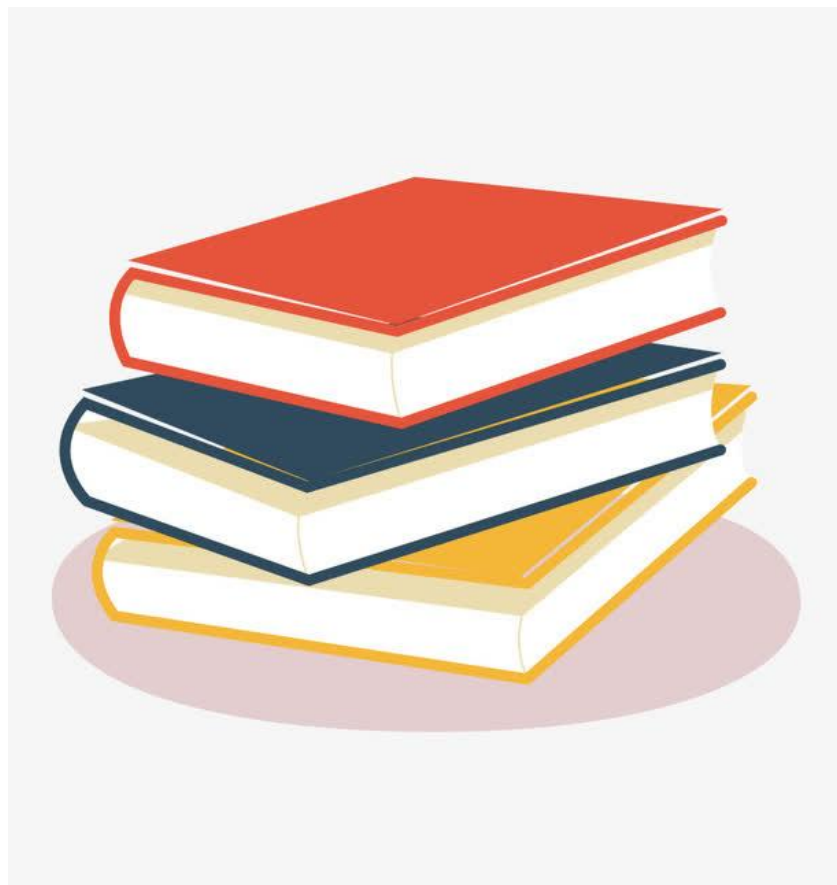


# Amazon Best-Selling Books

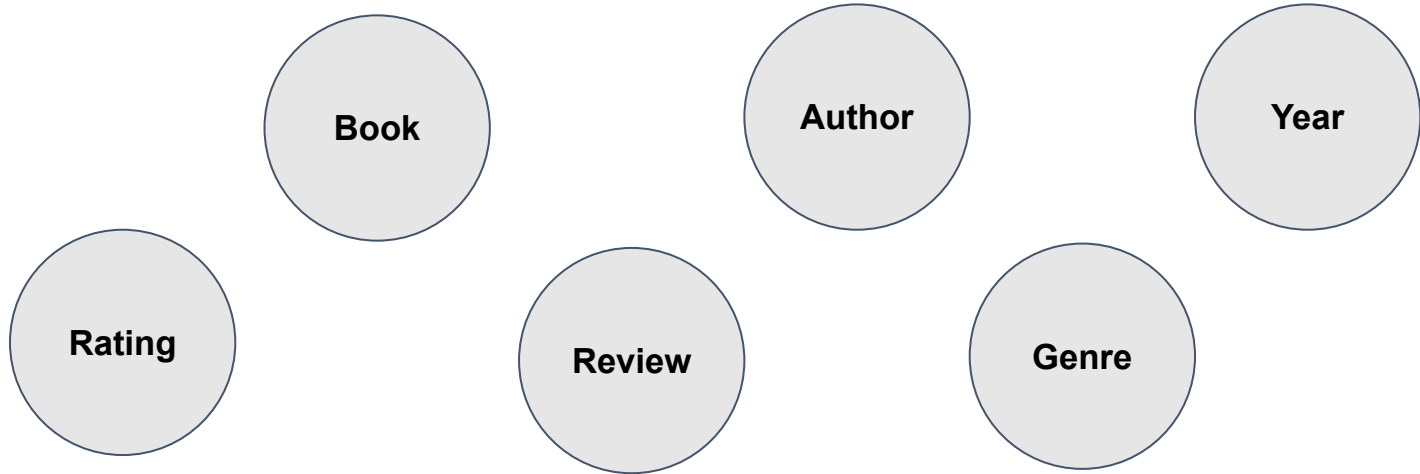
---

Group 1104  
Yuewei Wang, Yongyi Xu



# Domain

The summary of Amazon top 50 bestselling books from 2009 to 2019.



