1. Collect() retrieves all elements in a DataFrame as an array of row type to the driver node.
2. Collect() is an action hence it does not return a DataFrame instead, it returns data in an array to the driver. Once the data is in an array, you can use python for loop to process it further.
3. Collect() use it with small DataFrames. With big DataFrames it may result in out of memory error as its return entire data to single node(driver).

from pyspark.sql import \*  
from pyspark.sql.functions import \*  
  
*# Create a Spark session*spark = SparkSession.builder.appName("sample").getOrCreate()  
  
myData = [(1,'Harsha',2000),  
 (2,'Mokshit',3000)]  
  
mySchema = ['id','name','salary']  
  
df = spark.createDataFrame(myData,mySchema)  
  
df.show()  
  
list\_rows = df.collect()  
  
print(list\_rows)  
print(list\_rows[0])  
print(list\_rows[0][1])

**Output:**

+---+-------+------+

| id| name|salary|

+---+-------+------+

| 1| Harsha| 2000|

| 2|Mokshit| 3000|

+---+-------+------+

[Row(id=1, name='Harsha', salary=2000), Row(id=2, name='Mokshit', salary=3000)]

Row(id=1, name='Harsha', salary=2000)

Harsha