

Analyzing The Reebok Feed Files

All the analysis that will be presented below was done on 7 batches of files. A “batch”, represents a set of files between two different dates (inclusive of both dates).

Here is the list of the 7 batches that were analyzed -

- Batch 1 -> 26th of April to 1st of May
- Batch 2 -> 1st of May to 8th of May
- Batch 3 -> 8th of May to 15th of May
- Batch 4 -> 15th of May to 22nd of May
- Batch 5 -> 22nd of May to 5th of June
- Batch 6 -> 5th of June to 12th of June
- Batch 7 -> 12th of June to 22nd of June

Process of Data Analysis:

- For each batch, the items in that batch were coupled with the first time they appeared in that particular batch.
- For each item in the batch, the date corresponding to when it was uploaded onto the FINDMINE database was obtained. The difference in these dates is the **ingestion difference**.
- The differences between the dates was calculated and the average for all these differences for each batch was calculated. This difference for each item will be called an “ingestion difference”.
- A csv file with each item and its corresponding ingestion difference was generated.
- This data was then processed to generate another csv file where the first column is the ingestion difference and the second column is the number of files with the corresponding ingestion difference.
- Using the second csv file, a bar graph was generated with relation to the data in the file.

Analysis:

Batch 1 (26th of April to 1st of May):

- Average ingestion time (in days) - -22.59 days
This means, on an average, items were ingested 22.5 days before they first appeared in a Reebok feed file.
- Number of items ingested before and at the day of obtaining feed files - 1600

Breakdown:

