos datos es OpenTable, una red social que permite a la comunidad interactuar con los restaurantes de la zona. A través de OpenTable, los usuarios pueden realizar reservas, obtener recompensas y leer y escribir comentarios y reseñas sobre los establecimientos. (Mendiola, 2013). Los datos obtenidos, cuando se procesan y filtran adecuadamente, pueden proporcionar una gran cantidad de información valiosa. Esta información puede ser utilizada para mejorar las experiencias de los

turistas y para impulsar el crecimiento y la mejora continua de los establecimientos de comida locales. (Puh & Bagić Babac, 2023)

El presente trabajo tiene como objetivo emplear *web scraping*, minería de texto y análisis de sentimientos basado en *machine learning* para analizar las reseñas encontradas en OpenTable en la ciudad de Rehoboth Beach.

El análisis de sentimientos permitirá interpretar de manera más efectiva la información proporcionada por los clientes en sus reseñas y comentarios en línea, lo que contribuirá a fundamentar la toma de decisiones para optimizar las estrategias de marketing en el sector restaurantero, ayudando así a enfrentar los desafíos que aún persisten en la industria.

2. Objetivo general

Diseñar un modelo de análisis de sentimientos basado en machine learning y procesamiento de lenguaje natural (PLN) para las reseñas de los principales restaurantes en Rehoboth Beach.

2.1. Objetivos específicos

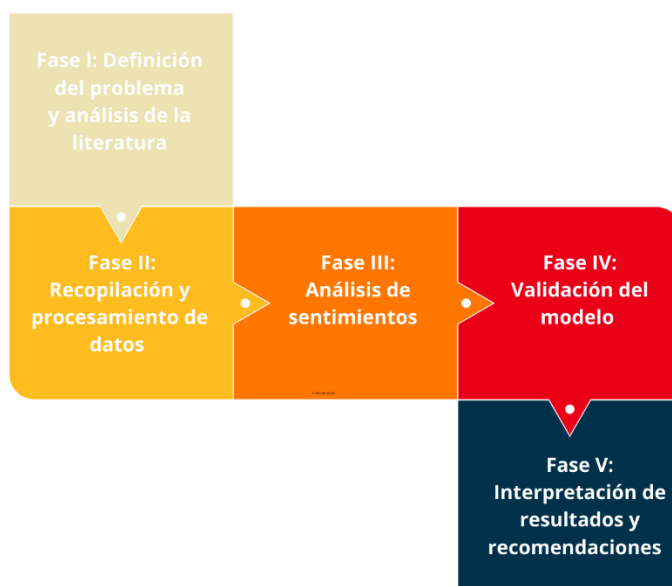
- Realizar una revisión de la literatura relacionada con el análisis de sentimientos, el procesamiento de lenguaje natural (PLN) y el machine learning aplicado a reseñas de restaurantes y turismo.
- Realizar la recopilación y almacenamiento de reseñas de los principales restaurantes de Rehoboth Beach a través de técnicas de web scraping y minería de datos.
- Seleccionar y adaptar un modelo de análisis de sentimientos pre-entrenado, utilizando técnicas de procesamiento de lenguaje natural (PLN), para clasificar las reseñas recopiladas de los principales restaurantes de Rehoboth Beach.
- Evaluar la precisión y eficacia del modelo de análisis de sentimientos diseñado mediante pruebas y métricas pertinentes.
- Proponer recomendaciones y estrategias de marketing para los restaurantes de Rehoboth Beach, basadas en los resultados del análisis de sentimientos, con el objetivo de mejorar la gestión y contribuir al impulso del turismo gastronómico en la ciudad

3. Metodología

La metodología de este proyecto se divide en 5 fases: Definición del problema y análisis de la literatura, recopilación y procesamiento de datos, análisis de sentimientos, validación del modelo y por último la interpretación de resultados y recomendaciones.

Figura 1

Estructura metodológica de la investigación



3.1. Fase I: Definición del problema y análisis de la literatura

En la primera fase, se identifica el problema que motiva el proyecto y se definen los objetivos correspondientes. Posteriormente, se construye una ecuación de búsqueda que será utilizada en la base de datos Scopus para seleccionar los documentos relevantes que puedan aportar a esta investigación. Además, se realiza una búsqueda de literatura gris con el objetivo de recopilar estudios previos relacionados con el análisis de sentimientos, procesamiento de lenguaje natural y

machine learning aplicado a reseñas de restaurantes y turismo, entre otras aplicaciones similares. Este análisis permite identificar las técnicas, algoritmos, modelos y enfoques más utilizados que servirán como base para el presente proyecto.

3.2. Fase II: Recopilación y procesamiento de datos

En esta fase se recopilaron las reseñas de los principales restaurantes de la ciudad de Rehoboth Beach a través de la plataforma OpenTable, empleando técnicas de web scraping. Para ello, se utilizan las herramientas Selenium y BeautifulSoup, las cuales permiten extraer las reseñas directamente desde la plataforma de manera automatizada. Los datos obtenidos son sometidos a un riguroso proceso de limpieza que incluye la eliminación de caracteres especiales, corrección de errores ortográficos, eliminación de duplicados y normalización del texto.

3.3. Fase III: Análisis de sentimientos

En esta fase se lleva a cabo el análisis de las opiniones de los clientes utilizando técnicas de procesamiento de lenguaje natural (NLP) aplicadas a las reseñas de restaurantes en Rehoboth Beach, previamente recolectadas. El análisis de sentimientos se realiza utilizando la librería TextBlob, que permite calcular la polaridad de los comentarios en un rango de -1 (muy negativo) a 1 (muy positivo). El sistema de clasificación de sentimientos se estructura en cinco categorías: "muy positivo", "positivo", "neutral", "negativo" y "muy negativo". Este etiquetado es automatizado, asignando a cada reseña una categoría según su polaridad. A diferencia de los enfoques tradicionales basados en aprendizaje automático, este método se fundamenta en la polaridad léxica de TextBlob, lo que facilita una clasificación precisa y eficiente sin necesidad de entrenar un modelo supervisado.

3.4. Fase IV: Validación del modelo

La validación del modelo de análisis de sentimientos se realiza en dos etapas. Primero, se divide el conjunto de datos en entrenamiento y prueba para que el modelo aprenda a identificar patrones y luego se evalúa en datos no vistos, utilizando métricas como precisión, sensibilidad y especificidad. En la segunda etapa, se comparan los resultados del modelo con las calificaciones de los usuarios (1 a 5 estrellas) en aspectos como "comida", "servicio" y "ambiente", asegurando que las predicciones reflejen las opiniones reales de los consumidores.

3.5. Fase V: Interpretación de resultados y recomendaciones

En la siguiente fase, se lleva a cabo el análisis de los resultados obtenidos del proceso de clasificación de opiniones de los clientes. Esto implica examinar la distribución de las opiniones clasificadas en positivas, negativas y neutras, así como identificar patrones y tendencias significativas en las reseñas. Además, se evalúa la efectividad del modelo de clasificación en la precisión de la categorización de las opiniones. Con base en los resultados del análisis, se proponen recomendaciones y estrategias operacionales y de marketing que ayuden a mejorar la experiencia del cliente y la percepción de los restaurantes de Rehoboth Beach

4. Marco de referencia

4.1. Marco de antecedentes

Este marco de antecedentes se basa en una búsqueda de investigaciones previas en el análisis de sentimientos aplicado a reseñas de restaurantes y aplicaciones similares. El objetivo de esta revisión es examinar las principales técnicas y enfoques utilizados en la literatura académica, lo que servirá como base para el desarrollo del trabajo actual.

El proceso de búsqueda bibliográfica se inició en la biblioteca virtual de la Universidad Industrial de Santander y se extendió a otras bases de datos y recursos académicos relevantes. Aunque los estudios específicos sobre el sector de hostelería y turismo fueron limitados, se encontraron numerosos trabajos relacionados con el análisis de sentimientos que proporcionaron diferentes enfoques y metodologías. A pesar de esta limitación, estos recursos fueron de gran utilidad para el estudio, ya que permitieron analizar diversas metodologías, enfoques y herramientas aplicadas en otros contextos. A continuación, se presentará un resumen de los hallazgos más relevantes obtenidos durante la búsqueda, destacando las tendencias, metodologías y resultados más significativos identificados en la literatura existente.

El trabajo de investigación que se toma como base para el desarrollo del presente trabajo se titula “Segmentación de clientes del sector turístico de Santander mediante el uso de análisis de sentimientos” Gomez Vargas (2023) propone una segmentación de clientes en el sector turístico de Santander. Para recopilar datos, específicamente reseñas de hoteles, la autora utiliza la plataforma TripAdvisor. Luego, lleva a cabo un análisis de sentimientos basado en polaridad. Para este fin, emplea dos métodos diferentes: la biblioteca de TextBlob y el modelo transformer BERT.

Posteriormente, se clasifican y segmentan los clientes utilizando el modelo DAMO-NLP-SG/zero-shot-classify-SSTuning-base. Este modelo analiza las reseñas según distintas etiquetas, como "limpieza", "decoración", "ubicación", entre otras. Finalmente, con los resultados obtenidos y tras limpiar los datos, se procede a clasificar los tipos de turistas según sus características principales destacando la diversidad de preferencias entre los tipos de viajeros. La autora ofrece un enfoque interesante al utilizar el análisis de sentimientos y la minería de texto como herramientas para potenciar la toma de decisiones estratégicas en el sector hotelero de Santander.

En su tesis, Pineda, (2021) se enfoca en emplear técnicas de aprendizaje automático para prever las fluctuaciones de los precios de las acciones en el mercado financiero colombiano. Para esto, utiliza tanto datos estructurados provenientes de la Bolsa de Valores de Colombia como datos no estructurados obtenidos de noticias financieras relacionadas con el índice Colcap, Ecopetrol y Bancolombia. Pineda recopila estos datos utilizando técnicas de web scrapping y luego lleva a cabo un análisis de sentimientos utilizando diversos enfoques, que incluyen algoritmos de Regresión Logística y lexicones como SenticNet y LoughranMcDonald, además de herramientas como VADER y Textblob. Este trabajo busca ser una referencia para futuras investigaciones sobre modelos predictivos y el uso de técnicas de aprendizaje automático y procesamiento de lenguaje natural en el ámbito económico y empresarial.

Torres Samboni, (2015), se centra en comprender las opiniones y percepciones sobre el posconflicto colombiano mediante análisis de sentimientos con herramientas de minería de texto, utilizando principalmente Knime, una plataforma de análisis de datos de código abierto. Se seleccionaron fuentes de información relevantes en Twitter y periódicos nacionales, y se extrajo la información utilizando diversos modelos de flujo en Knime y software adicional para la

extracción de datos de Twitter. Para la etiquetación de sentimientos, se empleó el Spanish Emotion Lexicon (SEL), Se implementó un aprendizaje supervisado para la etiquetación, utilizando el SEL para identificar palabras positivas y negativas en los textos, y posteriormente se interpretaron los resultados. La investigación proporciona una herramienta que ofrece información valiosa y precisa mediante el análisis de sentimientos logrando entender mejor la problemática postconflicto en Colombia.

El artículo “Comparación de métodos de análisis de sentimientos en comunidades de habla hispana” Moreno Sandoval et al. (2022) presentan un análisis comparativo de diferentes enfoques de aprendizaje automático para estudiar la subjetividad colectiva en las redes sociales, centrándose en los modos de pensar, actuar y sentir de las comunidades de habla hispana. Se evalúan modelos clásicos de aprendizaje automático (machine learning), técnicas de *deep learning* con *word embedding* y modelos basados en lexicones utilizando un corpus de Twitter en español proveniente de la competencia TASS2019. Los resultados muestran que el enfoque basado en deep learning supera a los otros métodos en precisión. Finalmente, los autores sugieren como trabajo futuro el entrenamiento de modelos de aprendizaje profundo con embeddings pre-entrenados específicos del español de cada país, lo que podría mejorar aún más la precisión al considerar las particularidades lingüísticas de cada región.

4.2. Marco teórico

4.2.1. Inteligencia artificial

Alan Turing dio el primer paso en 1956 para establecer si una máquina era considerada inteligencia artificial al plantear la pregunta “Puede una máquina pensar?”. La prueba de Turing

es un método para evaluar la capacidad de una máquina para exhibir un comportamiento inteligente similar al humano. Consiste en una interacción entre un juez humano y dos participantes: una máquina y un ser humano. El juez solo puede comunicarse con los participantes a través de texto. Si el juez no puede distinguir cuál de los participantes es la máquina y cuál es el humano basándose únicamente en sus respuestas, se considera que la máquina ha pasado la Prueba de Turing y se le atribuye inteligencia artificial (Blakemore, 2023). Dicho esto, podemos decir que la inteligencia artificial (IA) se refiere a la capacidad de las máquinas para imitar el pensamiento humano y comportarse de manera similar a las personas.

Xu et al. (2021) plantean que la inteligencia artificial tiene 3 capas de aplicación principales

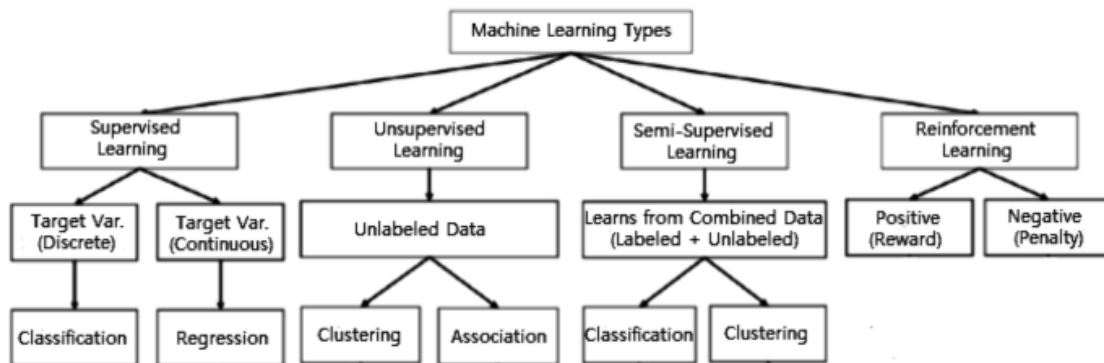
- Capa de visión: Otorga a las máquinas la capacidad de visión, audición, etc. Por ejemplo, identificar imágenes y reconocer elementos del habla.
- Capa cognitiva: se encarga de habilidades más complejas como el pensamiento inductivo, el razonamiento y la adquisición de conocimientos, utilizando recursos como el procesamiento del lenguaje natural (NLP), grafos de conocimiento y aprendizaje en curso.
- Capa de toma de decisiones: La inteligencia artificial puede tomar decisiones ideales, como planificación automática, sistemas expertos y asistencia para la toma de decisiones.

4.2.2. Machine learning

Arthur Samuel, en la década de 1950, definió el aprendizaje automático, conocido como "machine learning" en inglés, como “Un subcampo de la inteligencia artificial que permite a las computadoras aprender sin ser programadas explícitamente”.

Las computadoras se alimentan con grandes cantidades de datos y se les instruye sobre cómo procesar esa información mediante algoritmos. Estos algoritmos les ayudan a identificar patrones en los datos, lo que les permite desarrollar un modelo de cómo funciona el sistema en cuestión. (Tabb et al., 2021) .El algoritmo típico de aprendizaje automático supervisado consta de tres componentes:

- Un proceso de decisión: interpreta los datos para encontrar que tipo de patrón está buscando encontrar el algoritmo.
- Una función de error: evalúa la precisión de las predicciones al compararlas con ejemplos conocidos.
- Un proceso de actualización: ajusta el modelo en función de los errores para mejorar su desempeño futuro. (Berkeley School of Information, 2020)

Figura 2*Técnicas de aprendizaje automático*

Nota. Adaptado de (Sarker, 2021)

Existen diversas técnicas de aprendizaje automático las cuales según su capacidad de adaptación se usan para abordar un problema específico y extraer información útil de los datos para construir aplicaciones inteligentes

En el **aprendizaje supervisado**, un modelo de machine learning aprende a asociar entradas con salidas utilizando ejemplos de conjuntos de datos que están etiquetados, esto con el fin de que la máquina adquiera la capacidad de hacer predicciones. (Shinde, 2024) Las tareas supervisadas más comunes son la "clasificación", que divide los datos en categorías, y la "regresión", que ajusta los datos a una curva. (Sarker, 2021). Algunos ejemplos de aprendizaje supervisado son la filtración de correo electrónico, la evaluación de crédito y el reconocimiento de voz.

En el **aprendizaje no supervisado** no requiere la intervención humana ya que los algoritmos no usan datos etiquetados u organizados para indicar cómo tendría que clasificar la nueva información, sino que tienen que encontrar cómo clasificarlas ellos mismos. (Rouhiainen,

2018). Se utiliza para extraer características generativas, identificar tendencias y estructuras, agrupaciones en resultados y propósitos exploratorios.

El **aprendizaje semisupervisado** combina métodos supervisados y no supervisados, operando tanto con datos etiquetados como no etiquetados. En contextos donde los datos etiquetados son escasos y los no etiquetados son abundantes, el aprendizaje semisupervisado resulta útil. Su objetivo es mejorar las predicciones más allá de lo que podría lograrse utilizando solo datos etiquetados. Se aplica en áreas como traducción automática, detección de fraude y clasificación de texto. (Sarker, 2021)

El **aprendizaje por refuerzo** permite que los agentes de software y las máquinas aprendan a través de la “prueba y error”, utilizando la retroalimentación de sus propias experiencias. Se fundamenta en un sistema de recompensas y castigos, y su meta es hallar un modelo de acción idóneo que maximice la recompensa acumulativa o que minimice el riesgo total del agente. (Bhatt, 2018) Este tipo de aprendizaje ayuda a mejorar la eficiencia de sistemas robóticos, vehículos autónomos, juegos de computadora, entre otros.

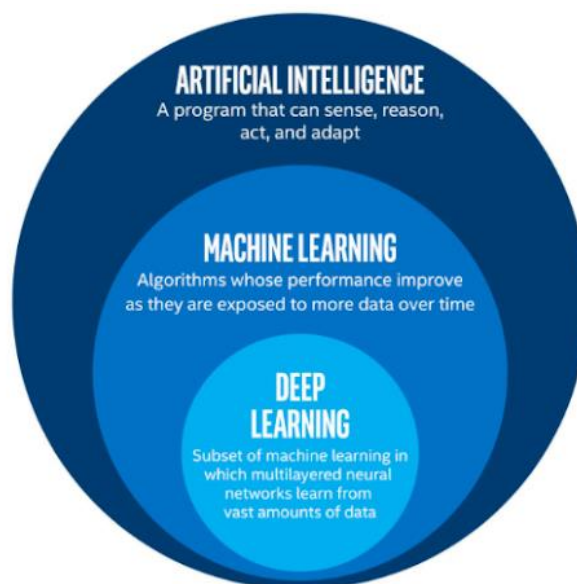
4.2.3. Deep learning

El aprendizaje profundo (deep learning) es una rama del aprendizaje automático (machine learning) que se inspira en la estructura del cerebro humano y en las teorías sobre su funcionamiento. Los algoritmos de aprendizaje profundo buscan obtener conclusiones similares a las que los humanos podrían obtener al analizar datos de manera continua y con una estructura lógica definida. Para lograrlo, utilizan una estructura de algoritmos de múltiples capas conocida como redes neuronales artificiales.

La principal diferencia entre el machine learning y el deep learning radica en la forma en que manejan los datos y aprenden de ellos. En el machine learning, se necesita un paso de extracción de características donde se identifican y seleccionan manualmente las características relevantes de los datos. Por otro lado, en el deep learning el paso de extracción de características ya es parte del proceso que ocurre en una red neuronal artificial. (Oppermann, 2023)

Figura 3

AI vs. machine learning vs. deep learning



Nota. Tomado de: What Is Deep Learning and How Does It Work? (Oppermann, 2023)

4.2.4. Procesamiento de lenguaje natural

El procesamiento del lenguaje natural (PLN o NPL por sus siglas del inglés Natural Language Processing), implica usar el lenguaje humano para interactuar con las computadoras, permitiendo que estas entiendan y respondan a las frases y oraciones que se les proporcionan. Esto facilita la creación de programas que pueden realizar tareas relacionadas con el lenguaje, como la

traducción automática o el análisis de sentimientos. Además, el PLN ayuda a desarrollar modelos que nos permiten entender mejor cómo los seres humanos procesan y utilizan el lenguaje. (Revista De investigación De Sistemas E Informática, 2009).

