

# InnoTech Solutions

09.25.2023

---

## Equipo

Francisco García

Fabian Gonzalez

Federico Molina

Ronnie Fuertes

## Entendimiento de la situación actual

Amazon ofrece una amplia gama de productos, lo que puede hacer que sea un desafío para los usuarios encontrar lo que están buscando. Además, muchos usuarios no están al tanto de la totalidad de la oferta de productos en la plataforma.

Dada su prominencia como uno de los gigantes del comercio electrónico a nivel global, Amazon está explorando nuevas estrategias para impulsar sus ventas y mantener su liderazgo en el mercado estadounidense. Para lograr este objetivo, han optado por colaborar con InnoTech Solutions, nuestra empresa reconocida por su capacidad para concebir soluciones innovadoras y generar resultados revolucionarios.

Esta asociación busca abordar de manera proactiva los desafíos que Amazon enfrenta en la actualidad, centrándose en optimizar la experiencia del usuario y potenciar las ventas a través de soluciones tecnológicas avanzadas y estratégicas.

## Objetivo General

Desarrollar e implementar un sistema de recomendación de productos que optimice las ventas de Amazon a través de la extracción de insights a partir de datos históricos, así como la creación de un modelo de recomendaciones personalizadas.

## Objetivos específicos

1. Analizar y procesar reseñas de clientes para extraer información relevante y valiosa para la toma de decisiones de la empresa (utilizando técnicas de Procesamiento de Lenguaje Natural - NLP).
2. Crear un sistema de recomendación de compras futuras basado en análisis de reseñas, información de productos frecuentemente comprados juntos y puntuaciones de productos.
3. Diseñar una arquitectura de datos escalable para el proyecto, permitiendo la ingesta eficiente de nuevos datos a la base de datos.
4. Desarrollar un dashboard para los ejecutivos de la empresa que muestre datos relevantes para la toma de decisiones estratégicas.
5. Implementar un conjunto de data pipelines para el procesamiento y almacenamiento eficiente de grandes cantidades de datos.

## Alcance del proyecto

El objetivo principal de este proyecto es desarrollar un sistema de recomendación de productos para ofrecer a los usuarios que pueda ayudar a mejorar las ventas de la empresa Amazon. El sistema utilizará algoritmos avanzados de aprendizaje automático para analizar el historial de compras, las preferencias y el comportamiento de navegación de los clientes con el fin de proporcionar recomendaciones precisas y relevantes de productos.

## Entregables Principales del Proyecto

1. Diseño y desarrollo del sistema de recomendación.
2. Integración del sistema en la plataforma de Amazon.
3. Pruebas exhaustivas del sistema para garantizar su precisión y rendimiento.
4. Documentación detallada del sistema, incluyendo manuales de usuario y especificaciones técnicas.

## Condiciones de Cumplirse Antes de Aceptar los Entregables

Los entregables serán aceptados cuando se cumplan las siguientes condiciones:

1. El sistema de recomendación debe demostrar una mejora significativa en la precisión de las recomendaciones en comparación con el sistema existente.
2. El sistema debe estar completamente integrado en la plataforma de Amazon sin interrupciones en la funcionalidad existente.
3. Las pruebas deben mostrar que el sistema es estable y puede manejar la carga de usuarios esperada.
4. La documentación debe ser completa y comprensible para los usuarios y el equipo de soporte técnico.

## Exclusiones del Trabajo Realizado

No se realizarán las siguientes actividades como parte de este proyecto:

1. Desarrollo de nuevas características importantes para la plataforma de Amazon que no estén relacionadas con la recomendación de productos.
2. Mantenimiento o actualización de sistemas o aplicaciones existentes que no estén directamente relacionados con el sistema de recomendación.

## Supuestos

Durante la planificación de este proyecto, se consideran los siguientes supuestos:

1. Se tendrán los datos de historial de compras y navegación necesarios para entrenar y probar los algoritmos de recomendación.
2. Los recursos técnicos y humanos requeridos estarán disponibles según lo programado.
3. No habrá cambios significativos en los objetivos o requisitos del proyecto una vez iniciado.


