

Stejara Dinulescu

13 January 2022

Assignment 1 – MySQL Query

Concept

The COVID-19 Pandemic introduced various health and lifestyle challenges, consequences, and disturbances, affecting the daily lives and routines of all people in some capacity. I am interested in investigating the patterns of library material *check-ins* (i.e. requiring that one checks out and brings back the item) from the SPL database in the two years before the COVID-19 pandemic (01/01/2018-12/31/2019) compared to the two years after the start of the COVID-19 pandemic (01/01/2020- 12/31/2021). Furthermore, I am interested in how these trends vary by Dewey Classification, which is a system that organizes and classifies subject matter in the SPL database. The research questions and hypotheses detailed below were chosen based on my personal thoughts when examining the Dewey Classification system in detail from the following citation –

<https://www.oclc.org/content/dam/oclc/dewey/resources/summaries/deweysummaries.pdf>.

Discrete and continuous pictures of the queried data were investigated to provide a greater understanding of check-in trends in the SPL over the past four years.

Research Question 1:

- A. Does the number of SPL check-ins increase or decrease after the start of the pandemic?

I hypothesize that in general, check-in count across all Dewey Classification categories will decrease as a result of the COVID-19 pandemic (due to socialization restrictions).

- B. What is the percent change of check-ins, from the period before the pandemic compared to the period after?

Rate of change = (Pre-COVID Total - Post-COVID Total) / Pre-COVID Total

I anticipate that a percent change above 25% will occur.

- C. What is the most popular Dewey Classification ID checked-in during these time periods?

I anticipate that the Medicine and Health Dewey Classification (610) will be the most popular category checked-in post-covid, as this public health crisis may have led to an increase in desire to learn more about the virus and how to treat symptoms. I anticipate that Art and Recreation (700-799) will be the most popular category checked-in pre-covid from the SPL, as most people utilize libraries either for recreation or for research purposes.

Research Question 2:

- A. What is the distribution of check-in counts across Dewey Classification General Categories before COVID-19, compared to the years after?
- B. Which Dewey category has the highest percent change of check-ins between the two time periods?
- C. How does frequency of check-ins change over time within each of the 10 general Dewey categories?

Query Code and Results

Query 1:

MySQL query code:

Pre-COVID:

```
SELECT DISTINCT deweyClass, count(*)  
from spl_2016.inraw  
where year(cin) >= '2018' AND year(cin) <= '2019'  
group by deweyClass order by count(*) DESC;
```

Post-COVID:

```
SELECT DISTINCT deweyClass, count(*)  
from spl_2016.inraw  
where year(cin) >= '2020' AND year(cin) <= '2021'  
group by deweyClass order by count(*) DESC;
```

- A) What is the check-in count number across ALL Dewey categories during the chosen pre-COVID time period versus post-COVID period (discrete)?

Total Check-in Count between 2018-2019 for ALL Dewey Categories: 4,790,063
- The top Dewey Category was 782.42166, with a count of 253494

Total Check-in Count between 2020-2021 for ALL Dewey Categories: 2,987,779
- The top Dewey Category was 782.42166, with a count of 136665

Therefore, total check-in count **decreased** as a result of the COVID-19 pandemic.

- B) What is the percent change of check-ins from pre-COVID to post-COVID time periods?

Total check-in count decreased by a rate of
 $(4,790,063 - 2,987,779) / 4,790,063 = 37.63\%$.

- C) Which discrete Dewey Catalog ID has the highest check-in count before and after the start of the pandemic?

The most popular category of check-ins in both pre- and post-COVID time periods is the general principles and form of secular vocal music (**782.421**), which is in the Arts and Recreation general category.

Query 2:

- A) What is the check-in count number across the big 10 general Dewey categories during the chosen pre-COVID time period versus post-COVID period (discrete)?

MySQL query code:

Pre-COVID:

```
SELECT DISTINCT LEFT(deweyClass, 1), count(*)
from spl_2016.inraw
where year(cin) >= '2018' AND year(cin) <= '2019'
group by LEFT(deweyClass,1) order by count(*) DESC;
```

Post-COVID:

```
SELECT DISTINCT LEFT(deweyClass, 1), count(*)
from spl_2016.inraw
where year(cin) >= '2020' AND year(cin) <= '2021'
group by LEFT(deweyClass,1) order by count(*) DESC;
```

Results

The most popular big 10 Dewey category, both before and after the pandemic, is Arts and Recreation (**700**). Another interesting observation is that the *order of the* frequency of check-ins is preserved before and after the pandemic (although check-in frequency is not the same). The least popular Dewey category is Language (**400**).