El procesamiento del lenguaje natural (PLN) abarca varias tareas tradicionales, entre ellas:

- Desarrollo de sistemas de diálogo: Creación de sistemas conversacionales que mejoran la precisión de la comunicación, ya sea mediante chat o reconocimiento de voz.
- Extracción y recuperación de información: Desarrollo de métodos informáticos para localizar documentos con datos textuales no estructurados.
- Traducción automática o semiautomática: Técnicas para crear programas que traducen texto de un idioma a otro, usando métodos estadísticos o neuronales. (Báez et al., 2022)

4.2.4.1. TextBlob. TextBlob es una biblioteca de Python diseñada para el procesamiento de datos textuales. Ofrece una interfaz sencilla para abordar diversas tareas de procesamiento del lenguaje natural (PLN), como el etiquetado de partes del discurso, la extracción de frases nominales, el análisis de sentimientos, la clasificación y más. La biblioteca de lexico de TextBlob es denominada `en-sentiment.xml`, un archivo XML que registra los distintos significados de una misma palabra, identificados por diferentes códigos. Este archivo almacena un total de 2919 registros recopilados por la biblioteca. TextBlob utiliza un enfoque simple para calcular el sentimiento de una palabra. En lugar de asignar un valor fijo a cada palabra, TextBlob busca esa palabra en su biblioteca de léxico, donde puede haber varias entradas para la misma palabra, cada una con diferentes polaridades y subjetividades. Por ejemplo, la palabra "bueno" podría tener múltiples entradas con diferentes grados de positividad.

De esta forma, cuando TextBlob encuentra una palabra en el texto, busca todas las entradas asociadas a esa palabra y calcula un promedio de sus polaridades y subjetividades. Esto le permite obtener una evaluación más precisa del sentimiento de esa palabra en el contexto del texto en el que se encuentra. Además, TextBlob también considera palabras que modifican o negan el sentimiento, como "muy" o "no", para proporcionar una evaluación aún más precisa del texto. (Insight-Group, n.d.)

4.2.5. Word embedding

Word embedding es una técnica de procesamiento del lenguaje natural (PLN) que se utiliza para representar palabras en un espacio vectorial continuo. Esta técnica asigna a cada palabra un vector de números reales, lo que permite capturar tanto las relaciones semánticas como las sintácticas entre las palabras. (Brownlee, 2019) No obstante, no consideran el valor emocional o

de sentimiento de cada palabra que se encuentra en un documento.(Giatsoglou et al., 2017) La investigación en la selección de características de word embedding cobró impulso en 2013, destacando principalmente tres algoritmos: Word2Vec, GloVe y FastText. (Mutinda et al., 2023).

4.2.5.1. Transformer. El modelo Transformer es una arquitectura revolucionaria en el campo del procesamiento del lenguaje natural creado como una alternativa innovadora a las arquitecturas de redes neuronales recurrentes tradicionales, como las LSTM (Long Short-Term Memory), un tipo de red neuronal recurrente (RNN) diseñada específicamente para modelar secuencias de datos y capturar dependencias a largo plazo.

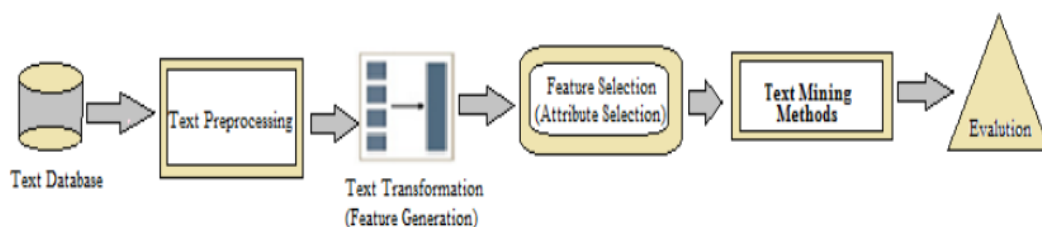
En lugar de depender de la recursividad para capturar relaciones de dependencia en secuencias de datos, el Transformer utiliza mecanismos de atención para enfocarse en partes específicas de la secuencia durante el procesamiento. Esto permite al modelo capturar relaciones de largo alcance de manera más efectiva y paralelizar el proceso de entrenamiento, lo que resulta en una mayor eficiencia y capacidad de escalabilidad. Un mecanismo de atención se refiere a la capacidad del modelo para enfocarse en partes específicas de una secuencia de entrada durante el proceso de codificación o decodificación. En lugar de procesar toda la secuencia de manera uniforme, el modelo puede asignar diferentes pesos o niveles de atención a diferentes partes de la secuencia, lo que le permite concentrarse en los elementos más relevantes para la tarea en cuestión. Esto permite al modelo capturar relaciones de largo alcance y manejar de manera efectiva secuencias de longitud variable. (Vaswani et al., 2017)

4.2.6. Minería de texto

La minería de texto, también conocida como minería de datos textuales, es el proceso de descubrir conocimiento útil y patrones significativos y/o no evidentes a partir de grandes colecciones de datos textuales no estructurados. Utiliza técnicas de procesamiento del lenguaje natural (NLP) para extraer información valiosa e ideas de documentos, correos electrónicos, publicaciones en redes sociales y otros textos no estructurados. (Kapilparshi, 2023)

Figura 4

Flujo general del proceso de minería de datos



Nota. Tomado de (Patel & Soni, 2012)

4.2.6.1. Pre-procesamiento de datos. Son las etapas de preparación y limpieza de los datos textuales antes de aplicar técnicas de análisis de texto

- Tokenización: es el proceso de dividir un texto en partes más pequeñas (tokens), como palabras o frases, eliminando espacios y signos de puntuación. (S & R, 2016)
- Eliminación de *stopwords*: Consiste en remover palabras que inciden muy poco o no contienen información útil, como artículos, pronombres, conjunciones, preposiciones, etc.

- Stemming: Consiste en eliminar los sufijos de las palabras para reducirlas a su raíz o forma base. Por ejemplo, convierte "playing", "played" y "plays" en la raíz "play".(Garcia Morgado, 2016)

4.2.6.2. Transformación de texto. Convertir un documento de texto en un modelo de bolsa de palabras (bag of words) o en una notación de modelo de espacio vectorial (Vector Space Model).

- Bolsa de Palabras (Bag of Words): La bolsa de palabras es una manera sencilla de representar un texto en forma numérica. Consiste en crear un "saco" que contiene todas las palabras únicas presentes en el texto, y luego contar cuántas veces aparece cada una de estas palabras en el texto.(Topper, 2023)
- Modelo de espacio vectorial (Vector Space Model): Representa los documentos como vectores en un espacio multidimensional. Cada dimensión corresponde a una palabra del vocabulario total del corpus y la posición de un documento en este espacio está determinada por las frecuencias de las palabras que contiene. (Garcia Morgado, 2016)

4.2.6.3. Selección de características/atributos. En esta fase, se eliminan las características que no son relevantes para la minería de datos. Esto permite trabajar con conjuntos de datos más pequeños, realizar menos cálculos y reducir el espacio de búsqueda necesario.(Patel & Soni, 2012)

4.2.6.4. Métodos de minería de texto. En esta etapa se aplican diferentes métodos de minería de texto tales como:

- **Clustering:** El clustering agrupa documentos similares en categorías sin tener etiquetas predefinidas. Es una técnica de aprendizaje no supervisado donde el objetivo es dividir un conjunto de datos en grupos (o clusters) de tal manera que los elementos dentro de un mismo grupo sean más similares entre sí que con los elementos de otros grupos.
- **Clasificación:** La clasificación asigna etiquetas predefinidas a los documentos. Es una técnica de aprendizaje supervisado donde el modelo se entrena con un conjunto de datos etiquetados. Luego, el modelo puede predecir la etiqueta de nuevos documentos. (Gorini, n.d.)
- **Análisis de temas:** Es una técnica de aprendizaje automático que clasifica y comprende grandes volúmenes de texto, asignando etiquetas o categorías basadas en el tema o asunto de cada documento. (Meddeb & Romdhane, 2022)

4.2.6.5. Interpretación o Evaluación. En esta fase se evalúan e interpretan los resultados mediante el cálculo de métricas como precisión, recall, exactitud, medida F, entre otras.

4.2.7. Selenium.

Selenium es una herramienta de código abierto para la automatización de pruebas de aplicaciones web en diferentes navegadores. Está principalmente desarrollada en Java y es compatible con varios lenguajes de programación. (Savage, 2024). Se utiliza principalmente para pruebas automáticas de aplicaciones web, pero también es muy útil para realizar web scraping (extraer datos de páginas web). Con Selenium, puedes controlar un navegador (como Chrome o

Firefox) a través de scripts, simulando interacciones de un usuario real, como hacer clic en botones, llenar formularios, o navegar entre páginas. (Chinmayee Deshpande, 2024)

4.2.8. Análisis de sentimientos.

En el análisis de sentimientos, también conocido como minería de opiniones, el propósito radica en identificar las percepciones y valoraciones expresadas por individuos a través de texto escrito. El término "sentimiento" engloba la idea de las emociones, experiencias personales y actitudes hacia determinado tema u objeto. Comparado con las herramientas analíticas convencionales que consideran que las pequeñas discrepancias entre dos textos apenas afectan su significado, el análisis de sentimientos es mucho más susceptible a incluso las diferencias más mínimas en el texto. (Adedoyin-Olowe et al., 2014). El Análisis de Sentimientos (SA) se clasifica en tres niveles principales:

- **Nivel de documento:** Se basa en la suposición de que todo el texto contiene ideas sobre un solo tema. A menudo, esta suposición es incorrecta, ya que un documento puede abordar múltiples temas.
- **Nivel de frase:** Asume que cada oración comunica solo un tema. Sin embargo, es común que una sola oración contenga múltiples temas o puntos de vista opuestos. La polaridad del sentimiento se determina para la oración completa.
- **Nivel de aspecto:** Busca determinar la polaridad del sentimiento expresado para cada aspecto específico mencionado en el texto. Permite una investigación más detallada utilizando los datos complementarios de la reseña. (Londhe & Rao, 2022)

A lo largo de la última década, se han empleado diversas técnicas para llevar a cabo el análisis de sentimientos abarcando desde enfoques basados en lexicón hasta métodos de aprendizaje automático (machine learning) y técnicas de aprendizaje profundo (deep learning). (Jurafsky & Martin, 2024)

En el análisis de sentimientos basado en léxicos, se evalúa cada palabra de un texto según si tiene un significado positivo o negativo según un léxico predefinido. Después, se calcula el sentimiento general del texto sumando estos valores y aplicando alguna operación de agrupación, como calcular el promedio de los valores de las palabras. (Puh & Bagić Babac, 2023). Entre los lexicones más comunes se encuentra el SentiWordNet el cual asigna puntuaciones de positividad, negatividad y neutralidad a palabras en un texto. Estas puntuaciones se basan en un modelo léxico que asigna valores de sentimiento a las palabras según su significado en el contexto (Roul, 2021). Otro lexicón utilizado para el análisis de sentimientos es VADER, (Valence Aware Dictionary and Sentiment Reasoner), el cual está especialmente diseñado para detectar los sentimientos expresados en las redes sociales. VADER no solo proporciona un puntaje de polaridad, sino que también nos indica cuán positiva o negativa es una opinión. (Bajaj, 2024)

Por otro lado, algunas de las técnicas más populares de machine learning en el análisis de sentimientos son: El método de Naive Bayes el cual se basa en el teorema de Bayes. En el contexto del análisis de sentimientos, el método Naive Bayes calcula la probabilidad de que un texto sea positivo, negativo o neutral dadas las palabras o características específicas presentes en él. (Gomez Vargas, 2023); Support vector machine (SVM) es un algoritmo utilizado para la clasificación, cuyo objetivo es encontrar un límite de decisión óptimo entre las diferentes clases de datos en un espacio de características, de manera que pueda predecir con precisión la clase a la que pertenecen nuevas

instancias de datos. (Shi & Li, 2011); y arboles de decisión los cuales utilizan reglas de decisión basadas en características de los datos para clasificar instancias en categorías o predecir valores de una variable objetivo. (Dueñas Quesada, 2020)

5. Revisión de la literatura

En la primera fase de investigación se realiza una revisión de la literatura con el objetivo de examinar los modelos de análisis de sentimientos aplicados a las reseñas de restaurantes y turismo. Para ello, se realizó una búsqueda en diversas fuentes; una de las fuentes son las bases de datos Scopus y ScienceDirect, conocidas por su amplia cobertura de revistas científicas y académicas. El proceso de revisión bibliográfica también involucró la búsqueda activa en repositorios institucionales, bibliotecas digitales y otros recursos bibliográficos especializados. Se realizaron análisis bibliográficos para identificar estudios relevantes que aborden específicamente los modelos de análisis de sentimientos en el contexto de las reseñas de restaurantes y turismo. Asimismo, se consideraron documentos de literatura gris, como informes técnicos, tesis no publicadas y otros recursos relevantes que podrían proporcionar información adicional sobre el tema.

5.1. Análisis bibliométrico

Para comenzar, se realizó una búsqueda utilizando el motor de Google, lo que permitió acceder a una amplia gama de fuentes de literatura gris. Estas incluyeron páginas web de instituciones académicas, como la Universidad Industrial de Santander, y otras universidades nacionales e internacionales. Además, se exploraron sitios web de conferencias y seminarios, así como blogs y redes sociales científicas como ResearchGate. Esto facilitó la obtención de informes técnicos, tesis y otros documentos no convencionales que podrían contener datos pertinentes para la investigación en cuestión. Se utilizó Google Scholar para agilizar la búsqueda de información científica y acceder eficientemente a artículos académicos relacionados con el análisis de

sentimientos en reseñas de restaurantes y en el ámbito del turismo. Cabe destacar que, aunque no se estableció un rango específico de años de publicación para obtener información a través de literatura gris, se dio mayor relevancia a la información obtenida en los últimos seis años.

Por otro lado, se utilizó la plataforma ScienceDirect, reconocida por su amplio catálogo de artículos académicos revisados por pares, para complementar y enriquecer aún más la información obtenida. Esto permitió acceder a investigaciones rigurosamente evaluadas y publicadas en revistas científicas de renombre, ampliando así el espectro de la revisión bibliográfica y asegurando la inclusión de estudios relevantes y actualizados.

La elección de Scopus como plataforma principal para la búsqueda de documentos académicos fue fundamental. Se utilizó una ecuación de búsqueda específica para obtener los resultados más relevantes. Posteriormente, se aplicaron filtros para excluir áreas de estudio y palabras clave no pertinentes al tema de investigación. Además, se estableció un rango de años de publicación de 2019 a 2024 para garantizar la obtención de información actualizada, especialmente relevante en el contexto post-pandemia. Este proceso meticuloso llevó a identificar 44 artículos académicos seleccionados, asegurando la calidad y relevancia de los documentos obtenidos para el estudio en cuestión.

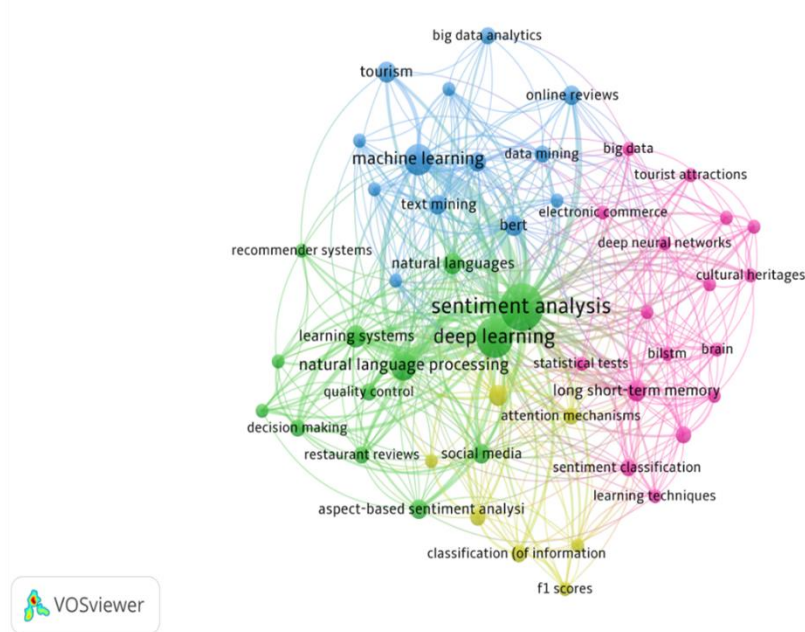
Figura 5

Ecuación de búsqueda

TS= ((NLP OR "deep learning" OR "machine learning") AND ("sentiment analysis")
AND (restaurants OR food OR tourism OR culinary OR hospitality) AND (opinions OR reviews
OR comments))

Figura 6

Mapa de correlación de palabras clave

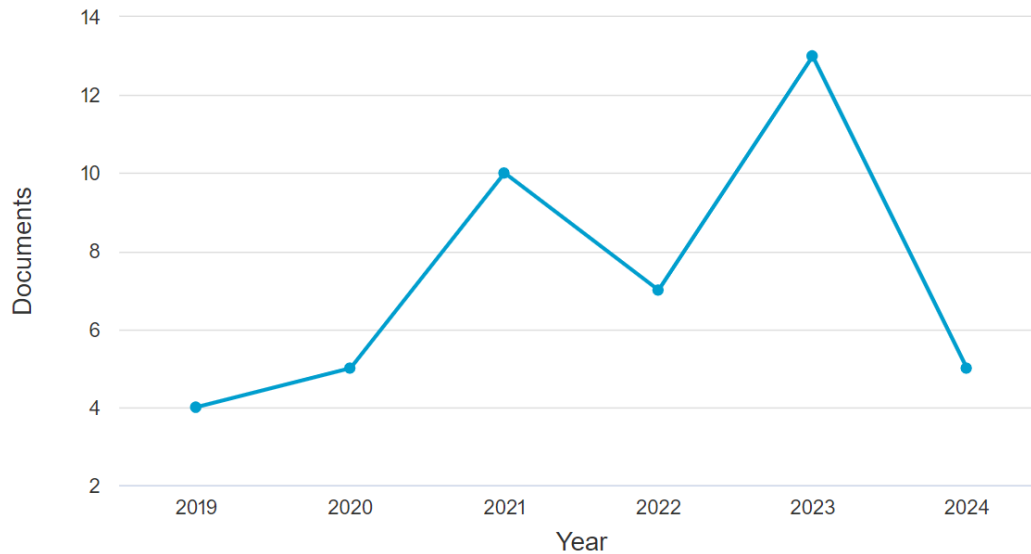


Nota. Figura adaptada de VosViewer

En la figura se presenta el análisis de correlación de palabras que se llevó a cabo utilizando el software VOSviewer. Este análisis permitió identificar las relaciones y asociaciones entre diferentes términos clave presentes en los artículos revisados. Los resultados revelaron que ciertas palabras, como "sentiment analysis", "deep learning", "machine learning", "text mining", "tourism" y "online reviews", destacan significativamente en la literatura revisada. Esto sugiere que estos conceptos están estrechamente relacionados y son prominentes en el ámbito del análisis de sentimientos para reseñas de restaurantes y turismo.

Figura 7

Distribución de documentos por año.

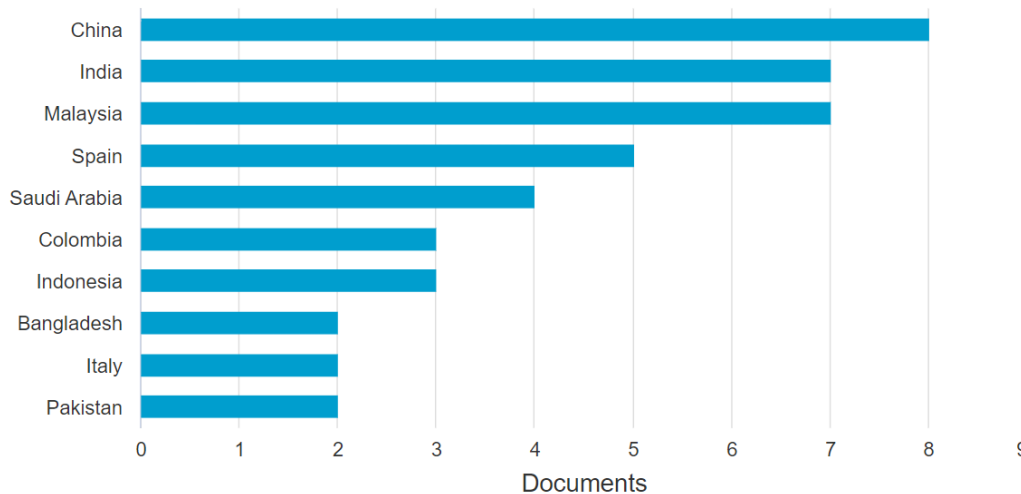


Nota. Figura adaptada del buscador Scopus

El análisis por años de la plataforma Scopus reveló una tendencia variable en el número de documentos publicados sobre análisis de sentimientos en el sector turístico y de restaurantes desde 2019 hasta 2024. Aunque hubo fluctuaciones en la cantidad de publicaciones a lo largo de estos años, se observa una tendencia general al alza. En particular, el año 2023 destacó como el año con más documentos publicados hasta la fecha, con un total de 13 documentos. Este aumento progresivo en la producción de documentos sugiere que aún existe un creciente interés y actividad en la investigación sobre análisis de sentimientos en este ámbito. Además, indica que en los años posteriores es probable que continúe esta tendencia al alza, con una mayor cantidad de investigaciones y avances en este campo específico.

