

```

import os
os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-8-openjdk-amd64"
os.environ["SPARK_HOME"] = "/content/spark-3.1.1-bin-hadoop3.2"

!apt-get install openjdk-8-jdk-headless -qq > /dev/null
!wget -q http://archive.apache.org/dist/spark/spark-3.1.1/spark-3.1.1-bin-hadoop3.2.tgz
!tar xf spark-3.1.1-bin-hadoop3.2.tgz
!pip install -q findspark

import warnings
warnings.filterwarnings('ignore')

import findspark
findspark.init()
from pyspark.sql import SparkSession
spark = SparkSession.builder.master("local[*]").getOrCreate()
spark.conf.set("spark.sql.repl.eagerEval.enabled", True) # Property used to format output
spark

```

SparkSession - in-memory**SparkContext**[Spark UI](#)

Version

v3.1.1

Master

local[*]

AppName

pyspark-shell

spark

SparkSession - in-memory**SparkContext**[Spark UI](#)

Version

v3.1.1

Master

local[*]

AppName

pyspark-shell

```

from pyspark.sql import Row
from pyspark.sql.types import StructField, StructType, StringType, LongType

```

```
t = StructType(
```

```
[
    StructField(name="Sex", dataType=StringType(), nullable=True),
    StructField(name="Class", dataType=StringType(), nullable=True),
    StructField(name="Survived", dataType=LongType(), nullable=False),
    StructField(name="Died", dataType=LongType(), nullable=False),
]
)
```

```
r = [
    Row('Children', 'First', 6, 0),
    Row('Children', 'Second', 24, 0),
    Row('Children', 'Third', 27, 52),
    Row('Men', 'First', 57, 118),
    Row('Men', 'Second', 14, 154),
    Row('Men', 'Third', 75, 387),
    Row('Men', 'Crew', 192, 693),
    Row('Women', 'First', 140, 4),
    Row('Women', 'Second', 80, 13),
    Row('Women', 'Third', 75, 387),
    Row('Women', 'Crew', 20, 3),
]
```

```
parallelizeRows = spark.sparkContext.parallelize(r)
```

```
df1 = spark.createDataFrame(parallelizeRows, t)
```

```
df1.show()
```

```
+-----+-----+-----+-----+
|      Sex|  Class|Survived|Died|
+-----+-----+-----+-----+
|Children| First|      6|   0|
|Children|Second|     24|   0|
|Children| Third|     27|  52|
|      Men| First|     57| 118|
|      Men|Second|     14| 154|
|      Men| Third|     75| 387|
|      Men|  Crew|    192| 693|
|    Women| First|    140|   4|
|    Women|Second|     80|  13|
|    Women| Third|     75| 387|
|    Women|  Crew|     20|   3|
+-----+-----+-----+-----+
```

```
df1=df1.filter(df1.Class != "Crew")
```

```
df1.show()
```

```
+-----+-----+-----+-----+
|      Sex|  Class|Survived|Died|
```

| | Sex | Class | Survived | Died |
|----------|--------|-------|----------|------|
| Children | First | 6 | 0 | |
| Children | Second | 24 | 0 | |
| Children | Third | 27 | 52 | |
| Men | First | 57 | 118 | |
| Men | Second | 14 | 154 | |
| Men | Third | 75 | 387 | |
| Women | First | 140 | 4 | |
| Women | Second | 80 | 13 | |
| Women | Third | 75 | 387 | |

```
res1=df1.Survived+df1.Died
```

```
df1=df1.withColumn('total_people', res1)
```

```
df1.show()
```

| | Sex | Class | Survived | Died | total_people |
|----------|--------|-------|----------|------|--------------|
| Children | First | 6 | 0 | 6 | |
| Children | Second | 24 | 0 | 24 | |
| Children | Third | 27 | 52 | 79 | |
| Men | First | 57 | 118 | 175 | |
| Men | Second | 14 | 154 | 168 | |
| Men | Third | 75 | 387 | 462 | |
| Women | First | 140 | 4 | 144 | |
| Women | Second | 80 | 13 | 93 | |
| Women | Third | 75 | 387 | 462 | |

```
df1=df1.drop('total_people')
```

```
df1.show()
```

| | Sex | Class | Survived | Died |
|----------|--------|-------|----------|------|
| Children | First | 6 | 0 | |
| Children | Second | 24 | 0 | |
| Children | Third | 27 | 52 | |
| Men | First | 57 | 118 | |
| Men | Second | 14 | 154 | |
| Men | Third | 75 | 387 | |
| Women | First | 140 | 4 | |
| Women | Second | 80 | 13 | |
| Women | Third | 75 | 387 | |

```
df1.where("(Survived*100)/(Survived+Died)>80").show()
```

↗

| | Sex | Class | Survived | Died |
|---|----------|--------|----------|------|
| ↗ | Children | First | 6 | 0 |
| | Children | Second | 24 | 0 |
| | Women | First | 140 | 4 |
| | Women | Second | 80 | 13 |

✓ 0s completed at 12:11 PM

● ✕