Brianna Griffin

MAT 265

Frequency Pattern Mining

For this presentation, I am focusing on the first book and movie of The Hunger Games trilogy. I have found that checkouts of both the book and DVD copy of the movie have been popular at the Seattle Public Library over time. Due to this prevalence, I will be analyzing the popularity and trends in checkouts of each over time. I am also looking at the difference between tracking barcodes and itemNumber for the novel and movie during a set period of time in order to see if there are any discrepancies in the library's tracking and data categorization system. Finally, I want to look at the lives of the many copies of the Hunger Games books. I want to see when the first time it was checked out in comparison to the last time it was checked out, measuring the time in between and how long the physical copy has been used for.
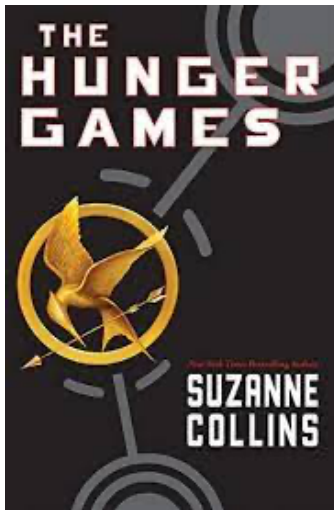
II.   Queries & Results

- First, I queried to see the total number of books and movies that the library has.
  - Query # 1
  - CSV file = number_of_books_and_movies_at_library_over_time

So, we see that there have been 719 total items of The Hunger Games book and movie at the SPL.

- Now, let's look at the popularity of the movie and the book and the movie over time. I will query for the number of checkouts for both the movie and the book per month. Following, I created a stacked bar chart to analyze the results and the relationship in trends between the book and movie of *The Hunger Games.*
  - Query # 2
  - CSV file = trend_over_time & hunger_games_stacked_bar_chart

The book became very popular, increasing in checkouts through 2012, but once the movie was available at the library, the number of checkouts for the movie exceeded that of the book. Once

the height of the popularity ended around 2017 the number of checkouts for the movie and the book became steadily equal.

- ● Trend in number of checkouts of *The Hunger Games* (Book):
    - ○ Query # 3
    - ○ CSV file = book_seperate & book_chart

The maximum of 420 monthly checkouts occurs in May of 2012 while the minimum of 2 monthly checkouts occurs in September 2022.

- ● Trend in the number of checkouts of "The Hunger Games" (Movie):
    - ○ Query # 4
    - ○ CSV file = movie_seperate & movie_chart

The maximum of 381 monthly checkouts occurs in April 2013 while the minimum of 2 monthly checkouts occurs in February and April of 2022.

- - Next, I wanted to track the number of checkouts by itemNumber and barcode and see if there were any discrepancies between the results. I will just look at a set period of time to be able to closely analyze any possible differences. I will choose to do October 2012 as the time period. Recall that ItemNumber is each item's unique identifier assigned upon arrival while the barcode number is the unique number to be scanned by the library.

- ● First I will look at *The Hunger Games* the book:
    - ○ Query # 5 (book by itemNumber) & Query # 6 (book by barcode)
    - ○ CSV file = item_number_oct2022, barcode_oct2022, item_book_chart

From the table and graphical outputs it is seen that tracking by barcode and itemNumber achieves the same numerical results with the data. There are no known discrepancies between the two.

- ● We see the same results with "The Hunger Games" the movie:
    - ○ Query # 7 (movie by itemNumber) & Query # 8 (movie by itemNumber)
    - ○ CSV file = item_number_movie, barcode_movie, item_movie_chart

Next, let's look at the main topic which is the life of *The Hunger Games* books at the Seattle Public library.

- - There have been 482 distinct copies of the book at SPL since 2012.

- ● Next, I will query for the number of checkouts per copy.
    - ○ Query # 9
    - ○ CSV file = num_COs_per_copy

From this query we see that the maximum number of checkouts for a copy of the book is 92 while 5 books have only been checked out only once. There could possibly be copies with 0 checkouts, but we do not know about these due to the nature of the data set.

- I will now search for the first and last time each copy of the book was checked out.
  - Query # 10
  - CSV file = first_and_last_book, first_and_last_viz, first_and_last_viz2

From the first visualization it is clear that books that were first checked out in 2009-2010 generally had a large difference between their first checkout date and last checkout date per physical copy. Then, as the years go by there are a larger frequency of books with a small number of months between first and last checkout. It can also be seen that the number of months in between starts to increase again in recent years.