Figura 8

Distribución de publicaciones según su país o territorio

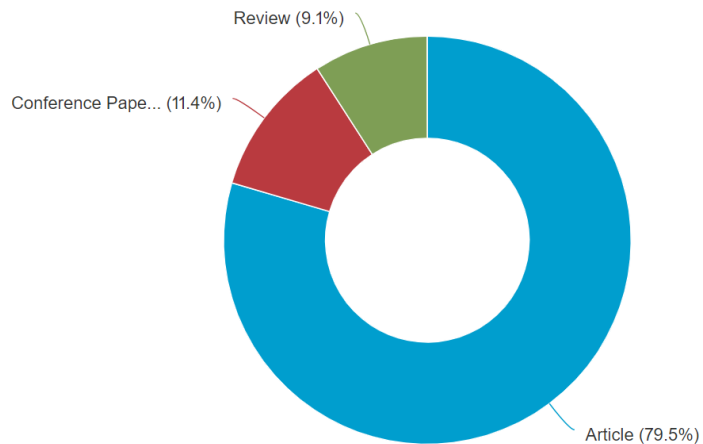


Nota. Figura adaptada del buscador Scopus

Al analizar los aportes de países, se identificó que China encabeza la lista de países con mayor contribución de documentos en análisis de sentimientos para el sector turístico y de restaurantes, con 8 documentos. Le siguen de cerca India y Malasia, cada uno con 7 documentos respectivamente. Sorprendentemente, Colombia figura como el único país americano representado, con 3 documentos publicados en 2021. Estos hallazgos indican que los países asiáticos, como China, India y Arabia Saudita e Indonesia, lideran el campo de investigación en esta área específica, demostrando un alto nivel de interés y actividad en la producción de conocimiento relacionado con el análisis de sentimientos en el contexto del turismo y la restauración.

Figura 9

Proporción de documentos según su tipo

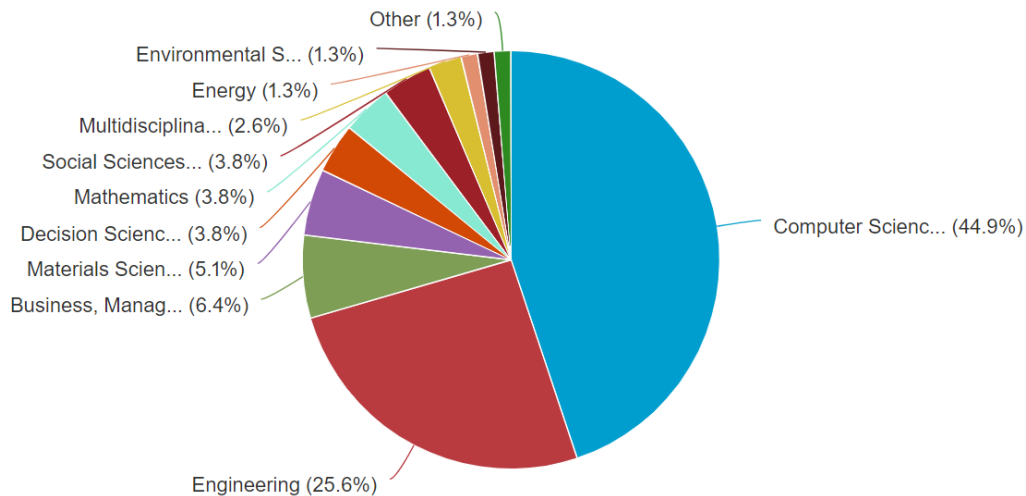


Nota. Figura adaptada del buscador Scopus

Según el análisis del tipo de documentos, se encontró que la mayoría son artículos, representando el 79.5% del total. Luego, un 11.4% corresponden a ponencias, y finalmente, los artículos de revisión constituyen un 9.1% del total de documentos.

Figura 10

Proporción de documentos según su área de investigación.



Nota. Figura adaptada del buscador Scopus

Finalmente, en la figura 6 se observa que la distribución de publicaciones según las áreas de investigación es liderada por el área de Ciencias Computacionales, con un 44.9%, seguida por Ingeniería, con un 25.6%. Estas dos áreas son las principales, mientras que áreas como Negocios, Administración y Contabilidad, Ciencias de la Decisión y Ciencias Sociales representan un porcentaje menor, con un 6.4% y 3.8%, respectivamente, en comparación con las dos áreas principales.

5.2. Análisis preliminar de la literatura

En los últimos años, ha surgido un creciente interés en la aplicación de la inteligencia artificial, técnicas de machine learning y deep learning, y análisis de sentimientos en el sector turístico y de la hospitalidad. La convergencia de la tecnología y estas industrias ha dado lugar a avances significativos, impulsados por el acceso a datos proporcionados por los clientes, lo que contribuye al mejoramiento general de la experiencia brindada. (Neiva et al., 2021)

Los avances tecnológicos y los aportes académicos han posibilitado el desarrollo de herramientas que contribuyen al análisis predictivo de la demanda en estas industrias (Muñoz et al., 2023a) así como a la mejora en la gestión y optimización de las ofertas disponibles. Además, se han desarrollado sistemas de recomendación turísticos basados en emociones, lo que representa una innovación significativa en la personalización de las experiencias turísticas (Santamaria-Granados et al., 2021). Estos avances tienen el potencial de transformar la forma en que se interactúa con los clientes y se ofrecen servicios en el sector turístico y de la hospitalidad.

Investigaciones recientes han proporcionado evidencia sobre el impacto positivo de las calificaciones y la valencia del sentimiento de las reseñas en línea en la rentabilidad de los servicios turísticos. (Bigne et al., 2023). En un estudio de síntesis realizado por (Yang et al., 2019) que abarcó 25 estudios en la industria turística y hotelera, se encontró que tanto las calificaciones de los clientes como la cantidad de reseñas en línea afectan el desempeño financiero de los hoteles y restaurantes. Dado el crecimiento exponencial en la cantidad de reseñas en línea generadas por los usuarios, se hace cada vez más necesario el procesamiento automatizado de estos grandes volúmenes de datos. Esto se debe a que resulta impracticable para los humanos leer y analizar todas estas reseñas por sí mismos. (Puh & Bagić Babac, 2023).

El análisis de sentimientos (SA) es una disciplina dentro del procesamiento del lenguaje natural (PLN) que utiliza la minería de texto y tecnologías relacionadas para categorizar el texto subjetivo en clases de opiniones, emociones u otras categorías relevantes (Mutinda et al., 2023). Esta área de estudio busca comprender cómo las personas perciben y sienten acerca de diversos temas, como productos, servicios o empresas.

Desde principios de los años 2000, el procesamiento de lenguaje natural (PLN) ha experimentado un crecimiento dinámico, especialmente en el campo del análisis de sentimientos (Shaeali et al., 2020). En los últimos diez años, se han empleado diversas técnicas para analizar sentimientos, incluyendo enfoques basados en lexicones, machine learning y deep learning. (Jurafsky & Martin, 2024)

En el análisis de sentimientos basado en lexicones, cada palabra en una oración tiene un valor semántico y una intensidad que determinan si la oración es positiva o negativa. Esto se logra mediante el uso de un léxico predefinido que clasifica las palabras en categorías positivas y negativas. (Babac & Podobnik, 2016). Las técnicas basadas en lexicones son simples, pero dependen en gran medida del etiquetado manual del texto. (Giatsoglou et al., 2017)

Khoo & Johnkhan (2018) presentan un nuevo lexicón de sentimientos llamado WKWSCI Sentiment Lexicon y lo comparan con cinco lexicones existentes en términos de su efectividad para categorizar sentimientos en diferentes tipos de textos. Se evaluó la precisión de estos lexicones utilizando conjuntos de datos de reseñas de productos y titulares de noticias. Los resultados mostraron que los léxicos WKWSCI, Multi-perspective Question Answering (MPQA) Subjectivity Lexicon, Hu & Liu Opinion Lexicon y Semantic Orientation Calculator (SO-CAL) lexicón son igualmente efectivos para categorizar sentimientos en reseñas de productos y el lexicón WKWSCI fue el más preciso para la categorización de sentimientos en titulares de noticias. En su estudio, Khan & Baharudin (2011) emplearon el SentiWordNet para asignar una puntuación a cada término en una oración, y la clasificación general de la oración se determinó a partir de la suma total de las puntuaciones individuales de los términos. Sin embargo, una limitación identificada en este enfoque es que ciertas palabras pueden tener la misma orientación sentimental,

pero negarse entre sí, lo que podría resultar en una clasificación incorrecta de la oración. Shaeali et al. (2020) en su artículo “Customer reviews analytics on food delivery service in social media” revisaron 53 distintas investigaciones y concluyeron que el enfoque basado en lexicón sobresale en la predicción de confiabilidad del análisis de sentimientos por encima de otras técnicas basadas en machine learning, NLP, Support Vector Machine (SVM) y minería de texto.

Las técnicas basadas en lexicones, aunque efectivas, enfrentan una limitación al aplicarse al análisis de sentimientos debido a que la mayoría de estos están contruidos con un enfoque muy general (Puh & Bagić Babac, 2023). El machine learning complementa este enfoque al capturar patrones más complejos y contextuales en el lenguaje. En este sentido, la combinación de ambos enfoques emerge como una estrategia poderosa para un análisis de sentimientos más preciso y completo. (Rezaeinia et al., 2019)

En un estudio realizado por (Bagherzadeh et al., 2021) se propuso superar el método estándar de análisis de sentimientos basado en lexicones mediante la creación de dos nuevos lexicones (uno ponderado y otro manual). Estos fueron validados utilizando métricas de precisión de clasificación en reseñas de TripAdvisor. Los resultados demostraron un rendimiento superior al de un método basado en el lexicón de SentiWordNet y a un algoritmo de aprendizaje automático Naïve Bayes en la clasificación de sentimientos. Por otro lado, Kang et al. (2012) abordaron también la limitación del lexicón de sentimientos SentiWordNet para abarcar y analizar las reseñas de restaurantes. Presentaron un algoritmo Naïve Bayes mejorado combinado con un nuevo lexicón de sentimientos que redujo la brecha entre la precisión de las clasificaciones positivas y negativas de las reseñas.

El método Naïve Bayes, ampliamente utilizado en machine learning para clasificar textos, ofrece resultados precisos y eficientes. No obstante, tiene limitaciones, como su sensibilidad a la selección de características. Una gran cantidad de características puede afectar negativamente su rendimiento de clasificación. (Fatah et al., 2024)

Existen otras técnicas de machine learning que son ampliamente utilizadas en el análisis de sentimientos, por ejemplo, Muñoz et al. (2023b) llevan a cabo un estudio para predecir el turismo en municipios afectados por el conflicto colombiano, específicamente determinando si los turistas pernoctan o no. Para lograr esto, emplean algoritmos de clasificación binaria, como Support Vector Machine (SVM), árboles de decisión y bosques aleatorios los cuales son implementados utilizando la biblioteca sklearn de Python.

Aunque los métodos previos han funcionado bien, recientemente el enfoque de aprendizaje profundo, conocido como deep learning en inglés, ha ganado terreno en el análisis de sentimientos y otras tareas de procesamiento de lenguaje natural. (Puh & Bagić Babac, 2023). El aprendizaje profundo, o deep learning, maneja grandes volúmenes de datos y emplea redes neuronales artificiales para analizar patrones de texto de manera más eficiente (Umarani et al., 2021) Una de las técnicas más utilizadas en este ámbito son los word embeddings. Varios autores resaltan la efectividad de esta técnica, ya que ofrece diversas ventajas en comparación con la representación de bolsa de palabras, como una dimensionalidad menor y una mayor coherencia semántica entre palabras similares en el espacio vectorial(Rezaeinia et al., 2019)

El modelo Bidirectional Encoder Representations from Transformers (BERT) es una arquitectura de deep learning que utiliza técnicas de word embedding, y ha ganado prominencia gracias a su enfoque bidireccional y mecanismo de atención, sobresaliendo en el análisis de

sentimientos (Prattasha et al., 2022). Varios investigadores respaldan la superioridad de BERT sobre otros modelos de procesamiento del lenguaje natural.

Cao et al. (2023) emplearon el modelo BERT para analizar los sentimientos de 17,555 opiniones en línea en chino sobre nueve lugares turísticos en Melaka, Malasia, evidenciando que el enfoque de aprendizaje profundo basado en BERT mejora la predicción de los sentimientos. Asimismo, se destacó la superioridad de BERT en el análisis de sentimientos de consumidores de productos en línea (Chang & Zhu, 2023) y de reseñas de restaurantes en Portugues (Branco et al., 2024). Por otro lado, en su estudio de análisis de sentimientos en bengalí, Prattasha et al. (2022) compararon el modelo BERT con diversas técnicas de word embeddings como Word2Vec, GloVe y fastText, así como con algoritmos clásicos de machine learning, concluyendo que el modelo BERT supera a estas técnicas. Por último, Sayeed et al. (2023) exploraron y evaluaron las aplicaciones del modelo BERT analizando sentimientos en campos, como el comercio electrónico, la agricultura, la puntuación de ensayos automatizados, Twitter y Google Play, concluyendo una vez más que el modelo BERT supera a otros modelos en varias tareas de comprensión del lenguaje.

Aunque BERT ha demostrado un rendimiento superior en comparación con otras técnicas, presenta limitaciones que requieren abordarse. Mutinda et al. (2023) se centran en enfrentar los desafíos asociados con la alta dimensionalidad de los vectores y la falta de información semántica en los mismos. Proponen un modelo denominado LeBERT, que combina lexicones de sentimientos, N-gramas, BERT y CNN. Este enfoque se evalúa en tres conjuntos de datos distintos: reseñas de productos de Amazon, reseñas de películas de IMDb y reseñas de restaurantes en Yelp. Los resultados experimentales demuestran que la combinación de lexicones de sentimientos y N-

gramas con el algoritmo de incrustación BERT genera vectores de palabras más representativos, lo que resulta en una mejora del rendimiento predictivo del modelo de análisis de sentimientos en comparación con otros modelos de word embedding existentes.

6. Datos

En esta sección se describe el proceso utilizado para la recolección y tratamiento de los datos fundamentales para este trabajo. El objetivo principal fue obtener los nombres de los restaurantes y sus respectivas reseñas a través de la plataforma en línea OpenTable. Esta información se utilizó como base para el análisis de sentimientos y la categorización de las opiniones de los clientes, proporcionando un panorama detallado sobre la experiencia en cada restaurante.

El análisis se realizó sobre una muestra de 38 restaurantes y se decidió limitar la recolección de datos a 10 páginas de reseñas por restaurante en lugar de extraer todas las reseñas (que pueden superar las 2000) por varias razones:

- **Variedad sobre Cantidad:** Una muestra representativa de reseñas captura suficientes opiniones diversas sin necesidad de procesar miles de reseñas redundantes. Más datos no siempre aportan más valor.
- **Optimización de Recursos:** Con hardware limitado, procesar grandes volúmenes de datos sería ineficiente y lento. Limitar la recolección reduce el costo computacional y optimiza el tiempo de análisis.
- **Impacto Mínimo en Restaurantes con Pocas Reseñas:** En restaurantes con menos de 90 reseñas, extraer más datos no mejora el análisis. Para restaurantes con más de 2000 reseñas, los resultados se estabilizan con una muestra.
- **Simplicidad en el Código:** Limitar el número de páginas evita sobrecomplicar el código, haciéndolo más eficiente y fácil de ajustar en futuros análisis.

El primer paso en el proceso consiste en configurar un entorno que permita controlar un navegador web de forma automática. Para ello, se utiliza Selenium, una herramienta popular que facilita la interacción con páginas web como si se tratara de un usuario real. Esto incluye la apertura de una página específica, la navegación entre secciones y la simulación de acciones como hacer clic en enlaces o desplazarse por la página.

Figura 11

Extracción de datos: importación de librerías

```
from selenium import webdriver
from selenium.webdriver.chrome.service import Service as ChromeService
from webdriver_manager.chrome import ChromeDriverManager
from bs4 import BeautifulSoup
import csv
import time

driver = webdriver.Chrome(service=ChromeService(ChromeDriverManager().install()))
driver.get(main_page_url)
```

Nota. Figura adaptada de Visual Studio Code

Este código configura un navegador para que abra automáticamente la URL correspondiente a los restaurantes de Rehoboth Beach. Esto es como automatizar la navegación hacia la página web que contiene la información que se desea extraer. Muchas páginas web no cargan todo su contenido de una vez; a menudo, elementos adicionales se muestran conforme el usuario se desplaza hacia abajo en la página (técnica conocida como *lazy loading*). Por lo tanto, es crucial implementar un mecanismo que permita desplazarse hasta el final de la página para asegurarse de que toda la información relevante esté visible y lista para ser extraída.

Figura 12

Apertura y desplazamiento dentro del navegador

```
def progressive_scroll(driver, scroll_pause_time=1, scroll_increment=900):
    last_height = driver.execute_script("return document.body.scrollHeight")
    while True:
        driver.execute_script(f"window.scrollTo(0, {scroll_increment});")
        time.sleep(scroll_pause_time)
        new_height = driver.execute_script("return document.body.scrollHeight")
        if new_height == last_height:
            break
        last_height = new_height

progressive_scroll(driver)
```

Nota. Figura adaptada de Visual Studio Code

El código simula el desplazamiento hacia abajo de la página de forma progresiva, similar a usar la rueda del ratón para ver más contenido. La pausa (*sleep*) entre desplazamientos permite que nuevos elementos, como los enlaces a los restaurantes, se carguen correctamente.

Una vez que la página ha cargado completamente su contenido, el siguiente paso es recopilar los enlaces de cada restaurante listado. Este paso es clave para obtener más detalles específicos sobre cada establecimiento, como su nombre, dirección y reseñas.

Figura 13*Extracción de links*

```
main_soup = BeautifulSoup(main_page_source, 'html.parser')
restaurant_links = main_soup.select('a.qCITanV81-Y-[href*="/r/"]')

for url in unique_links:
    driver.get(url)
    soup = BeautifulSoup(page_source, 'html.parser')
    restaurant_name_tag = soup.find(class_='E-vwXONV9nc-')
    restaurant_name = restaurant_name_tag.text.strip() if restaurant_name_tag else "No"
```

Nota. Figura adaptada de Visual Studio Code

Aquí, el código localiza todos los enlaces a restaurantes mediante su clase CSS, navega a cada uno de ellos y extrae el nombre del restaurante. Este proceso es como anotar la dirección web de cada restaurante y luego ingresar a cada uno para obtener detalles específicos.

Una vez recopilada la información, el siguiente paso es organizarla en un formato adecuado para su posterior análisis. Los datos extraídos, como nombres de restaurantes, direcciones, reseñas y calificaciones se almacenan en archivos CSV, un formato de tabla comúnmente utilizado que permite una fácil manipulación y análisis en herramientas como Excel o Python.

Figura 14*Almacenamiento de base de datos*

```
with open(csv_file_path, "w", newline="", encoding="utf-8") as file:
    writer = csv.writer(file)
    writer.writerow(["Restaurant Name", "Review Text"])
    writer.writerows(reviews)
```

Nota. Figura adaptada de Visual Studio Code

Este fragmento guarda la información obtenida en un archivo CSV, similar a una hoja de cálculo. Los datos estructurados de esta manera permiten un análisis más profundo y comparativo, facilitando la identificación de patrones y tendencias clave dentro del sector gastronómico

7. Análisis de sentimientos

7.1. Análisis de sentimientos según su polaridad

En el análisis de datos textuales, el análisis de sentimientos según su polaridad es una técnica clave que permite clasificar textos en categorías como positivo, negativo, o neutral. Este proceso implica la aplicación de modelos de inteligencia artificial que interpretan el tono emocional de las palabras, frases o reseñas completas. A continuación, se detalla cómo se lleva a cabo este análisis, desde la carga de los datos hasta la clasificación y almacenamiento de los resultados, proporcionando una visión clara del sentimiento expresado en cada texto.

7.1.1. Importación de librerías

El primer paso del proceso consiste en la importación de dos librerías clave para la manipulación de archivos y el análisis de sentimientos: CSV y TextBlob.

- **CSV:** Esta librería es empleada para gestionar archivos CSV (Comma Separated Values), los cuales contienen los datos de las reseñas estructurados en formato tabular. Este tipo de archivo facilita el almacenamiento y manipulación de grandes volúmenes de datos textuales.
- **TextBlob:** Es una herramienta de procesamiento de lenguaje natural (NLP, por sus siglas en inglés) que permite analizar el texto y calcular su polaridad, es decir, determinar si el sentimiento expresado es positivo, negativo o neutral. TextBlob devuelve un valor numérico que oscila entre -1 y 1, donde -1 representa un sentimiento completamente negativo, 0 representa neutralidad, y 1 un sentimiento completamente positivo.