## Restricciones o Limitaciones del Proyecto

Las siguientes restricciones y limitaciones se aplican al proyecto:

1. El presupuesto y el tiempo asignados son limitados.
2. El sistema de recomendación se basará únicamente en datos proporcionados por Amazon y no incluirá fuentes de datos externas.
3. El proyecto no garantiza un aumento específico en las ventas o ingresos de Amazon, ya que esto dependerá del comportamiento individual de los clientes.

## Indicadores Clave de Desempeño (KPI's)

Nombre del KPI	Descripción	Formula	Periodicidad	Objetivo
Tasa de Satisfacción del Cliente	Mide el porcentaje de clientes satisfechos con un producto o servicio.	(Se calcula dividiendo el número de reseñas positivas ) / (Total de reseñas)	Mensual	Conseguir una tasa de satisfacción del cliente de al menos el 70%
Puntuación Promedio de Reseñas por Categoría	Calcula el promedio de puntuaciones de reseñas para productos dentro de cada categoría.	(Suma de todas las puntuaciones de reseñas para productos en una categoría) / (Cantidad total de reseñas en esa categoría)	Mensual	Conseguir una puntuación promedio de al menos 4 puntos por categoría.
Productos con Mayor Cantidad de Reseñas Útiles	Identifica los productos que han recibido la mayor cantidad de reseñas útiles por parte de otros clientes.	(Suma de todas las reseñas útiles en una categoría) / (Cantidad total de reseñas en esa categoría)	Mensual	Obtener al menos 30 productos con reseñas de al menos 4 puntos.
Productos con Mayor Cantidad de Compras Adicionales	Indica qué productos tienden a ser comprados junto con otros productos	Encontrar los productos que son comprados junto con mayor frecuencia	Annual	Conseguir al menos 10 pares de productos que sean comprados juntos.
Productos con Mayor Cantidad de Visualizaciones Adicionales	Indica qué productos tienden a ser visualizados junto con otros productos.	Encontrar los productos que son visualizados junto con mayor frecuencia	Anual	Conseguir al menos 10 pares de productos que sean vistos juntos.



Tasa de Respuesta de la Venta	Indica qué porcentaje de visualizaciones de productos se convierten en ventas.		Trimestral	
Incremento de Ventas	Compara las ventas del período actual con las del período anterior. Es esencial para evaluar el desempeño y el éxito de las estrategias de ventas implementadas.		Trimestral	

## Solución propuesta

### Repositorio en Github

[https://github.com/fransindi/recomendacion\\_amazon](https://github.com/fransindi/recomendacion_amazon)

### Tecnologías utilizadas

Tecnología/Aplicación/Servicio	Descripción
GitHub	Plataforma de control de versiones y colaboración en el desarrollo de software
Git	Sistema de control de versiones distribuido
Google Docs, Sheets, Slides	Herramientas de creación y edición de documentos
Python	Lenguaje de Programación
Scikit-Learn	Frameworks de Machine Learning
Pandas	Biblioteca de código abierto de Python diseñada para la manipulación y análisis de datos
PySpark	Biblioteca de código abierto de Python que procesar y analizar grandes conjuntos de datos de manera distribuida
Numpy	Biblioteca de código abierto en Python que proporciona soporte para arreglos multidimensionales y funciones matemáticas
SpaCy	Biblioteca de procesamiento de lenguaje natural (PLN) de código abierto en Python
Visual Studio Code	Entorno de desarrollo de código abierto desarrollado por Microsoft
AWS s3	Servicio de almacenamiento en la nube
AWS Glue	Servicio de ETL completamente administrado que facilita la preparación y carga de datos para el análisis
AWS Glue Data Catalog	Repositorio centralizado que almacena metadatos y esquemas de datos para permitir la búsqueda y acceso a los datos de manera eficiente

AWS Redshift	Data warehouse diseñado para ejecutar consultas de análisis de datos complejas en grandes conjuntos de datos
AWS SageMaker	Servicio para desarrollo, entrenamiento y despliegue de modelos de aprendizaje automático (machine learning) a escala.
AWS QuickSight	Servicio de inteligencia empresarial (BI) que permite crear y compartir visualizaciones de datos interactivas y paneles de control
Cloudwatch	Servicio de monitoreo y observabilidad que proporciona métricas y registros para las aplicaciones y recursos de AWS

## Metodología de Trabajo

### 1. Metodología de Gestión de Proyectos

Se emplea la metodología Agile, específicamente el marco de trabajo Scrum, para la gestión del proyecto. Esto implica ciclos de desarrollo iterativos y colaborativos, con entregas incrementales y flexibilidad para adaptarse a cambios en los requisitos.

### 2. Roles y Responsabilidades

Se asignan roles específicos dentro del equipo:

- *Scrum Master*: Encargado de facilitar el proceso Scrum y remover obstáculos para el equipo.
- *Product Owner*: Representante del cliente, responsable de definir y priorizar el backlog de productos.
- *Desarrolladores*: Encargados de la implementación técnica del sistema de recomendación.
  - Analista de Datos: Francisco García
  - Ingeniero de Datos: Fabian Gonzalez, Ronnie Fuertes
  - Científico de Datos: Federico Molina



Nombre del desarrollador	Rol en el proyecto	Responsabilidades
Fabian Gonzalez Ronnie Fuertes	Ingeniero de Datos	Realizar una selección cuidadosa del stack tecnológico, eligiendo las herramientas y tecnologías adecuadas que permitan cumplir con los requisitos y objetivos del proyecto, optimizando así la eficiencia y el rendimiento del sistema. Implementar medidas para garantizar la disponibilidad de datos limpios y estructurados, asegurando que la información utilizada esté completa, actualizada y libre de errores, lo que permitirá tomar decisiones informadas basadas en datos confiables y precisos.
Federico Molina	Científico de Datos	Desarrollar un modelo de machine learning altamente preciso y eficiente que utilice técnicas de recomendación para predecir con exactitud las preferencias individuales de los usuarios, con el objetivo de ofrecer recomendaciones personalizadas y mejorar su experiencia. Crear un modelo predictivo avanzado utilizando técnicas de machine learning para realizar pronósticos precisos sobre el crecimiento futuro de los negocios, permitiendo tomar decisiones estratégicas basadas en datos y maximizar las oportunidades de crecimiento y rentabilidad.
Francisco García	Analista de Datos	Desarrollar una interfaz interactiva que brinde la posibilidad de explorar de manera intuitiva los resultados del análisis de datos en las reseñas de los usuarios. La finalidad es facilitar la visualización y comprensión de los insights obtenidos, permitiendo una interacción dinámica con la información para una toma de decisiones más informada.