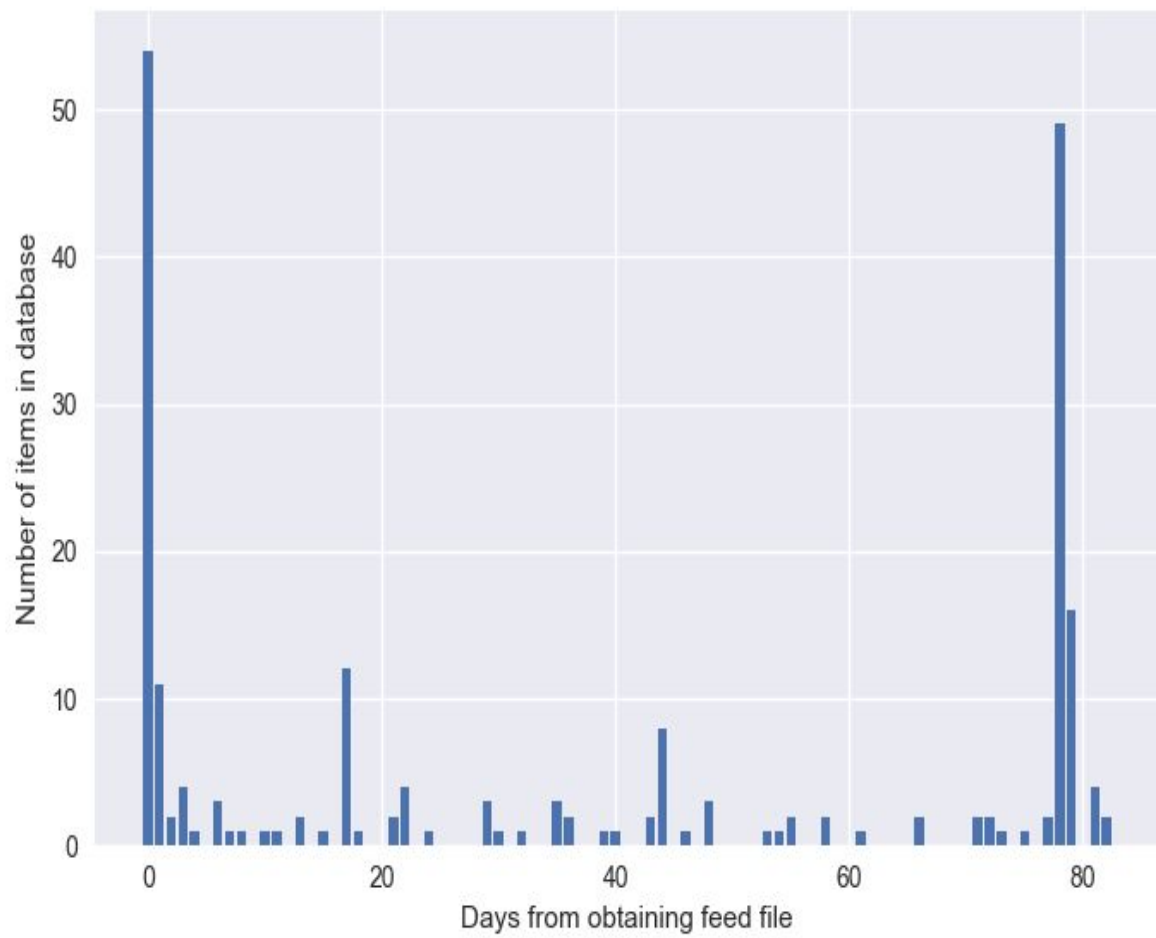
Ingestion Difference, Number of files

-50,4
-35,2
-34,46
-33,1061
-32,111
-31,201
-22,1
-11,101
-10,12
-9,5
-8,2
-7,1
-6,7
-5,1
-2,1
-1,4
0,40
1,11
2,2
3,4
4,1
6,3
7,1
8,1
10,1
11,1
13,2
15,1
17,12
18,1
21,2
22,4

24,1
29,3
30,1
32,1
35,3
36,2
39,1
40,1
43,2
44,8
46,1
48,3
53,1
54,1
55,2
58,2
61,1
66,2
71,2
72,2
73,1
75,1
77,2
78,49
79,16
81,4
82,2

Graphical representation of the data presented above:

Note - Bar line at zero is only for representational purposes and doesn't exactly equal the number of files ingested before and at the time of obtaining the csv files.



Batch 2 (1st of May to 8th of May):

- Average ingestion time (in days) - -26.87 days
This means, on an average, items were ingested 26.87 days before they first appeared in a Reebok feed file.
- Number of items ingested before and at the day of obtaining feed files - 1647

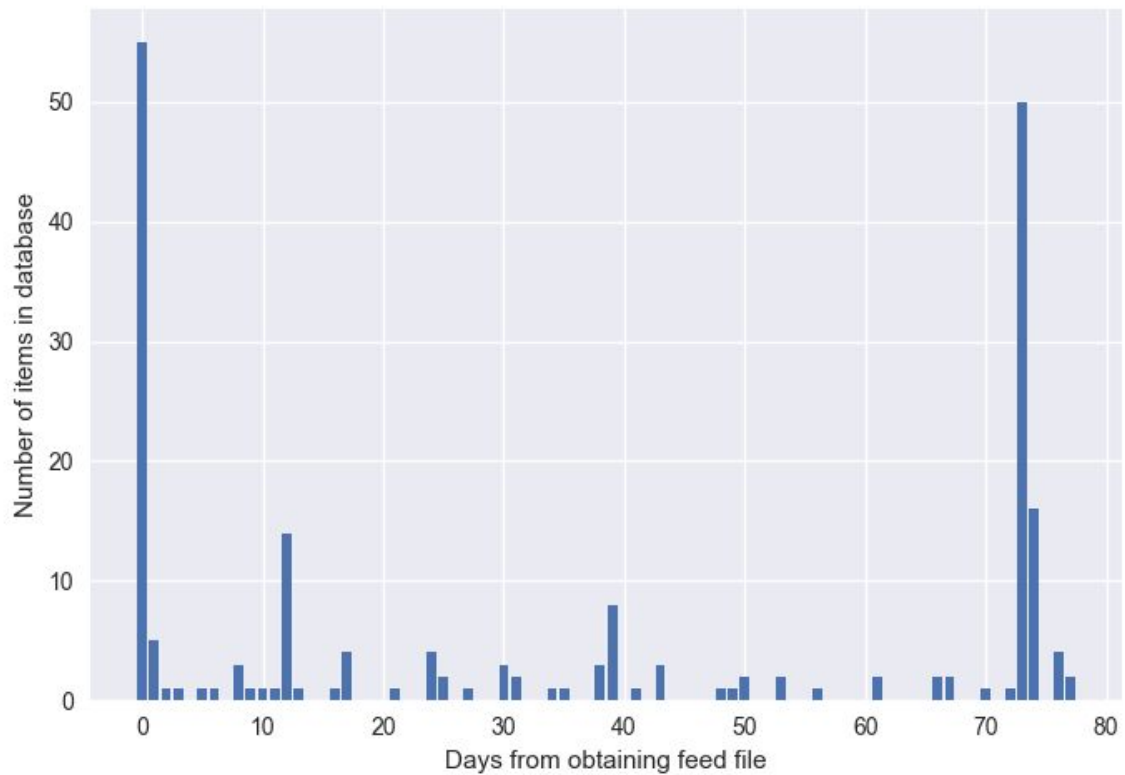
Breakdown:**Ingestion Difference, Number of files**

