(1) WhatsApp | Home Page - Select or create a | createrdd - Jupyter Notebook | Learning Apache Spark with Py

localhost:8888/notebooks/createrdd.ipynb#By-using-createDataFrame(-)-function

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

In [1]:
```
import pyspark
```

In [2]:
```
from pyspark.sql import SparkSession
```

In [3]:
```
spark =SparkSession.builder.appName("FirstProgram").getOrCreate()
```

In [4]:
```
spark
```

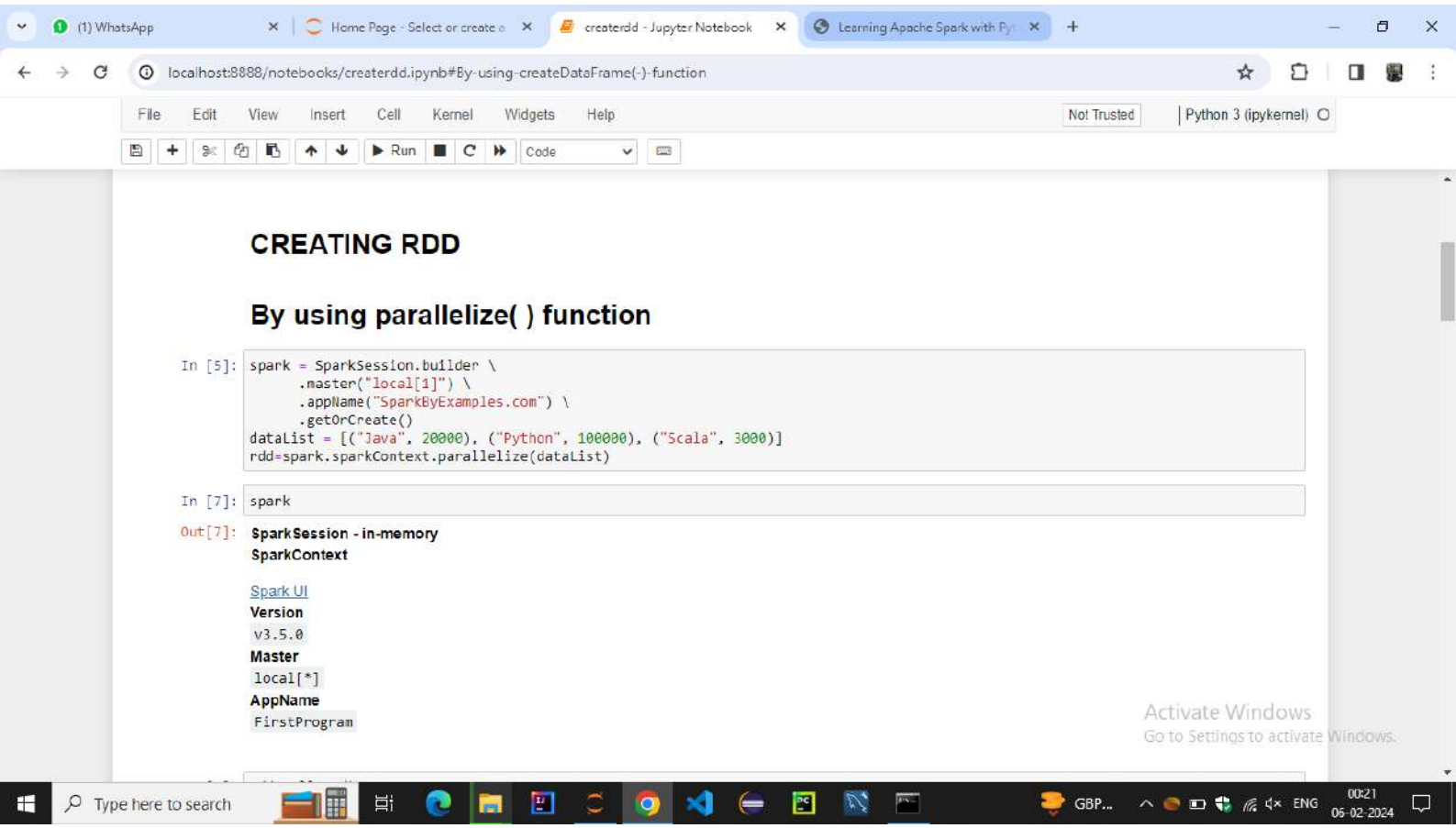
Out[4]:
SparkSession - in-memory
SparkContext