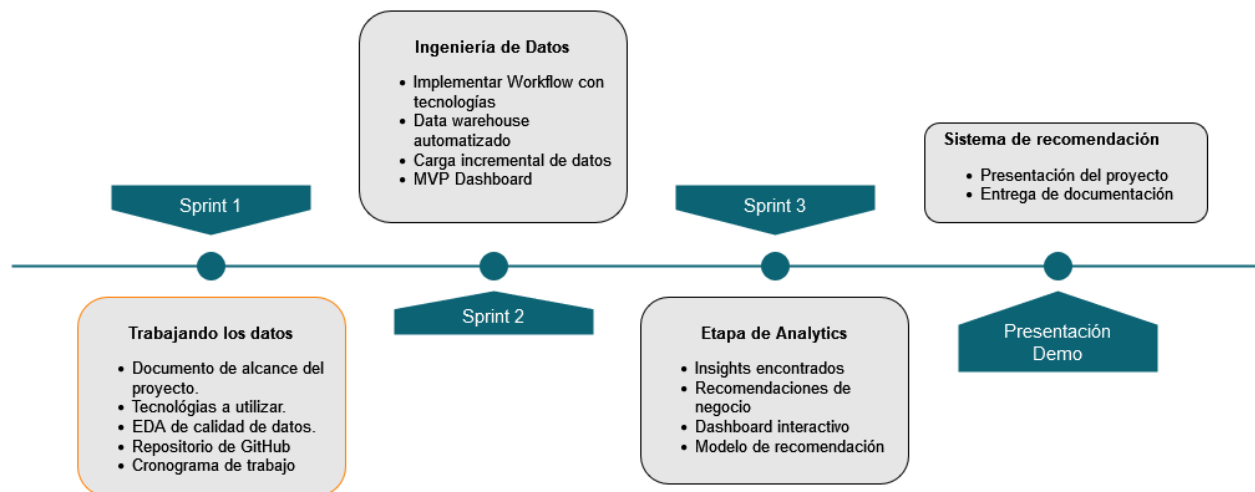
### 3. Procesos y Actividades

- *Planificación del proyecto:* Al inicio del primer sprint, se realiza una reunión de planificación para definir las tareas a realizar en el sprint.
- *Reuniones Diarias:* Se llevarán a cabo reuniones cortas diarias para compartir actualizaciones y resolver impedimentos.
- *Revisión del Sprint:* Al finalizar cada sprint, se presentarán los resultados al equipo y al cliente para recibir retroalimentación.
- *Retrospectiva del Sprint:* Se realizará una reunión al final de cada sprint para evaluar el proceso y proponer mejoras.

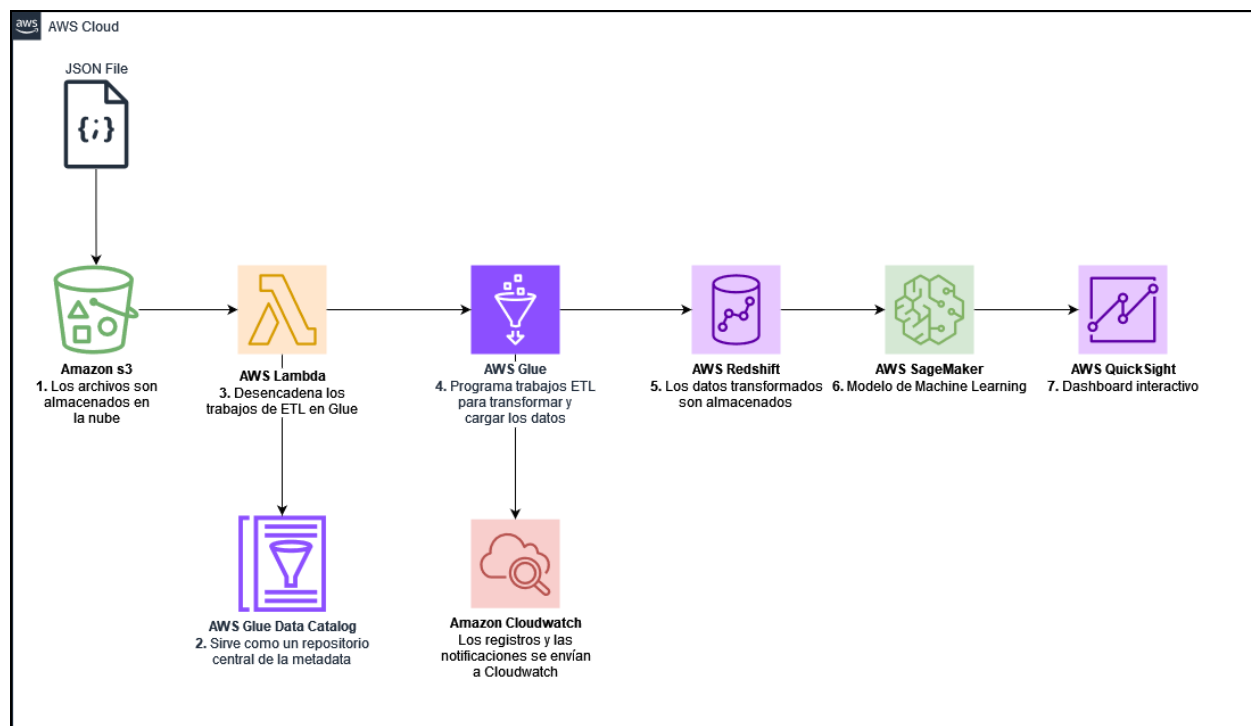
### 4. Planificación y Programación

Se utilizará un diagrama de Gantt para visualizar y gestionar las tareas y su progreso en tiempo real. Las tareas se asignan con estimaciones de tiempo para su ejecución.