-55,4
-41,4
-39,45
-38,1058
-37,110
-36,203
-27,1
-20,1
-17,1
-16,100
-15,12
-14,5
-13,2
-12,1
-11,7
-7,1
-6,1
-5,8
-4,7
-3,2
-2,4
-1,2
0,68
1,5
2,1
3,1
5,1
6,1
8,3
9,1
10,1
11,1
12,14
13,1

16,1
17,4
21,1
24,4
25,2
27,1
30,3
31,2
34,1
35,1
38,3
39,8
41,1
43,3
48,1
49,1
50,2
53,2
56,1
61,2
66,2
67,2
70,1
72,1
73,50
74,16
76,4
77,2

Graphical representation of the data presented above:

Note - Bar line at zero is only for representational purposes and doesn't exactly equal the number of files ingested before and at the time of obtaining the csv files.



Batch 3 (8th of May to 15th of May):

- Average ingestion time (in days) - -33.08 days
Number of items ingested before and at the day of obtaining feed files - 1582

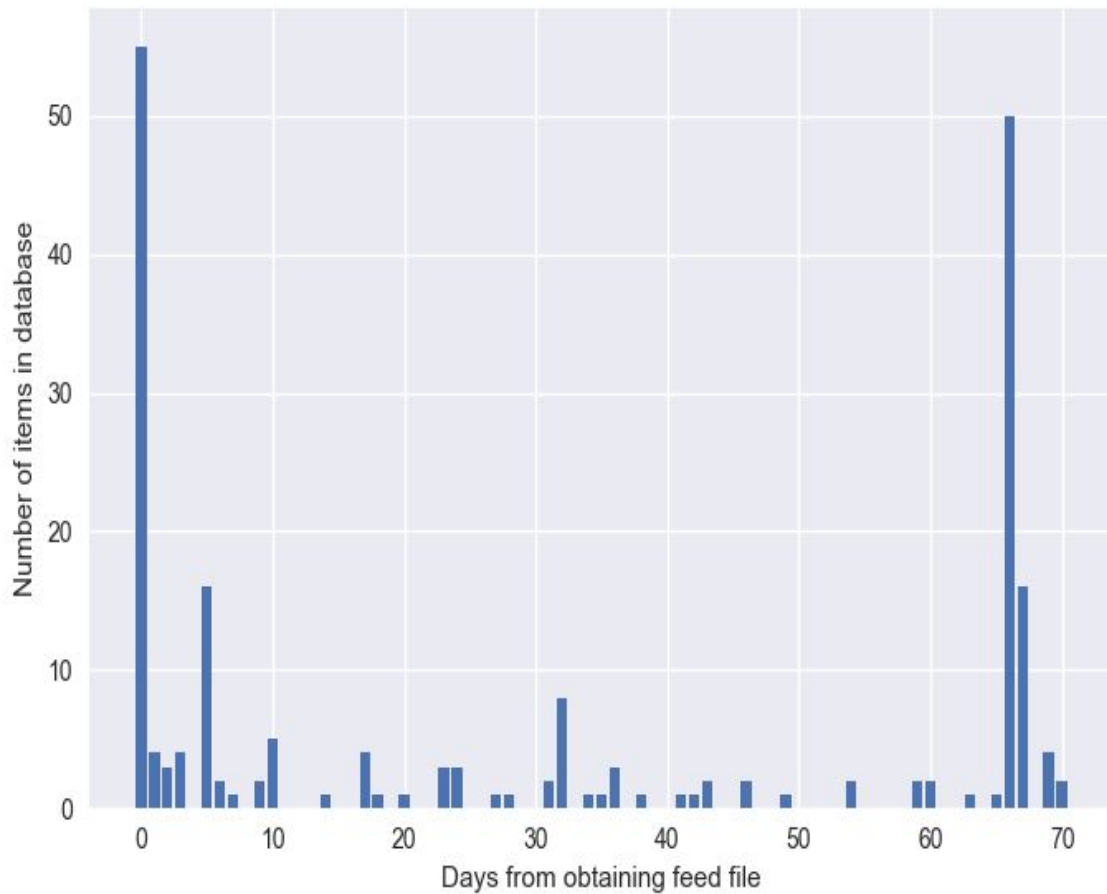
Breakdown:**Ingestion Difference, Number of files**

-62,3
-51,1
-50,2
-46,44
-45,1003
-44,105
-43,195
-34,1
-25,1
-23,99
-22,12
-21,5
-20,2
-19,1
-18,7
-14,1
-13,1
-12,7
-11,6
-10,2
-9,4
-8,1
-7,41
-6,2
-5,6
-4,8
-3,12
-2,1
-1,2
0,7
1,4
2,3
3,4
5,16
6,2
7,1

9,2
10,5
14,1
17,4
18,1
20,1
23,3
24,3
27,1
28,1
31,2
32,8
34,1
35,1
36,3
38,1
41,1
42,1
43,2
46,2
49,1
54,2
59,2
60,2
63,1
65,1
66,50
67,16
69,4
70,2

Graphical representation of the data presented above:

Note - Bar line at zero is only for representational purposes and doesn't exactly equal the number of files ingested before and at the time of obtaining the csv files.



Batch 4 (15th of May to 22nd of May):

- Average ingestion time (in days) - -39.44 days
Number of items ingested before and at the day of obtaining feed files - 1622

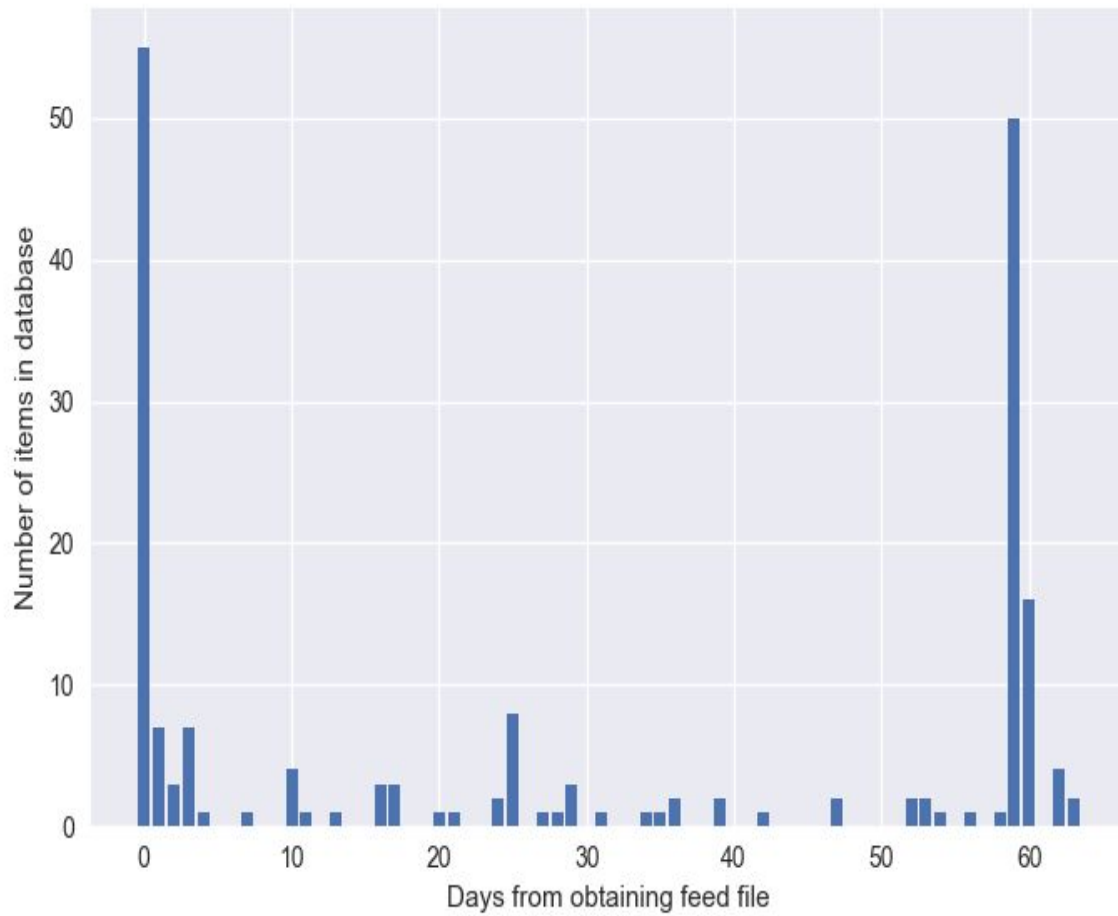
Breakdown:**Ingestion Difference, Number of files**

-69,3
-57,1
-56,1
-55,1
-54,2
-53,41
-52,1006
-51,103
-50,195
-41,1
-31,1
-30,97
-29,12
-28,5
-27,2
-26,1
-25,7
-23,1
-21,1
-20,1
-19,7
-18,6
-17,2
-16,4
-15,1
-14,40
-13,2
-12,6
-11,7
-10,11
-8,1
-7,3
-5,1
-4,3
-3,2
-2,17

-1,3
0,24
1,7
2,3
3,7
4,1
7,1
10,4
11,1
13,1
16,3
17,3
20,1
21,1
24,2
25,8
27,1
28,1
29,3
31,1
34,1
35,1
36,2
39,2
42,1
47,2
52,2
53,2
54,1
56,1
58,1
59,50
60,16
62,4
63,2

Graphical representation of the data presented above:

Note - Bar line at zero is only for representational purposes and doesn't exactly equal the number of files ingested before and at the time of obtaining the csv files.



Batch 5 (22nd of May to 5th of June):

- Average ingestion time (in days) - -42.57 days
Number of items ingested before and at the day of obtaining feed files - 1732

