MySQL Basic Queries Examples to Explore the Seattle Library Metadata
https://www.mat.ucsb.edu/~g.legrady/academic/courses/23w259/23w259.html

---
ID - The numeric ID of each entry in the database – increments by 1

```
// Returns  first 100 entries in the database based on checkouts
SELECT * from spl_2016.outraw order by id limit 100

// Returns  most recent 100 entries in the database
SELECT * from spl_2016.outraw order by id DESC limit 100

// looks at number value in title
SELECT id, itemtype, title from spl_2016.outraw where title > 0 limit 100
```

ItemNumber – The assigned value to an item when it enters the Seattle Database – is incremental but there are gaps in the system

```
//ItemNumbers titles in the database
SELECT DISTINCT
   (itemNumber), title
FROM
   spl_2016.outraw
   Order by ItemNumber
LIMIT 200

// ItemNumbers that have values less then 100
SELECT DISTINCT
   (itemNumber), title
FROM
   spl_2016.inraw
   where itemNumber < 100
limit 100

//Which items have the most number of copies -
select
count(itemNumber) as copies, bibNumber
from spl_2016.inraw
group by bibNumber
order by copies desc;

//First occurrence of each itemNumber
select itemNumber, min(cout) as first_cout
from spl_2016.outraw
group by itemNumber;
```

BibNumber - Assigned by the Library of Congress

```
// Get the titles, itemtypes for these items // try without itemtypes
SELECT DISTINCT
```

```
   (title), bibNumber, itemtype
FROM
   spl_2016.inraw
WHERE
   bibNumber IN ('3030520' , '3489506', '2469502')

// Subjects (keywords) from a specific item
SELECT * FROM spl_2016.subject where spl_2016.subject.bibnumber = 3428193

//Will return 4 vertical columns of various bibnumbers and the total count for each
SELECT bibNumber, itemType, title, COUNT(bibNumber) AS Counts FROM spl_2016.inraw WHERE title
LIKE '%Blade Runner%' GROUP BY bibNumber , itemType , title ORDER BY Counts DESC
```

Cout – Short for checkout , time and date when an item has been recorded to have been checkedout

```
//Get how long items with “Batman” have been checked out
SELECT
   cout, cin, title, TIMESTAMPDIFF(HOUR, cout, cin) / 24.0
FROM
   spl_2016.inraw
WHERE
   (MONTH(cout) > 1)
      AND title LIKE '%batman%';

// Get checkouts between January and April, 2007
SELECT
   *
FROM
   spl_2016.outraw
WHERE
   YEAR(cout) = 2007 AND MONTH(cout) >= 1
      AND MONTH(cout) <= 4;
```

Cin – The timestamp for the return of the item into the Seattle library database

```
//Longest borrowed
SELECT
title, itemtype,
DATEDIFF(cin, cout) AS DateDiff, cin, cout
FROM spl_2016.inraw
WHERE cout > '2005-01-01'
order by DateDiff desc
limit 30;

//shortest by default
SELECT
title,
DATEDIFF(cin, cout) AS DateDiff, cin, cout
FROM spl_2016.inraw
WHERE cout > '2005-01-01'
order by DateDiff
limit 10;
```

CollCode – Collection Code, a combination of item location and what it is:
https://data.seattle.gov/Community/Integrated-Library-System-ILS-Data-Dictionary/pbt3-ytbc

```
// Only select interactions from the Central Library which has the code cen in its collcode
SELECT
   *
FROM
   spl_2016.outraw
WHERE
   YEAR(cout) = 2007 AND MONTH(cout) >= 1
      AND MONTH(cout) <= 4;
      where collcode like '%cen%'
limit 10;
```

ItemType – What is the media of the item checkedout or checkedin

```
//how many itemtypes are there
SELECT DISTINCT
itemType
FROM
spl_2016.itemType
LIMIT 1000

// any itemtype that has bk at the end of its description
SELECT distinct title, itemtype FROM spl_2016.outraw WHERE itemtype like '%bk' limit 100
```

Barcode – Unique RF tag on each item for tracking purposes

```
//Get the different barcodes for Mary Poppins
SELECT distinct
   barcode, title
FROM
   spl_2016.inraw
   where title LIKE '%Mary Poppins%'
LIMIT 300
```

