

We've recently made an accessibility improvement to the community and therefore posts without any content are no longer allowed. Please use the spoiler feature or add a short message in the message body in order to submit your weekly challenge.

2022-05-26 Updates: Email: If you're not seeing emails be delivered from the Community, please check your spam and mark the Community emails as not junk. Thank you for your patience.



SIGN IN



Free Trial

Weekly Challenge

Solve the challenge, share your solution and summit the ranks of our Community!

Also available in | Français | Português | Español | 中文

IDEAS WANTED

We're actively looking for ideas on how to improve Weekly Challenges and would love to hear what you think!

[SUBMIT FEEDBACK](#)

[Weekly Challenge](#)

Challenge #116: A Symphony of Parsing Tools!



ChristineB
Alteryx Alumni (Retired)

A solution for last week's challenge has been posted [here](#)!

The NY Philharmonic, as various ensembles, has been performing for audiences around the world for over 175 years! Wow! This week's Challenge asks you to parse the data for each of their programs (not for the past 175 years...that file was HUGE!) from 2011 - 2017. For each program, identify the concert information (Date, Location, Time, etc), as well as the pieces played during that program and the solo performers (if applicable). Note: the posted solution has removed records representing an intermission.



[challenge_116_start_file.v2yxml.yxmd](#)

[challenge_116_solution.yxmd](#)

[Data Analysis](#) [Developer](#) [Difficult](#) [Join](#) [Parse](#) [Preparation](#) [Transform](#)

Share

3 LIKES

[Reply](#)



mmongeon
8 - Asteroid

I think there is an issue with the output provided, it doesn't match what I see in the XML.
See the Performance Date in the screenshot below:


Record #	Program ID	Piece ID	Ensemble	Season	Location	Venue	Event Type	Performance Date	Time	Conductor	Composer
1	11629	5618*	New York Philharmonic	2011-12	Manhattan, NY	Avery Fisher Hall	Saturday Matinee	2011-01-01	2:00PM	Paul	Schubert, Franz
2	11646	5552*	New York Philharmonic	2011-12	Manhattan, NY						
3	11639	5729*	New York Philharmonic	2011-12	Manhattan, NY						
4	11971	7424*	Musicians from the New York Philharmonic	2011-12	Manhattan, NY						
5	11757	10525*	NY Philharmonic Ensembles	2011-12	Manhattan, NY						
6	11757	7201*	NY Philharmonic Ensembles	2011-12	Manhattan, NY						
7	11757	7194*	NY Philharmonic Ensembles	2011-12	Manhattan, NY						
8	11757	7207*	NY Philharmonic Ensembles	2011-12	Manhattan, NY						
9	11754	9003-7	New York Philharmonic	2011-12	Montreal, CANADA						
10	11683	1390*	New York Philharmonic	2011-12	Manhattan, NY						
11	11633	7421	New York Philharmonic	2011-12	Manhattan, NY						
12	11755	1178*	New York Philharmonic	2011-12	Montreal, CANADA						
13	11887	7339*	Musicians from the New York Philharmonic	2011-12	Manhattan, NY						
14	11647	4072*	New York Philharmonic	2011-12	Manhattan, NY						
15	10981	7163*	NY Philharmonic Ensembles	2011-12	Manhattan, NY						
16	10681	5656*	NY Philharmonic Ensembles	2011-12	Manhattan, NY						
17	10681	7164*	NY Philharmonic Ensembles	2011-12	Manhattan, NY						

This site uses different types of cookies, including analytics and functional cookies (its own and from other sites). To change your cookie settings or find out more, [click here](#). If you continue browsing our website, you accept these cookies.

[Reject](#)

[I AGREE](#)

[LEARN MORE](#)



ChristineB


Alteryx Alumni (Retired)

Hmmm...let me investigate! Thanks for letting me know [@mmongeon](#)

Share

 0 LIKES

Reply



ChristineB


Alteryx Alumni (Retired)

Good catch [@mmongeon](#)! I did something silly with my DateTime tool. The start file has been updated with the correct (at least until someone catches something else!) start file.

