

```
In [1]: from pyspark.sql import SparkSession
```

```
In [2]: spark=SparkSession.builder.appName('Dataframe').getOrCreate()
```

```
In [3]: spark
```

**Out[3]: SparkSession - in-memory
SparkContext**

[Spark UI \(http://Arnak:4040\)](http://Arnak:4040)

Version

v3.5.1

Master

local[*]

AppName

Dataframe

```
In [5]: spark.read.csv(r"D:\NIT\FEBRUARY\26 feb (time series and spark)\sample1.csv")
```

```
In [6]: master1
```

Out[6]: DataFrame[Name: string, age: string, Experience: string, Salary: string]

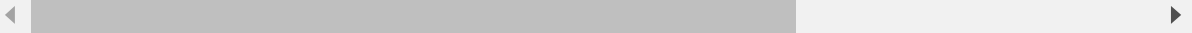
```
In [7]: master1.show()
```

```
+-----+-----+-----+-----+
|  Name|age|Experience|Salary|
+-----+-----+-----+-----+
|   jack| 31|        10| 30000|
|   alex| 30|         8| 25000|
|caroline| 29|         4| 20000|
|   paul| 24|         3| 20000|
|  sandra| 21|         1| 15000|
|casandra| 23|         2| 18000|
+-----+-----+-----+-----+
```

```
In [9]: master1.printSchema()
```

```
root
|-- Name: string (nullable = true)
|-- age: string (nullable = true)
|-- Experience: string (nullable = true)
|-- Salary: string (nullable = true)
```

```
In [10]: master2=spark.read.csv(r"D:\NIT\FEBRUARY\26 feb (time series and spark)\sample
```



```
In [11]: master2
```

```
Out[11]: DataFrame[Name: string, age: int, Experience: int, Salary: int]
```

```
In [12]: master2.printSchema()
```

```
root
|-- Name: string (nullable = true)
|-- age: integer (nullable = true)
|-- Experience: integer (nullable = true)
|-- Salary: integer (nullable = true)
```

```
In [13]: type(master2)
```

```
Out[13]: pyspark.sql.dataframe.DataFrame
```

```
In [14]: master2.head(3)
```

```
Out[14]: [Row(Name='jack', age=31, Experience=10, Salary=30000),
Row(Name='alex', age=30, Experience=8, Salary=25000),
Row(Name='caroline', age=29, Experience=4, Salary=20000)]
```

```
In [15]: master2.show()
```

```
+-----+---+-----+-----+
|   Name|age|Experience|Salary|
+-----+---+-----+-----+
|   jack| 31|         10| 30000|
|   alex| 30|          8| 25000|
|caroline| 29|          4| 20000|
|   paul| 24|          3| 20000|
|  sandra| 21|          1| 15000|
|casandra| 23|          2| 18000|
+-----+---+-----+-----+
```

```
In [16]: master2.select(['Name', 'Experience']).show()
```

```
+-----+-----+
|   Name|Experience|
+-----+-----+
|   jack|        10|
|   alex|         8|
|caroline|         4|
|   paul|         3|
|  sandra|         1|
|casandra|         2|
+-----+-----+
```

```
In [17]: master2.dtypes
```

```
Out[17]: [('Name', 'string'), ('age', 'int'), ('Experience', 'int'), ('Salary', 'int')]
```

```
In [18]: master2.describe().show()
```

```
+-----+-----+-----+-----+-----+
|summary|  Name|          age|      Experience|      Salary|
+-----+-----+-----+-----+-----+
|  count|     6|           6|           6|           6|
|   mean|  NULL|26.33333333333332|4.66666666666667|21333.333333333332|
| stddev|  NULL|4.179314138308661|3.559026084010437| 5354.126134736337|
|    min| alex|          21|           1|      15000|
|    max|sandra|          31|          10|      30000|
+-----+-----+-----+-----+-----+
```

```
In [19]: master2=master2.withColumn('Experience After 2 year',master2['Experience']+2)
```

```
In [20]: master2.show()
```

```
+-----+-----+-----+-----+-----+
|   Name|age|Experience|Salary|Experience After 2 year|
+-----+-----+-----+-----+-----+
|   jack| 31|        10| 30000|                12|
|   alex| 30|         8| 25000|                10|
|caroline| 29|         4| 20000|                 6|
|   paul| 24|         3| 20000|                 5|
|  sandra| 21|         1| 15000|                 3|
|casandra| 23|         2| 18000|                 4|
+-----+-----+-----+-----+-----+
```

```
In [22]: master2=master2.drop('Experience After 2 year')
```

In [23]: master2.show()

Name	age	Experience	Salary
jack	31	10	30000
alex	30	8	25000
caroline	29	4	20000
paul	24	3	20000
sandra	21	1	15000
casandra	23	2	18000

In [25]: master2.withColumnRenamed('Name', 'New Name').show()

New Name	age	Experience	Salary
jack	31	10	30000
alex	30	8	25000
caroline	29	4	20000
paul	24	3	20000
sandra	21	1	15000
casandra	23	2	18000

In []: