1.Find the single longest and shortest borrow time

The longest and shortest book borrowing time among all entries.
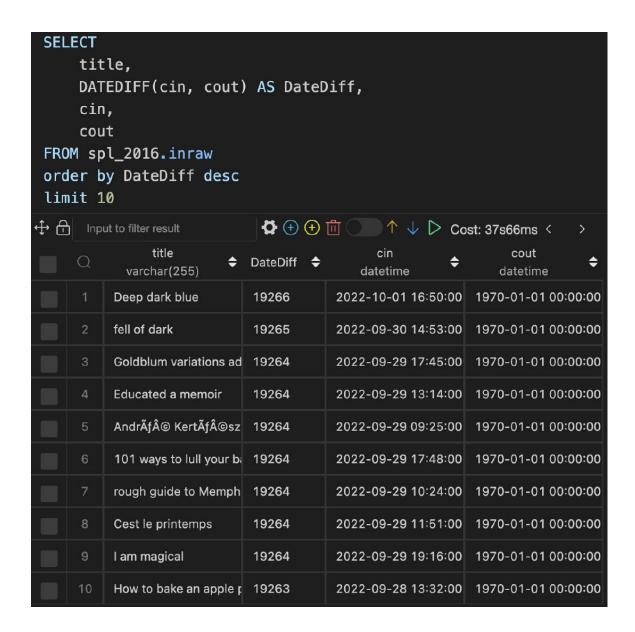
```
1   SELECT title,DATEDIFF(cin, cout) AS DateDiff
2   FROM spl_2016.inraw
3   order by DateDiff desc
4   limit 10;
```

Result:

title	DateDiff
Deep dark blue	19266
fell of dark	19265
Goldblum variations adventures of Jeff Goldblum across the known and unknown universe	19264
Educated a memoir	19264
AndrÃƒÂ© KertÃƒÂ©sz postcards from Paris	19264
101 ways to lull your baby to sleep	19264
rough guide to Memphis Minnie queen of the country blues	19264
Cest le printemps	19264
I am magical	19264
How to bake an apple pie	19263

Problem:

After doing a simple math, we may find out that 19266/365 > 52 yrs, which is out of range (dates back to 1970). After listing the checkin, checkout column, we get the result as below, which is obviously weird.

```
SELECT
    title,
    DATEDIFF(cin, cout) AS DateDiff,
    cin,
    cout
FROM spl_2016.inraw
order by DateDiff desc
limit 10
```

		title varchar(255)	DateDiff	cin datetime	cout datetime
	1	Deep dark blue	19266	2022-10-01 16:50:00	1970-01-01 00:00:00
	2	fell of dark	19265	2022-09-30 14:53:00	1970-01-01 00:00:00
	3	Goldblum variations ad	19264	2022-09-29 17:45:00	1970-01-01 00:00:00
	4	Educated a memoir	19264	2022-09-29 13:14:00	1970-01-01 00:00:00
	5	AndrÃfÂ© KertÃfÂ©sz	19264	2022-09-29 09:25:00	1970-01-01 00:00:00
	6	101 ways to lull your ba	19264	2022-09-29 17:48:00	1970-01-01 00:00:00
	7	rough guide to Memph	19264	2022-09-29 10:24:00	1970-01-01 00:00:00
	8	Cest le printemps	19264	2022-09-29 11:51:00	1970-01-01 00:00:00
	9	I am magical	19264	2022-09-29 19:16:00	1970-01-01 00:00:00
	10	How to bake an apple p	19263	2022-09-28 13:32:00	1970-01-01 00:00:00

Solution:

Since we collect data from 2005 (known), we can add a filter of check out date after year 2005.

```
1  SELECT
2      title,
3      DATEDIFF(cin, cout) AS DateDiff,
4      cin,
5      cout
6  FROM spl_2016.inraw
7  WHERE cout > '2005-01-01'
8  order by DateDiff desc
9  limit 10;
```

We get: some people actually borrowed the book for a interval of almost 15 years!

		title varchar(255)	DateDiff	cin datetime	cout datetime
	1	MÃƒÂ©lodie in E minor	5377	2021-05-13 16:51:00	2006-08-23 15:30:00
	2	Inoi mir sovetskie zapis	5254	2020-10-07 11:35:00	2006-05-20 13:52:00
	3	To the River Plate and I	5235	2019-08-27 14:39:00	2005-04-27 16:11:00
	4	Black people who mad	5151	2022-03-03 12:11:00	2008-01-25 12:31:00
	5	Negro and white unite	5099	2022-01-08 16:11:00	2008-01-23 13:40:00
	6	Annual report of the Bo	5034	2019-03-07 10:41:00	2005-05-25 18:41:00
	7	Owners versus players	4913	2022-01-10 12:41:00	2008-07-29 15:46:00
	8	Unions in transition ent	4886	2022-01-08 16:11:00	2008-08-23 11:49:00
	9	Velvet be bop kente clo	4883	2020-03-01 12:16:00	2006-10-18 11:15:00
	10	Three trios for flute vio	4705	2021-05-23 08:12:00	2008-07-05 12:28:00

Same applies to the shortest borrow time, except it's in increasing order (default).

```
#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;
```