Share

 1 LIKE

Reply



joe_yang

6 - Meteoroid

Hi,

This could just be my lack, but I am wondering why Program ID 11640 was excluded from the Output when the Performance date landed on 2011-09-07? See image below:


```
<?xml version="1.0" encoding="utf-8"?>

<programs>
  <program>
    <id>7358870b-65c8-43d5-ab56-514bde52db88-0.1</id>
    <programID>11640</programID>
    <orchestra>New York Philharmonic</orchestra>
    <season>2011-12</season>
    <concertInfo>
      <eventType>Non-Subscription</eventType>
      <location>Manhattan, NY</location>
      <venue>Avery Fisher Hall</venue>
      <date>2011-09-07T04:00:00Z</date>
      <time>7:30PM</time>
    </concertInfo>
    <concertInfo>
      <eventType>Non-Subscription</eventType>
      <location>Manhattan, NY</location>
      <venue>Avery Fisher Hall</venue>
      <date>2011-09-08T04:00:00Z</date>
      <time>7:30PM</time>
    </concertInfo>
    <worksInfo>
      <work ID="5733">
        <composerName>Bernstein, Leonard</composerName>
        <workTitle>WEST SIDE STORY (WITH FILM)</workTitle>
        <conductorName>Newman, David</conductorName>
      </work>
      <work ID="0">
        <interval>Intermission</interval>
      </work>
    </worksInfo>
  </program>
  <program>
    <id>e0d79f25-6340-4369-a528-962bb03eabd7-0.1</id>
    <programID>11878</programID>
    <orchestra>NYP Presentation</orchestra>
    <season>2011-12</season>
```

Share

 2 LIKES

Reply



mmongeon

8 - Asteroid

Here is my solution.
I'm actually getting more data than was provided in the given Output, but, from the spot checks I've performed, I think my additional records are valid.

↳ Spoiler


 challenge_116_mongeon_complete.yxmd



Share

 3 LIKES

Reply



ChristineB

Alteryx Alumni (Retired)

Another good catch [@joe_yang](#)! I've updated the start file again. Also, I think I found a point of discrepancy: I opted to remove the records containing "Intermission" in the column "interval". That means I removed records that [@mmongeon](#)'s solution includes (and that are included in the original xml file). I'm loving all the intense data investigation this Challenge requires (especially on under-caffeinated Mondays....)!

Share

 3 LIKES

Reply

I didn't get exactly the same output. Saw that there were some changes and that Intermissions were to be excluded. Couldn't face going back again. My solution is messy and I should have summarized earlier in the process.

Spoiler

@mmongeon solution is prettier and almost certainly performs better.

challenge_116_start_file-ISV.yxmd

Share

1 LIKE

Reply

terry10

11 - Bolide

My XML parsing revealed that each program could have more than one concert, so my final counts were different from the challenge output. (e.g. Program 11633 has 3 concert dates and the 3 pieces were played at each concert, so I ended up with 9 records.)

Record #	programID	orchestra	Location	Venue	Performance Date	Time
1	11633	New York Philharmonic	Manhattan, NY	Avery Fisher Hall	2011-10-05	7:30PM
2	11633	New York Philharmonic	Manhattan, NY	Avery Fisher Hall	2011-10-06	7:30PM
3	11633	New York Philharmonic	Manhattan, NY	Avery Fisher Hall	2011-10-07	11:00AM

3 concert dates for Program 11633

programID	orchestra	season	eventType	Location	Venue	Performance Date	Time	Piece ID
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-05	7:30PM	1905*
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-06	7:30PM	1905*
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-07	11:00AM	1905*
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-05	7:30PM	51880*
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-06	7:30PM	51880*
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-07	11:00AM	51880*
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-05	7:30PM	743*
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-06	7:30PM	743*
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-07	11:00AM	743*

My output for program 11633

Record #	Program ID	Piece ID	Ensemble	Season	Location	Venue	Event Type	Performance Date	Time
1	11633	1905*	New York Philharmonic	2011-12	Manhattan, NY	Avery Fisher Hall	Subscription Season	2011-10-05	7:30PM
2	11633	51880*	New York Philharmonic	2011-12	Manhattan, NY	Avery Fisher Hall	Subscription Season	2011-10-05	7:30PM
3	11633	743*	New York Philharmonic	2011-12	Manhattan, NY	Avery Fisher Hall	Subscription Season	2011-10-05	7:30PM

Solution Output for program 11633

challenge_116_terry10.yxmd

Share

2 LIKES

Reply

Natasha

9 - Comet

Here is my solution. I agree with @terry10 that there might be multiple instances of concertInfo, though the solution seems to keep only the first one so I did the same.

Spoiler

challenge_116_Natasha.yxmd

Share

7 LIKES

Reply