7.1.2. Carga del archivo CSV

En esta etapa, se especifica la ruta del archivo CSV que contiene las reseñas a analizar. El archivo se carga en modo lectura ('r') y se configura con una codificación adecuada, como 'latin-1', para asegurar el manejo correcto de caracteres especiales. Este archivo actúa como la base de datos de las reseñas de los clientes, estructurado en filas que representan cada reseña y columnas que contienen diversos atributos, como el contenido del texto y su fecha.

7.1.3. Estructura para almacenar los resultados

Antes de proceder al análisis de sentimientos, se prepara una estructura que permita almacenar los resultados del análisis. Para ello, se define una lista vacía denominada *sentiment_results*, donde se registrará la polaridad y categoría de cada reseña procesada.

Figura 15

Definición de conjunto de etiquetas

```
sentiment_results = []

sentiment_labels = {
    'muy positivo': (0.6, 1.0),
    'positivo': (0.2, 0.6),
    'neutral': (-0.2, 0.2),
    'negativo': (-0.6, -0.2),
    'muy negativo': (-1.0, -0.6)
}
```

Nota. Figura adaptada de Visual Studio Code

Además, se establece un conjunto de etiquetas de sentimiento *sentiment_labels* () que permiten clasificar las reseñas en categorías específicas. Estas etiquetas corresponden a rangos

de polaridad, como "muy positivo", "positivo", "neutral", "negativo" y "muy negativo", proporcionando una visión más granular de los sentimientos expresados en las reseñas.

7.1.4. Función para asignar etiquetas de sentimiento

Se define una función denominada *assign_label()* cuya función es asignar una etiqueta de sentimiento adecuada a cada reseña en función de la polaridad calculada por TextBlob.

Figura 16

Definición de función assign_label ()

```
def assign_label(polarity):  
    for label, (lower, upper) in sentiment_labels.items():  
        if lower <= polarity < upper:  
            return label  
    return 'unknown'
```

Nota. Figura adaptada de Visual Studio Code

Por ejemplo, una reseña con una polaridad de 0.7 se clasifica como "muy positivo", mientras que una polaridad cercana a 0 se clasifica como "neutral". Esta función permite categorizar de manera coherente las reseñas, facilitando así el análisis de grandes volúmenes de datos.

7.1.5. Leer las reseñas y realizar el análisis

El archivo CSV se procesa fila por fila, extrayendo el contenido textual de cada reseña. Una vez extraído el texto, se emplea TextBlob para calcular su polaridad.

Figura 17*Ciclo de categorización*

```
with open(csv_file_path, "r", encoding="latin-1") as file:
    reader = csv.DictReader(file)
    for row in reader:
        review_text = row['Review Text']
        blob = TextBlob(review_text)
        sentiment = blob.sentiment.polarity
        sentiment_label = assign_label(sentiment)
        sentiment_results.append([review_text, sentiment_label])
```

Nota. Figura adaptada de Visual Studio Code

El valor de polaridad devuelto por TextBlob, que varía entre -1 y 1, es luego procesado mediante la función `assign_label()`, que clasifica la reseña en una de las categorías predefinidas según el rango de polaridad correspondiente. Este proceso permite obtener una representación clara y estructurada del sentimiento dominante en cada reseña.

7.2. Análisis de sentimientos según su categoría

La clasificación de sentimientos en categorías específicas es un componente crucial del análisis de datos textuales, ya que permite desglosar los textos en temas clave para comprender mejor la opinión expresada. Para realizar este análisis, se emplea el modelo pre-entrenado "DAMO-NLP-SG/zero-shot-classify-SSTuning-base", el cual identifica y clasifica los principales temas presentes en cada reseña, como "comida", "servicio" o "ambiente". A continuación, se describe detalladamente el proceso que sigue este modelo para clasificar los textos en las categorías mencionadas y cómo se almacenan los resultados para su análisis posterior.

7.2.1. Carga de herramientas y configuración del modelo

El primer paso consiste en importar las librerías y herramientas necesarias para ejecutar el análisis de sentimientos:

- **Pandas:** Esta librería es esencial para la lectura y manipulación de datos estructurados, como tablas, de manera eficiente.
- **Transformers:** Permite el uso de modelos pre-entrenados de análisis de texto, incluyendo el modelo que clasifica el texto en diferentes categorías.
- **Torch:** Se utiliza para manejar los cálculos intensivos que se requieren en el procesamiento de datos, distribuyéndolos entre el CPU o el GPU según disponibilidad.
- **String y random:** Son herramientas básicas de Python utilizadas para la manipulación de texto y para mezclar datos de manera aleatoria cuando sea necesario.

7.2.2. Carga del modelo y el tokenizer

En este paso se cargan dos componentes fundamentales para el análisis:

- **Tokenizer:** Esta herramienta descompone el texto en fragmentos o "tokens" que el modelo puede procesar y entender.
- **Modelo:** El modelo de clasificación, "DAMO-NLP-SG/zero-shot-classify-SSTuning-base", utiliza los tokens generados para determinar a qué categoría pertenece cada texto. Este modelo preentrenado permite clasificar reseñas según los temas predominantes, como "comida", "servicio" o "ambiente". La interacción entre el modelo y el tokenizer es esencial para que el sistema interprete y clasifique correctamente las reseñas, actuando como el "núcleo" del análisis de sentimientos.

7.2.3. Configuración del dispositivo

El siguiente paso es definir el dispositivo en el que se ejecutarán los cálculos:

- **CUDA (GPU):** Si se dispone de una GPU, se utilizará para ejecutar los cálculos de manera más rápida y eficiente.
- **CPU:** Si la GPU no está disponible, el código se ejecutará en la CPU, aunque a una velocidad menor.

Seleccionar el dispositivo adecuado garantiza que el procesamiento de las reseñas se realice de la manera más eficiente posible, optimizando el tiempo de ejecución.

7.2.4. Función para procesar y analizar el texto

Se implementa una función que procesa y analiza el contenido de cada reseña: El texto se organiza y etiqueta de manera que el modelo pueda analizarlo adecuadamente. Posteriormente el modelo genera una predicción que indica cuál es la categoría que mejor describe el sentimiento expresado en la reseña, por ejemplo, "comida" o "servicio". Además, el modelo proporciona las probabilidades asociadas a cada categoría, permitiendo evaluar la confianza de la predicción.

Figura 18

Interpretación del lenguaje natural: Tokenizer

```
def check_text(model, text, list_label, shuffle=False):
    list_label_new = list_label + [tokenizer.pad_token] * (20 - len(list_label))
    text = f'{list_label_new} {tokenizer.sep_token} {text}'

    model.to(device).eval()
    encoding = tokenizer([text], truncation=True, max_length=512, return_tensors='pt')
    item = {key: val.to(device) for key, val in encoding.items()}
    logits = model(**item).logits
    logits = logits[:, 0:len(list_label)]
    probs = torch.nn.functional.softmax(logits, dim=-1).tolist()
    predictions = torch.argmax(logits, dim=-1).item()

    return predictions, probs
```

Nota. Figura adaptada de Visual Studio Code

Esta función es esencial para que el modelo pueda "leer" y entender las reseñas, y para que realice las predicciones de manera estructurada y coherente.

7.2.5. Lectura y categorización las reseñas

Una vez configurado el modelo, se procede a leer el archivo CSV que contiene las reseñas de los clientes. Para cada reseña:

- Se imprime el número de reseña para llevar un control del análisis.
- La función *check_text()* se utiliza para analizar el contenido de la reseña, asignándole una categoría basada en los resultados del modelo.
- Los resultados, que incluyen el texto de la reseña, la categoría asignada y las probabilidades asociadas, se almacenan para su uso posterior.

Figura 19

Ciclo de identificación y asignación de categoría según su probabilidad

```
data = pd.read_csv('rehoboth_beach_reviews_E.csv', encoding='latin-1')

reviews = []
labels = []
probabilities_list = []

for index, row in data.iterrows():
    review_text = row['Review Text']
    print(f'Analizando review {index + 1}...')
    prediction, probabilities = check_text(model, review_text, list_label)
    label = list_label[prediction]

    reviews.append(review_text)
    labels.append(label)
    probabilities_list.append(probabilities)
```

Nota. Figura adaptada de Visual Studio Code

Este proceso permite realizar el análisis categórico de cada reseña de manera sistemática y eficiente, facilitando la identificación de temas recurrentes y áreas de mejora.

7.2.6. Almacenamiento de resultados

Finalmente, se crea un nuevo archivo que contiene los resultados del análisis:

- **DataFrame:** Los resultados del análisis (incluyendo el texto de la reseña, la categoría asignada y las probabilidades) se organizan en una tabla.
- **Guardar en CSV:** Esta tabla se guarda en un archivo CSV titulado 'Rehoboth Beach_L.csv', el cual puede ser utilizado para análisis posteriores o compartido con otros investigadores.

Figura 20*Almacenamiento de resultados*

```
df = pd.DataFrame({'Review Text': reviews, 'Label': labels})
for i, label in enumerate(list_label):
    df[label + ' Probability'] = [probs[i] for probs in probabilities_list]

df.to_csv('Rehoboth Beach_L.csv', index=False)
print('¡Análisis completado y resultados guardados en el archivo CSV!')
```

Nota. Figura adaptada de Visual Studio Code

El almacenamiento de los resultados en un archivo CSV garantiza que los datos estén accesibles para futuros análisis o revisiones, y facilita su manipulación por otras personas o equipos.

8. Validación del modelo

La validación de modelos de clasificación y análisis de sentimientos es un paso crucial para garantizar que los modelos implementados representen con precisión la realidad observada en los datos. En este proyecto, se ha empleado la técnica de división de datos para separar el conjunto de datos en dos partes: una para entrenamiento y otra para prueba. Posteriormente, se realiza una validación basada en datos reales, comparando las predicciones del modelo con las calificaciones directas proporcionadas por los usuarios en plataformas de reseñas. De este modo, es posible evaluar la precisión y eficacia de los modelos implementados en la clasificación de categorías como "comida", "servicio" y "ambiente", y en la polaridad de los sentimientos.

8.1. División de Datos (Data Splitting)

El proceso comenzó con la división del conjunto de datos en dos partes: conjunto de entrenamiento y conjunto de prueba. El conjunto de entrenamiento fue utilizado para ajustar los modelos de clasificación, permitiendo que estos aprendieran a identificar patrones en los datos y realizar predicciones en las categorías y polaridades establecidas. Una vez entrenado el modelo, el conjunto de prueba se empleó para evaluar su rendimiento, midiendo métricas clave como la precisión, la sensibilidad y la especificidad.

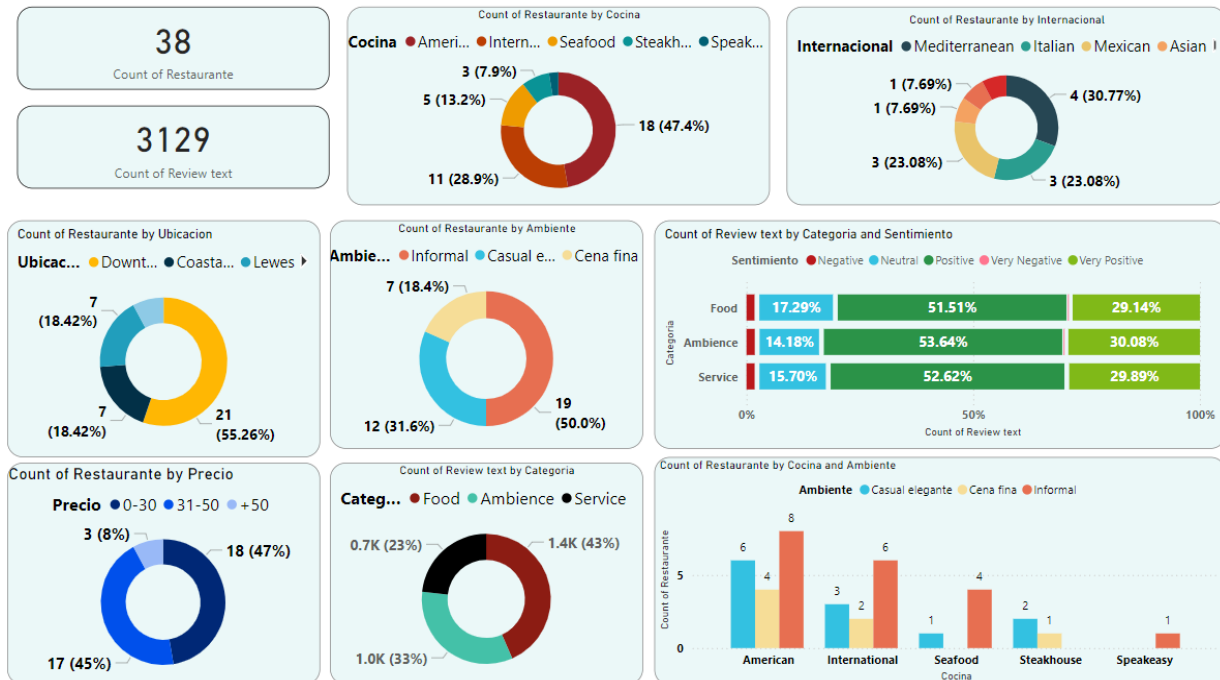
Esta técnica asegura que los modelos puedan generalizar adecuadamente a nuevos datos y no se sobre ajusten a los datos de entrenamiento. La capacidad del modelo para hacer predicciones precisas en el conjunto de prueba es fundamental para determinar su aplicabilidad en datos no vistos.

8.2. Validación con Datos Reales

Tras la etapa de prueba, se procedió a una validación más cercana al mundo real, comparando los resultados obtenidos por el modelo con las puntuaciones y calificaciones proporcionadas por los consumidores en la plataforma de reseñas. Las puntuaciones de 1 a 5 otorgadas por los usuarios fueron utilizadas como referencia para validar la polaridad de los sentimientos, mientras que las calificaciones en categorías como "comida", "servicio" y "ambiente" fueron comparadas con las asignaciones automáticas realizadas por el modelo. Esta comparación permitió evaluar si el modelo lograba capturar con precisión las opiniones e intenciones expresadas por los consumidores, proporcionando así una validación práctica y significativa de su desempeño en un entorno real.

9. Resultados

Tras realizar el análisis de sentimientos y la categorización de las reseñas según las principales características de los restaurantes, se recopiló toda la información en un archivo CSV. Para garantizar la calidad y precisión del análisis, se llevó a cabo una depuración de los datos en Excel, corrigiendo errores y haciendo los ajustes necesarios en la base de datos. Este proceso de limpieza fue crucial para asegurar la integridad y coherencia de los datos, permitiendo así una importación fluida y precisa en Power Query. Una vez que los datos estuvieron correctamente depurados, se importaron a Power Query, una herramienta que facilita la transformación y estructuración de la información de acuerdo con los requisitos específicos del análisis. Power Query permitió realizar ajustes adicionales en la estructura y formato de los datos, proporcionando un control más detallado para optimizar la organización y preparación de los mismos para el análisis final. Posteriormente, se emplearon las capacidades de visualización de Power BI Desktop para desarrollar un tablero interactivo y visualmente atractivo. Este tablero permite a los usuarios explorar y analizar los datos de manera dinámica, ofreciendo una representación clara y accesible de los resultados. Gracias a las funciones interactivas de Power BI, fue posible identificar patrones, tendencias y conclusiones clave de manera eficiente.

Figura 21*Análisis de los restaurantes de Rehoboth Beach*

Nota. Figura adaptada de Power BI

Se extrajeron un total de 3,129 reseñas que proporcionaron una visión detallada de la experiencia de los clientes, destacando los aspectos positivos y negativos de cada establecimiento durante su visita. A través del uso de modelos de análisis de sentimientos y técnicas de procesamiento de lenguaje natural (PLN), se logró identificar la actitud de los consumidores en relación con los principales aspectos evaluados en la plataforma web de Opentable.

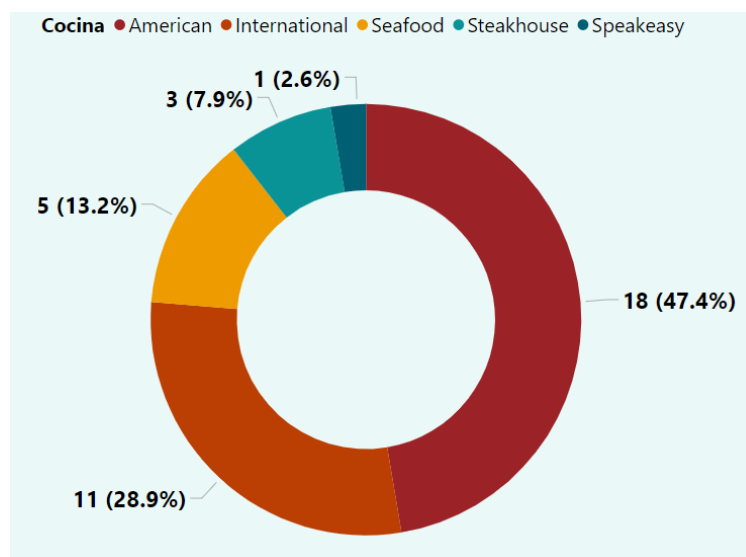
Para la clasificación de los restaurantes, se seleccionaron cuatro categorías clave: tipo de cocina, ubicación, rango de precios y ambiente. Estas categorías permiten una segmentación precisa de los establecimientos, facilitando un análisis más detallado de las percepciones de los

clientes. Al capturar los factores más influyentes en la experiencia del cliente, estas categorías ofrecen una visión completa del rendimiento de cada restaurante en función de sus características particulares.

9.1. Tipo de cocina

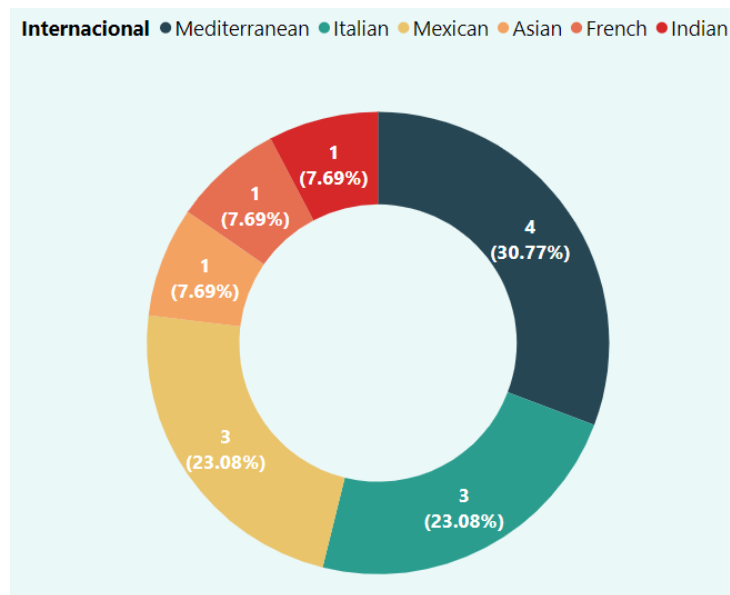
Figura 22

Clasificación de restaurantes según el tipo de cocina



Nota. Figura adaptada de Power BI

La cocina americana es la más predominante en la región con el 47.4% del total de restaurantes en Rehoboth Beach. El restante 52.6% se distribuye entre la cocina internacional, seafood (pescados y mariscos), steakhouse (especializada en cortes de carne), y speakeasy, en ese orden.

Figura 23*Clasificación de restaurantes internacionales*

Nota. Figura adaptada de Power BI

Sin embargo, cuando se desglosa respecto a los distintos tipos de comida internacional, se identifica que la comida mediterránea es la más popular, seguida por la asiática y la mexicana. De esto podemos concluir que, aunque la comida internacional está relativamente bien representada en el área, no se observa una gran diversidad en términos de la variedad de cocinas internacionales ofrecidas.

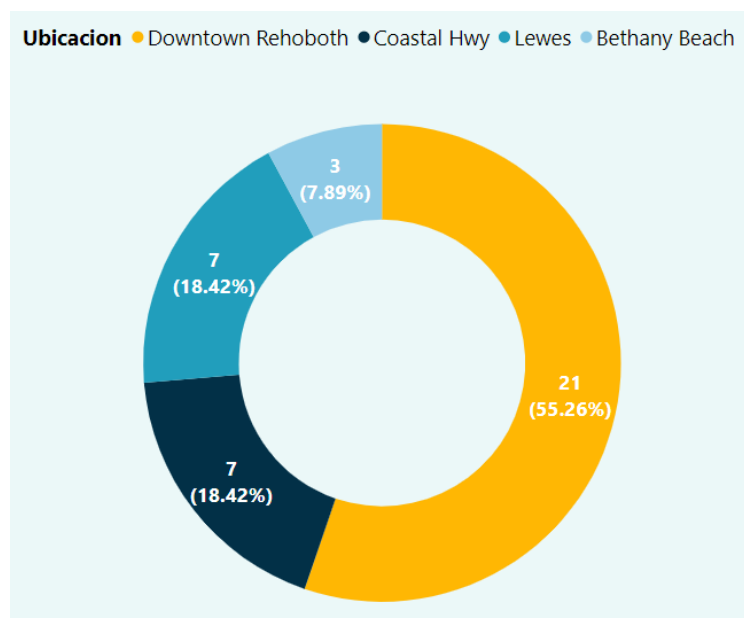
La distribución demográfica de Rehoboth Beach podría explicar este patrón, dado que el 93.1% de la población es blanca, seguida por un 2.8% de hispanos y un 1.5% de otras razas, que incluyen asiáticos, afroamericanos y otros grupos étnicos. Esta demografía influye en las preferencias culinarias y la demanda de determinados tipos de cocina en la región.

