databricksRDD Exploration

(https://databricks.com)

```
from pyspark import SparkContext


# Create a SparkContext
sc = SparkContext("local", "RDD Exploration")
```

```
ValueError: Cannot run multiple SparkContexts at once; existing SparkContext(app=Databricks Shell, master=local[8]) created by __init__ at /dat
abricks/python_shell/dbruntime/spark_connection.py:127
```

```
# Create an RDD from a list
data = [1, 2, 3, 4, 5]
rdd = sc.parallelize(data)


# Perform some basic operations on the RDD
# 1. Count the number of elements
count = rdd.count()
print("Number of elements:", count)
```

```
Number of elements: 5
```

```
# 2. Sum all elements
total_sum = rdd.sum()
print("Sum of all elements:", total_sum)
```

```
Sum of all elements: 15
```

```
# 3. Calculate the mean
mean = total_sum / count
print("Mean of elements:", mean)
```

```
Mean of elements: 3.0
```

```
# 4. Find the maximum and minimum elements
max_element = rdd.max()
min_element = rdd.min()
print("Maximum element:", max_element)
print("Minimum element:", min_element)
```

```
Maximum element: 5
Minimum element: 1
```

```
# 5. Filter elements greater than 3
filtered_rdd = rdd.filter(lambda x: x > 3)
print("Elements greater than 3:", filtered_rdd.collect())
```

```
Elements greater than 3: [4, 5]
```

```
# 6. Map operation to square each element
squared_rdd = rdd.map(lambda x: x*x)
print("Squared elements:", squared_rdd.collect())
```

```
Squared elements: [1, 4, 9, 16, 25]

Sum of elements using reduce: 15
```