

# Modelo de satisfacción al Cliente

## Integrantes:

Sergio Agreda  
Daniel Sanchez  
Ludwing Young



# De Reseñas a Reputación

## Pipeline automatizado parara Trustpilot

Un sistema integral que transforma opiniones en línea en inteligencia competitiva accionable.



### Scrapping

Extracción automatizada



### Limpieza

Validación y normalización



### Analisis

Análisis de sentimiento



### Dashboard

Exploración interactiva

# Motivación de la Investigación



**Benchmark de Reputación**



**Detección de Problemas**



**Seguimiento Temporal**

# Validación Académica

Esta metodología se alinea con la vanguardia de la investigación, utilizando Transformers para cuantificar matices del lenguaje que las calificaciones por estrellas no capturan.



## Fundamento Teórico

Estudios como Le et al. (2024) demuestran que los modelos BERT-BiGRU predicen la satisfacción del cliente al combinar análisis de sentimiento con metadatos.



## Superioridad de Transformers

Investigaciones como Aftan & Shah (2023) validan que los modelos Transformer (AraBERT) superan a CNN/RNN tradicionales en la clasificación de satisfacción.



## Hacia el Futuro

La investigación de Teichert & Shah (2026) explora LLMs para convertir reseñas en constructos teóricos, señalando el camino a seguir.

# Categorías y Empresas Objetivo

Una muestra representativa de 40 empresas para un análisis comparativo sólido.



## Tiendas de Ropa

Incluyendo análisis detallados de marcas como ZARA y TWOTHIRDS para mostrar el potencial comparativo del dashboard.