Data reference: <https://www.kaggle.com/sootersaalu/amazon-top-50-bestselling-books-2009-2019>

# Questions

## **1. BookType & Popularity?**

How are the booktype and popularity related. Are fiction books always more popular than non-fiction books? How are the readers' feedback and popularity related?

## **2. Bestseller for many years means higher rating?**

For the books that are bestsellers for multiple years, do they always have higher ratings than the books who only be the bestseller for one year? Are these books always having non-decreasing reviews or reader ratings?

## **3. BookType & Authors?**

For the authors who have multiple published books and are bestsellers, do they always focus on writing single genre types?

# Results

## Query 1

**Genre**

**Rating**

**Review**

- **Fiction more popular**
- **user rating 4.8+**
- **Same winner**

All the books in the top 20 books:

bid	bookname	genre	userrating	review
23	Rush Revere and the Brave Pilgrims: Time-Travel Adventures with Exceptional Americans (1)	Fiction	4.9	7150
31	Hamilton: The Revolution	Non Fiction	4.9	5867
34	Harry Potter and the Chamber of Secrets: The Illustrated Edition (Harry Potter, Book 2)	Fiction	4.9	19622
93	Jesus Calling: Enjoying Peace in His Presence (with Scripture References)	Non Fiction	4.9	19576
101	Wrecking Ball (Diary of a Wimpy Kid Book 14)	Fiction	4.9	9413
105	Oh, the Places You'll Go!	Fiction	4.9	21834
137	Dog Man: Fetch-22: From the Creator of Captain Underpants (Dog Man #8)	Fiction	4.9	12619
144	Harry Potter and the Sorcerer's Stone: The Illustrated Edition (Harry Potter, Book 1)	Fiction	4.9	10052
169	Dog Man: For Whom the Ball Rolls: From the Creator of Captain Underpants (Dog Man #7)	Fiction	4.9	9089
181	The Wonderful Things You Will Be	Fiction	4.9	8842
198	Brown Bear, Brown Bear, What Do You See?	Fiction	4.9	14344
210	Goodnight, Goodnight Construction Site (Hardcover Books for Toddlers, Preschool Books for Kids)	Fiction	4.9	7038
211	Strange Planet (Strange Planet Series)	Fiction	4.9	9382
243	Dog Man: Lord of the Fleas: From the Creator of Captain Underpants (Dog Man #5)	Fiction	4.9	5470
250	The Very Hungry Caterpillar	Fiction	4.9	19546
257	Unfreedom of the Press	Non Fiction	4.9	5956
281	The Magnolia Story	Non Fiction	4.9	7861
346	Harry Potter and the Goblet of Fire: The Illustrated Edition (Harry Potter, Book 4) (4)	Fiction	4.9	7758
350	Last Week Tonight with John Oliver Presents A Day in the Life of Marlon Bundo (Better Bundo Book, LGBT Children's Book)	Fiction	4.9	11881
351	Dog Man: Brawl of the Wild: From the Creator of Captain Underpants (Dog Man #6)	Fiction	4.9	7235

(20 rows)

The number of Fiction and non-fiction books (f and nf) among these 20 books:

f	nf
16	4

(1 row)

The number of Fiction and non-fiction books (f and nf) among these yearly top 1 books:

f	nf
11	3

(1 row)

bid	year	userrating	review	genre
179	2009	4.8	13871	Fiction
94	2009	4.8	13871	Fiction
267	2009	4.8	13871	Fiction
213	2010	4.8	29673	Non Fiction
206	2010	4.8	29673	Non Fiction
93	2011	4.9	19576	Non Fiction
105	2012	4.9	21834	Fiction
105	2013	4.9	21834	Fiction
105	2014	4.9	21834	Fiction
105	2015	4.9	21834	Fiction
105	2016	4.9	21834	Fiction
105	2017	4.9	21834	Fiction
105	2018	4.9	21834	Fiction
105	2019	4.9	21834	Fiction

(14 rows)