Pre-COVID:		4790063		Post-COVID:		2987779	
	7	1049068			7	536062	
	6	445560			6	303606	
	3	318135			3	218277	
	9	268868			9	153719	
	5	195665			5	139780	
	8	130238			8	86918	
	1	108992			1	73632	
	2	64520			2	36238	
	0	50319			0	27088	
	4	28723			4	12663	

B) Rate of change of check-ins:

7: 48.90% 6: 31.86% 3: 31.39% 9: 42.83% **5: 28.56%**
 8: 33.26% 1: 32.44% 2: 43.83% 0: 46.17% **4: 55.91%**

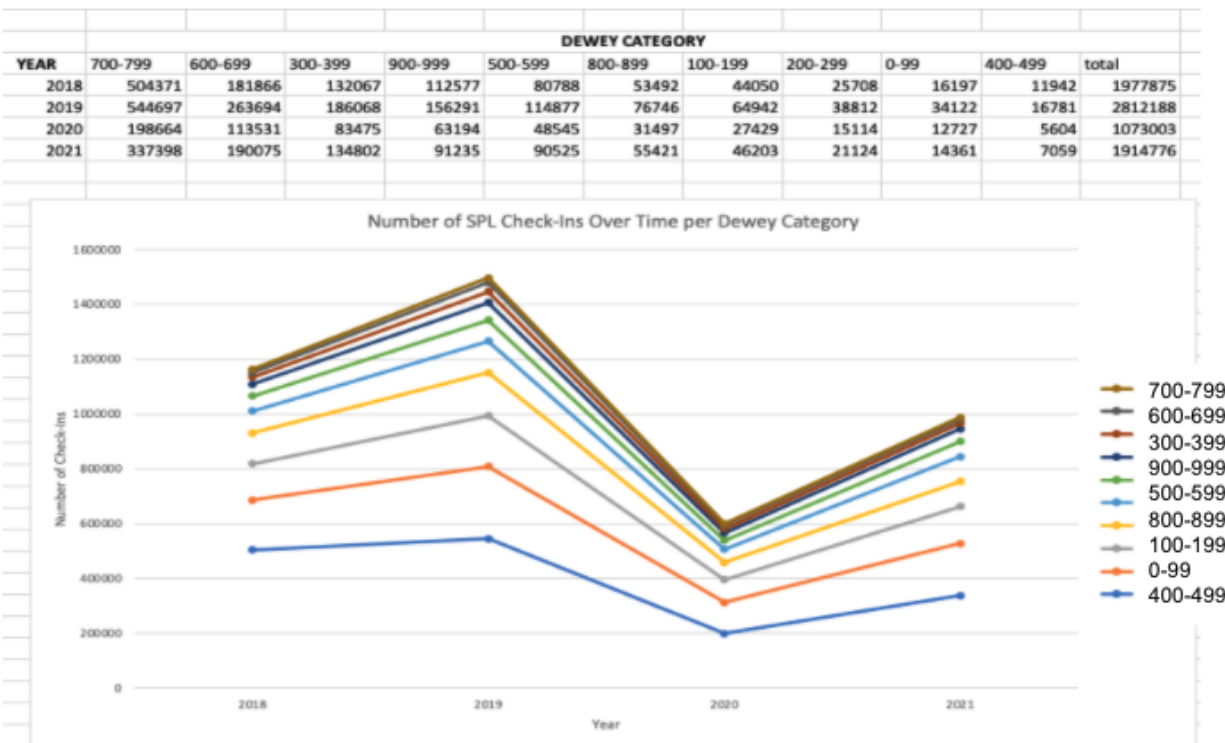
The highest rate of change (i.e. decrease in check-ins) is in Dewey category 4, which is the Language category. The lowest rate of change (i.e. decrease in check-ins) is in Dewey category 5, which is the Science category.

C) How does check-in count change over time (between the years 2018 to 2021 inclusive, taking each day as a data point) across the four specified categories (continuous)?

MySQL query code:

```
SELECT LEFT(deweyClass, 1), year(cin), count(*)
from spl_2016.inraw
where year(cin) >= '2018' AND year(cin) <= '2021'
GROUP BY year(cin), LEFT(deweyClass, 1)
ORDER BY year(cin), count(*);
```

Results:



From this display of the data, we can clearly see the abrupt decrease in check-ins due to the onset of the pandemic, with a bit of an increase back up again in 2021. All Dewey categories showed roughly similar trends across the four years.

Conclusion

Through these queries and subsequent presentation of the data, we can see how much of an effect the pandemic has had over check-in activity at the SPL between the years of 2018 and 2021. Obtaining both discrete and continuous metrics enabled a more comprehensive picture of check-in trends. In general, check-ins at the SPL decreased after the start of the COVID-19 pandemic. This is believed to be due to the significant changes in lifestyle and added challenges for many people as a direct result of the pandemic. A deeper investigation was conducted into check-in counts across the 10 general Dewey classification categories, as well as metrics for change over time. While Arts and Recreation was the most popular Dewey category before and after the start of the COVID-19 pandemic, the highest rate of change in check-ins occurred in the least popular Dewey category, which was the Language category.