Spark UI
Version
`v3.5.0`
Master
`local[*]`
AppName
`FirstProgram`

CREATING RDD

By using parallelize( ) function

Type here to search GBP... ENG 00:21 06-02-2024


(1) WhatsApp
Home Page - Select or create a
createrdd - Jupyter Notebook
Learning Apache Spark with Py
localhost:8888/notebooks/createrdd.ipynb#By-using-createDataFrame(-)-function
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
CREATING RDD
By using parallelize( ) function
In [5]:
spark = SparkSession.builder \
.master("local[1]") \
.appName("SparkByExamples.com") \
.getOrCreate()
dataList = [("Java", 20000), ("Python", 100000), ("Scala", 3000)]
rdd=spark.sparkContext.parallelize(dataList)
In [7]:
spark
Out[7]:
SparkSession - in-memory
SparkContext
Spark UI
Version
v3.5.0
Master
local[*]
AppName
FirstProgram
Activate Windows
Go to Settings to activate Windows.
Type here to search
GBP...
ENG
00:21
06-02-2024

(1) WhatsApp | Home Page - Select or create a | createrdd - Jupyter Notebook | Learning Apache Spark with Py

localhost:8888/notebooks/createrdd.ipynb#By-using-createDataFrame(-)-function

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code

Out[8]: [('Java', 20000), ('Python', 100000), ('Scala', 3000)]

Read dataset from .csv file

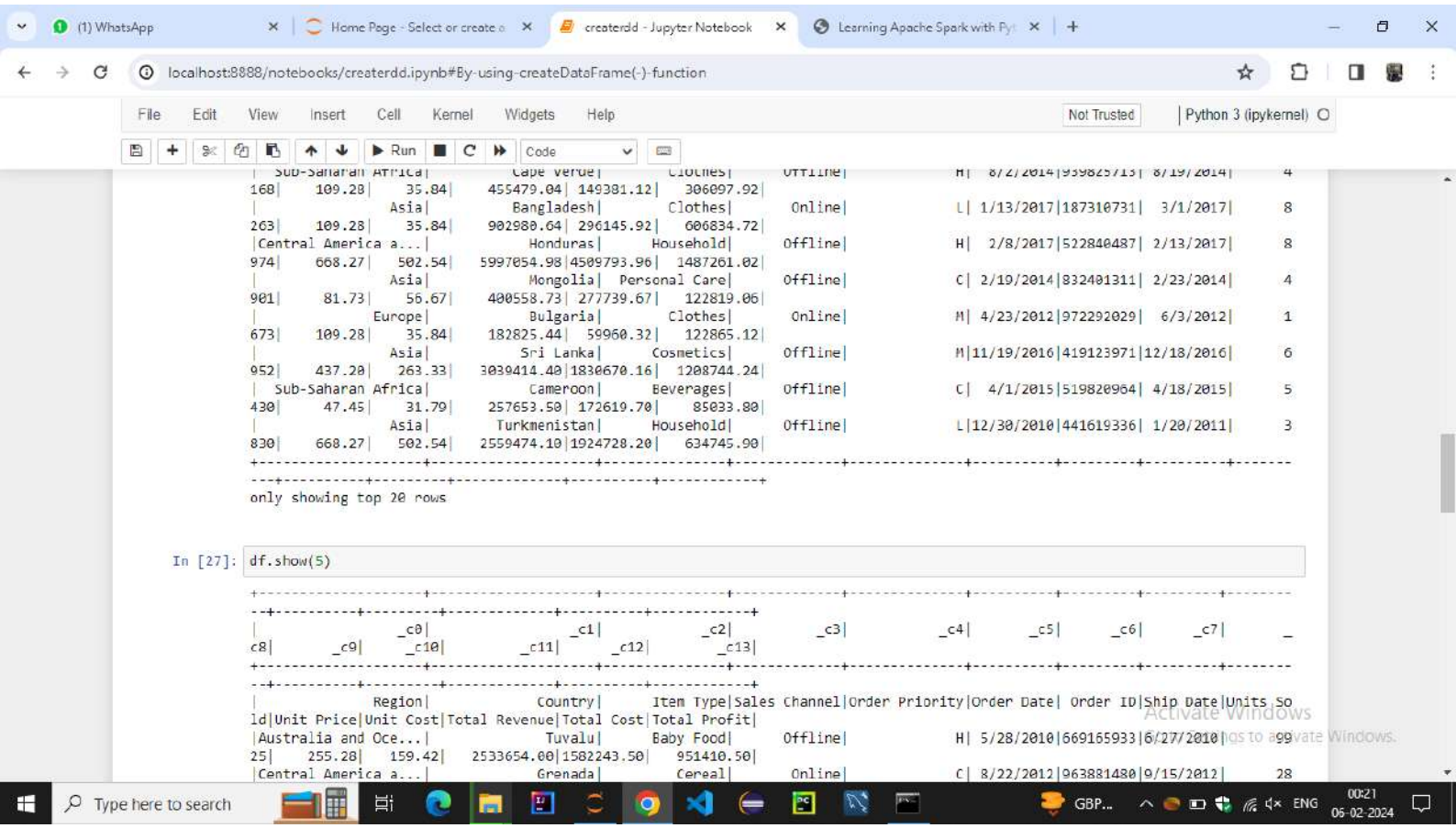
In [9]:
```
spark = SparkSession.builder \
      .master("local[1]") \
      .appName("SparkByExamples.com") \
      .getOrCreate()

df=spark.read.csv("Desktop/SalesRecords.csv")
```