# Results

## Query 2

**Rating**

**Years**

**multiple year bestselling books**

- **non-decreasing user rating**
- **non-decreasing reviews**
- **not significant popularity advantages**

```
csc343h-wangyuew=> select count(distinct bid) from MultiyearInfo;
count
-----
    106
(1 row)
```

```
csc343h-wangyuew=> select count(distinct bid) from SingleyearInfo;
count
-----
    255
(1 row)
```

### Userrating comparison between multi & single

bid	num	freq
351	234	0.92
198	234	0.92
210	234	0.92
250	234	0.92
105	234	0.92
93	234	0.92
181	234	0.92
15	177	0.69
67	177	0.69
12	177	0.69

(10 rows)

bid	num	freq
334	3	0.01
9	2	0.01
118	2	0.01
66	3	0.01
6	3	0.01
296	5	0.02
260	5	0.02
86	17	0.07
245	17	0.07
108	17	0.07

(10 rows)

**num:** # of single-year-best sellers have lower ratings

**freq:** num / all 255 single-year-best sellers

# Results

Genre

Query 3

Author

```
[csc343h-wangyuew=> select count(*) from author;
```

```
count
```

```
-----  
    248  
(1 row)
```

```
[csc343h-wangyuew=> select count(*) from answer;
```

```
count
```

```
-----  
    53  
(1 row)
```

aid	authorname	ficnum	nonficnum
4	DK	1	1
19	Scholastic	1	1



```
[csc343h-wangyuew=> select * from written where aid = 4 or aid = 19;
```

aid	bid
4	81
4	35
19	339
19	155

(4 rows)

```
[csc343h-wangyuew=> select * from feedback where bid = 81 or bid = 35 or bid = 339 or bid = 115;
```

bid	year	userrating	review
35	2015	4.6	5360
81	2014	4.5	2586
115	2011	4.2	2094
339	2015	4.7	3564

(4 rows)



[back](#)

# Summative Results



The diagram consists of three circles arranged horizontally on a light gray background. Each circle contains a label. Above the middle circle, there is a small, light gray rectangular box. The circles are connected by faint lines, suggesting a network or flow. The labels are 'Bookstore', 'Authors', and 'Reader' from left to right.

**Bookstore**

**Authors**

**Reader**

# Challenges

- **Phase1:**

topic, data choosing; schema design...

- **Thoughtful queries generation:**

different roles; different factors; how to address all into one big picture

# Lessons

## preliminary check for designing schema and cleaning data

```
csc343h-xuyongy1=> select * from book;
csc343h-xuyongy1=> create view t1 as
csc343h-xuyongy1-> select name, genre from data;
CREATE VIEW

|csc343h-xuyongy1=> select * from
|csc343h-xuyongy1-> t1 c1, t1 c2
|csc343h-xuyongy1-> where c1.name = c2.name and c1.genre != c2.genre;
  name | genre | name | genre
-----+-----+-----
(0 rows)

|csc343h-wangyuew=> select d1.name
|csc343h-wangyuew-> from data d1, data d2
|csc343h-wangyuew-> where d1.name=d2.name and d1.author!=d2.author;
  name
-----
(0 rows)
```

# Lessons

## connection between queries supports a big picture

- **2019NationalDiff(country, rank, GDP, socialSupport, health, freedom, corruption, generosity)**

A tuple in this relation represents the factors that qualify the happiness score and happiness rank of each country in 2019.

- Author (AID, authorname)

*The author of an Amazon's Top 50 bestselling book.*

Attribute	Description	Type
AID	the ID of the author	int
authorname	the name of the author	TEXT

- Written (AID, BID)

*Indicates which author wrote which book.*

FOREIGN KEY (BID) REFERENCES Book(BID),

FOREIGN KEY (AID) REFERENCES Author(AID)

Attribute	Description	Type
AID	the ID of the author	int
BID	the ID of the book	int

# Questions?



# References

- Bookpic:  
[https://www.google.com/url?sa=i&url=https%3A%2F%2F599ku.com%2Fsc%2F11026379.html&psig=AOvVaw1dkPG3XWU9ychV-dxjXdIc&ust=1638223524005000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCNj\\_1bKIvPOCFQAAAAAdAAAAABAc](https://www.google.com/url?sa=i&url=https%3A%2F%2F599ku.com%2Fsc%2F11026379.html&psig=AOvVaw1dkPG3XWU9ychV-dxjXdIc&ust=1638223524005000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCNj_1bKIvPOCFQAAAAAdAAAAABAc)
- Data reference:  
<https://www.kaggle.com/sootersaalu/amazon-top-50-bestselling-books-2009-2019>