

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

Out[1]: **SparkSession - in-memory**
SparkContext

[Spark UI \(http://DESKTOP-5NN8I6U:4040\)](http://DESKTOP-5NN8I6U:4040)

Version

v3.5.1

Master

local[*]

AppName

Filter_operations

```
In [4]: 1 df=spark.read.csv(r'C:\Users\Neha\Downloads\NARESH IT DATA SCIENCE\MAY\
        ◀────────────────────────────────────────▶▶
```

```
In [5]: 1 df.filter("Salary <= 20000").show()
```

```
+-----+-----+-----+-----+
|   Name|age|Experience|Salary|
+-----+-----+-----+-----+
|caroline| 29|         4| 20000|
|   paul| 24|         3| 20000|
|  sandra| 21|         1| 15000|
|casandra| 23|         2| 18000|
+-----+-----+-----+-----+
```

```
In [6]: 1 df.filter("Salary >= 20000").show()
```

```
+-----+-----+-----+-----+
|   Name| age|Experience|Salary|
+-----+-----+-----+-----+
|   jack|  31|        10| 30000|
|   alex|  30|         8| 25000|
|caroline| 29|         4| 20000|
|   paul| 24|         3| 20000|
|   dan|NULL|        NULL| 40000|
|  NULL| 34|        10| 38000|
+-----+-----+-----+-----+
```

```
In [7]: 1 df.filter("Salary > 20000").show()
```

```
+-----+-----+-----+-----+
|Name| age|Experience|Salary|
+-----+-----+-----+-----+
|jack|  31|        10| 30000|
|alex|  30|         8| 25000|
| dan|NULL|        NULL| 40000|
|NULL|  34|        10| 38000|
+-----+-----+-----+-----+
```

```
In [8]: 1 df.filter("Salary < 20000").show()
```

```
+-----+-----+-----+-----+
|   Name|age|Experience|Salary|
+-----+-----+-----+-----+
| sandra| 21|         1| 15000|
|casandra| 23|         2| 18000|
+-----+-----+-----+-----+
```

```
In [10]: 1 df.filter("Salary <= 20000").select(['Name', 'age']).show()
```

```
+-----+-----+
|   Name|age|
+-----+-----+
|caroline| 29|
|   paul| 24|
|   sandra| 21|
|casandra| 23|
+-----+-----+
```

```
In [11]: 1 df.filter("Salary <= 20000").select(['Name', 'age', 'Salary']).show()
```

```
+-----+-----+-----+
|   Name|age|Salary|
+-----+-----+-----+
|caroline| 29| 20000|
|   paul| 24| 20000|
|   sandra| 21| 15000|
|casandra| 23| 18000|
+-----+-----+-----+
```

```
In [12]: 1 df.filter("Salary == 20000").select(['Name', 'age', 'Salary']).show()
```

```
+-----+-----+-----+
|   Name|age|Salary|
+-----+-----+-----+
|caroline| 29| 20000|
|   paul| 24| 20000|
+-----+-----+-----+
```

In [21]: 1 df.filter((df["Salary"] < 2000) & (df["Salary"] > 5000)).show()

```
+-----+-----+-----+-----+
|Name|age|Experience|Salary|
+-----+-----+-----+-----+
+-----+-----+-----+-----+
```

In [24]: 1 df.filter((df["Salary"] < 20000) | (df["Salary"] > 50000)).show()

```
+-----+-----+-----+-----+
|   Name|age|Experience|Salary|
+-----+-----+-----+-----+
|  sandra| 21|         1| 15000|
|casandra| 23|         2| 18000|
+-----+-----+-----+-----+
```

In [25]: 1 df.filter((df["Salary"] < 2000) | (df["Salary"] > 5000)).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|
|     dan|NULL|        NULL| 40000|
|   NULL| 34|         10| 38000|
+-----+-----+-----+-----+
```

In [27]: 1 df.filter(~(df["Salary"] <= 2000)).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|
|     dan|NULL|        NULL| 40000|
|   NULL| 34|         10| 38000|
+-----+-----+-----+-----+
```

In []: 1

