

Base de datos reviews Amazon

I. DESCRIPCIÓN

Este proyecto se centra en las reviews de Amazon, explorando la integración de bases de datos relacionales y no relacionales. El trabajo abarca el diseño y carga de datos en MySQL y MongoDB respectivamente, además de aplicaciones Python para acceso, visualización y análisis. Se explora la similitud entre usuarios, la relación entre usuarios y artículos utilizando Neo4J, y se expande con nuevas visualizaciones y la relación con Machine Learning para predicciones de puntuaciones de productos.

II. DESARROLLO

Este proyecto consta de los siguientes elementos de desarrollo de *software*:

- 5 ficheros Python entre los que están distribuidas todas las funciones y código necesario para la ejecución de este
- Carpeta con imágenes para algunas visualizaciones
- Librerías usadas: PIL, pymongo, pymysql, dash, dcc, html, pandas, plotly, wordcloud, os, numpy, json, random, GraphDatabase.

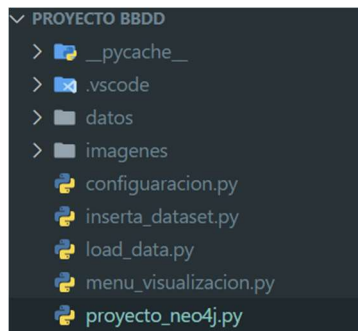


Figura 1. Distribución de módulos

III. RESULTADOS

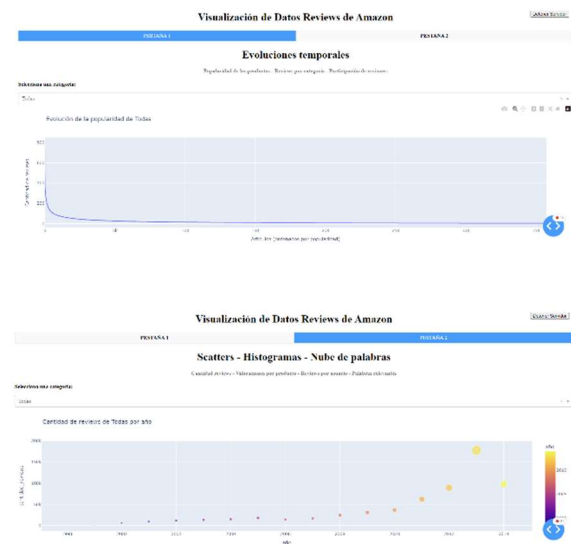
Integración de bases de datos: La estructura relacional de MySQL se aprovechó para mantener la integridad de los datos y facilitar consultas complejas, mientras que MongoDB proporcionó flexibilidad para almacenar datos semi-estructurados y no relacionales.

reviews_proyecto	
<ul style="list-style-type: none"> DigitalMusic GroceryFood MusicalInstruments ToysGames VideoGames 	
	<pre> _id: 10000002 reviewText: "A classically-styled and introverted album, Memory of Trees is a master..." summary: "Enya at her most elegant" reviewTime: "2001-06-03" </pre>
	<pre> _id: 10000003 reviewText: "I never thought Enya would reach the sublime heights of Evacuee or Nar..." summary: "The best so far" reviewTime: "2003-07-14" </pre>

reviewID	reviewerID	asin	overall	unixReviewTime	min_helpful	max_helpful	category
10000001	A3EBH4CZ06V2A4	5555991584	5	1158019200	3	3	DigitalMusic
10000002	A2PWAXJG90JXV	5555991584	5	991526400	0	0	DigitalMusic
10000003	A3BRL0X2T4Q9F	5555991584	5	1058140800	2	2	DigitalMusic
10000004	A2K38AUJ7G0X0	5555991584	5	957312000	1	1	DigitalMusic
10000005	A1A5P0I0H7H0X	5555991584	4	1200528000	1	1	DigitalMusic
10000006	A3P49WD75WHAG5	5555991584	5	1059523200	62	65	DigitalMusic
10000007	A3O90GID71SEGG	5555991584	3	959385600	1	5	DigitalMusic
10000008	A3EJYJC25OJVKK	5555991584	5	1040428800	5	5	DigitalMusic
10000009	A1DAEVOH9NR6C7	5555991584	5	1167868800	4	4	DigitalMusic
10000010	A33TRNCQK4LU07	5555991584	5	1052697600	12	12	DigitalMusic
10000011	AWY3EPK0UJ1WV	5555991584	5	978393600	2	3	DigitalMusic
10000012	A1SCJW0M9Q3W39K	5555991584	4	1360454400	1	1	DigitalMusic
10000013	A14BTJR9H9VNLJ1	5555991584	5	1375142400	1	1	DigitalMusic
10000014	A2AQOZQ3W7NVVOK	5555991584	5	1108425600	12	13	DigitalMusic
10000015	A1BXAS3M3AJOKL	5555991584	4	1114473600	2	2	DigitalMusic

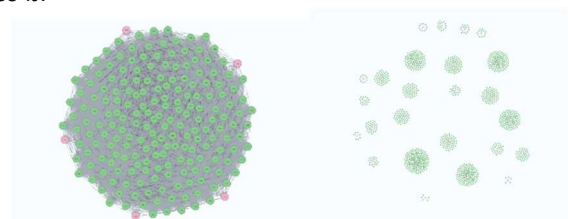
Figuras 2 y 3. Estructura tablas SQL y archivos MongoDB

Visualización dinámica: La creación de un dashboard interactivo con Dash permite a los usuarios explorar y analizar los datos de manera dinámica. Las diferentes pestañas del dashboard facilitan la organización y presentación de las visualizaciones, mejorando la experiencia del usuario.



Figuras 4 y 5. Pestañas 1 y 2 del dashboard

Integración con Neo4J: Se calculan y visualizan similitudes entre usuarios basadas en sus reviews mediante el índice de Jaccard. También se muestran enlaces entre usuarios y artículos, incluyendo propiedades como la puntuación y la fecha de las reviews. Además, se identifican usuarios que han revisado múltiples tipos de artículos y se representan en Neo4J.



Figuras 5 y 6. Algunos resultados de neo4j