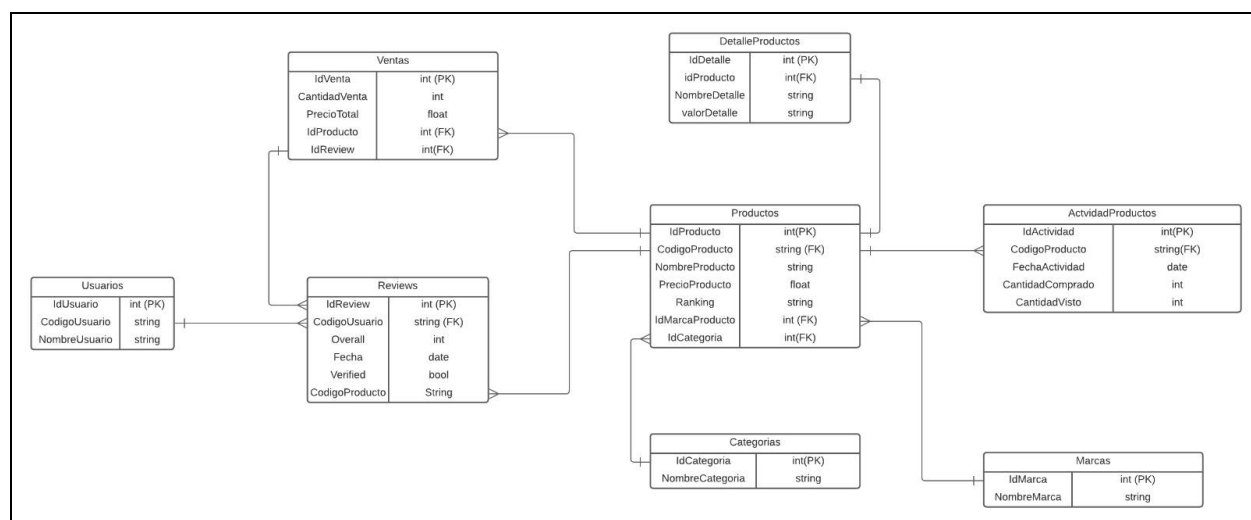
### Flujo de trabajo del proyecto



## Trabajo en la nube



## Arquitectura Data warehouse (DER)



## Cronograma del Proyecto

SPRINT 1			FECHA DE INICIO	% COMPLETADO DE LA TAREA	25/09 al 29/09				
					SEMANA 1				
					L	M	X	J	V
Puesta en marcha del proyecto									
1	Entendimiento de las necesidades	25/09/23	100 %						
2	Creación de repositorio en GitHub	25/09/23	100 %						
3	Documentar alcance, objetivos y entregables	26/09/23	100 %						
4	Seleccionar dataset a utilizar	25/09/23	100 %						
5	Definir columnas importantes para los KPIs, su calidad y metadata	26/09/23	100 %						
6	Realizar EDA de dataset seleccionados	26/09/23	100 %						
7	Selección de KPIs	25/09/23	100 %						
8	Realización del DER	27/09/23	100 %						
9	Seleccionar Stack Tecnológico	28/09/23	40 %						
10	Definir entidad-relación de los datasets a utilizar	27/09/23	100 %						
11	Revisar servicios web de Amazon	25/09/23	100 %						
12	Probar servicios web de Amazon	26/09/23	30 %						
13	Crear y actualizar diagrama de Gantt	26/09/23	100 %						
14	Crear documento de propuesta de proyecto	26/09/23	100 %						
15	Estructurar presentación 'Sprint 1'	28/09/23	100 %						
16	Presentación Demo semanal	29/09/23	0 %						