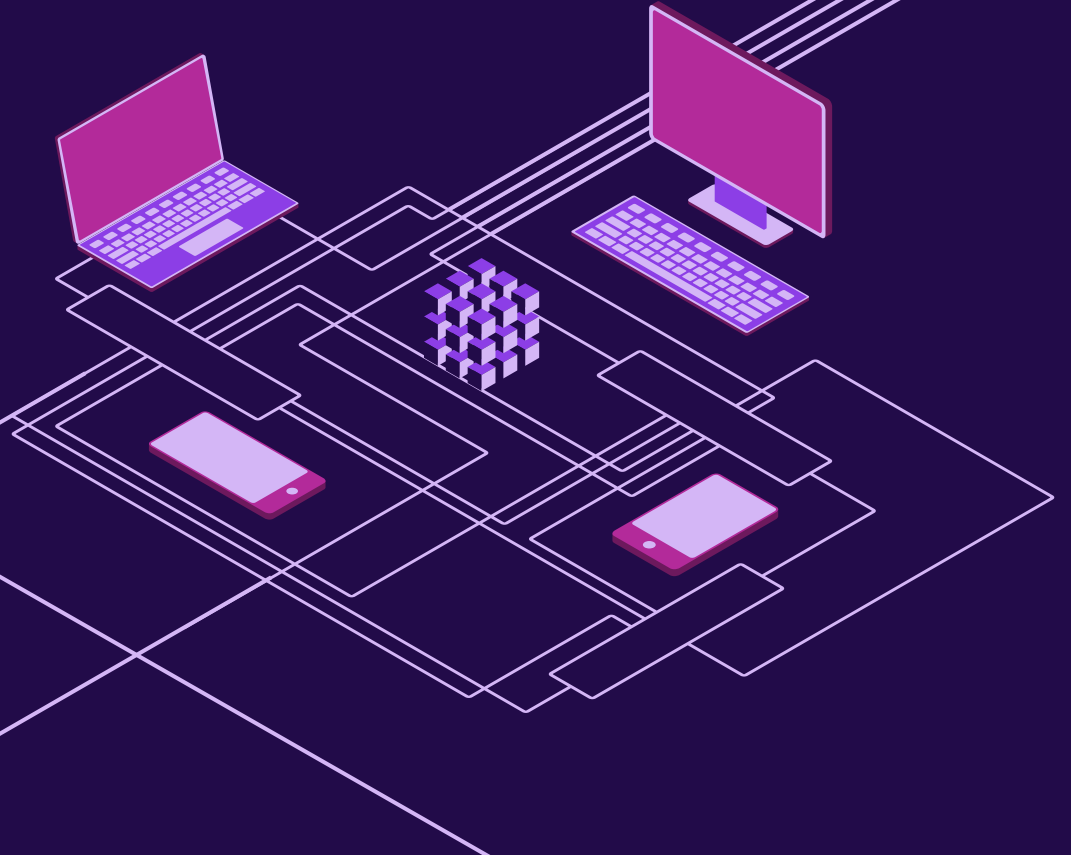


Book-Crossing

CC5212 - Procesamiento Masivo de Datos

Sebastián Aguilera - Cinthya Robles - Bryan Ortiz



CONTENIDOS

01

Overview

Resumen del proyecto.

04

Resultados

Lo que obtuvimos

02

Data

Datos utilizados, fuente, etc.

05

Conclusiones

Concluyendo conclusiones

03

Metodología

El proceso del trabajo.



OVERVIEW

El objetivo principal de este proyecto es crear un script en Pig Latin que permita obtener un ranking de autores, basados en el promedio de los ratings de sus libros.





OVERVIEW

Generando tuplas del estilo:

Position	Author	<BestScoredBooks>	BestScore	Avg. Score	#Votes
----------	--------	-------------------	-----------	------------	--------

1. La posición en el ranking se decide primero por su Avg. Score y, en caso de empate, por el #Votes.
2. Sólo son considerados autores con más de 5 votos en total.