Por otro lado, los restaurantes de mariscos (*seafood*) son más populares que los restaurantes de BBQ y los especializados en cortes de carne (*steakhouse*) debido a la ventaja geográfica de estar cerca de la costa, la cultura local centrada en el consumo de mariscos frescos, y las preferencias de los visitantes, que tienden a buscar experiencias gastronómicas auténticas y frescas cuando visitan una ciudad costera como Rehoboth Beach.

9.2. Ubicación

Figura 24

Clasificación de restaurantes según la ubicación



Nota. Figura adaptada de Power BI

Según la gráfica, podemos observar que los restaurantes en el área de Rehoboth Beach se distribuyen en cuatro localidades principales: el centro de la ciudad (downtown), la autopista Coastal Highway, y las zonas de Lewes y Bethany.

El centro de la ciudad (downtown) alberga la mayor concentración de restaurantes, con 21 locales registrados en la muestra. Esta alta densidad se debe a su ubicación estratégica como el principal centro turístico de la ciudad, donde se encuentra el popular paseo marítimo (boardwalk). Muchos turistas prefieren comer en restaurantes cerca de la playa, lo que convierte al downtown en un lugar privilegiado para los negocios gastronómicos.

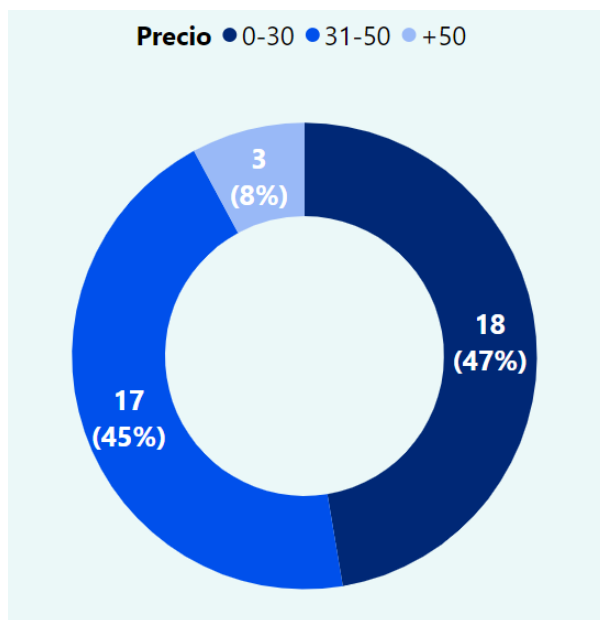
En segundo lugar, se encuentran los restaurantes ubicados a lo largo de la Coastal Highway, una vía clave para los visitantes que llegan a la ciudad en automóvil. Estos restaurantes suelen atraer a aquellos que buscan opciones más accesibles o menos concurridas que las del centro de la ciudad, pero que aún desean una experiencia de calidad.

Finalmente, Lewes y Bethany, dos vecindarios conocidos por su tranquilidad y su rica historia y cultura, tienen una menor cantidad de restaurantes en comparación con las otras áreas. Su atractivo radica en ofrecer experiencias gastronómicas más relajadas y auténticas, que combinan bien con el entorno histórico y cultural de estos vecindarios.

9.3. Rango de precios

Figura 25

Clasificación de restaurantes según el rango de precios.



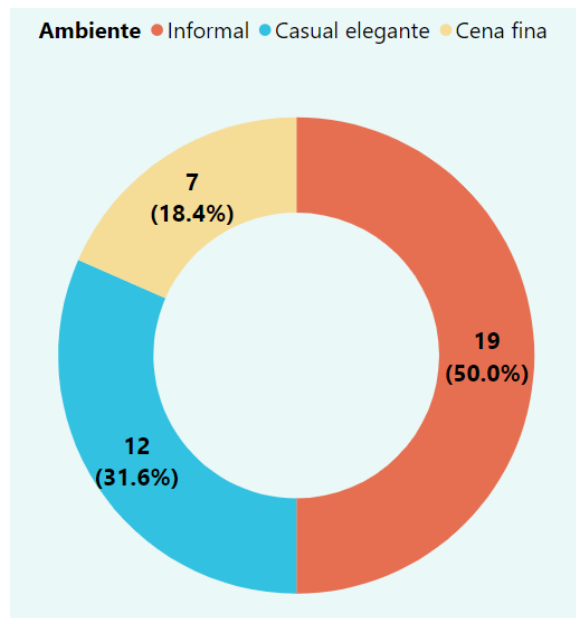
Nota. Figura adaptada de Power BI

Al analizar la distribución de los restaurantes en Rehoboth Beach según el rango de precios, la mayoría se concentran en los segmentos de precios accesibles. Los restaurantes en el rango de \$0-\$30 predominan, seguidos por aquellos en el rango de \$31-\$50. Esto indica que la oferta gastronómica en la región está orientada principalmente a una clientela que busca opciones asequibles y de rango medio. Solo una pequeña cantidad de establecimientos pertenecen al grupo de más de \$50, lo que sugiere que los restaurantes de alta gama representan una minoría, probablemente dirigidos a un nicho específico de clientes que buscan experiencias gastronómicas más exclusivas.

9.4. Tipo de ambiente

Figura 26

Clasificación de restaurantes según el tipo de ambiente



Nota. Figura adaptada de Power BI

En cuanto a la clasificación de los restaurantes por tipo de ambiente, los resultados muestran una clara preferencia por los espacios más informales. Los restaurantes de ambiente informal son los más numerosos, con 19 establecimientos, caracterizados por una atmósfera relajada y sin códigos de vestimenta estrictos.

El siguiente grupo son los restaurantes de ambiente casual elegante, con 12 locales. Estos ofrecen una experiencia más sofisticada, a menudo con un código de vestimenta semiformal, lo que los convierte en opciones ideales para ocasiones especiales o cenas más formales, sin llegar a ser demasiado exclusivos.

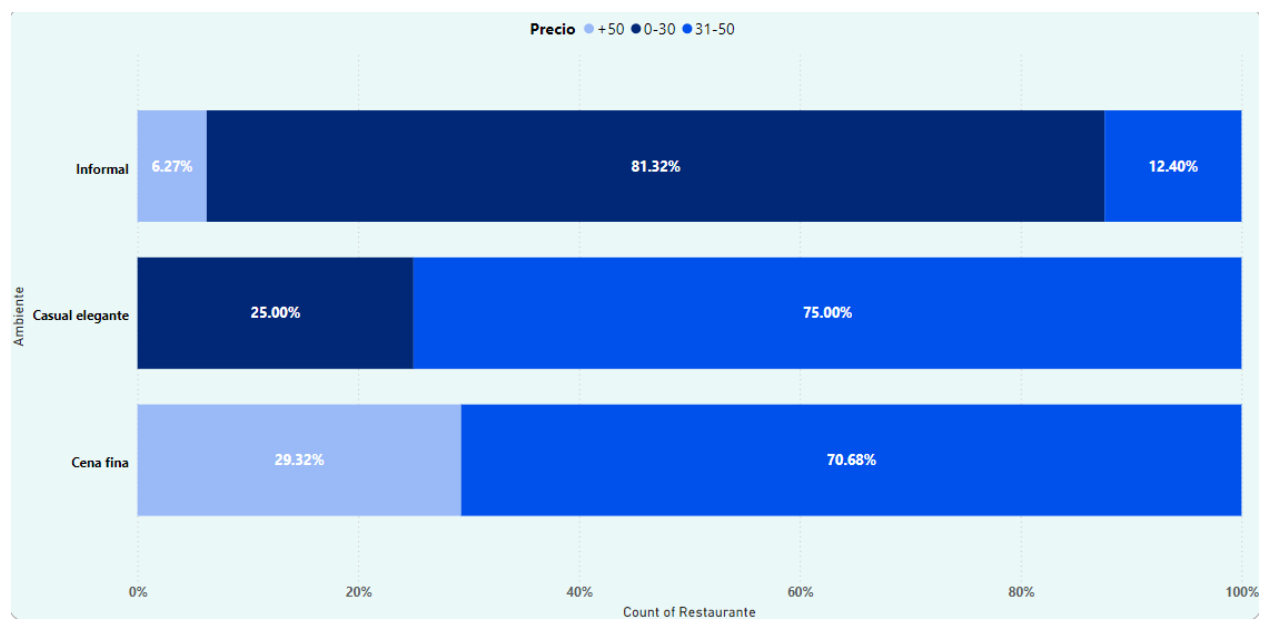
Por último, hay 7 restaurantes de cena fina. Estos representan el segmento más exclusivo, con un ambiente refinado y un servicio más formal, donde los códigos de vestimenta y la

atención al detalle son prioritarios. Este tipo de restaurantes está dirigido a clientes que buscan una experiencia gastronómica de lujo y es ideal para cenas importantes o celebraciones especiales.

9.5. Precio vs tipo de ambiente

Figura 27

Análisis de la relación precio vs tipo de ambiente



Nota. Figura adaptada de Power BI

El análisis de la relación entre el rango de precios y el tipo de ambiente en los restaurantes de Rehoboth Beach revela patrones interesantes que destacan la accesibilidad respecto a la diversidad gastronómica de la ciudad.

- **Restaurantes de ambiente informal:** La mayoría de estos restaurantes se encuentran en el rango de precios de \$0-\$30, representando un 81.32%. Un número más reducido de estos establecimientos cae en el rango de \$31-\$50, y solo un restaurante informal excede

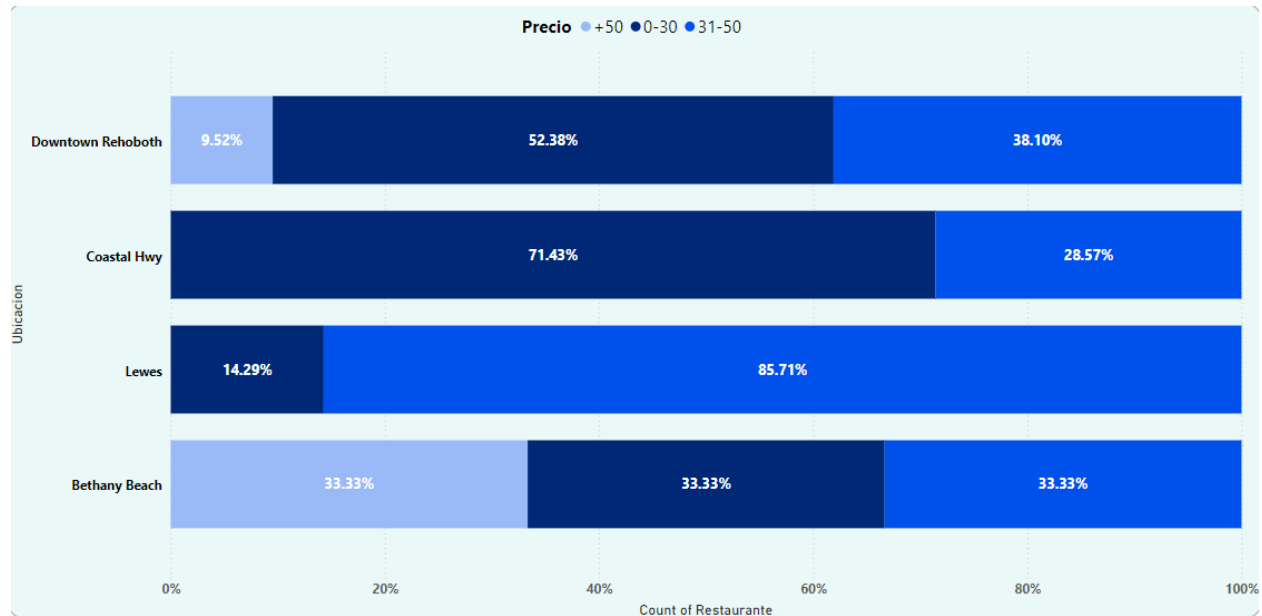
los \$50. Esto indica que los restaurantes informales en Rehoboth Beach están mayoritariamente orientados a una clientela que prioriza la accesibilidad y el confort, sin sacrificar la calidad de la experiencia gastronómica.

- **Restaurantes de ambiente casual elegante:** El 75% de estos restaurantes se sitúan en el rango de precios medio, entre \$31 y \$50. Sin embargo, también se encuentran algunos en el rango más accesible de \$0-\$30. Curiosamente, ninguno de los restaurantes de ambiente casual elegante supera los \$50. Esto sugiere que, en Rehoboth Beach, los restaurantes que ofrecen una experiencia más sofisticada y semiformal aún mantienen precios accesibles para un público más amplio, lo que los convierte en opciones viables para turistas y locales que buscan una cena más refinada sin incurrir en costos elevados.
- **Restaurantes de cena fina:** Como podría suponerse, los restaurantes de cena fina, conocidos por su exclusividad y ambiente de lujo, no presentan precios en el rango de \$0-\$30. Sin embargo, lo que destaca es que la mayoría de estos restaurantes se encuentran en el rango de \$31-\$50, lo cual contradice la expectativa de precios significativamente más altos en este tipo de establecimientos. Esto sugiere que, aunque Rehoboth Beach ofrece experiencias gastronómicas de alta gama y ambiente exclusivo, los precios siguen siendo razonables y accesibles en comparación con otros destinos turísticos de lujo.

9.6. Precio vs ubicación

Figura 28

Análisis de la relación precio vs ubicación



Nota. Figura adaptada de Power BI

En el análisis de la relación entre precio y ubicación, se observan diferencias significativas en la distribución de precios según la zona de Rehoboth Beach y sus alrededores. En el área de downtown Rehoboth Beach, la distribución de restaurantes según el rango de precios es bastante diversa, lo cual es comprensible dado que se trata del principal centro turístico de la ciudad. Aquí, los visitantes pueden encontrar una amplia variedad de opciones gastronómicas, desde locales más económicos hasta establecimientos de alta gama. Sin embargo, los restaurantes en el rango de precios de \$0-\$30 predominan, seguidos por aquellos en el rango de \$31-\$50 y una menor cantidad en el segmento de más de \$50. Esto sugiere que, aunque downtown ofrece experiencias culinarias para todo tipo de presupuesto, la mayoría de los visitantes tienden a optar por opciones más asequibles.

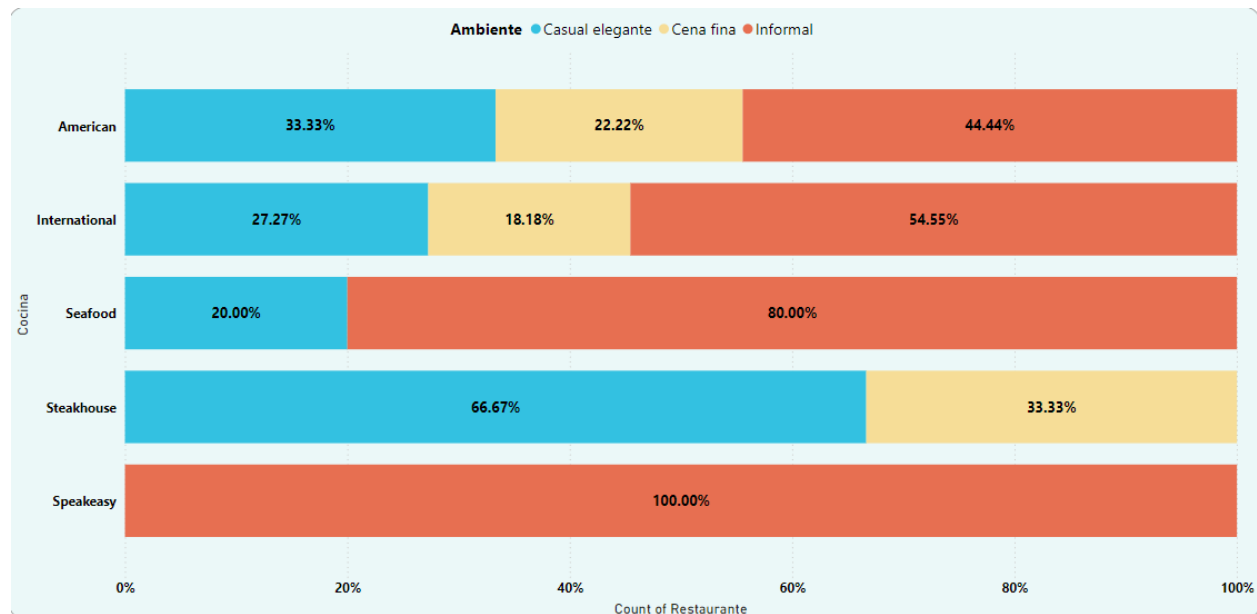
En contraste, en las zonas de Lewes y Bethany, la mayoría de los restaurantes se encuentran en el rango de \$31-\$50, lo que indica que estas áreas tienden a ser más costosas. Esto puede atribuirse a varios factores, como el entorno más exclusivo y el perfil demográfico de estos pueblos, que suelen atraer a residentes y visitantes con un mayor poder adquisitivo. Estas localidades, a menudo asociadas con áreas de retiro, ofrecen una atmósfera más cómoda y relajada, lo que justifica precios más elevados.

Por otro lado, los restaurantes ubicados a lo largo de la Coastal Highway muestran una mayor concentración en el rango de \$0-\$30, lo que los convierte en una opción más accesible para el público. Esta tendencia podría explicarse por el hecho de que la Coastal Highway no es un área tan turística como el downtown o las zonas más exclusivas de Lewes y Bethany. Al no contar con la ventaja de estar en una ubicación privilegiada cerca de la playa o en zonas históricas, los restaurantes en la autopista compensan con precios más bajos.

9.7. Tipo de cocina vs tipo de ambiente

Figura 29

Análisis de la relación tipo de ambiente vs tipo de cocina



Nota. Figura adaptada de Power BI

Los restaurantes de cocina americana e internacional en Rehoboth Beach ofrecen una mayor diversidad de ambientes, aunque predomina el estilo informal. En cuanto a los restaurantes de mariscos, la propuesta sigue una línea similar, ya que aproximadamente el 80% de los establecimientos de esta categoría también se clasifican como informales, lo que sugiere que este tipo de cocina se asocia con un ambiente más relajado y cercano al cliente, probablemente en línea con el estilo costero de la región.

Por otro lado, los restaurantes de steakhouse destacan por ofrecer un ambiente casual elegante y de cena fina, gracias a su propuesta gastronómica basada en cortes de carne de alta calidad y platos sofisticados. Esta categoría es la única que no incluye restaurantes con un

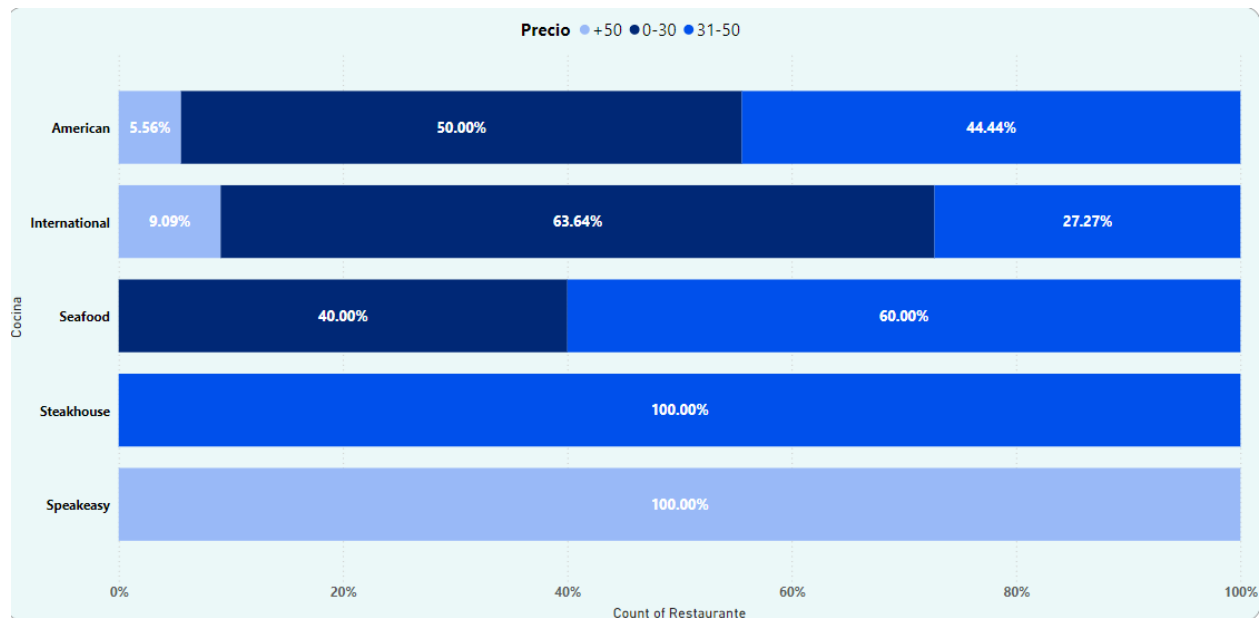
ambiente informal, lo que refuerza su enfoque en ofrecer una experiencia más exclusiva y refinada.

