

I'm interested in sci fi movies. So I'd like to do a study on the movie: "Star Trek Into Darkness", which is published in 2013.

First, I'd like to know how many copies are purchased by the library. First, to understand what the dataset looks like, I made the following query:

```
Select *
from spl_2016.inraw
where title='Star Trek Into Darkness'
```

The result is the following (the query takes ~1 seconds):

#	id	itemNumber	bibNumber	cout	cin	collcode	itemtype	barcode	title	callNumber	deweyClass	subj
1	65237646	4920932	2928226	2013-09-12 13:46:00	2013-09-13 17:09:00	cadvd	acdvd	0010079420179	Star trek Into darkness	DVD STAR TR		NULL
2	65241808	4920951	2928226	2013-09-12 19:27:00	2013-09-14 12:02:00	nadvd	acdvd	0010079419726	Star trek Into darkness	DVD STAR TR		NULL
3	65242458	4921013	2928226	2013-09-13 13:10:00	2013-09-14 12:13:00	nadvd	acdvd	0010079420898	Star trek Into darkness	DVD STAR TR		NULL
4	65249064	4921021	2928226	2013-09-13 17:40:00	2013-09-14 14:44:00	nadvd	acdvd	0010079420716	Star trek Into darkness	DVD STAR TR		NULL
5	65252460	4920957	2928226	2013-09-12 14:57:00	2013-09-14 15:51:00	nadvd	acdvd	0010079419783	Star trek Into darkness	DVD STAR TR		NULL
6	65260001	4920992	2928226	2013-09-14 11:42:00	2013-09-15 13:39:00	nadvd	acdvd	0010079420633	Star trek Into darkness	DVD STAR TR		NULL

So the dataset spl\_2016.inraw is a big chart which shows the check-in and check-out record of each item in the library. It has 12 columns, cout stands for check-out time, and cin stands for check-time. As we can see cin is always later than cout, so how long each items are borrowed from one record can be computed by cin - cout.

Next I'd like to know how many copies are purchased by the library..

```
SELECT title, count(title) As total
from (
    SELECT title, itemNumber,itemType, COUNT(itemNumber) AS Counts
    from spl_2016.outraw
    where title = 'Star Trek Into Darkness'
    group by itemNumber, itemType
) sub
where title = 'Star Trek Into Darkness'
group by title
```

The result is : (the query less than 1 seconds)

#	title	total
1	Star trek Into darkness	179

It appears the library bought total 179 copies of “Star Trek Into Darkness”. The sub query shows how many times each copies are borrowed (The item Number is unique, and the library can have multiple copies of same movie).

The barcode of should also be unique for each item. This can be confirmed by changing itemNumber in the above query into barCode. The query returns:

#	title	total
1	Star trek Into darkness	179

As we can see the number 179 is same as the query using itemNumber. So we can conclude the barCode should also be unique. Otherwise the query made with barCode should be smaller than the query made with itemNumber.

To see if there are new versions of the item have been introduced, the following queries can be ran:

```
SELECT title, bibNumber,itemType, COUNT(bibNumber) AS Counts
from spl_2016.inraw
where title = 'Star Trek Into Darkness'
group by bibNumber, itemType
```

The result is (the query takes less than 1 seconds):

Result Grid   Filter Rows: 					Export:  Wrap Cell Content: 	
#	title	bibNumber	itemType	Counts		
1	Star trek Into darkness	2928226	acdvd	4768		

So there is only one version of the movie “Star Trek Into Darkness”. And the movie has been borrowed total 4768 times.

Besides this movie, several other sci-fi movies are released around same time, and I’d like to see if there are correlations between how often they are borrowed. I select other three movies: Cloud Atlas, gravity and 'prometheus. I divide the time into months, and retrieve how many times the movie is borrowed in one month.

The query is the following:

```
SELECT YEAR(cout) AS Year, MONTH(cout) AS Month,
SUM(CASE
```

WHEN title = 'Star Trek Into Darkness' Then 1

ELSE 0 END) AS 'Star Trek Into Darkness',

SUM(CASE

WHEN title = 'Cloud Atlas' Then 1

ELSE 0 END) AS 'Cloud Atlas',

SUM(CASE

WHEN title = 'gravity' Then 1

ELSE 0 END) AS 'gravity',

SUM(CASE

WHEN title = 'prometheus' Then 1

ELSE 0 END) AS 'prometheus'

FROM spl\_2016.inraw

WHERE

(itemtype = 'acdv' OR itemtype = 'acvhs')

AND YEAR(cout) >= '2006'

GROUP BY MONTH(cout), YEAR(cout)

ORDER BY YEAR(cout) , MONTH(cout)

The results is (takes about 20 seconds.)

#	Year	Month	Star Trek Into Darkness	Cloud Atlas	gravity	prometheus
3	2012	11	0	0	0	208
4	2012	12	0	0	0	186
5	2013	1	0	0	0	212
6	2013	2	0	0	0	186
7	2013	3	0	0	0	200
8	2013	4	0	0	0	208
9	2013	5	0	150	0	200
0	2013	6	0	221	0	214
1	2013	7	0	200	0	211
2	2013	8	0	217	0	184
3	2013	9	184	210	0	157
4	2013	10	252	216	0	170
5	2013	11	238	211	0	158
6	2013	12	249	188	0	125
7	2014	1	250	173	0	137
8	2014	2	226	186	122	103
9	2014	3	240	177	333	120
0	2014	4	208	162	323	102
1	2014	5	183	97	318	104
2	2014	6	216	98	303	71
3	2014	7	189	87	279	72
4	2014	8	160	88	258	56
5	2014	9	137	102	272	54
6	2014	10	134	77	271	60
7	2014	11	127	75	235	45
8	2014	12	118	92	228	55
9	2015	1	112	86	199	42

As we can see all four movies are borrowed from 2013.

The graph which shows the trend of these four movies:

