

MAT259 Project 1 Report  
**MySQL Query & Knowledge Discovery**  
Lijuan Cheng  
Winter Quarter 2022

Last year, I watched two film adaptations of the famous novels , “Little Women” and “Dune”, which have the same leading actor “Timothée Chalamet”. So I am very interested in the correlation between the movie releasing activity and the check out behavior of the original novel. Hence, I designed the first query to search the check out numbers of these two novels from 2006 to 2021 to observe the variation trend. Furthermore, I also noticed that the population diversity of the Seattle metropolitan area keeps growing in the last decade. Especially more and more Asian people move to Seattle due to the increasing job opportunities in tech companies. Therefore, the second query in my project is to investigate the reading interest of history knowledge about different continents, which can help us to understand the correlation between reading behavior and population composition.

1. Research Question: What is the impact of film adaptation on original novel borrowing? Does it relate to the film genre?

As we all know, the film adaptation will affect the reading behavior of the original novel. But I want to further study if the different film genres will moderate this effect. Hence, I chose two latest films with various genres to do this research.

Little Women premiered at the Museum of Modern Art in New York City on December 7, 2019, and was released theatrically in the United States on December 25, 2019, by Sony Pictures Releasing. The film received critical acclaim with particular praise for Gerwig's screenplay and direction as well as the performances of the cast, and grossed \$218 million worldwide[1].

Dune (titled onscreen as Dune: Part One) is a 2021 American epic science fiction film and was originally scheduled for a late 2020 release, but it was delayed by the COVID-19 pandemic. The film premiered at the 78th Venice International Film Festival on September 3, 2021 and then was released in United States theaters and streaming on HBO Max on October 21, 2021. It received critical acclaim for its direction, visuals, scope, and ambition, and has grossed \$397 million worldwide[2].

Table 1 shows the important film details of these two movies to better understand the difference between them.

**Table 1: The important film details of “Little Women” and “Dune”**

Film	Release Date	Box Office	Film Genre	Version
Little Women	December 7, 2019	\$218 million	Coming-of-age period drama film	Seventh adaptation
Dune	September 3, 2021	\$397 million	Epic science fiction film	Third adaptation

### 1.1 MySQL Query 1

In this query, I chose the year and yearly check out counting of “Little Women” and “Dune” as the output columns. And I grouped and ordered the results by ascending year. The query took 64 seconds in total and the codes are as follows.

```
//Execution time : 64 seconds
SELECT
  YEAR(cout) AS Years,
  COUNT(IF(LOWER(title) = 'little women',
    1,
    NULL)) AS 'Little Women',
  COUNT(IF(LOWER(title) = 'dune', 1, NULL)) AS 'Dune'
FROM
  spl_2016.outraw
WHERE
  YEAR(cout) < 2022
GROUP BY Years
ORDER BY Years
```

1.2 Data results:

Table 2 shows the search results of query 1.

**Table 2: The yearly aggregate check out of “Little Women” and “Dune”**

Years	Little Women	Dune
2006	654	425
2007	541	404
2008	669	<b>463</b>
2009	687	436
2010	600	341
2011	513	275
2012	466	359
2013	429	420
2014	482	392
2015	412	407
2016	364	293
2017	290	386
2018	712	231
2019	1567	327
2020	1250	132
2021	1201	336

1.3 Analysis of results:

As you can see from table 2, the greatest check out of “Little Women” was 712 in 2018, but that of “Dune” was 463 in 2008. And you can see the dramatic increase of “Little Women” in 2019, the check out was suddenly more than double to 1567. Even in 2020, although the Covid-19 caused the check out of most library materials to decline, the “Little Women” still remained the high value at 1250. So the film adaptation definitely improved the reading behavior of “Little Women”. And this effect lasted for three years until the end of 2021.

On the other hand, you can't see the same trend in the reading habit of “Dune”. Though the film was released in September 2021, actually the filming took place from March to July 2019 and originally scheduled for a late 2020 release. So in 2019 and 2020, there should be a lot of news and promotional campaigns. However, there was no significant increase in the check out of “Dune” from 2019 to 2021, and the lowest number 132 in 2020 must be attributed to the Covid-19. Despite the check out in 2021 raising to 336, it is still lower than the greatest number 463 in 2008. Therefore, the film adaptation of “Dune” also facilitated the reading behavior of the original novel to some extent, but not so significantly as the effect of “Little Women”.