## Bancos






## Seguros

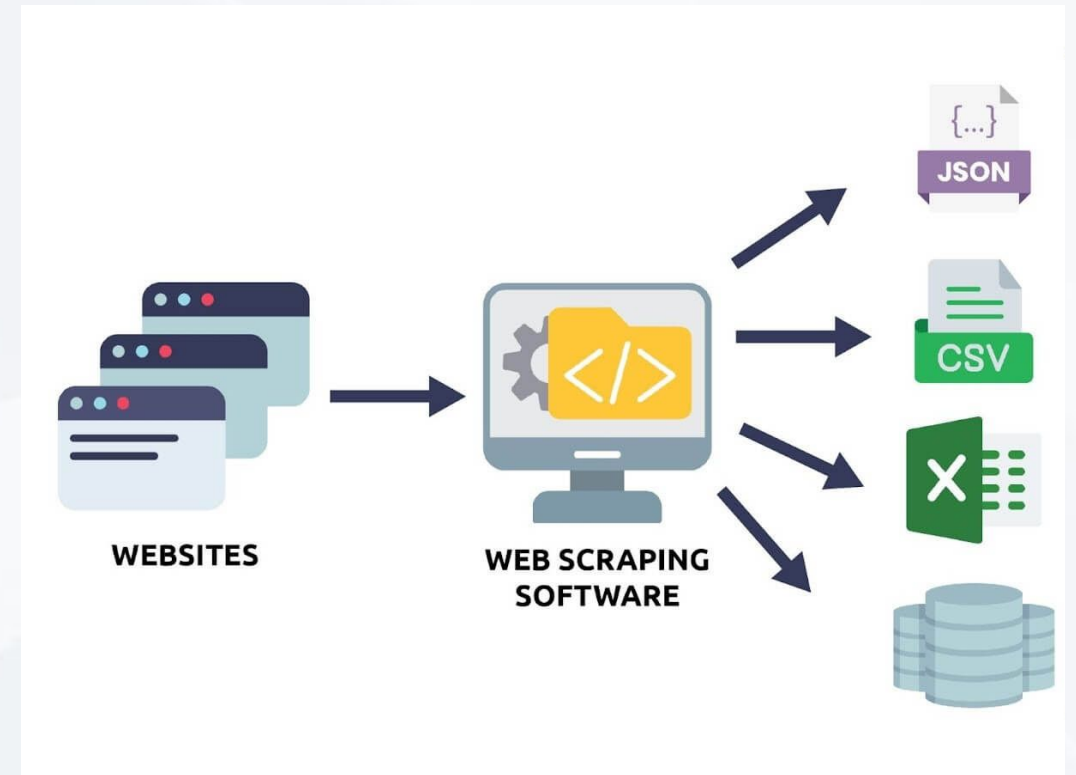


## Concesionarios de Autos

**Dataset Total:** 1492 Reseñas - Última reseña: 14/11/2025 a las 11:59

# Arquitectura de Scraping

-  **Selenium + WebDriver Manager:** Este scraper automatiza la extracción de datos desde Trustpilot utilizando Selenium y WebDriver Manager.
-  Obtiene información clave de las empresas (nombre, ubicación, puntuación y sitio web).
-  El proceso es estable, reproduce la navegación real del usuario y captura los elementos de forma eficiente.



# Flujo de Trabajo de Refinamiento

Una arquitectura de datalake de tres capas para un procesamiento de datos claro y controlado.



## ZONA DE INGRESO

Datos crudos (CSV, JSON) tal como se extraen.



## ZONA REFINADA

Datos procesados, limpiados y validados.



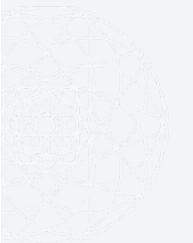
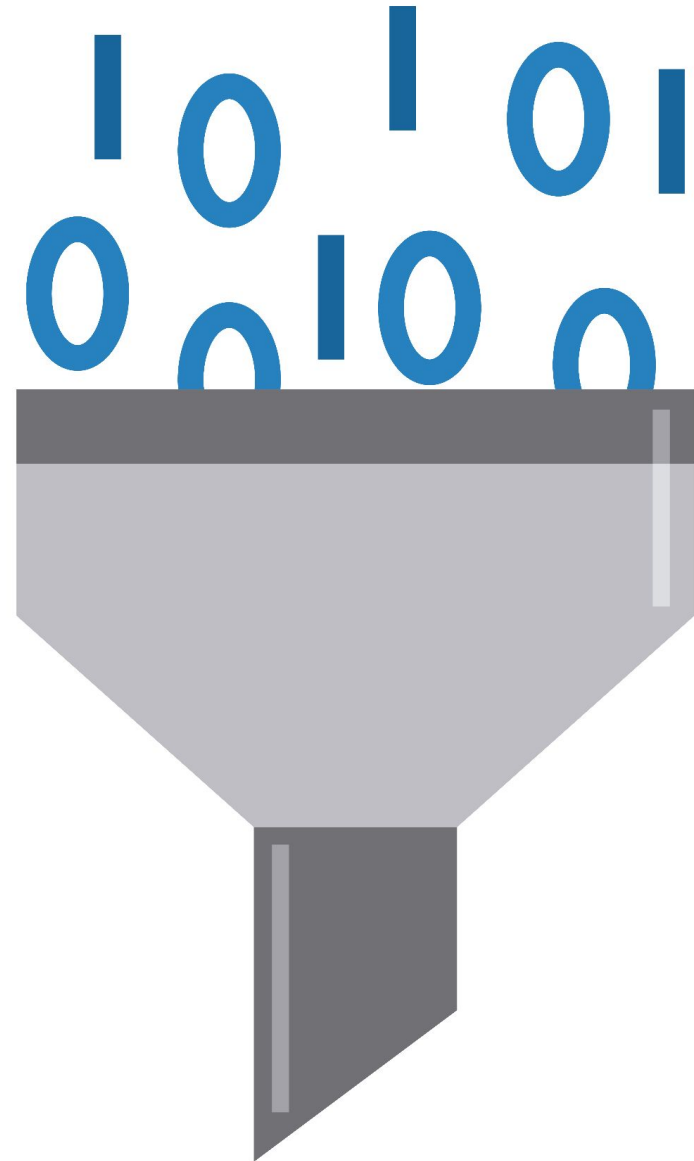
## ZONA DE CONSUMO

Dashboard y datos listos para el análisis.

# Pasos de Limpieza y Validación

Un proceso riguroso que garantiza la calidad del dato, transformando información cruda en un conjunto limpio, coherente y plenamente confiable para su análisis.