Breakdown:**Ingestion Difference, Number of files**

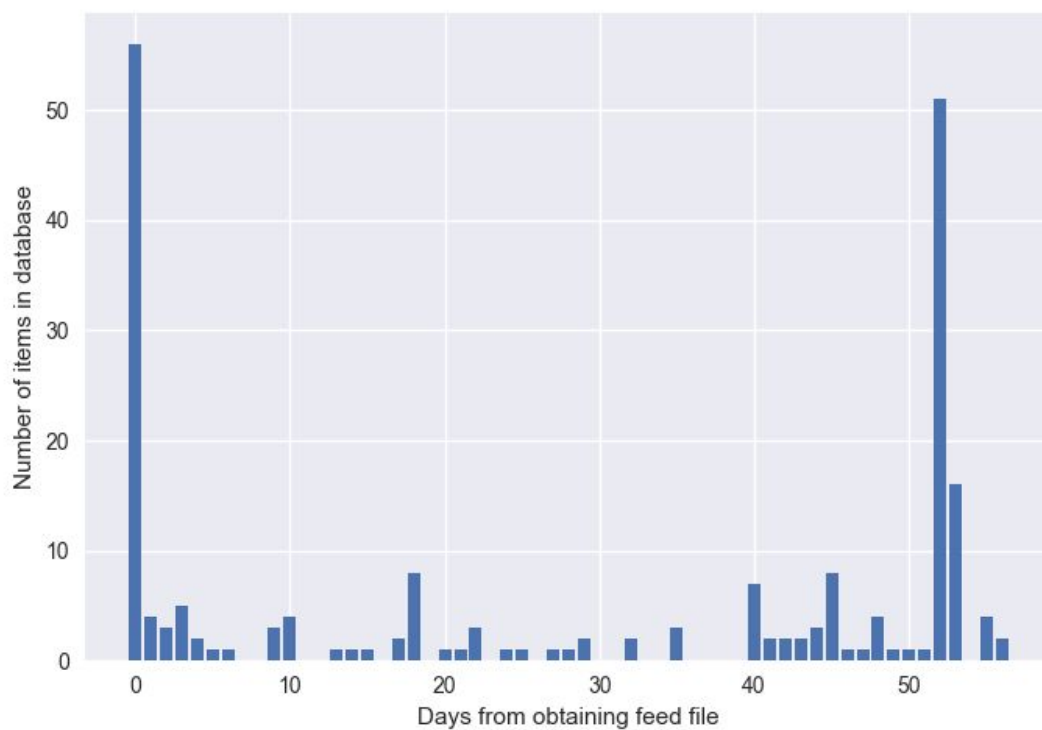
-76,3
-71,2
-70,1
-66,2
-65,1
-63,1
-62,1
-61,4
-60,41
-59,999
-58,100
-57,195
-48,1
-46,1
-39,1
-38,1
-37,98
-36,12
-35,5
-34,3
-33,2
-32,7
-28,2
-27,1
-26,7
-25,6
-24,2
-23,4
-22,1
-21,39
-20,2
-19,6
-18,7
-17,11
-15,1
-14,3

-12,1
-11,3
-10,2
-9,17
-8,2
-7,6
-6,5
-5,7
-4,16
-3,6
-2,2
-1,26
0,66
1,4
2,3
3,5
4,2
5,1
6,1
9,3
10,4
13,1
14,1
15,1
17,2
18,8
20,1
21,1
22,3
24,1
25,1
27,1
28,1
29,2
32,2
35,3
40,7
41,2
42,2
43,2
44,3
45,8
46,1

47,1
48,4
49,1
50,1
51,1
52,51
53,16
55,4
56,2

Graphical representation of the data presented above:

Note - Bar line at zero is only for representational purposes and doesn't exactly equal the number of files ingested before and at the time of obtaining the csv files.



Batch 6 (5th of June to 12th of June):

- Average ingestion time (in days) - -54.02 days
Number of items ingested before and at the day of obtaining feed files - 1686

