

```
In [1]: from pyspark.sql import SparkSession
spark=SparkSession.builder.appName('Aggregate').getOrCreate()
```

```
In [2]: spark
```

**Out[2]: SparkSession - in-memory  
SparkContext**

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

**Version**

v3.5.1

**Master**

local[\*]

**AppName**

Aggregate

```
In [3]: master2=spark.read.csv(r"D:\NIT\FEBRUARY\26 feb (time series and spark)\sample
```

```
In [4]: master2.show()
```

Name	Departments	salary
jack	Data Science	10000
alex	ML	5000
caroline	AI	4000
jack	AI	4000
sandra	Data Science	3000
jack	NLP	20000
alex	AI	10000
kathy	ML	5000
andrew	Data Science	10000
jack	NLP	2000

```
In [5]: master2.printSchema()
```

```
root
|-- Name: string (nullable = true)
|-- Departments: string (nullable = true)
|-- salary: integer (nullable = true)
```

```
In [6]: master2.groupBy('Name').sum().show()
```

```
+-----+-----+
|   Name|sum(salary)|
+-----+-----+
|   jack|      36000|
| andrew|      10000|
| sandra|       3000|
|   alex|      15000|
|  kathy|       5000|
|caroline|       4000|
+-----+-----+
```

```
In [7]: master2.groupBy('Name').avg().show()
```

```
+-----+-----+
|   Name|avg(salary)|
+-----+-----+
|   jack|      9000.0|
| andrew|     10000.0|
| sandra|      3000.0|
|   alex|      7500.0|
|  kathy|      5000.0|
|caroline|      4000.0|
+-----+-----+
```

```
In [8]: master2.groupBy('Departments').sum().show()
```

```
+-----+-----+
|Departments|sum(salary)|
+-----+-----+
|          NLP|      22000|
|          AI|      18000|
|          ML|      10000|
|Data Science|      23000|
+-----+-----+
```

```
In [9]: master2.groupBy('Departments').mean().show()
```

```
+-----+-----+
|Departments|    avg(salary)|
+-----+-----+
|          NLP|      11000.0|
|          AI|       6000.0|
|          ML|       5000.0|
|Data Science|  7666.666666666667|
+-----+-----+
```

```
In [10]: master2.groupBy('Departments').count().show()
```

```
+-----+-----+
| Departments|count|
+-----+-----+
|           NLP|    2|
|           AI |    3|
|           ML |    2|
|Data Science|    3|
+-----+-----+
```

```
In [11]: master2.agg({'Salary': 'sum'}).show()
```

```
+-----+
|sum(Salary)|
+-----+
|      73000|
+-----+
```

```
In [ ]:
```