Finalmente, los restaurantes con estilo speakeasy se caracterizan por una atmósfera privada y relajada, diseñada para crear una experiencia acogedora y personalizada para los clientes. A pesar de que se considera informal, la exclusividad y el aire de misterio de estos espacios los hace únicos, ofreciendo una experiencia diferenciada dentro de la categoría de ambientes informales.

9.8. Tipo de cocina vs precio

Figura 30

Análisis de la relación tipo de cocina vs precio



Nota. Figura adaptada de Power BI

La comida americana y la internacional tienen una mayor variedad de precios, permitiendo atraer a un público más diverso. La mayor parte de los restaurantes de comida

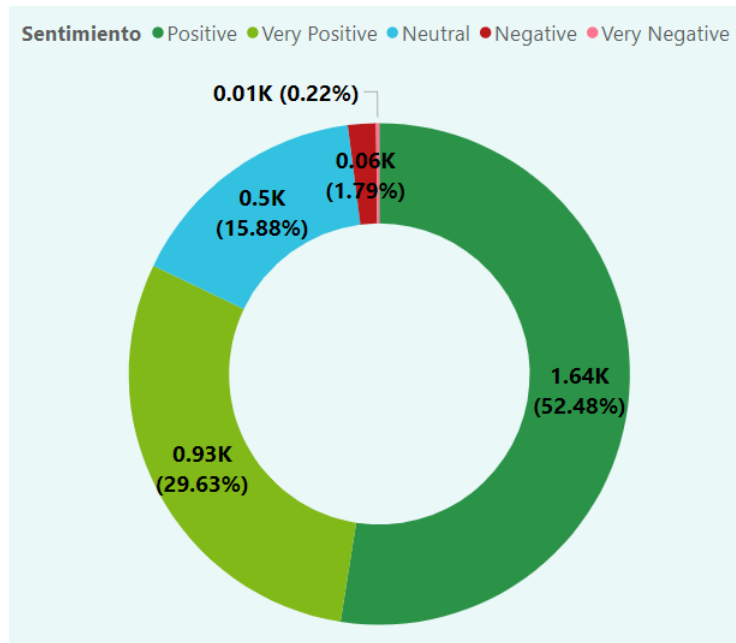
americana (50%) se encuentran en el rango de precios más bajo (\$0-\$30), lo que indica que esta categoría se asocia con opciones más accesibles. Por otro lado, los restaurantes de pescados y mariscos (seafood), steakhouse y speakeasy muestran una segmentación más clara hacia precios medios y altos. El enfoque en la experiencia de calidad y la exclusividad es especialmente notable en los restaurantes tipo speakeasy y steakhouse los cuales se encuentran dentro del rango de precios de (\$31-\$50) y +\$50, mientras que los de mariscos tienden a cubrir un rango más amplio, aunque con menor presencia en el segmento de lujo.

9.9. Análisis de polaridad

Una vez definido el perfil de los restaurantes de la muestra obtenida en la ciudad de Rehoboth Beach, se pasa a realizar un análisis de las reseñas aplicando un análisis de polaridad. Este análisis permite identificar las fortalezas, debilidades y áreas de mejora para cada restaurante. Se recopilaron un total de 3,129 reseñas, y para el análisis de polaridad se establecieron cinco categorías: muy positivas, positivas, neutrales, negativas y muy negativas.

Figura 31

Análisis general de polaridad de las reseñas



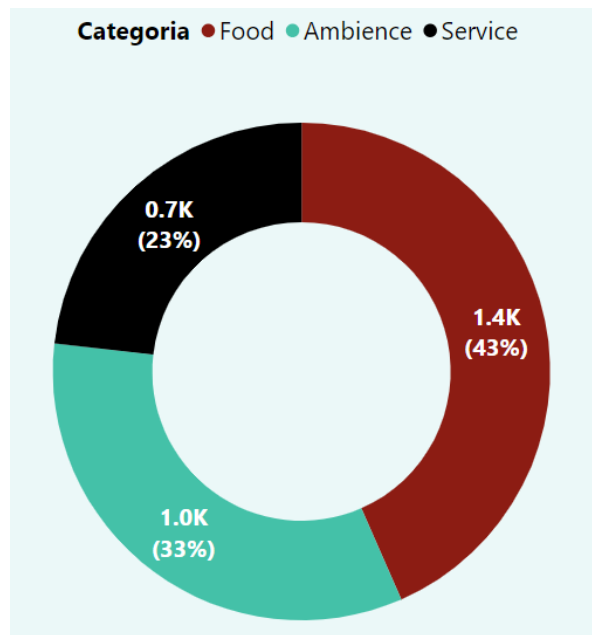
Nota. Figura adaptada de Power BI

Según la gráfica presentada, se puede observar que las reseñas positivas y muy positivas constituyen el 82.10% del total, lo que refleja un alto nivel de satisfacción general entre los clientes. Las reseñas neutrales representan un 15.88%, lo que sugiere que, aunque los comensales no tuvieron una experiencia negativa, tampoco hubo aspectos que destacaran significativamente. Finalmente, las reseñas negativas y muy negativas el porcentaje es de 2.01% del total, lo que indica que las experiencias desfavorables son mínimas.

9.9.1. Aspectos clave

Figura 32

Clasificación de reseñas según de los aspectos clave



Nota. Figura adaptada de Power BI

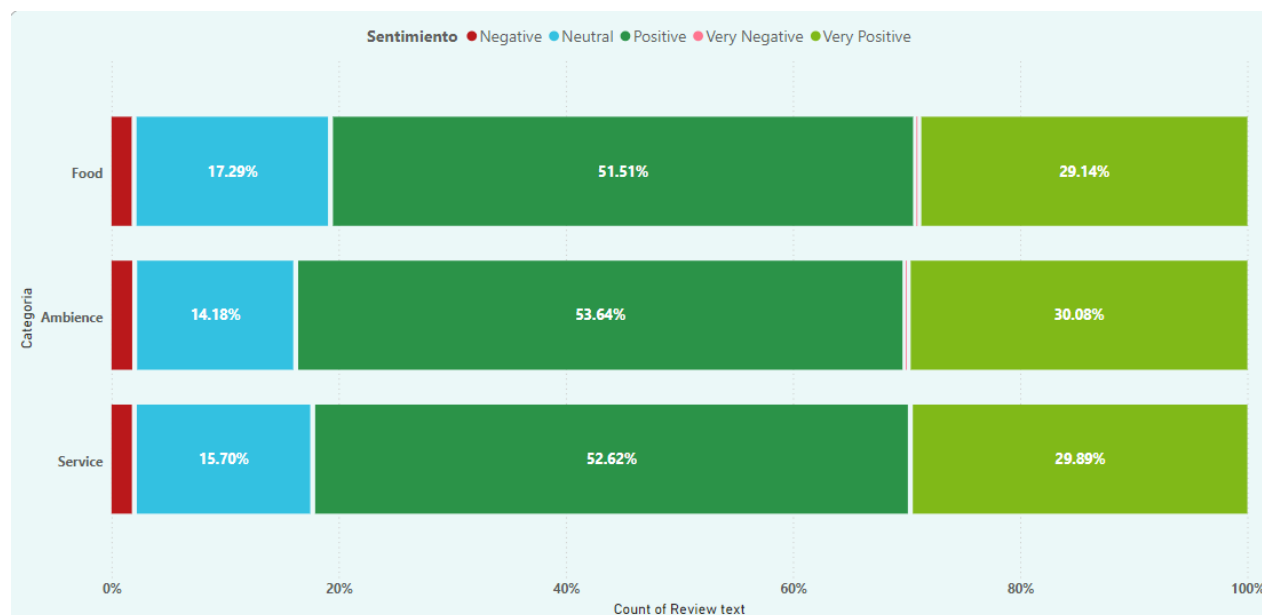
Al realizar un análisis más detallado, se identifican tres aspectos clave que los clientes evalúan cuando visitan un restaurante: la comida, el ambiente y el servicio. De acuerdo con los resultados, la comida es el aspecto más frecuentemente evaluado, ya que concentra la mayor cantidad de reseñas. Esto es comprensible, dado que la calidad de los platos es uno de los factores más decisivos en la satisfacción de los comensales.

El ambiente, entendido como la atmósfera general del restaurante (incluyendo la decoración, iluminación, música y nivel de formalidad), es el segundo aspecto más comentado. Los clientes valoran no solo la comida, sino también la experiencia que rodea su visita, destacando aquellos lugares donde se cuida el entorno y la comodidad. Por último, el servicio es

el tercer aspecto evaluado, lo que podría indicar que, aunque es importante, puede no ser tan decisivo como la calidad de la comida o el ambiente en la percepción global de la experiencia gastronómica en Rehoboth Beach.

Figura 33

Análisis de polaridad según aspectos clave



Nota. Figura adaptada de Power BI

En la gráfica que muestra el porcentaje de reseñas para cada factor evaluado, se observa que el ambiente es el aspecto más débil en términos de satisfacción del cliente. Aunque la diferencia con otros factores no es significativa, el ambiente recibe el menor porcentaje de reseñas positivas y muy positivas, con un 67.8%, y el mayor porcentaje de reseñas negativas y muy negativas, con un 2.11%. Esto sugiere que, aunque el ambiente no es el factor más comentado, las críticas negativas son más frecuentes en esta categoría.

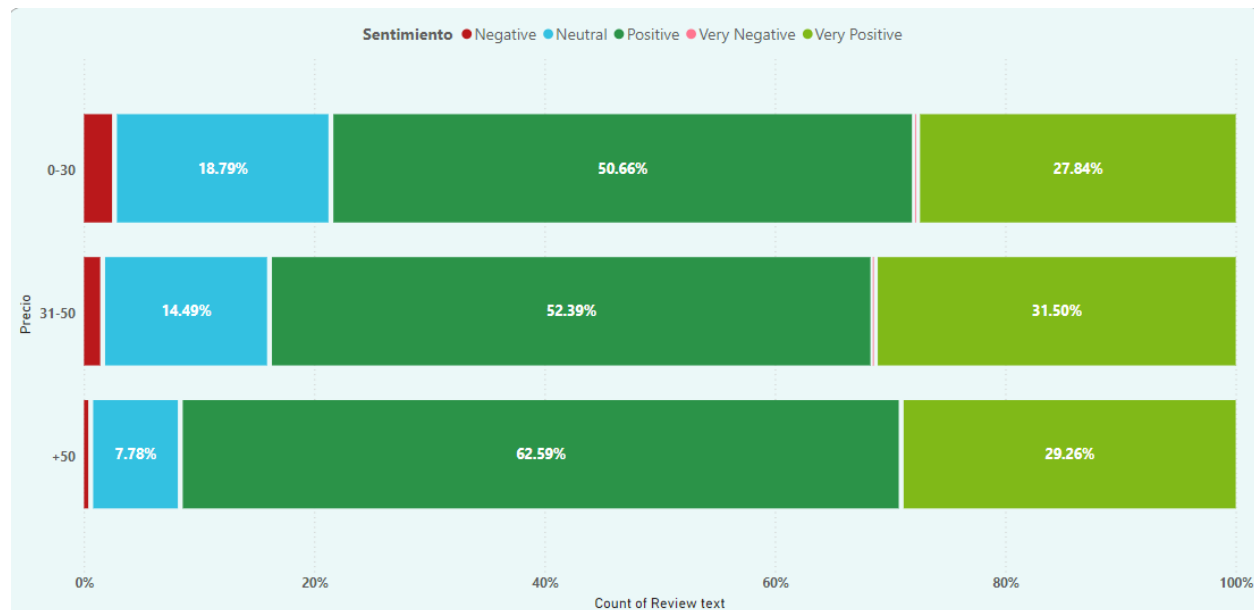
Por otro lado, el servicio es el factor con el menor porcentaje de reseñas negativas y muy negativas, con un 1.79%, y el segundo mayor porcentaje de reseñas positivas y muy positivas,

alcanzando un 68.32%. Aunque el servicio es el aspecto menos comentado en comparación con otros factores, su alta valoración y baja crítica negativa indican que, en general, es percibido de manera muy positiva por los clientes.

Estos resultados indican que, mientras el ambiente presenta desafíos significativos que deben abordarse, el servicio se destaca como una fortaleza clave y es generalmente bien recibido por los clientes. Mejorar el ambiente podría incrementar la satisfacción general, mientras que mantener y reforzar la calidad del servicio ayudará a consolidar las percepciones positivas actuales.

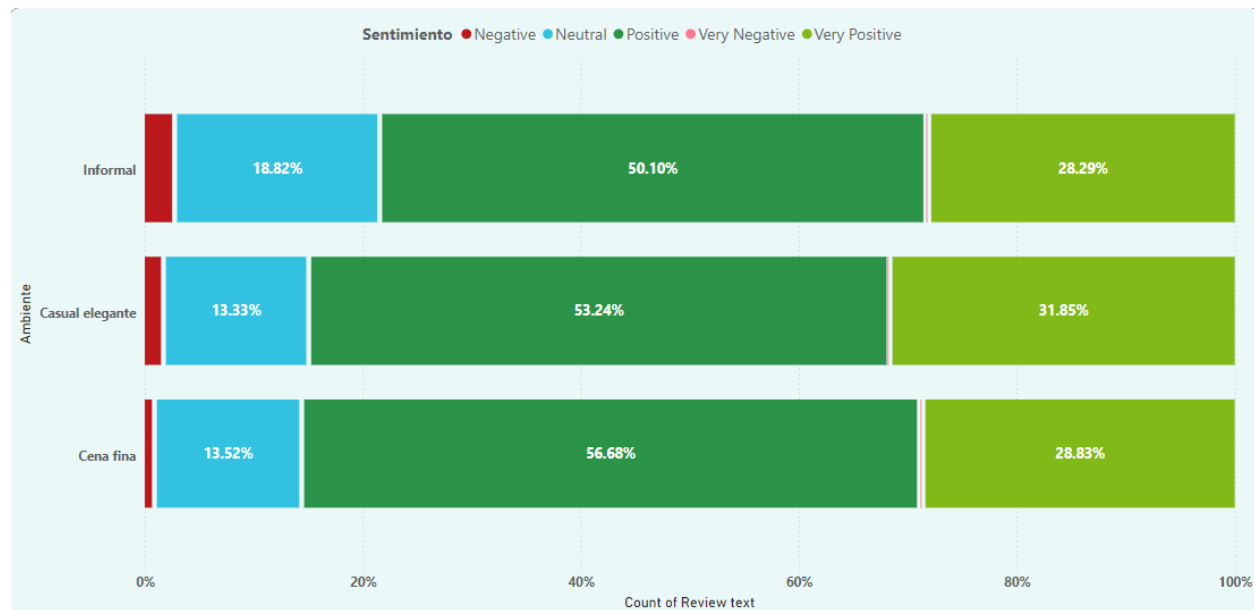
9.9.2. Rango de precios y tipo de ambiente

Al realizar el análisis de polaridad de las reseñas, clasificándolas según el rango de precios y el ambiente, se identifican patrones clave que ofrecen una visión interesante sobre la satisfacción del cliente en los restaurantes de Rehoboth Beach.

Figura 34*Análisis de polaridad según rango de precios*

Nota. Figura adaptada de Power BI

Los restaurantes en el rango de precios de \$0-\$30 presentan el menor porcentaje de reseñas positivas y muy positivas, y el mayor porcentaje de reseñas negativas y muy negativas. Esto sugiere que los restaurantes de menor precio suelen tener más dificultades para satisfacer a los clientes en comparación con aquellos en rangos de precios más altos. Esto puede deberse a que, en general, los restaurantes de menor precio a menudo no pueden invertir tanto en aspectos que impactan directamente en la experiencia del cliente, como la calidad del servicio, la decoración o el ambiente. Sin embargo, el precio más accesible puede compensar parcialmente estas deficiencias. Por otro lado, los restaurantes en el rango de precios superior (+\$50) tienden a recibir mejores calificaciones en general. Esto puede justificarse por una mayor atención al detalle y un enfoque más cuidado en la experiencia del cliente, características típicas de los establecimientos que se posicionan en el segmento de precios más altos.

Figura 35*Análisis de polaridad según rango de precios y tipo de ambiente*

Nota. Figura adaptada de Power BI

De manera similar, el análisis de polaridad según el ambiente revela patrones consistentes. Los restaurantes informales reciben las calificaciones más bajas en comparación con los restaurantes de cena fina, que son los mejor calificados. Esto podría deberse a que los restaurantes de cena fina suelen ofrecer una experiencia más sofisticada y detallada, contribuyendo a una mayor satisfacción del cliente. En contraste, los restaurantes informales, aunque accesibles y cómodos, pueden no ofrecer el mismo nivel de detalle y exclusividad, lo que se refleja en sus calificaciones más bajas.

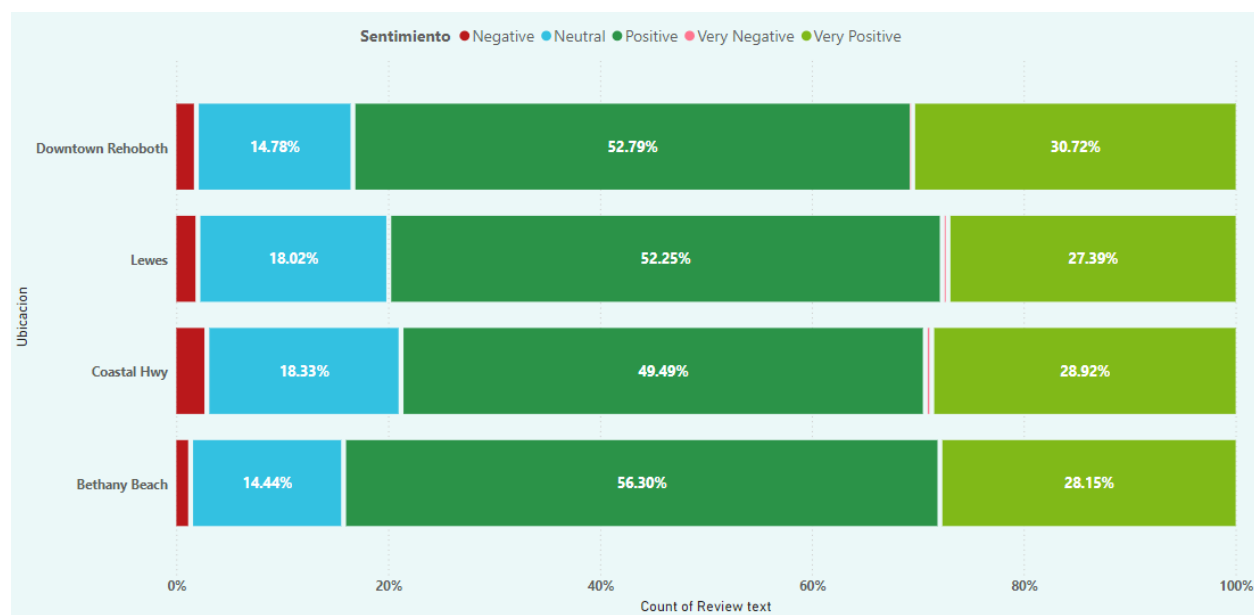
A pesar de las calificaciones más bajas para los restaurantes informales y de menor precio, es relevante destacar que las reseñas positivas en general superan a las neutras y negativas. Esto indica que, en general, los restaurantes en Rehoboth Beach están haciendo un buen trabajo en satisfacer a sus clientes y que el destino sigue siendo altamente apreciado por los

visitantes de la costa este. Esta tendencia subraya el éxito general del sector gastronómico en la región y su capacidad para mantener una buena reputación a pesar de las variaciones en el rango de precios y el ambiente.