- Now, I want to look closer at copies that are left on the shelves for a long period of time. Let's look at books obtained in 2012 when the book was most popular and query for copies where the difference between the last checkout and the first checkout was less than a year. This means that they were only used for a year and then proceeded to stay on the shelves after that.
  - Query # 11
  - CSV file = 2012_books_under1year

We see that there are 136 of them! There is even one copy that has a date difference of 0 meaning that it was only checked out once: could be never returned to the library or just a book on the shelves for a very long period of time. It makes sense that so many of the books that the library obtained due to *The Hunger Games' increasing* popularity went unused after a year due to its phasing out of popularity.

- Now, let's take a look at some of these books in more detail. I will choose 5 of them randomly. These are 5 books that were obtained by the SPL in 2012 and only used for less than a year.
  - NOTES: I replaced the itemNumbers with letters A,B,C,D and E by a series of case-when statements for simplicity sake.
  - Query #12
  - CSV file = 2012_books_under1year_detail

From the CSV output table it can be noted that:
- ➔ Copy A:
  - ◆ First Checked Out: **1/17/2012**
  - ◆ Last Returned: **10/20/2012**
  - ◆ Number of Times Checked out: **7**
- ➔ Copy B:
  - ◆ First Checked Out: **2/21/2012**

- ◆ Last Returned: **11/28/2012**
- ◆ Number of Times Checked out: **10**
- ➔ Copy C:
    - ◆ First Checked Out: **2/21/2012**
    - ◆ Last Returned: **8/8/2012**
    - ◆ Number of Times Checked out: **7**
- ➔ Copy D:
    - ◆ First Checked Out: **1/3/2012**
    - ◆ Last Returned: **11/21/2012**
    - ◆ Number of Times Checked out: **6**
- ➔ Copy E:
    - ◆ First Checked Out: **2/1/2012**
    - ◆ Last Returned: **12/1/2012**
    - ◆ Number of Times Checked out: **8**

Each of the 5 copies were checked out a decent amount of times within the year that they were being used, but after that they remained on the shelf.

- ● Next, I wanted to see if the library slowed down on getting new copies of the book in the coming years after 2012 given that there were so many (136) that were only checked out during a time period of less than a year. So, I queried for the amount of new copies per year of *The Hunger Games* book.
    - ○ Query # 13
    - ○ CSV file: new_copies_per_year_bk

There were 219 new copies obtained in 2012 and 0 in 2013. It makes sense that there were 0 new copies of *The Hunger Games* in 2013 since it seems that the library had a lot of extra copies on their shelves from the 2012 hype around the trilogy series.

- ● Now I am interested in selecting the date of the first checkout (i.e. the date near which the library obtained the copy) for the copies that have the longest shelf life. This means that the difference between the last checkout and first checkout is the greatest. They have been at SPL for the longest.
    - ○ NOTE: in the output table the column date_diff is measured in days
    - ○ Query # 14
    - ○ CSV file: dates_longest_used

The longest used copy has been used going on 10 years at the SPL! Most of the books on the table are books that were first checked out in 2012.

- ● I will do the same thing now, but looking at books with the shortest shelf life. This means that the difference between their first checkout date and last checkout date is the smallest amongst all of the copies of *The Hunger Games*.

- Once again date_diff is measured in days
- Query # 15
- CSV file: dates_shortest_used

5 copies were only checked out once. Otherwise, the other copies listed were only in use for around a month or less than a month. These instances happened in a variety of years and it wasn't constant throughout the data points.

III. <mark>Conclusion & References</mark>

    A. Given the time period that the book and movie of the Hunger Games was released, it makes sense that it was popular. The dystopian fiction novel touches on themes such as the effect of technology, economic inequality, and strict government reign. This was during the time of the economic recession, threat of terrorism, and the rise of technology that faced society. Many turned to the novel and the book as a way to resonate with issues they were facing in a fictional sense.

    B. After this hype of the book and movie there has since been a decline in its number of checkouts and popularity. This could be due to many readers and watchers solely interested in the outer plot line (i.e. the love story) and not the literature and thematics for what they truly are. One of the main themes that Suzanne Collins writes about is the main character, a female lead, who is able to overtake and overpower strong government forces. While this is going on, there is also a love story going on involving other main characters. Once viewers and readers know what happens with the love story they are unlikely to want to watch or read the story again.

    C. In conclusion, my analysis gave a lot of insight on how the shelves of a library work. It was interesting to look at this for a book in which the hype was mostly during a 2 year period. We saw that there was an increase in the amount of copies of books and checkouts during this period. After, a lot of these copies remained on the shelves of the library or have even since left.

- https://videolibrarian.com/articles/essays/looking-back-on-the-hunger-games/
- https://www.forbes.com/sites/scottmendelson/2022/03/24/why-the-hunger-games-vanished-from-the-pop-culture-conversation/?sh=334eb28e558b