

books

Paula Boks

Enjoy the good-read!

January 2025

BOOK CLUB

POEM

02

BOOK CLUB

FICTION

03

BOOK CLUB

ROMANCE

04

BOOK CLUB

BIOGRAPHY

05

BOOK CLUB

YOUNG ADULT

06

BOOK CLUB

MYSTERY

07

BOOK CLUB

SCIENCE FICTION

08

Data

CONTENT

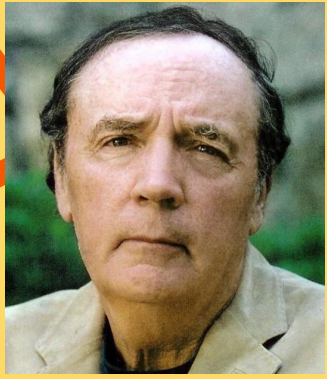
- ★ **goodreads.com**
- ★ Dataset from kaggle
- ★ 25 columns and > **52,000 rows** on books
- ★ one row = one book
- ★ Columns on books, ratings, prices, genre...

STATE

- ★ bookId = **unique identifier**
- ★ Lots of missing values
- ★ Faulty dtypes
- ★ Some letters not permitted in **utf-8**
- ★ Coded in lists

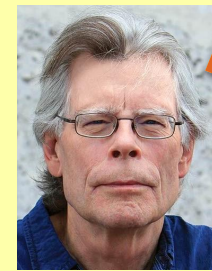
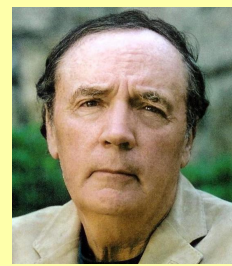


Key Findings: Authors I



1

Authors II



Number of Book Titles per Author	
main_author	
Nora Roberts	81
James Patterson	79
Agatha Christie	73
Stephen King	72
Terry Pratchett	58
Erin Hunter	54
Karen Kingsbury	53
Carolyn Keene	53
Meg Cabot	51
Alex Archer (pseudonym)	50
Name: count, dtype: int64	

Number of Awards per Author	
main_author	
Stephen King	97
Neil Gaiman	75
China Miéville	69
Suzanne Collins	62
Kate DiCamillo	62
Terry Pratchett	61
Judy Blume	53
Ursula K. Le Guin	52
Laurie Halse Anderson	52
Louise Penny	50
Name: number_awards, dtype: int64	



Most Successful Book



WHO KNOWS
THE
AUTHOR?

Genres

Genre (count)

genre	
Fiction	5001
Fantasy	4068
Young Adult	2512
Nonfiction	2008
Historical Fiction	1862
Mystery	1605
Romance	1417
Science Fiction	1063
Classics	868
History	757

Name: count, dtype: int64

Genre by Rating (mean)

genre	
Alchemy	4.65
Baha I	4.62
Dinosaurs	4.57
Detective	4.56
Geography	4.55
Gender Studies	4.50
Ghost Stories	4.50
Folk Tales	4.49
Gaming	4.48
Urban	4.46

Name: rating, dtype: float64



Genre by Price (mean)	
genre	
19th Century	285.93
Comic Book	229.84
Apocalyptic	173.48
Art Design	172.51
Alchemy	97.81
Semiotics	80.18
Diary	75.93
Portugal	75.06
Fashion	73.50
Drama	64.33
Name: price, dtype: float64	



... more

Book Format (count)	
bookFormat	
Paperback	20669
Hardcover	10174
Mass Market Paperback	2325
Kindle Edition	447
ebook	67
Audio CD	60
Unknown Binding	51
Trade Paperback	48
Audiobook	21
Library Binding	18
Name: count, dtype: int64	



Book Title by Price	
title	
I See by My Outfit	898.64
Margin of Safety: Risk-Averse Value Investing Strategies for the Thoughtful Investor	867.05
V/Crying of Lot 49/Gravity's Rainbow	811.04
Men's Garments, 1830-1900: Guide to Pattern Cutting and Tailoring	796.14
Viper's Creed	791.43
Progeny	777.35
Children of Time	711.55
The Oxford English Dictionary (20 Volume Set)	653.73
Revolutionary Letters	631.99
Fantasy Pieces in Callot's Manner	570.41
Name: price, dtype: float64	



