

< Weekly Challenge

Challenge #116: A Symphony of Parsing Tools!



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.

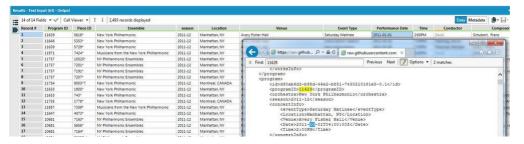


"They'll be expecting adagio. Go with the allegro."





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:



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

Hmmm...let me investigate! Thanks for letting me know @mmongeon





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.





joe_yang

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"?>
<season>2011-12</season>
<concertInfo>
        </concertInfo>
        </work>
</work ID="0*">
     <interval>Intermission</interval>
        </work>
</worksInfo>

<
```







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





Share 3 LIKES Reply

ChristineB

Alteryx Alumni (Retired)

Another good catch @joe_vang! 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

 $\underline{@mmongeon} \ solution \ is \ prettier \ and \ almost \ certainly \ performs \ better.$







My XML parsing revealed that each program could have more that one concert, so my final counts were different from the challenge output. (e.g. Program 11633 has 3 concert date× 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	Ш
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-05	7:30PM	1905*	E
1633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-06	7:30PM	1905*	E
1633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-07	11:00AM	1905*	E
1633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-05	7:30PM	51880*	E
1633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-06	7:30PM	51880*	E
1633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-07	11:00AM	51880*	E
1633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-05	7:30PM	743*	E
1633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-06	7:30PM	743*	E
11633	New York Philharmonic	2011-12	Subscription Season	Manhattan, NY	Avery Fisher Hall	2011-10-07	11:00AM	743*	E

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	Gi
2	11633	51880*	New York Philharmonic	2011-12	Manhattan, NY	Avery Fisher Hall	Subscription Season	2011-10-05	7:30PM	Gi
3	11633	743*	New York Philharmonic	2011-12	Manhattan, NY	Avery Fisher Hall	Subscription Season	2011-10-05	7:30PM	Gil

Solution Output for program 11633







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

D Spoiler





< **1** 2 3 ... 10 →

La