9.9.3. Ubicación

Figura 36

Análisis de polaridad según la ubicación



Nota. Figura adaptada de Power BI

En el análisis de polaridad según la ubicación, se observa que los restaurantes situados en la Coastal Highway son los peor calificados, con un porcentaje de reseñas negativas del 3.26% y solo 78.41% de reseñas positivas. Esto está en línea con análisis previos que indican que la mayoría de los restaurantes en esta zona tienen un ambiente predominantemente informal y precios en el rango de \$0 a \$30. Estas características, como se ha evidenciado en estudios anteriores, tienden a recibir las puntuaciones más bajas, por lo que resulta lógico que los

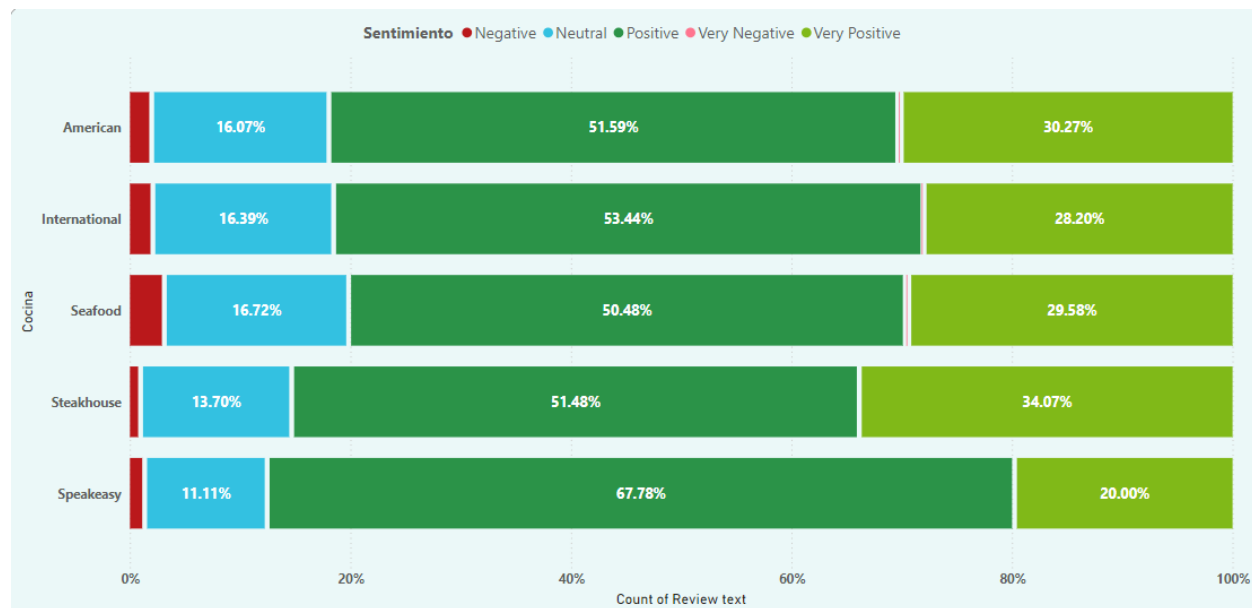
establecimientos de la Coastal Highway tengan un desempeño menos favorable en términos de reseñas y satisfacción del cliente.

En contraste, la localidad mejor puntuada es Bethany, con un 84.45% de reseñas positivas y apenas un 1.11% de negativas. Esto puede atribuirse a factores como un ambiente más refinado y relajado, así como a una oferta gastronómica de mayor calidad y precios más elevados, que suelen ser mejor valorados por los clientes. Además, Bethany es conocida por atraer a un público con mayor poder adquisitivo, lo que podría explicar la tendencia hacia mejores calificaciones.

9.9.4. Tipo de comida

Figura 37

Análisis de polaridad según el tipo de comida



Nota. Figura adaptada de Power BI

El análisis de polaridad de las reseñas por tipo de cocina revela que los restaurantes de mariscos (seafood) son los peor calificados, con el porcentaje más bajo de reseñas positivas

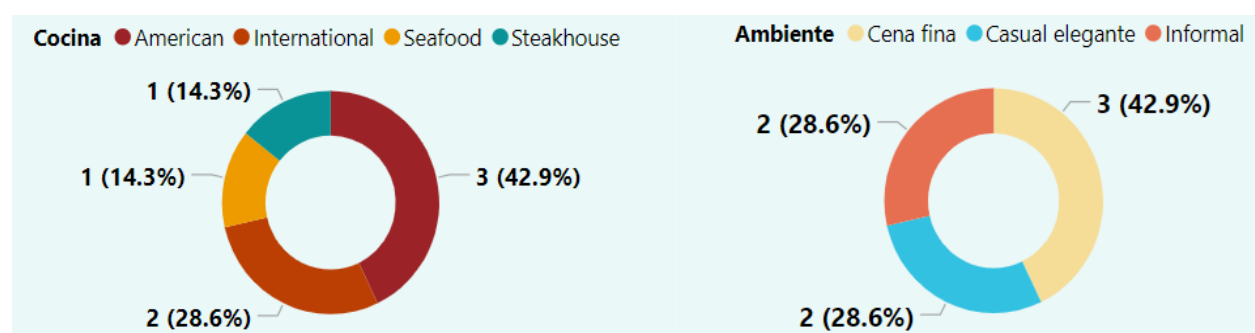
(80.06%) y el mayor número de reseñas negativas (3.21%). Les siguen los restaurantes de cocina internacional y americana, mientras que los mejor puntuados son los restaurantes tipo steakhouse y los nuevos speakeasy.

Esta tendencia puede explicarse en parte porque muchos restaurantes de mariscos en Rehoboth Beach suelen tener un ambiente informal y precios de rango medio, características que, como se observó en análisis anteriores, tienden a asociarse con calificaciones más bajas. En contraste, los restaurantes tipo steakhouse y los speakeasy ofrecen una experiencia más exclusiva, con un ambiente cuidadosamente diseñado y precios más altos, lo que se refleja en mejores calificaciones. La atención al detalle, tanto en el servicio como en la atmósfera, juega un papel fundamental en la satisfacción del cliente, lo que favorece a estos tipos de establecimientos. Además, la novedad de los speakeasy, con su enfoque en ofrecer una experiencia única y temática, puede haber contribuido a su alto nivel de aceptación.

9.10. Restaurantes mejor calificados

Figura 38

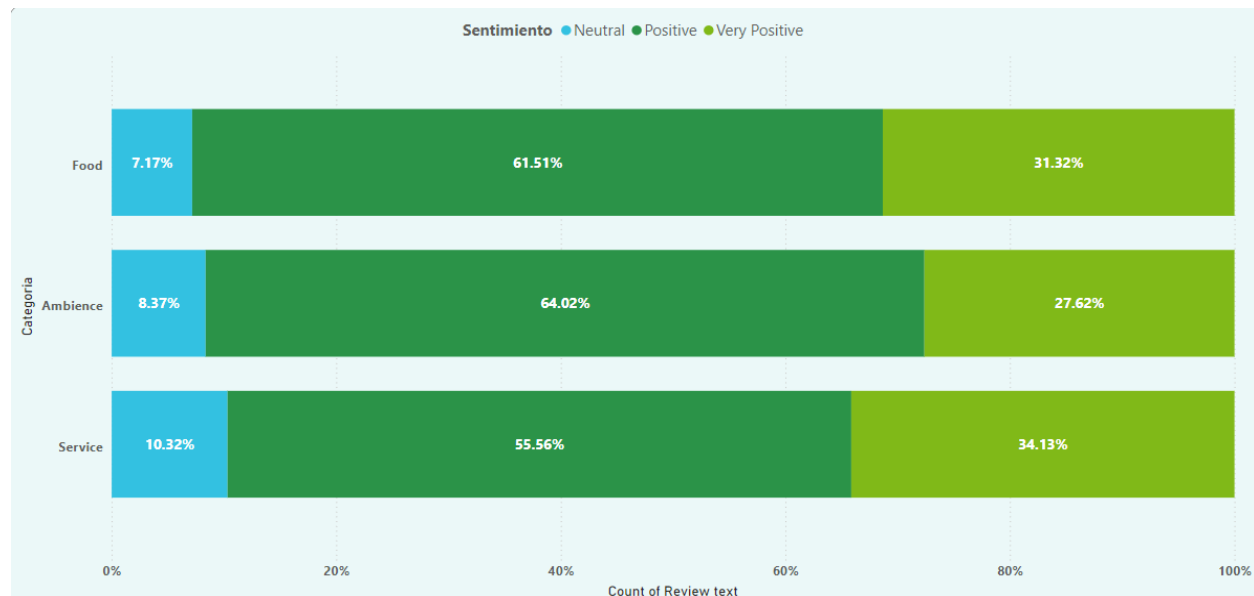
Distribución de los mejores restaurantes según tipo de comida y ambiente



Nota. Figura adaptada de Power BI

Figura 39

Análisis de polaridad de los mejores restaurantes según aspectos clave



Nota. Figura adaptada de Power BI

Se analizaron las reseñas de los siete restaurantes mejor puntuados con el fin de identificar sus fortalezas y determinar en qué aspectos sobresalen frente a la competencia. Los resultados indican que los tipos de cocina que dominan entre estos restaurantes son la americana, internacional, steakhouse y mariscos, siendo la comida americana la que lidera en popularidad.

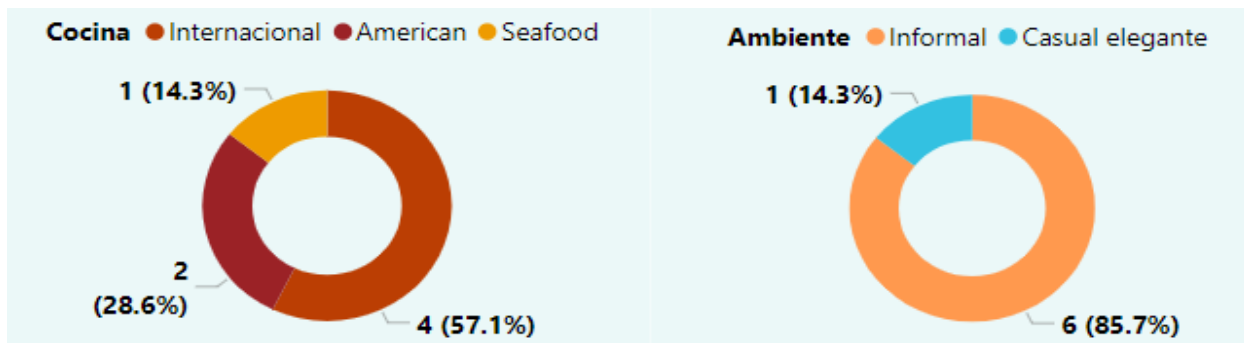
En cuanto al ambiente, el estilo de cena fina destaca como el preferido por los comensales, seguido por un empate entre el ambiente casual elegante y el informal. Esto no es sorprendente, ya que análisis previos mostraron que el ambiente de cena fina tiende a recibir las mejores calificaciones debido a la atención al detalle y la calidad de la experiencia que ofrecen.

Un aspecto relevante es que ninguno de estos restaurantes cuenta con reseñas negativas, lo cual es un indicio claro de una alta satisfacción general por parte de los clientes. En términos de las categorías más valoradas, la comida es el factor que acumula más reseñas positivas, seguida por el ambiente, y finalmente el servicio.

9.11. Análisis de los restaurantes peor calificados

Figura 40

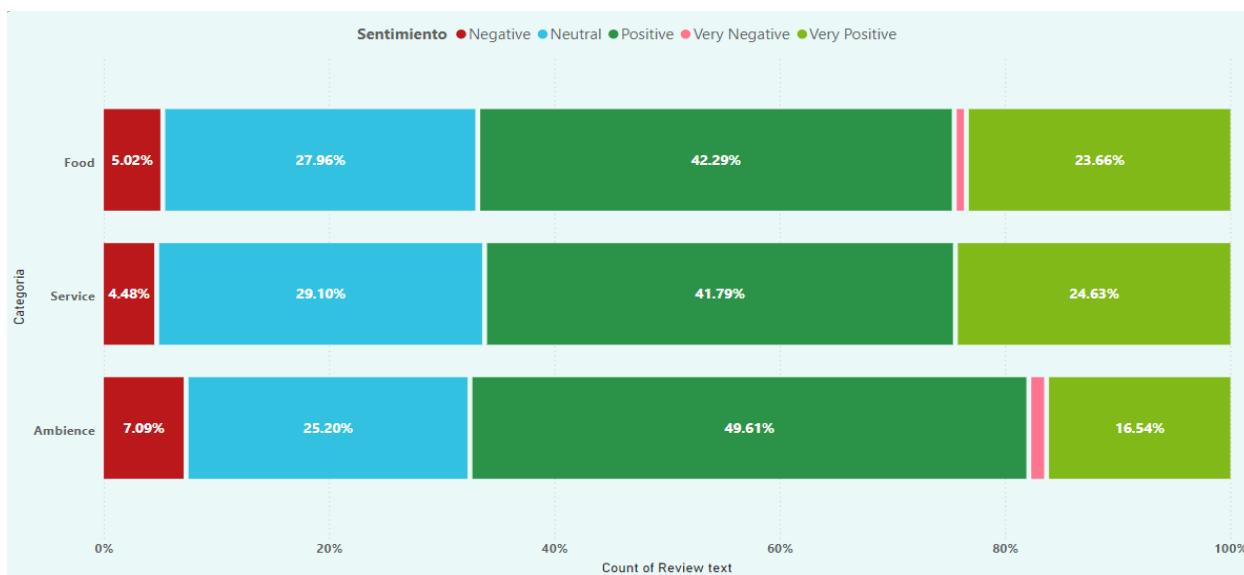
Distribución de los peores restaurantes según tipo de comida y ambiente



Nota. Figura adaptada de Power BI

Figura 41

Análisis de polaridad de los peores restaurantes según aspectos clave



Nota. Figura adaptada de Power BI

Al realizar un análisis de los siete restaurantes peor evaluados, se identificaron diversas áreas de mejora y debilidades que destacan frente a los mejores valorados, revelando

oportunidades para elevar su calidad. Los restaurantes de comida internacional son los que presentan las calificaciones más bajas, seguidos de los restaurantes de mariscos y comida americana. Este resultado sugiere que la oferta de comida internacional no está generando el impacto deseado, lo que podría deberse a una falta de autenticidad en la propuesta gastronómica o a la incapacidad de vender una experiencia cultural convincente a un mercado mayoritariamente blanco. La falta de conexión con el público objetivo y la posible escasa diferenciación frente a opciones locales más tradicionales podrían estar contribuyendo a esta tendencia.

En términos de tipos de ambiente, como era de esperarse, la mayoría de estos restaurantes tienen un ambiente informal, con solo un restaurante clasificado como casual elegante, y ninguno de ellos pertenece al segmento de cena fina. Esto refuerza la conclusión de que los restaurantes con ambientes más cuidados y formales tienden a tener mejor aceptación entre los clientes. La informalidad parece estar asociada con una mayor propensión a factores que generan insatisfacción, como el ruido excesivo, la presencia de niños y una menor atención al detalle en el servicio y la presentación.

Finalmente, se observa que la categoría que recibe más reseñas negativas es el ambiente, lo cual concuerda con lo discutido arriba. La falta de un entorno acogedor o bien diseñado puede afectar negativamente la experiencia del cliente, sobre todo en un mercado turístico donde los visitantes buscan una experiencia más completa que solo buena comida.

10. Estrategias de marketing

Los resultados obtenidos en la fase de análisis revelaron patrones, tendencias y áreas de mejora en los restaurantes de Rehoboth Beach. A partir de estos hallazgos, se han desarrollado estrategias y propuestas de marketing dirigidas a optimizar tanto la experiencia del cliente como la percepción que los comensales tienen de los establecimientos. Estas estrategias buscan no solo mejorar el rendimiento operativo y comercial de los restaurantes, sino también atraer a más turistas a la región, reforzando su fidelización a largo plazo.

10.1. Marketing de eventos

Como se concluyó anteriormente, los restaurantes de cocina americana son los más populares en Rehoboth Beach, impulsados por factores demográficos y la demanda de los turistas. Además, los restaurantes de mariscos superan en popularidad a los steakhouse y a los restaurantes de BBQ. Dado que Rehoboth Beach es un destino turístico de playa, muchos visitantes buscan disfrutar de una experiencia gastronómica auténtica que refleje el entorno costero, y la cultura y la identidad culinaria del país por esto los restaurantes de mariscos y los de comida americana son una opción perfecta para cumplir con estas expectativas. Reforzar la percepción de estos restaurantes no solo es esencial para mantener su popularidad, sino también para aumentar el atractivo turístico de la región al ofrecer a los visitantes una propuesta culinaria genuina y representativa del lugar.

Las estrategias de marketing basadas en eventos aplicadas a restaurantes son tácticas promocionales que utilizan eventos especiales como una forma de atraer clientes, mejorar la visibilidad del restaurante, fortalecer la lealtad de los clientes y, en última instancia, aumentar las ventas. Estas estrategias aprovechan la oportunidad de crear experiencias gastronómicas

memorables, involucrar a los clientes de manera activa y resaltar la propuesta de valor del restaurante a través de actividades en vivo o temáticas. (Mayday, 2024)

Una estrategia efectiva para promocionar los restaurantes de mariscos y de cocina americana en Rehoboth Beach es la implementación de un evento gastronómico que conecte a los visitantes con la cultura culinaria local. Se propone un organizar un tour gastronómico en el que los asistentes compren un boleto que incluye un mapa de los restaurantes participantes, ofreciendo degustaciones de platos destacados según el tipo de cocina. En un tour de mariscos, por ejemplo, podrían servirse crab cakes, langosta o almejas, mientras que, en uno de cocina americana, hamburguesas, hot dogs o pretzels. Los participantes podrán recorrer los restaurantes, descubrir la variedad de platos y votar por su favorito en base a sabor, presentación y experiencia general, utilizando una aplicación móvil o formularios en papel. El restaurante con la mejor calificación recibirá un reconocimiento público, aumentando su visibilidad.

Los previos análisis nos indican que los restaurantes de cocina internacional enfrentan desafíos que podría deberse a una falta de autenticidad en la propuesta gastronómica o a la incapacidad de vender una experiencia cultural convincente a un mercado mayoritariamente blanco. Para abordar esto se propone:

10.2. Marketing cultural

El marketing cultural, o marketing multicultural, se centra en identificar y aprovechar diversas costumbres sociales, tales como tradiciones, símbolos, festividades y eventos, para construir conexiones auténticas y significativas con el público. (ESIC, 2024) En Rehoboth Beach, donde la oferta de restaurantes internacionales es limitada, es crucial destacar los restaurantes

existentes para incrementar la demanda de este tipo de gastronomía. El marketing cultural puede ser especialmente efectivo en este contexto mediante estrategias como:

- **Eventos culturales temáticos:** Organizar eventos que celebren festividades o tradiciones de las culturas representadas en los restaurantes. Estos eventos pueden incluir degustaciones especiales, espectáculos en vivo o actividades interactivas que proporcionen una experiencia inmersiva y auténtica.
- **Menús especiales durante celebraciones internacionales:** Ofrecer menús exclusivos y promociones durante fechas festivas importantes de diferentes culturas. Esta estrategia no solo realza la autenticidad de la oferta gastronómica, sino que también atrae la atención de los residentes y visitantes interesados en experimentar sabores internacionales.
- **Colaboraciones con comunidades culturales locales:** Asociarse con organizaciones y grupos culturales para realizar actividades conjuntas, como festivales gastronómicos o talleres culinarios. Estas colaboraciones pueden aumentar la visibilidad de los restaurantes y fomentar un mayor interés en la cocina internacional.

10.3. Marketing de Proximidad

Los restaurantes en Coastal Highway enfrentan calificaciones más bajas, lo que puede reflejar una falta de fidelización entre los clientes. Para abordar esta situación, el marketing de proximidad puede ser una estrategia efectiva para atraer nuevos clientes y mantener a los existentes.

El marketing de proximidad utiliza tecnologías como Bluetooth, GPS y Wi-Fi para enviar mensajes y promociones personalizadas a los dispositivos móviles de los usuarios que se encuentren cerca del restaurante. (Calvo, 2021). Esta técnica es ideal para áreas con alta

circulación de turistas, como Coastal Highway, ya que permite captar la atención de los viajeros que están en tránsito por la costa este de Estados Unidos. Algunas estrategias que se proponen son;

- **Ofertas por Tiempo Limitado:** Implementar promociones que se activen cuando los clientes estén cerca del restaurante. Por ejemplo, enviar una notificación de descuento especial o una oferta de menú del día a los dispositivos móviles de los turistas que se encuentren a pocos metros del restaurante.
- **Incentivos de Parada Rápida:** Dado que muchos de los restaurantes en Coastal Highway tienen un ambiente informal y precios accesibles, la propuesta puede resaltar la conveniencia de una comida rápida y sabrosa para los viajeros en movimiento. Ofrecer un incentivo adicional, como un postre gratuito o una bebida con la compra de un plato principal, puede ser una forma efectiva de motivar a los turistas a detenerse.

10.4. Marketing digital

El downtown de Rehoboth Beach es, sin duda, el área más turística de la ciudad, y donde se concentra la mayor cantidad de restaurantes. En este contexto, el marketing digital se presenta como una herramienta sumamente eficaz para atraer a más turistas y resaltar el atractivo y los puntos fuertes de estos establecimientos, ya que permite conectar con clientes potenciales en plataformas que frecuentan a diario. Realizar una gestión online efectiva es fundamental para maximizar el impacto.

- **Optimización de la información online:** Es vital asegurarse de que los restaurantes estén completamente optimizados en plataformas como Google My Business, con descripciones atractivas, fotos de alta calidad, horarios actualizados y la ubicación exacta

en el mapa. Esto facilita que los turistas encuentren fácilmente los establecimientos y decidan visitarlos.