DATA

```
graph TD;
    DATA[DATA] --- Usuarios;
    DATA --- Libros;
    DATA --- Ratings;
```

278K

Usuarios

User-ID
Location
Age

271K

Libros

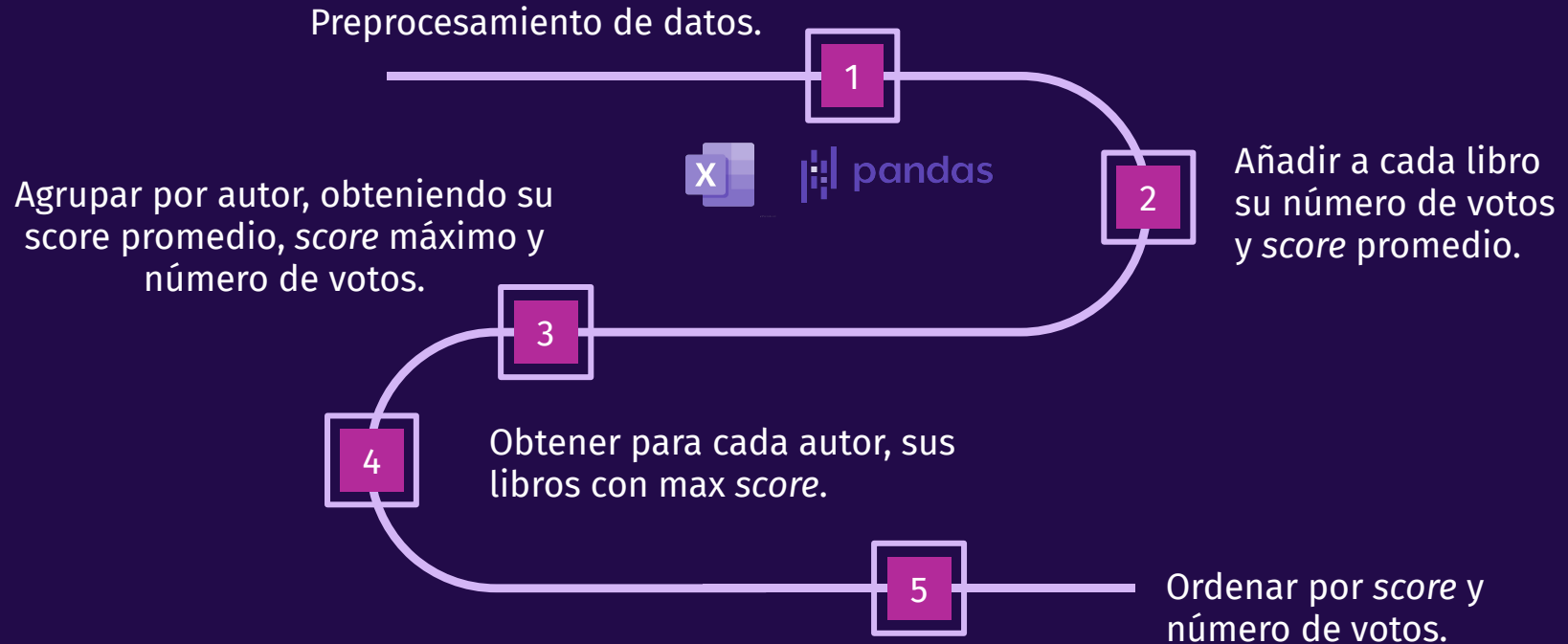
ISBN
Book-Title
Book-Author
Year-Of-Publication
Publisher

1,14M

Ratings

User-ID
ISBN
Rating

METODOLOGÍA



RESULTADOS



1	Michiro Ueyama	{{(ZOIDS: Chaotic Century, Vol. 11 # 2003), (ZOIDS: Chaotic Century, Vol. 10 # 2002), ... } }	10.0	10.0	13
2	Pamela E. Apkarian-Russell	{{(Postmarked Yesteryear: 30 Rare Holiday Postcards # 2001)}}	10.0	10.0	11
3	Wataru Yoshizumi	{{(Marmalade Boy #2 # 2002), (Marmalade Boy (Book 3) # 2002), ... } }	10.0	10.0	11

No son resultados muy concluyentes

Hice todo ese trabajo por nala



CONCLUSIONES

Los votos se concentran en muy pocos libros, es decir, gran parte de los votos (buenos o malos) va a los mismos títulos y autores.

Tomando en cuenta que los datos procesados son viejos (2004), los resultados no son muy representativos para la actualidad, mas el script utilizado puede funcionar para data más actual.



THANKS!

Con esto concluye nuestra presentación. Gracias!