In view of the two movies are very popular and obtain a lot of attention from the public, but the effects on the reading behavior of the original novels are very different. If we analyze the box office and the popularity of online discussion, maybe “Dune” is much better than “Little Women”. But the variation trend of check out in the Seattle public library is on the contrary of the popularity tendency. Hence, I guess this will be mainly attributed to the difference of film genres and novel types. The novel of “Little Women” can almost cover all of the age groups and have the everlasting influence of various readers. In contrast, the novel “Dune” is science fiction which is loved by a small number of readers. Even if a large amount of audiences are fascinated by the wonderful movie, they won't read the original novel owing to the obscurity of science fiction.

2. Research Question: How does population demography affect the reading interest of history of different continents?

Seattle has historically had a mostly white population. While Seattle's percentage of white residents is lower than the United States as a whole and declining, it is still one of the whitest large cities in the U.S. From 1960 to 2010, the percentage of whites in Seattle has dropped from 91.6% to 66.5%. According to an American Community Survey, people who speak Asian languages at home account for 10% of the population, followed by Spanish at 4.5%. Seattle's foreign-born population has increased by 40% in the last ten years. Of the Asian population, 4.1% are Chinese with origins in mainland China, Hong Kong, and Taiwan[3].

According to the most recent ACS, the racial composition of Seattle was[3]:

**Table 3: Seattle Population by Race**

Race	Population	Percentage
White	487486	67.30%
Asian	111795	15.43%
Black or African American	53112	7.33%
Two or More Races	49680	6.86%

<b>Some Other Race</b>	16628	2.30%
<b>American Indian and Alaska Native</b>	3687	0.51%
<b>Native Hawaiian and Other Pacific Islander</b>	1917	0.26%

Hence, I want to know if the population demography will affect the reading behavior of history of the residents' origins.

### 2.1 MySQL Query 2:

In this query, I chose the year and the check out numbers of history of five main continents as the output columns. And I grouped and ordered the results by ascending year. The query took 28 seconds in total and the codes are as follows.

//Execution time: 28 seconds

```

SELECT
    YEAR(cout) AS Years,
    COUNT(IF(deweyClass >= 940 AND deweyClass <= 949, 1, NULL)) AS 'History of
Europe',
    COUNT(IF(deweyClass >= 950 AND deweyClass <= 959, 1, NULL)) AS 'History of Asia',
    COUNT(IF(deweyClass >= 960 AND deweyClass <= 969, 1, NULL)) AS 'History of
Africa',
    COUNT(IF(deweyClass >= 970 AND deweyClass <= 979, 1, NULL)) AS 'History of North
America',
    COUNT(IF(deweyClass >= 980 AND deweyClass <= 989, 1, NULL)) AS 'History of
South America'
FROM
    spl_2016.outraw
WHERE
    deweyClass > 900 AND YEAR(cout) < 2022
GROUP BY Years
ORDER BY Years

```

### 2.2 Data Results :

Table 4 shows the search results of query 2.

**Table 4: The yearly check out of history of five main continents**

<b>Years</b>	<b>History of Europe</b>	<b>History of Asia</b>	<b>History of Africa</b>	<b>History of North America</b>	<b>History of South America</b>
<b>2006</b>	35231	22081	5118	53346	2400
<b>2007</b>	34114	21778	5926	51004	2168
<b>2008</b>	42054	28684	8233	66232	2743
<b>2009</b>	42350	27881	7181	67883	2754
<b>2010</b>	40896	24209	6256	56953	2793

2011	40774	21605	6007	52581	3016
2012	40135	18506	5638	49416	2571
2013	41935	18379	5044	51232	2672
2014	42043	15775	4133	46773	2178
2015	39842	14138	3597	42870	2005
2016	36789	13338	3339	40553	1868
2017	33548	11565	3188	39854	1470
2018	20903	8028	2111	32143	908
2019	29166	10509	2473	47664	1238
2020	8960	2950	693	15079	362
2021	15163	5377	1079	23017	485

### 2.3 Analysis of results:

As you can see from table 4, the ranking of check out numbers every year always keeps the same: history of North America, history of Europe, history of Asia, history of Africa and history of South America. The results are as expected, consistent with population proportions. In 2020, there was a dramatic decrease at the check out numbers compared with that of 2019. We can know that the Covid-19 indeed affected reading behavior significantly. Besides, we also can find that there was a continuous downward trend in the check out of the history field from 2011 to 2021. I guess the reason can be attributed to the development of electronic books, as it mentioned in the project description: "Electronic books have been in the collection since 2009 but these are not recorded in the database we receive".

### References:

[1] [https://en.wikipedia.org/wiki/Little\\_Women\\_\(2019\\_film\)](https://en.wikipedia.org/wiki/Little_Women_(2019_film))

[2] [https://en.wikipedia.org/wiki/Dune\\_\(2021\\_film\)](https://en.wikipedia.org/wiki/Dune_(2021_film))

[3] <https://worldpopulationreview.com/us-cities/seattle-wa-population>