BOOK CLUB

HELLO

01

BOOK CLUB

INTRO

02

Relationships

VARIABLES	CHI ²	CRAMER'S V	SIGNIFICANT & STRONG?
Author & Genre	0.00000	0.903	✓
Author & Rating	0.00000	0.751	✓
Genre & Rating	0,00000	0.118	✗
Language & Format	0.00000	0.11	✗

BOOK CLUB

NEXT MONTH

06

BOOK CLUB

CONTACT

07

BOOK CLUB

NEXT MEETING

08

01

02

Challenge: Data Format

MYSTERY

ROMANCE

	bookId	title	series	author	rating
6	170448.Animal_Farm	Animal Farm	NaN	George Orwell, Russell Baker (Preface), C.M. W...	3.95
7	11127.The_Chronicles_of_Narnia	The Chronicles of Narnia	The Chronicles of Narnia (Publication Order) #1-7	C.S. Lewis, Pauline Baynes (Illustrator)	4.26
8	30_J.R.R.Tolkien_4_Book_Boxed_Set	J.R.R. Tolkien 4-Book Boxed Set: The Hobbit and...	The Lord of the Rings #0-3	J.R.R. Tolkien	4.60
9	18405.Gone_with_the_Wind	Gone with the Wind	NaN	Margaret Mitchell	4.30

ALL GOOD!

A1						
	A	B	C	D		
1	bookId,"title","series","author","rating","description","lan					
2	2767052-the-hunger-games,"The Hunger Games","The Hu					
3	2.Harry_Potte a big surpris and the looming terror of the C					
4	2657.To_Kill_a_Mockingbird,"To Kill a Mockingbird","To k					
5	1885.Pride_and_Prejudice,"Pride and Prejudice","","Jane					
6	41865.Twilight,"Twilight","The Twilight Saga #1","Stephen					
7						
8	First, Edward was a vampire.					
9						
10	Second, there was a part of himâ€”and I didn't know how					
11						
12	And third, I was unconditionally and irrevocably in love v					
13						
14	Deeply seductive and extraordinarily suspenseful, Twilig					
15	19063.The_Book_Thief,"The Book Thief","","Markus Zusak					
16	170448.Animal_Farm,"Animal Farm","","George Orwell, R					

NOT SO GOOD

03

04

05

06

07

08

BOOK CLUB

BOOK CLUB

BOOK CLUB

BOOK CLUB

BOOK CLUB

BOOK CLUB

NON-FICTION

BIOGRAPHY

YOUNG ADULT

FANTASY

CLASSICS

HISTORY

Trying to repair...



```
books_sql_4 = books_sql_3.drop(index=19428)
books_sql_4.to_csv("books_sql_4.csv", index=False, encoding="utf-8", quotechar='')
```

```
with open("books_sql_4.csv", "rb") as file: # rb= binary modus, content is shown
    data = file.read()
    # shows the problematic special characters Location
    if b'\x8d' in data: # byte 0x8d was given by MySQL as the one which could not
        position = data.find(b'\x8d')
        print(f"Problematic special characters found at index {position}: {data[p
```

Problematic special characters found at index 19428: b'"402093.Sh_gun", "Sh\xc5\x8d'

```
print(19428 in books_sql_4.index) # should be false and is so
```

False

```
books_sql_4.to_csv("books_sql_4.csv", index=False, encoding="utf-8", quotechar='')
```

```
# Load dataframe again
reloaded_df = pd.read_csv("books_sql_4.csv")
print(reloaded_df.loc[reloaded_df.index == 19428]) # should be empty but is not
```

	bookId	title	rating	language	bookFormat	
19428	28677.The_Spirit_Ring	The Spirit Ring	3.63	English	Paperback	
	pages	bbeScore	price	genre	number_awards	main_author
19428	369	95	4.91	Fantasy	0	Lois McMaster Bujold

Paula Boks

thank you

BOOK CLUB

POEM

02

BOOK CLUB

FICTION

03

BOOK CLUB

ROMANCE

04

BOOK CLUB

BIOGRAPHY

05

BOOK CLUB

YOUNG ADULT

06

BOOK CLUB

MYSTERY

07

BOOK CLUB

SCIENCE FICTION

08