

Ming-Yu Lin

Assignment 1: Harry Potter Series

Introduction

The Harry Potter series is a popular series in 2000s. The book series started from 1997 and ended in 2007. The film series started from 2001 and ended in 2017. I am interested in the relationship between the numbers of check-out of book version between the numbers checkout of movie version. In specific, I want to know if the book will be checked out more times when a new movie is released or a DVD version is released. As a result, I just consider movies which were released after 2006.

Data retrieval

The first step was to check if there is any title including "Harry Potter".

```
SELECT
title, COUNT(title) AS COUNTS
FROM
spl_2016.inraw
WHERE
title LIKE "%Harry potter%"
GROUP BY title
ORDER BY COUNTS DESC;
```

10 17:34:09 SELECT title, COUNT(title) AS COUNTS FROM spl_2016.inraw WHERE title LIKE ... 132 row(s) returned

title	COUNTS
Harry Potter and the half blood prince	20194
Harry Potter and the Order of the Phoenix	18700
Harry Potter and the sorcerers stone	17727
Harry Potter and the ooblet of fire	17179
Harry Potter and the chamber of secrets	15775
Harry Potter and the prisoner of Azkaban	15206
Harry Potter and the deathly hallows	11994
Harry Potter and the Deathly Hallows Part 1	5847
Harry Potter and the deathly hallows Part 2	4114
Harry Potter and the cursed child Parts one and...	1180
Harry Potter and the sorcerers stone original m...	870
Harry Potter and the ooblet of fire original moti...	831
Harry Potter and the Order of the Phoenix origi...	788
Harry Potter and the chamber of secrets origina...	665
Harry Potter and the prisoner of Azkaban origin...	614

From the result of first query, I noticed that there were 132 different titles which mostly were different languages version and some soundtracks. Since most of records are English version, I just picked up specific titles in English version and distinguish them into book version and film version.

```

SELECT
bibNumber, title, itemtype, COUNT(bibNumber) AS COUNTS
FROM
spl_2016.inraw
WHERE
( title LIKE "Harry Potter and the Order of the Phoenix%" OR
title LIKE "Harry Potter and the half blood prince%" OR
title LIKE "Harry Potter and the deathly hallows%" ) AND
( itemtype LIKE "%dvd" OR itemtype LIKE "%bk" ) AND
title NOT LIKE "%soundtrack"
GROUP BY bibNumber, title, itemtype
ORDER BY title;

```

bibNumber	title	itemtype	COUNTS
2410427	Harry Potter and the deathly hallows	icbk	6323
2446851	Harry Potter and the deathly hallows	icbk	1144
2703893	Harry Potter and the Deathly Hallows Part 1	acdvd	5847
2751501	Harry Potter and the deathly hallows Part 2	acdvd	3643
3211483	Harry Potter and the deathly hallows Part 2	acdvd	471
2277368	Harry Potter and the half blood prince	icbk	5831
2613376	Harry Potter and the half blood prince	acdvd	8104
2304178	Harry Potter and the half blood prince	icbk	978
2176667	Harry Potter and the Order of the Phoenix	icbk	4929
2205013	Harry Potter and the Order of the Phoenix	icbk	462
2457339	Harry Potter and the Order of the Phoenix	acdvd	9356

From this query, I found that there would be two book copies and one dvd version for each title.

```

SELECT
title, itemtype, COUNT(title) AS COUNTS
FROM
spl_2016.inraw
WHERE
( title LIKE "Harry Potter and the Order of the Phoenix%" OR
title LIKE "Harry Potter and the half blood prince%" OR
title LIKE "Harry Potter and the deathly hallows%" ) AND
( itemtype LIKE "%dvd" OR itemtype LIKE "%bk" ) AND
title NOT LIKE "%soundtrack"
GROUP BY title, itemtype
ORDER BY title;

```

title	itemtype	COUNTS
Harry Potter and the deathly hallows	icbk	7467
Harry Potter and the Deathly Hallows Part 1	acdvd	5847
Harry Potter and the deathly hallows Part 2	acdvd	4114
Harry Potter and the half blood prince	acdvd	8104
Harry Potter and the half blood prince	icbk	6809
Harry Potter and the Order of the Phoenix	acdvd	9356
Harry Potter and the Order of the Phoenix	icbk	5391

From the above result, I correctly filter out results corresponded with my requirements which were English version books and DVDs. The final step was to check the numbers of checkout from 2006 to 2016. To make the result succinct, I would make the query based on Year.

```

SELECT
title, itemtype, COUNT(title) AS COUNTS,
SUM(CASE
WHEN (YEAR(cout) = 2005) THEN 1
ELSE 0
END) AS '2005',
SUM(CASE
WHEN (YEAR(cout) = 2006) THEN 1
ELSE 0
END) AS '2006',
SUM(CASE
WHEN (YEAR(cout) = 2007) THEN 1
ELSE 0
END) AS '2007',
SUM(CASE
WHEN (YEAR(cout) = 2008) THEN 1
ELSE 0
END) AS '2008',
SUM(CASE
WHEN (YEAR(cout) = 2009) THEN 1
ELSE 0
END) AS '2009',SUM(CASE
WHEN (YEAR(cout) = 2010) THEN 1
ELSE 0
END) AS '2010',
SUM(CASE
WHEN (YEAR(cout) = 2011) THEN 1
ELSE 0

```

```

END) AS '2011',
SUM(CASE
WHEN (YEAR(cout) = 2012) THEN 1
ELSE 0
END) AS '2012',
SUM(CASE
WHEN (YEAR(cout) = 2013) THEN 1
ELSE 0
END) AS '2013',
SUM(CASE
WHEN (YEAR(cout) = 2014) THEN 1
ELSE 0
END) AS '2014',
SUM(CASE
WHEN (YEAR(cout) = 2015) THEN 1
ELSE 0
END) AS '2015'
FROM
spl_2016.inraw
WHERE
( title LIKE "Harry Potter and the Order of the Phoenix%" OR
title LIKE "Harry Potter and the half blood prince%" OR
title LIKE "Harry Potter and the deathly hallows%" ) AND
( itemtype LIKE "%dvd" OR itemtype LIKE "%bk" ) AND
title NOT LIKE "%soundtrack"
GROUP BY title, itemtype
ORDER BY title;

```

title	itemtype	COUNTS	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Harrv Potter and the deathlv hallows	icbk	7467	0	0	1716	1094	1101	951	821	281	275	217	236
Harrv Potter and the Deathlv Hallows Part 1	acdvd	5847	0	0	0	0	0	0	2015	1472	837	444	275
Harrv Potter and the deathlv hallows Part 2	acdvd	4114	0	0	0	0	0	0	70	1851	906	454	237
Harrv Potter and the half blood prince	acdvd	8104	0	0	0	0	361	3179	1651	903	652	366	190
Harrv Potter and the half blood prince	icbk	6809	132	1065	1208	707	947	624	598	290	259	208	199
Harrv Potter and the Order of the Phoenix	acdvd	9356	0	0	191	2966	1813	1311	964	680	333	160	100
Harrv Potter and the Order of the Phoenix	icbk	5391	58	588	832	468	674	660	568	241	247	221	200

In conclusion, I found that DVDs were purchased and numbers of checkout of books would increase in the year when movie was released. Also, the numbers of checkout of book version would decrease badly and numbers of checkout of DVD version would increase on the very next year.