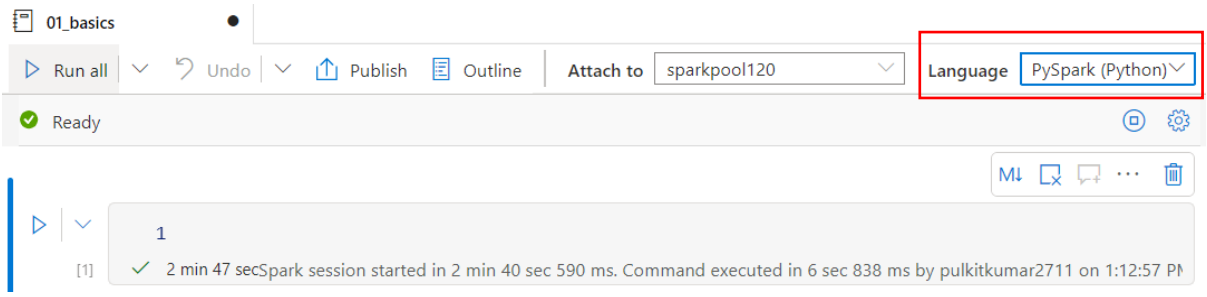
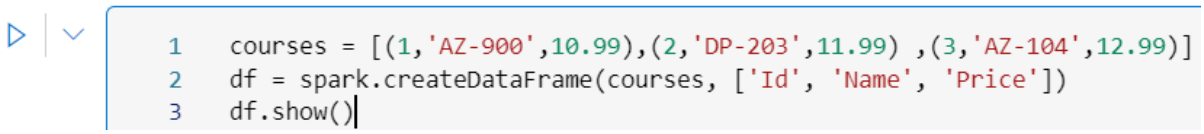
😊 Spark Pool with Python

In this lab, we're focusing on utilizing our Spark pool with Python instead of Scala. We start by switching the language to Python in our notebook and then proceed to write and execute Python code for our data analysis tasks. After running the code, we observe the output displayed in the form of a table, showing the results of our analysis. Specifically, we demonstrate how to convert course information into a DataFrame and then sort the data to obtain it in a sorted manner for better analysis and presentation.

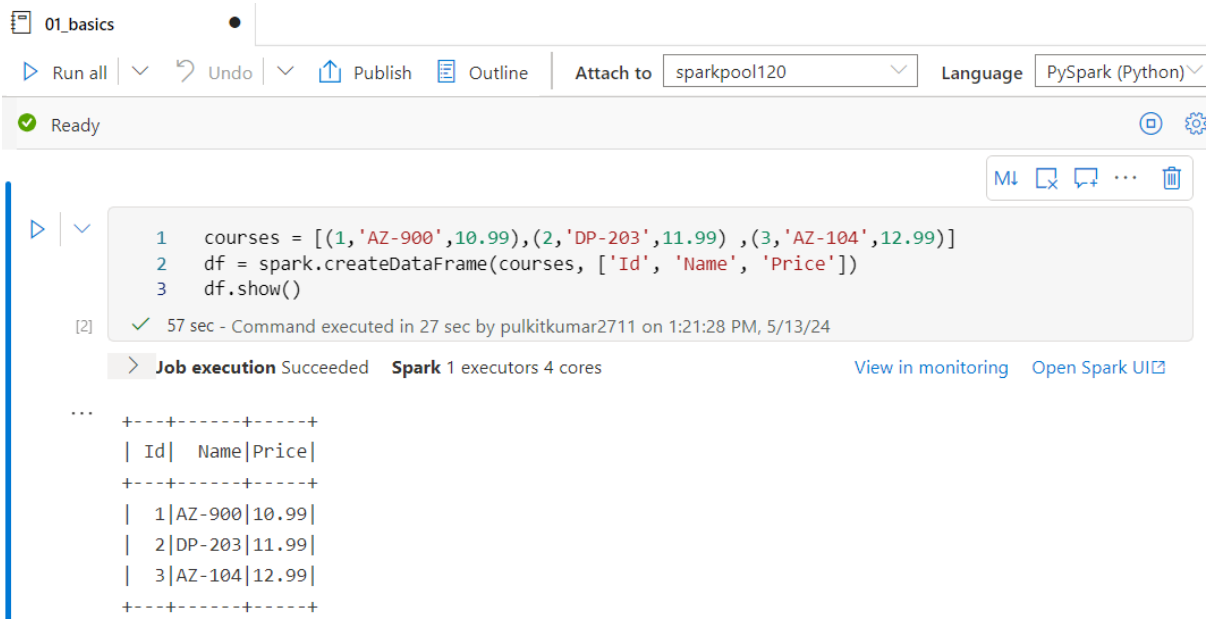
1. In this lab we are going to work on our spark pool but we use Python as our language instead of Scala.
2. Now delete your previous code and choose your language as Python.



3. After that you need to write your code and run it.



4. Below you can see that you got the output of your code in form of a table.



5. Our course, list of data or our course information in Python has now been converted onto a data frame.
6. Now you are going to add another cell to write more code but this time to sort your data. Below you can see that we ran our code and we got the data in sorted manner.

```
1  from pyspark.sql.functions import col
2
3  sorteddf=df.sort(col("Price").desc())
4  sorteddf.show()
```

[3] ✓ 1 sec - Command executed in 1 sec 852 ms by pulkitkumar2711 on 1:26:30 PM, 5/13/24

Job execution Succeeded **Spark** 1 executors 4 cores View in monitoring Open Spark UI

```
+---+------+-----+
| Id|  Name|Price|
+---+------+-----+
|  3|AZ-104|12.99|
|  2|DP-203|11.99|
|  1|AZ-900|10.99|
+---+------+-----+
```