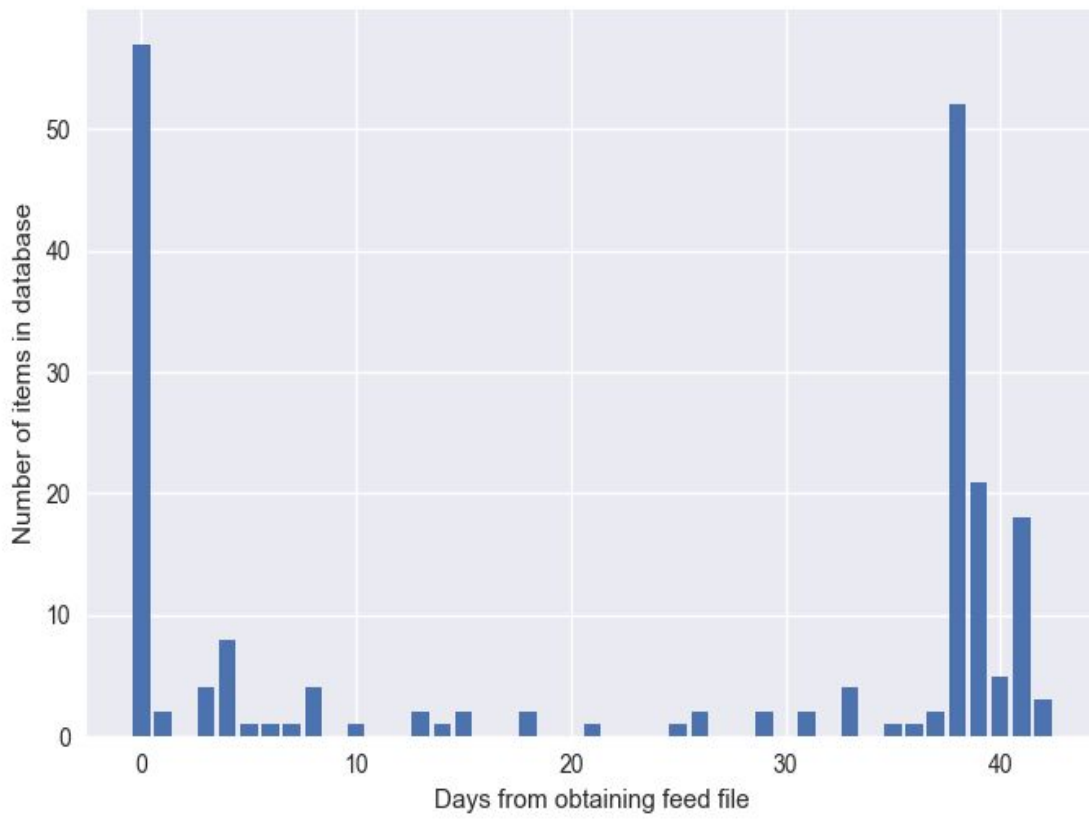
Breakdown:**Ingestion Difference, Number of files**

-90,3
-77,3
-76,4
-75,3
-74,38
-73,943
-72,91
-71,182
-62,1
-53,1
-52,1
-51,98
-50,12
-49,5
-48,3
-47,1
-46,7
-42,2
-41,1
-40,5
-39,6
-38,3
-37,4
-35,40
-34,2
-33,5
-32,7
-31,8
-28,3
-26,1
-25,2
-24,2
-23,16
-22,1
-21,5
-20,5

-19,7
-18,15
-17,6
-14,4
-12,3
-11,4
-7,7
-6,1
-5,4
-4,37
-3,17
-2,26
-1,28
0,13
1,2
3,4
4,8
5,1
6,1
7,1
8,4
10,1
13,2
14,1
15,2
18,2
21,1
25,1
26,2
29,2
31,2
33,4
35,1
36,1
37,2
38,52
39,21
40,5
41,18
42,3

Graphical representation of the data presented above:

Note - Bar line at zero is only for representational purposes and doesn't exactly equal the number of files ingested before and at the time of obtaining the csv files.



Batch 7 (12th of June to 22nd of June):

- Average ingestion time (in days) - -55.31 days
Number of items ingested before and at the day of obtaining feed files - 1838