- ✓ Conversión a minúsculas y eliminación de acentos.
- ✓ Homogeneización de formatos de fecha y zona horaria.
- ✓ Conversión de calificaciones a escala 1-5.
- ✓ Filtrado de duplicados y validación de campos clave.





# Enriquecimiento con Análisis de Sentimiento

Usando modelos Transformers de HuggingFace, vamos más allá de las estrellas. Clasificamos cada reseña y asignamos un puntaje de sentimiento para capturar matices que las calificaciones tradicionales pasan por alto.



## **Positivo**

Experiencias favorables

## **Neutro**

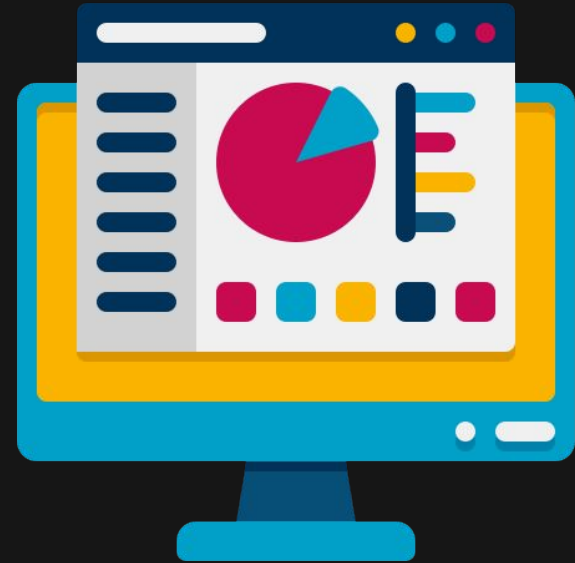
Comentarios sin mucho contexto

## **Negativo**

Experiencias negativas

# Diseño del Dashboard Interactivo

Una interfaz flexible que permite a los usuarios explorar los datos desde una visión general del sector hasta una queja individual, con todos los gráficos y KPIs que se actualizan dinámicamente en base a los filtros.



# Componentes Visuales Clave

Una suite de herramientas visuales para una exploración rápida y efectiva de los datos de reseñas.



## Indicadores Clave (KPIs)

Número de empresas, total de reseñas y calificación promedio para una referencia rápida.



## Dial de Sentimiento

Un indicador visual del sentimiento agregado.



## Tabla Detallada

Exploración granular de cada reseña.



## Nube de Palabras

Identifica temas recurrentes y palabras clave a vuelo de pájaro.

# Filtros Dinámicos

Por categoría, empresa, año y mes.

# Mapa de palabras

Palabras más frecuentes en las reseñas.

# Tabla de reseñas

Una visión general a cada reseña individual.

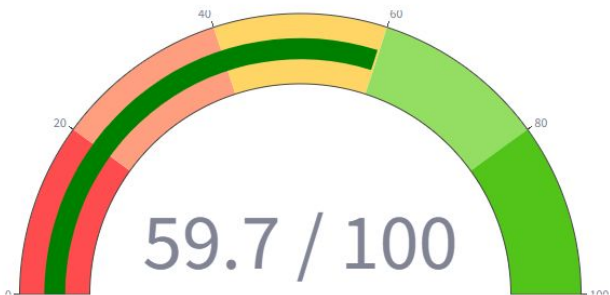
# Actualización en Vivo

Todos los componentes se sincronizan.