Title – The title of the item

```
//What is the activity of specific keyphrases in titles
SELECT
   YEAR(cout) AS years, COUNT(*) AS Activity
FROM
   spl_2016.inraw
WHERE
   title LIKE '%data science%'
GROUP BY year(cout)
LIMIT 20

//get titles with ‘happiness’ on a day with checkout timestamp that was out longer then

SELECT itemNumber, timestamp(cout), barcode, title, itemtype, TIMESTAMPDIFF(DAY, cout, cin) FROM
spl_2016.inraw WHERE DATE(cout) = '2008-02-20' AND title LIKE '%happiness%' AND
TIMESTAMPDIFF(DAY, cout, cin) > 30 ORDER BY TIMESTAMPDIFF(DAY, cout, cin) ASC;
```

---

callNumber – The assigned number for the location on the shelves, also includes Dewey code

```
// select dvds that do not have Dewey classification numbers
SELECT
   *
FROM
   spl_2016.inraw
   where deweyClass = '' and callNumber LIKE '%DVD%'
LIMIT 30
```

DeweyClass – The Dewey numerical classification for the item if its in the Dewey system

```
// how many dewey, non-dewey circulating per year since 2016
//when in case - https://www.w3schools.com/sql/sql_case.asp

SELECT Year(cout) AS year, SUM(CASE WHEN deweyClass = "" THEN 1 ELSE 0 END) AS nonDewey,
SUM(CASE WHEN deweyClass != "" THEN 1 ELSE 0 END) AS Dewey FROM spl_2016.inraw WHERE
YEAR(cout) >= 2016 GROUP BY year ORDER BY year
```

Subj – The tags assigned to each item when the item first enters the library database. This metadata is only available in the spl_2016.subject table and requires a join function if it is to be used with either inraw or outraw

```
//brings up all items that the subject has classified as ‘love’
SELECT * FROM spl_2016.subject WHERE spl_2016.subject.subject = ‘love’;

// get the number of checkouts where one of the tags is “love”
SELECT count(cout) FROM spl_2016.outraw, spl_2016.subject WHERE spl_2016.outraw.bibnumber =
spl_2016.subject.bibnumber AND spl_2016.subject.subject = 'love'
```

Additional MySQL Commands

Date

```
// Count how many on a specific day
SELECT count(cout) FROM spl_2016.outraw where DATE(cout) = '2019-12-09'

//get titles with the word happiness from a specific date SELECT * FROM spl_2016.outraw WHERE
DATE(cout) = '2012-08-16' AND title LIKE '%happiness%' LIMIT 10
```

Time

```
// how many hours dvds checked-out & returned on a specific date, 43 sec //substring removes (3) chars
from left
SELECT DISTINCT title, date(cout), TIMESTAMPDIFF(HOUR, cout, cin) / 24.0 AS duration FROM
spl_2016.inraw WHERE itemType = SUBSTRING(itemType, 3) = 'dvd' AND date(cin) = '2019-12-10'
ORDER by duration LIMIT 20000
```

Join – Do the search in two or more tables

```
//Returns titles and their  subject tags from inRaw and subject tables.
SELECT distinct
 (itemNumber), title, subject
FROM
 spl_2016.inraw,
 spl_2016.subject
WHERE
 spl_2016.inraw.bibnumber = spl_2016.subject.bibnumber
 AND spl_2016.inraw.title LIKE '%happy%'
LIMIT 100
```

```
// Filter out dates before 2005
SELECT checkOut, checkIn, title FROM spl_2016.transactions, spl_2016.title WHERE
spl_2016.transactions.bibnumber = spl_2016.title.bibnumber AND date(checkOut) > '2005' order by
checkOut limit 100
```

```
//Inner join between multiple tables
SELECT i.itemNumber, i.bibNumber, it.itemType, MIN(ir.cout), ir.title, b.deweyClass FROM
spl_2016.itemToBib AS i INNER JOIN spl_2016.itemType AS it ON i.itemNumber = it.itemNumber INNER
JOIN spl_2016.inraw AS ir ON i.itemNumber=ir.itemNumber INNER JOIN spl_2016.deweyClass AS b ON
i.bibNumber = b.bibNumber WHERE b.deweyClass > 0 GROUP BY i.itemNumber, i.bibNumber,
it.itemType, ir.title, b.deweyClass ORDER BY i.itemNumber
```

Count – example of using Count

```
// how many items checked out on two days using Count + if condition

SELECT YEAR(cout) AS years, COUNT(IF(DATE(cout) = '2018-12-09', 1, NULL)) AS '2018',
COUNT(IF(DATE(cout) = '2019-12-09', 1, NULL)) AS '2019' FROM spl_2016.outraw GROUP BY
year(cout);
```

