CSC343 Project Phase 3 Discussion

Yuewei Wang, Yongyi Xu

Query 1: How are the booktype and popularity related?

From the original dataset, the bestselling and popular level can be reflected by user ratings. Considering the sample size factor for reviews, it is more convincing to conclude a high rating book is popular with more reviews than with fewer reviews. Therefore, the popularity criterion combines user ratings as well as reviews numbers. If any two books have identical user rating scores, the book with more reviews will be considered more popular than the other book.

To measure the relationship between popularity and book type from more perspectives, Query1 discovered the popularity standard in two aspects: the most popular book for each year from 2009 to 2019 and the top 20 most popular books of the decade. The purpose of setting two different aspects is to include more comprehensive information. As demonstrated in Figure 1a) and 1b), book No. 105 was the top 1 popular book in consecutive years. Only having the top 20 most popular tables will result in missing this information. Besides, less distinct books appeared in the yearly top 1 table, which will make the query result within a relatively narrow scope. Combining the top 20 in a decade could provide the audience with a larger picture with more bestseller situations.

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\u0092s 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)

Figure 1a): The top 20 most popular books in Amazon from 2009 to 2019

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)

ji

Figure 1b): Yearly top 1 popular book in Amazon from 2009 to 2019. Note: it may have multiple books be top 1 popular due to the same user ratings and reviews numbers.

For evaluating how the book type affects the popularity, the comparison of numbers of fiction versus nonfiction from the above books was performed. As Figure 2 displayed, fiction remains more popular than non-fiction in both aspects (16 vs 4 among 20 books, 11 vs 3 among 14 bestsellings). Besides, combining the result from Figure 1, all top popular books have user ratings no less than 4.8. Thus, it could be preliminarily concluded that the fiction book with user ratings of no less than 4.8 is more likely to become a bestseller of the year according to the trend over the decade in Amazon.

```
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)
```

Figure 2: Fiction and non-fiction count of the top 20 most popular books and yearly top 1 popular book

Query 2: User rating for multiple-year-bestsellers versus single-year-bestsellers

Observed that some books had become bestsellers for more than one years (for instance Book No.105), therefore, all bestselling books can be divided into multiple-year-bestsellers (MYB) and single-year-bestsellers (SYB) . All MYBs beat at least one SYB in the aspect of user ratings. As Figure 3 showed, Query 2 first assessed the number of SYBs that each multiple-year-bestseller beat and corresponding occurrence frequency of defeated SYBs as winning percentage. Only seven of MYBs achieved a high winning percentage of 92%. For all MYBs, additional findings on user rating and review trends revealed no decrease in their ratings and reviews. Therefore, the results of Query 2 suggest that MYBs are not always more popular than SYBs, but MYBs always maintain non-decreasing user ratings and reviews.

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

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

csc343h-wangyuew=> select count(*) from HigherRatingOverview;
 count
-------
   106
(1 row)

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

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

Figure 3a): Summative counting information from Query 2

```
 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)
```

Figure 3b): Sample results of MYBs winning percentage

Query 3: Consistency in the book types of authors wrote multiple bestselling books

Query 3 investigated the consistency of book types of authors who had written more than one bestselling book in Figure 4. There are 53 such authors out of all 248, but only two authors had bestselling books in cross-genre types. Thus, it is rare that an author could have several bestsellers across genres. Most bestselling authors focus on one genre.

```
[csc343h-wangyuew=> select count(*) from author;
 count
-------
   248
(1 row)

[csc343h-wangyuew=> select count(*) from answer;
 count
-------
    53
(1 row)
```

Figure 4a): Summative counting information from Query 3

```
 aid | authorname | ficnum | nonficnum
-----+------------+--------+-----------
   4 | DK         |      1 |         1
  19 | Scholastic |      1 |         1
(2 rows)
```

Figure 4b): Query result in cross-genre types

For additional discovery of these cross-genre authors, Figure 5 demonstrated the user ratings and reviews of their books. Compared to the result of Query 1, these books are not having high user ratings and review numbers as those top popular books.

```
[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)
```

Figure 5: Additional query for Figure 4b) results for observing the user ratings and reviews

Conclusion:

Overall, Query 1 provides an overview of the criteria used to assess the popularity and the characteristics of the top bestsellers from the dataset. Most of the top bestselling books have a user rating of 4.8 or higher, and most of them are in the fiction genre. There exists the possibility that a bestselling book defeats others in consecutive years. However, there is no evidence showing that multiple-year-bestsellers have a more significant popularity advantage than single-year-bestsellers. Most bestselling authors focus on one genre of writing, fiction or non-fiction. For those authors who had multiple bestsellers in different genres, their books did not reach the same rating as the top bestsellers (4.8+).

As a result combining all queries, authors who have written multiple top bestsellers and focus on fiction genre have a better chance of becoming new bestsellers. From a bookseller's perspective, it may be more profitable to import books by such authors. As a reader, fiction books may be a good choice if the reader wants to read popular books.