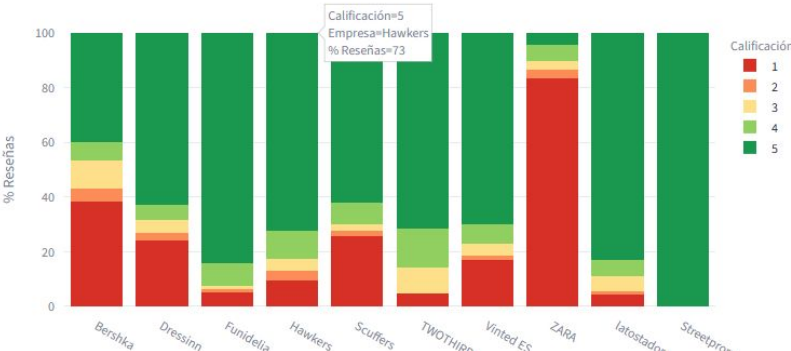
## Reseñas

	nombre	sentimiento	titulo	contenido
5578	Bershka	negativo	me habeis notificado que esta el pedido	me habeis notificado que esta el pedido en
5579	Bershka	positivo	llego un dia antes	llego un dia antes super buena calidad
5580	Bershka	positivo	todo muy sencillo rapidez	todo muy sencillo rapidez eficacia gracia
5581	Bershka	negativo	hice un pedido el sabadopague casi 4	hice un pedido el sabadopague casi 4 euros
5582	Bershka	neutral	hize un pedido online y nunca llego	sin contenido
5583	Bershka	negativo	el transporteentrega un desastre	el problema fue con el transporte y la entre
5584	Bershka	negativo	podria ser peor	todo fue bien hasta que llego la hora de esp
5585	Bershka	positivo	quickly and good	my jacket came very quickly and everythir
5586	Bershka	neutral	llego muy pronto y todo bien embalado	sin contenido
5587	Bershka	negativo	dentro de todo buen servicio	las botas parecian como mas grandes en la



## Distribución de calificaciones

Distribución porcentual de calificaciones por empresa



# TWOTHIRDS

Un líder en calidad y satisfacción del cliente en el sector de ropa. El análisis de 2025 revela una percepción altamente favorable y una lealtad de marca sólida.

80/100

Puntaje de Sentimiento

78%

Reseñas de 5 Estrellas

## Temas Clave Detectados:

Calidad, perfecto , pedido, sostenibilidad, tejidos.



# ZARA

Un caso que evidencia desafíos operativos y de servicio. El análisis de 2025 muestra una percepción desfavorable, revelando áreas críticas para la mejora.

**< 20/100**

## Puntaje de Sentimiento

# 67%

## Reseñas de 1 Estrella

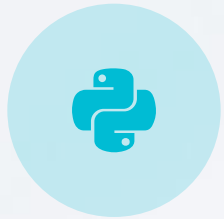
## Problemas Recurrentes:

Devoluciones, reembolsos, servicio al cliente , pedidos.



# Resultados y Entregables Clave

Un pipeline completo, desde la adquisición de datos hasta la entrega de una herramienta analítica funcional.



## **Notebook Reproducible**

Código de scraping y limpieza.



## **Dataset en JSON**

1.2M reseñas limpias y etiquetadas.



## **Etiquetas de Sentimiento**

Puntajes y confianza para cada reseña.



## **Dashboard Streamlit**

Herramienta de análisis interactiva.



# Impacto en el Negocio

Transformamos el análisis de reputación de un proceso manual y tardío en una ventaja competitiva ágil y basada en datos.



## Benchmark Competitivo

Los gerentes pueden comparar su posición frente a la competencia en tiempo real.



## Detección Temprana

El marketing puede detectar elogios o quejas emergentes para gestionar la marca.



## Monitoreo de KPIs

Las operaciones pueden monitorear indicadores clave casi en tiempo real.





# Escalabilidad y Ética

Esta solución está diseñada para crecer y adaptarse, manteniendo siempre un compromiso con la ética y el cumplimiento normativo.



## Escalabilidad Horizontal

Soporte para ingestión paralela de múltiples plataformas y ejecución programada con Airflow.



## Privacidad y Cumplimiento

Anonimización de datos GDPR y gestión segura de la información del usuario.



## Recolección Ética





Uso de limitación de tasa (rate limiting) y rotación de user-agents para respetar los términos del sitio.





# Mejoras Futuras

Una hoja de ruta para expandir las capacidades analíticas y el valor de la plataforma.

-  Análisis de sentimiento por aspecto (shipping, producto, soporte).
-  Mejora de modelos multilingües para jerga local.
-  Integración de metadatos estructurados (tiempos de entrega).
-  Uso de LLMs para generación de informes ejecutivos.

# Gracias