- **Uso de palabras clave locales:** En las páginas web, blogs o redes sociales de los restaurantes, es recomendable utilizar palabras clave locales como “mejores restaurantes en el downtown de Rehoboth Beach” o “restaurantes cerca del paseo marítimo de Rehoboth”. Estas palabras clave ayudan a que los restaurantes aparezcan en los primeros resultados cuando las personas buscan opciones de comida en las cercanías. Además, el uso estratégico de hashtags locales en redes sociales también incrementa la visibilidad de los restaurantes en búsquedas geolocalizadas.
- **Colaboración con influencers locales:** Trabajar con *influencers* que cuenten con seguidores en el área de Rehoboth Beach, o que sean populares entre los turistas que frecuentan la ciudad, es otra estrategia eficaz. Estos creadores de contenido pueden compartir sus experiencias en los restaurantes, generando interés y atrayendo a nuevos clientes a través de sus recomendaciones y publicaciones.

10.5. Marketing experiencial

El análisis de los restaurantes en Rehoboth Beach reveló que los establecimientos de ambiente informal enfrentan los mayores desafíos y reciben la mayor cantidad de críticas negativas. Mejorar la percepción del ambiente en estos restaurantes es crucial, ya que es el factor que más influencia tiene en la experiencia de los comensales. Para abordar este aspecto, una estrategia eficaz es implementar marketing experiencial, que permite transformar la experiencia del cliente, mejorando significativamente la percepción del ambiente. El marketing experiencial se enfoca en crear experiencias memorables y únicas para los clientes, más allá de solo ofrecer

comida. Este tipo de marketing busca generar una conexión emocional con los comensales a través de vivencias que los hagan sentirse parte del restaurante y que los motiven a regresar. (Ferreira, 2023)

En los restaurantes de ambiente informal, el enfoque está en crear experiencias relajadas, divertidas y accesibles, que conecten con un público más joven o casual. Algunos ejemplos incluyen:

- **Experiencia gastronómica interactiva:** Una experiencia gastronómica interactiva puede ser una excelente estrategia para atraer y fidelizar a los clientes en restaurantes informales. Ofrecer opciones de personalización, donde los comensales elijan los ingredientes y armen su propio plato, como pizzas, tacos o ensaladas, es una manera de involucrar activamente al cliente en su experiencia culinaria. Además, se puede incentivar la participación en concursos de redes sociales, invitando a los clientes a compartir fotos de sus platos o del ambiente del restaurante, premiando la imagen que obtenga más reacciones. Otra propuesta atractiva es la creación de desafíos de comida, donde los clientes compitan de manera amistosa para completar un plato grande o un menú especial, recompensando a los ganadores con descuentos o premios especiales. Estas dinámicas no solo mejoran la experiencia del cliente, sino que también fomentan la interacción social y el boca a boca digital.
- **Noches temáticas:** Ofrecer noches temáticas es una estrategia efectiva para crear experiencias únicas e inmersivas en restaurantes informales. Estas noches temáticas pueden variar ampliamente, adaptándose a los intereses y preferencias de los clientes. Por ejemplo: Organizar competencias de trivia sobre temas variados, como cultura general, deportes, o temas locales, fomenta la participación y el espíritu competitivo entre los

comensales. Además, se pueden ofrecer premios o descuentos a los equipos ganadores, aumentando la emoción y el atractivo del evento. Implementar noches de karaoke o noches de juegos de mesa o videojuegos en el restaurante puede atraer a grupos de amigos y familia y añadir una capa extra de diversión a la experiencia gastronómica.

11. Conclusiones

La integración de herramientas avanzadas como el procesamiento de lenguaje natural (PLN) y Power BI ha sido clave para extraer insights valiosos de las reseñas de los clientes. Gracias a la capacidad de analizar de manera automatizada grandes volúmenes de opiniones y visualizarlas de forma clara, fue posible comprender en mayor profundidad las percepciones y experiencias de los comensales. Esta combinación tecnológica no solo facilitó el análisis de datos, sino que también permitió identificar áreas de mejora y oportunidades de crecimiento para los restaurantes, orientando decisiones estratégicas tanto en marketing como en operaciones.

Además, el análisis de las reseñas reveló que la segmentación de las preferencias de los clientes está influenciada significativamente por factores como la ubicación y el precio. Los restaurantes ubicados en áreas exclusivas como Bethany y Lewes, con precios más altos, atraen a una clientela con mayor poder adquisitivo. Por otro lado, los restaurantes situados en zonas más accesibles y turísticas, como el downtown de Rehoboth Beach, tienden a ofrecer una gama de precios más diversa, atrayendo tanto a turistas como a residentes locales.

Un aspecto crucial que se destacó en el análisis es que los restaurantes que mejor reflejan la identidad culinaria local continúan siendo los más populares entre los visitantes. Esto resalta la importancia de mantener una oferta gastronómica auténtica, que no solo cumpla con las expectativas de los turistas, sino que también refuerce la imagen de la región como un destino culinario distintivo de la costa este. Promover la autenticidad gastronómica, a través de la calidad de los platos y la narrativa cultural que los acompaña, es una estrategia clave para atraer a más visitantes y mejorar la percepción de los restaurantes.

A pesar de los retos específicos que enfrentan algunos establecimientos, en general, el desempeño de los restaurantes en Rehoboth Beach es positivo. Las reseñas clasificadas como "positivas" superan con creces a las "negativas", lo que refuerza la reputación de la ciudad como uno de los destinos turísticos más atractivos de la costa este de los Estados Unidos.

12. Recomendaciones

- A medida que las empresas se vuelven más rigurosas en la protección de sus datos, las barreras para acceder a ellos aumentan. Por lo tanto, es crucial mantenerse al día con las actualizaciones en las técnicas de scraping, adaptando las herramientas y métodos empleados para cumplir con las normativas y superar las restricciones impuestas por las plataformas.
- Antes de iniciar cualquier proyecto, es recomendable evaluar las capacidades técnicas del equipo y los recursos con los que se cuenta. En algunos casos, puede ser necesario ajustar el alcance del proyecto para garantizar que los resultados obtenidos sean coherentes y factibles, evitando así sobrecargar los recursos disponibles o comprometer la calidad del análisis.
- Para obtener resultados más precisos y representativos, se sugiere trabajar con muestras más grandes en futuros estudios. Esto permitirá un análisis más robusto y evitará posibles sesgos en la interpretación de los datos, mejorando la validez de las conclusiones.
- Es importante realizar consultas con personas que conozcan el mercado o el sector que se está analizando. Esto ayudará a contextualizar los datos y asegurar que el análisis y los resultados reflejen la realidad del entorno. Sin este contexto, los resultados pueden quedarse en un plano meramente numérico, perdiendo su valor práctico.

Referencias Bibliográficas

- Adedoyin-Olowe, M., Gaber, M. M., & Stahl, F. (2014). A Survey of Data Mining Techniques for Social Media Analysis. *Journal of Data Mining & Digital Humanities*, 2014. <https://doi.org/10.46298/jdmdh.5>
- Babac, M. B., & Podobnik, V. (2016). A sentiment analysis of who participates, how and why, at social media sport websites: How differently men and women write about football. *Online Information Review*, 40(6), 814–833. <https://doi.org/10.1108/OIR-02-2016-0050>
- Báez, P., Arancibia, A. P., Chaparro, M. I., Bucarey, T., Núñez, F., & Dunstan, J. (2022). Procesamiento de lenguaje natural para texto clínico en español: el caso de las listas de espera en Chile. *Revista Médica Clínica Las Condes*, 33(6), 576–582. <https://doi.org/10.1016/j.rmclc.2022.10.002>
- Bagherzadeh, S., Shokouhyar, S., Jahani, H., & Sigala, M. (2021). A generalizable sentiment analysis method for creating a hotel dictionary: using big data on TripAdvisor hotel reviews. *Journal of Hospitality and Tourism Technology*, 12(2), 210–238. <https://doi.org/10.1108/JHTT-02-2020-0034>
- Bajaj, A. (2024). *Can Python understand human feelings through words? – A brief intro to NLP and VADER Sentiment Analysis*. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/06/vader-for-sentiment-analysis/>
- Berkeley School of Information. (2020). *What Is Machine Learning (ML)?* Berkeley School of Information. <https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/>
- Bhatt, S. (2018). *Reinforcement Learning 101*. Towards Data Science . <https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292>

- Bigne, E., Ruiz, C., Perez-Cabañero, C., & Cuenca, A. (2023). Are customer star ratings and sentiments aligned? A deep learning study of the customer service experience in tourism destinations. *Service Business*, 17(1), 281–314. <https://doi.org/10.1007/s11628-023-00524-0>
- Blakemore, E. (2023). *La nueva IA podría superar el famoso Test de Turing; este es el hombre que lo creó*. National Geographic. <https://www.nationalgeographic.es/ciencia/2023/03/alan-turing-test-inteligencia-artificial>
- Branco, A., Parada, D., Silva, M., Mendonça, F., Mostafa, S. S., & Morgado-Dias, F. (2024). Sentiment Analysis in Portuguese Restaurant Reviews: Application of Transformer Models in Edge Computing. *Electronics (Switzerland)*, 13(3). <https://doi.org/10.3390/electronics13030589>
- Brownlee, J. (2019). *What Are Word Embeddings for Text?* Machine Learning Mastery. <https://machinelearningmastery.com/what-are-word-embeddings/>
- Calvo, L. (2021). *¿Qué es el marketing de proximidad y cómo funciona? Definición y ejemplos*. GoDaddy. <https://www.godaddy.com/resources/es/marketing/marketing-de-proximidad>
- Cao, Z., Xu, H., & Teo, B. S. X. (2023). Sentiment of Chinese Tourists towards Malaysia Cultural Heritage Based on Online Travel Reviews. *Sustainability (Switzerland)*, 15(4). <https://doi.org/10.3390/su15043478>
- Chang, W., & Zhu, M. (2023). Sentiment analysis method of consumer comment text based on BERT and hierarchical attention in e-commerce big data environment. *Journal of Intelligent Systems*, 32(1). <https://doi.org/10.1515/jisys-2023-0025>
- Chinmayee, D. (2024). *An Introduction to Selenium WebDriver*. SimpliLearn. <https://www.simplilearn.com/tutorials/selenium-tutorial/what-is-selenium-webdriver>

- Delaware Sea Grant College. (2017). *The Contribution of The Coastal Economy to the State of Delaware*".
- Dueñas Quesada, J. M. (2020). *Aplicación de técnicas de Machine Learning a la ciberseguridad: Aprendizaje supervisado para la detección de amenazas web mediante clasificación basada en árboles de decisión*. Universitat oberta de Catalunya. <http://hdl.handle.net/10609/118166>.
- ESIC. (2024). *Marketing cultural: qué es, ejemplos y estrategias*. ESIC, Bussiness & Marketing School. <https://www.esic.edu/rethink/marketing-y-comunicacion/marketing-cultura-que-es-ejemplos-estrategias-c>
- Fatah, D. A., Rochman, E. M. S., Setiawan, W., Aulia, A. R., Kamil, F. I., & Su'ud, A. (2024). Sentiment Analysis of Public Opinion Towards Tourism in Bangkalan Regency Using Naïve Bayes Method. *E3S Web of Conferences*, 499. <https://doi.org/10.1051/e3sconf/202449901016>
- Ferreira, A. (2023). *Qué es el marketing experiencial: todas las claves y ejemplos*. InboundCycle. <https://www.inboundcycle.com/blog-de-inbound-marketing/que-es-el-marketing-experiencial-como-implementarlo-y-ejemplos>
- Garcia Morgado, M. (2016). PyText: Una Librería para la Minería de Textos Basada en Python. *Benemerita Universidad Autónoma de Puebla*. <https://repositorioinstitucional.buap.mx/server/api/core/bitstreams/18e6f2af-c0ad-4914-b0d2-827eb0dfe0a5/content>
- Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications*, 69, 214–224. <https://doi.org/10.1016/j.eswa.2016.10.043>

- Gomez Vargas, D. C. (2023). *SEGMENTACIÓN DE CLIENTES Segmentación de clientes del sector turístico de Santander mediante el uso de análisis de sentimientos*. Universidad Industrial de Santander.
- Gorini, M. (n.d.). Clasificación vs. clusterización: una explicación práctica. *BiSmart*. Retrieved May 19, 2024, from <https://blog.bismart.com/la-clasificaci%C3%B3n-y-la-clusterizaci%C3%B3n-una-explicaci%C3%B3n-pr%C3%A1ctica>
- Insight-Group. (n.d.). *Enfoques basados en léxico: ¿NLTK Vader o TextBlob?* Insight-Group. Retrieved May 16, 2024, from https://insight--group-github-io.translate.goog/MFIN7036/sentiment-analysis-lexicon-based-nv-or-tb.html?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es-419&_x_tr_pto=sc
- Jurafsky, D., & Martin, J. H. (2024). *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Third Edition draft*.
- Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5), 6000–6010. <https://doi.org/10.1016/j.eswa.2011.11.107>
- Kapilsparshi. (2023). *Difference Between Data Mining and Text Mining*. GeeksforGeeks. <https://www.geeksforgeeks.org/difference-between-data-mining-and-text-mining/>
- Khan, A., & Baharudin, B. (2011). Sentiment classification using sentence-level semantic orientation of opinion terms from blogs. *2011 National Postgraduate Conference*, 1–7. <https://doi.org/10.1109/NatPC.2011.6136319>

- Khoo, C. S. G., & Johnkhan, S. B. (2018). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, 44(4), 491–511. <https://doi.org/10.1177/0165551517703514>
- Londhe, A., & Rao, P. V. R. D. P. (2022). Dynamic Classification of Sentiments from Restaurant Reviews Using Novel Fuzzy-Encoded LSTM. In *International Journal on Recent and Innovation Trends in Computing and Communication* (Vol. 10, Issue 9, pp. 112–124). Auricle Global Society of Education and Research. <https://doi.org/10.17762/ijritcc.v10i9.5714>
- Mayday, M. (2024). *¿Qué es el marketing de eventos en los negocios? Guía estratégica y consejos*. ON24. <https://www.on24.com/es/blog/what-is-event-marketing-in-business-strategy-guide-tips/>
- Meddeb, A., & Romdhane, L. Ben. (2022). Using Topic Modeling and Word Embedding for Topic Extraction in Twitter. *Procedia Computer Science*, 207, 790–799. <https://doi.org/10.1016/j.procs.2022.09.134>
- Mendiola, J. (2013). “La cuenta por favor”, *OpenTable acaba con las agónicas esperas en el restaurante*. El Confidencial. https://www.elconfidencial.com/tecnologia/2013-08-06/la-cuenta-por-favor-opentable-acaba-con-las-agonicas-esperas-en-el-restaurante_14725/
- Moreno Sandoval, L. G., Pomares Quimbaya, A., Cruz Gutiérrez, C. E., García Pachón, J. F., & Vanegas Ramírez, D. F. (2022). *Comparación de métodos de análisis de sentimientos en comunidades de habla hispana*. 1–12. <https://doi.org/10.26507/paper.2367>
- Morris, R. (2023). *Delaware tourism booms back with record 28M visitors*. Delaware Business Times. <https://delawarebusinesstimes.com/news/delaware-tourism->

- Patel, F. N., & Soni, N. R. (2012). Text mining: A Brief survey. In *International Journal of Advanced Computer Research*.
- Pineda, J. D. M. (2021). Un modelo para la predicción del movimiento del precio de las acciones del mercado bursátil basado en un análisis de sentimiento y datos históricos de la BVC. In *Universidad Industrial de Santander*.
- Prottasha, N. J., Sami, A. A., Kowsher, M., Murad, S. A., Bairagi, A. K., Masud, M., & Baz, M. (2022). Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. *Sensors*, 22(11). <https://doi.org/10.3390/s22114157>
- Puh, K., & Bagić Babac, M. (2023). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights*, 6(3), 1188–1204. <https://doi.org/10.1108/JHTI-02-2022-0078>
- Revista De investigación De Sistemas E Informática. (2009). *Procesamiento de lenguaje natural*. Revista De investigación De Sistemas E Informática. <https://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/view/5923>
- Rezaeinia, S. M., Rahmani, R., Ghodsi, A., & Veisi, H. (2019). Sentiment analysis based on improved pre-trained word embeddings. *Expert Systems with Applications*, 117, 139–147. <https://doi.org/10.1016/j.eswa.2018.08.044>
- Rouhiainen, L. (2018). *Inteligencia artificial : 101 cosas que debes saber hoy sobre nuestro futuro*. Alienta.
- Roul, A. (2021). *Sentiment Analysis- Lexicon Models vs Machine Learning*. Medium. <https://medium.com/nerd-for-tech/sentiment-analysis-lexicon-models-vs-machine-learning-b6e3af8fe746>

S, V., & R, J. (2016). Text Mining: open Source Tokenization Tools – An Analysis. *Advanced Computational Intelligence: An International Journal (ACIJ)*, 3(1), 37–47.

<https://doi.org/10.5121/acii.2016.3104>

Santamaria-Granados, L., Mendoza-Moreno, J. F., & Ramirez-Gonzalez, G. (2021). Tourist recommender systems based on emotion recognition—a scientometric review. *Future Internet*, 13(1), 1–38. <https://doi.org/10.3390/fi13010002>

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. In *SN Computer Science* (Vol. 2, Issue 3). Springer.

<https://doi.org/10.1007/s42979-021-00592-x>

Sayed, M. S., Mohan, V., & Muthu, K. S. (2023). BERT: A Review of Applications in Sentiment Analysis. In *HighTech and Innovation Journal* (Vol. 4, Issue 2, pp. 453–462). Ital

Publication. <https://doi.org/10.28991/HIJ-2023-04-02-015>

Schultz, A. (2024). *10 Best Places to Visit in Delaware*. Touropia.

[https://www.touropia.com/best-places-to-visit-in-](https://www.touropia.com/best-places-to-visit-in-delaware/#:~:text=Billing%20itself%20as%20'The%20Nation's,neighboring%20Virginia%2C%20Maryland%20and%20Pennsylvania)

[delaware/#:~:text=Billing%20itself%20as%20'The%20Nation's,neighboring%20Virginia%2C%20Maryland%20and%20Pennsylvania](https://www.touropia.com/best-places-to-visit-in-delaware/#:~:text=Billing%20itself%20as%20'The%20Nation's,neighboring%20Virginia%2C%20Maryland%20and%20Pennsylvania)

Shaeali, N. S., Mohamed, A., & Mutalib, S. (2020). Customer reviews analytics on food delivery services in social media: A review. In *IAES International Journal of Artificial Intelligence* (Vol. 9, Issue 4, pp. 691–699). Institute of Advanced Engineering and Science.

<https://doi.org/10.11591/ijai.v9.i4.pp691-699>

Shi, H.-X., & Li, X.-J. (2011). A sentiment analysis model for hotel reviews based on supervised learning. *2011 International Conference on Machine Learning and Cybernetics*, 950–954.

<https://doi.org/10.1109/ICMLC.2011.6016866>

- Shinde, S. (2024). *What is Supervised Learning in Machine Learning? A Comprehensive Guide*. Emeritus. <https://emeritus.org/blog/ai-and-ml-supervised-learning/>
- Tabb, M., Delvisio, J., & Gawrylewsky, A. (2021). *What Is Machine Learning, and How Does It Work? Here's a Short Video Primer*. Scientific American. <https://www.scientificamerican.com/video/what-is-machine-learning-and-how-does-it-work-heres-a-short-video-primer/>
- Topper, N. (2023). *Bag of Words Model in NLP Explained*. Built In. <https://builtin.com/machine-learning/bag-of-words>
- Torres Samboni, L. A. (2015). ANÁLISIS DE SENTIMIENTOS SOBRE EL POSCONFLICTO COLOMBIANO UTILIZANDO HERRAMIENTAS DE MINERÍA DE TEXTO. *Escuela Colombiana de Ingeniería Julio Garavito*. <https://repositorio.escuelaing.edu.co/bitstream/handle/001/403/Torres%20Samboni%2c%20Lina%20Andrea%20-%202016.pdf?sequence=1&isAllowed=y>
- Umarani, V., Julian, A., & Deepa, J. (2021). Sentiment Analysis using various Machine Learning and Deep Learning Techniques. *Journal of the Nigerian Society of Physical Sciences*, 3(4), 385–394. <https://doi.org/10.46481/jnsps.2021.308>
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*.
- Welch, J. (2022). *Value of Tourism in Delaware*. https://assets.simpleviewinc.com/simpleview/image/upload/v1/clients/delaware/Value_of_Tourism_2022_bc573d08-aa9c-4094-8cd2-f6f70e789e26.pdf
- Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.-W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., ... Zhang, J. (2021).

Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, 2(4), 100179. <https://doi.org/10.1016/j.xinn.2021.100179>

Yang, M., Ren, Y., & Adomavicius, G. (2019). Understanding user-generated content and customer engagement on Facebook business pages. *Information Systems Research*, 30(3), 839–855. <https://doi.org/10.1287/isre.2019.0834>