Sum and **Case** - Aggregates according to consecutive years

```
//Number of cds checked out in each Dewey category between 2006 and 2019 – 87.391 sec SELECT
YEAR(cout) AS Years, SUM(CASE WHEN deweyClass > 000 AND deweyClass < 100 THEN 1 ELSE 0
END) AS D000_099, SUM(CASE WHEN deweyClass > 100 AND deweyClass < 200 THEN 1 ELSE 0
END) AS D100_199, SUM(CASE WHEN deweyClass > 200 AND deweyClass < 300 THEN 1 ELSE 0
END) AS D200_299, SUM(CASE WHEN deweyClass > 300 AND deweyClass < 400 THEN 1 ELSE 0
END) AS D300_399, SUM(CASE WHEN deweyClass > 400 AND deweyClass < 500 THEN 1 ELSE 0
END) AS D400_499, SUM(CASE WHEN deweyClass > 500 AND deweyClass < 600 THEN 1 ELSE 0
END) AS D500_599, SUM(CASE WHEN deweyClass > 600 AND deweyClass < 700 THEN 1 ELSE 0
END) AS D600_699, SUM(CASE WHEN deweyClass > 700 AND deweyClass < 800 THEN 1 ELSE 0
END) AS D700_799, SUM(CASE WHEN deweyClass > 800 AND deweyClass < 900 THEN 1 ELSE 0
END) AS D800_899, SUM(CASE WHEN deweyClass > 900 AND deweyClass < 1000 THEN 1 ELSE 0
END) AS D900_999 FROM spl_2016.inraw WHERE itemtype like '%cd' AND YEAR(cout) >= '2006' AND
YEAR(cout) <= '2019' GROUP BY YEAR(cout);
```

Rand – random sampling

This results in a sampling over 204819 recovered in 0.075 sec
If ordered randomly then time increases to 24.570 sec

```
// random sample on a specific day
SELECT
   *
FROM
   spl_2016.outraw
WHERE
   DATE(cout) = '2021-10-10'
      AND RAND() < .001;
```

SubString – Use selected parts of the string

```
//How many types of media items on a specific day 24.328 inraw, 24.894 outraw –

SELECT distinct SUBSTRING(itemType, 2) AS mediaType, COUNT(ItemNumber) AS Counts FROM
spl_2016.inraw Where date(cout) = '2019-12-10' GROUP BY mediaType ORDER BY Counts DESC
LIMIT 100

/https://www.w3schools.com/sql/func_mysql_left.asp //only uses the right three chars
SELECT distinct RIGHT(itemType, 3) AS mediaType, COUNT(ItemNumber) AS Counts FROM
spl_2016.inraw Where date(cout) = '2019-12-10' GROUP BY mediaType ORDER BY C
```

Select within Select –

```
// Select within Select, use of variables, t.yearCount.csv, 115.682 seconds
SELECT
   t.Year AS Year,
   t.Date AS Month,
   t.dewey AS dewey,
   COUNT(t.dewey) AS counts
FROM
   (SELECT
      YEAR(cout) AS Year,
         DATE_FORMAT(cout, '%m-%d') AS Date,
         CASE
            WHEN itemType LIKE '%bk' THEN 'book'
            ELSE 'non-book'
         END AS item,
         CASE
            WHEN deweyClass >= 0 AND deweyClass < 100 THEN '0'
            WHEN deweyClass >= 100 AND deweyClass < 200 THEN '100'
            WHEN deweyClass >= 200 AND deweyClass < 300 THEN '200'
            WHEN deweyClass >= 300 AND deweyClass < 400 THEN '300'
            WHEN deweyClass >= 400 AND deweyClass < 500 THEN '400'
            WHEN deweyClass >= 500 AND deweyClass < 600 THEN '500'
            WHEN deweyClass >= 600 AND deweyClass < 700 THEN '600'
            WHEN deweyClass >= 700 AND deweyClass < 800 THEN '700'
            WHEN deweyClass >= 800 AND deweyClass < 900 THEN '800'
            WHEN deweyClass >= 900 AND deweyClass < 1000 THEN '900'
            ELSE ''
         END AS dewey
   FROM
      spl_2016.inraw) t
WHERE
```

```
    t.item = 'book' AND t.dewey != ''
        AND t.Year > 2006
        AND t.Year < 2017
GROUP BY t.Year , t.Date , t.dewey
ORDER BY t.Year , t.Date
LIMIT 4000;
```