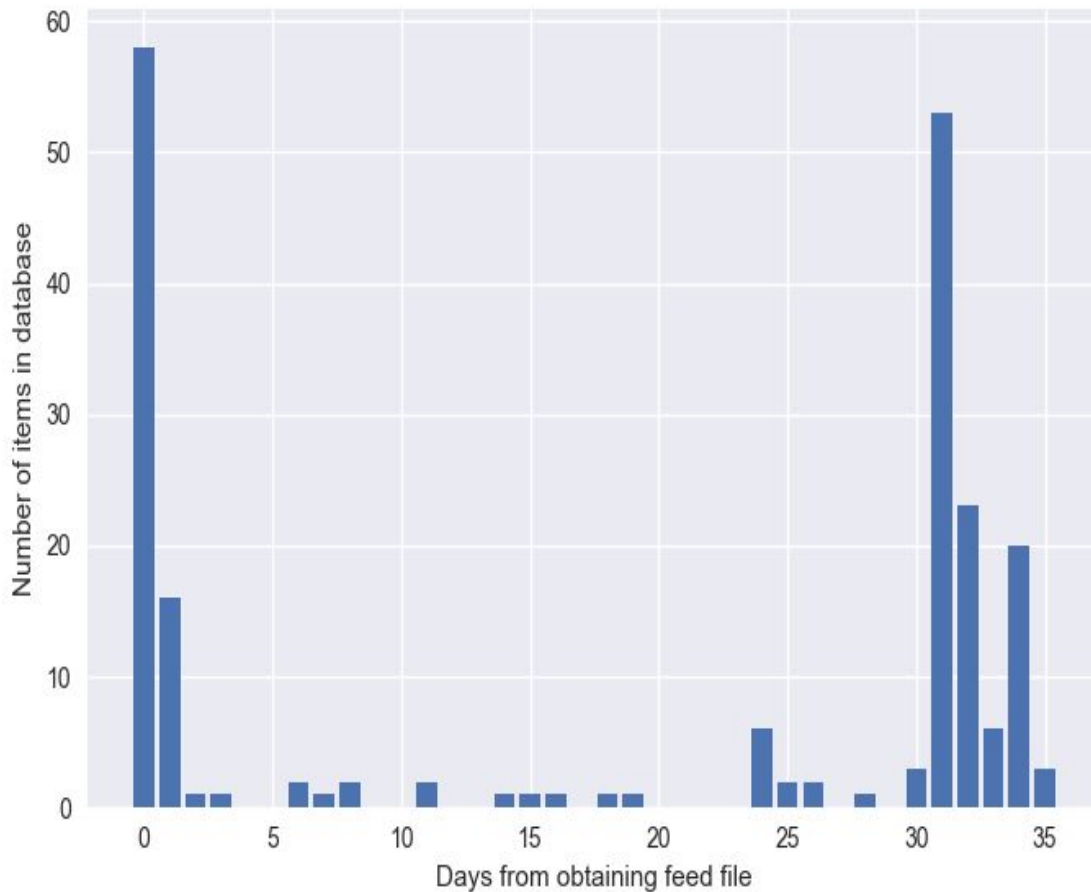
Breakdown:**Ingestion Difference, Number of files**

-97,3
-90,1
-88,1
-84,4
-82,3
-81,39
-80,928
-79,90
-78,181
-69,1
-58,98
-57,12
-56,6
-55,3
-54,1
-53,7
-49,2
-48,1
-47,5
-46,6
-45,2
-44,3
-42,39
-41,2
-40,5
-39,7
-38,8
-35,3
-33,2
-32,2
-31,2
-30,15
-29,1
-28,5
-27,5
-26,4
-25,15

-24,6
-21,4
-19,3
-18,4
-17,1
-16,1
-14,7
-13,1
-12,4
-11,37
-10,20
-9,27
-8,6
-7,9
-6,11
-5,9
-4,15
-3,10
-2,4
-1,92
0,55
1,16
2,1
3,1
6,2
7,1
8,2
11,2
14,1
15,1
16,1
18,1
19,1
24,6
25,2
26,2
28,1
30,3
31,53
32,23
33,6
34,20
35,3

Graphical representation of the data presented above:

Note - Bar line at zero is only for representational purposes and doesn't exactly equal the number of files ingested before and at the time of obtaining the csv files.



Questions:

If you have any questions with respect to this report, or if you have found any errors please email me - krish.chelikavada@findmine.com