Result: Fast mind-changers

title	DateDiff	cin	cout
Mortdecais endgame	0	2006-01-02 10:01:00	2006-01-02 10:01:00
hikers guide to the Hawaiian Islands	0	2006-01-02 10:57:00	2006-01-02 10:36:00
Children of Dune	0	2006-01-02 11:06:00	2006-01-02 10:28:00
lion of St Mark	0	2006-01-02 10:20:00	2006-01-02 10:07:00
Batman war games Act 3 End game	0	2006-01-02 11:06:00	2006-01-02 10:28:00
Ripleys game	0	2006-01-02 11:06:00	2006-01-02 10:28:00
Love medicine	0	2006-01-02 10:24:00	2006-01-02 10:22:00
dust roads of Monferrato	0	2006-01-02 10:20:00	2006-01-02 10:07:00
Always time to die	0	2006-01-02 09:30:00	2006-01-02 09:30:00
National treasure	0	2006-01-02 10:25:00	2006-01-02 10:24:00

We might see that some people immediately checked in after their check-out, probably don't feel like reading that book in few minutes.

2.What's the number of books of different dewey classes (Items of same bibNumber count once) ?

```
1  SELECT
2      deweyClass,
3      COUNT(bibNumber) as BookNumber
4  from deweyClass as D
5  group by D.deweyClass
6  order by BookNumber desc
7  limit 100;
```

Result:

deweyClass	BookNumber
	483794
782	47730
741	27619
641	18018
895	13539
796	11821
811	10137
973	9594
917	8164
784	7775
658	7521
306	7248
616	7218
940	6866
305	6822
613	6780
791	6608
891	6095
781	6058
789	5417
629	5403

Most of the books are not categorized. Among those which has an entry, top 3 classes are 782 (Vocal Music), 741 (Drawings) and 641 (Food & Drink), which is kind of surprising.

3.What's the number of items of different dewey classes (Items of same bibNumber count multiple times according to their number of copies)?

This is slightly different from Q2, which I think can better represent the popularity of a certain dewey category. However, due to the restriction of MySQL on functional dependency, I have to use nested query as below (rather than a non-nested version) :

```sql
select
    D.deweyClass,
    sum(copies) as itemNums
from (
        select
            deweyClass,
            bibNumber
        from
            deweyClass as class
        where
            class.deweyClass REGEXP '^[0-9]{3}$'
    ) as D
    inner join (
        select
            count(itemNumber) as copies,
            bibNumber
        from spl_2016.inraw
        where year(cout = 2007)
        group by
            bibNumber
    ) as A on D.bibNumber = A.bibNumber
group by D.deweyClass
order by itemNums desc
limit 10;
```

This should work conceptually, but due to the large amount of data and temporary access issue with the deweyClass table, I rewrite the query in the form as below. Note that the deweyClass (char 12) attribute differs a little bit from the one in deweyClass table (char 3). So I made a little modifications here. (Using Substring function). To make executing queries faster, I narrow down the time scope from 2022 to today.

```sql
select
    final.bigClass,
    sum(final.copies) as itemNums
from (
        select
            DISTINCT A.bibNumber,
```

```sql
                bigClass,
                copies
            from (
                    select
                        count(itemNumber) as copies,
                        bibNumber
                    from
                        spl_2016.inraw
                         where
                         cout > '2022-01-01'
                    group by
                        bibNumber
                ) as A
                inner join (
                    select
                        bibNumber,
                        SUBSTRING(deweyClass, 1, 3) as bigClass
                    from
                        spl_2016.inraw
                    where
                        deweyClass > ''
                        and cout > '2022-01-01'
                ) as D on D.bibNumber = A.bibNumber
        ) as final
    group by final.bigClass
    order by itemNums desc
    limit 100;
```

This tooks 11s to complete. Part of the table looks like this. And still, the categories with the most book copies are 782 (Vocal Music), 741 (Drawings) and 641 (Food & Drink).

782	98692
641	70912
741	70134
796	17966
305	16036
917	15358
616	15317
398	14382
306	14214
158	13590

If we make the category a little bit wider, we group them with their first dewey decimal, we get:

4, 0, 2 are Language, General and Religon respectively.

4	6840
0	16874
2	19043
8	49830
1	50025
5	84877
9	92974
3	116368
6	178671
7	288854

4.Which book have the most number of copies? (Top 10)

```
1  select
2      count(itemNumber) as copies,
3      bibNumber
4  from spl_2016.inraw
5  group by bibNumber
6  order by copies desc
```

Result:

	copies	bibNumber int
1	37681	3030520
2	18907	3489506
3	15611	2469502
4	15519	1205054
5	15436	2919580
6	14736	2542732
7	14073	2474843
8	13837	2482761
9	13795	2560429
10	13788	2474845
11	13740	2560415

Top 3 of the books are, Into the wild, SPL HotSpot connections, Headphones.

```sql
select
    distinct(title),
    bibNumber
from spl_2016.inraw
where
    bibNumber in (
        '3030520',
        '3489506',
        '2469502'
    )
```

title varchar(255)	bibNumber int
Into the wild	2469502
SPL HotSpot connectir	3030520
Headphones	3489506