In [10]:
```
df.show()
```

```
+--------------------+--------------------+--------------+-------------+--------------+----------+--------+----------+-------
---+----------+---------+-------------+-----------+------------+
|                 _c0|                 _c1|           _c2|          _c3|           _c4|       _c5|     _c6|       _c7|
_c8|       _c9|     _c10|         _c11|       _c12|         _c13|
+--------------------+--------------------+--------------+-------------+--------------+----------+--------+----------+-------
---+----------+---------+-------------+-----------+------------+
|              Region|             Country|     Item Type|Sales Channel|Order Priority|Order Date|Order ID| Ship Date|Units S
old|Unit Price|Unit Cost|Total Revenue|Total Cost|Total Profit|
|Australia and Oce...|              Tuvalu|     Baby Food|      Offline|             H| 5/28/2010|669165933| 6/27/2010|      9
925|    255.28|   159.42|   2533654.00|1582243.50|  951410.50|
|Central America a...|            Grenada|        Cereal|       Online|             C| 8/22/2012|963881480| 9/15/2012|      2
804|    205.70|   117.11|    576782.80| 328376.44|  248406.36|
|              Europe|              Russia|Office Supplies|      Offline|             L|  5/2/2014|341417157|  5/8/2014|      1
779|    651.21|   524.96|   1158502.59| 933903.84|  224598.75|
|  Sub-Saharan Africa|Sao Tome and Prin...|         Fruits|       Online|             C| 6/20/2014|514321792| 7/5/2014|      8
102|      9.33|     6.92|     75591.66|  56065.84|   19525.82|
|  Sub-Saharan Africa|              Rwanda|Office Supplies|      Offline|             L|  2/1/2013|115456712|  2/6/2013|      5
```


Type here to search | GBP... | ENG | 00:21 06-02-2024




```
| Sub-Saharan Africa|          Cape Verde|       Clothes|      Offline|          H|  8/2/2014|939825713| 8/19/2014|      4
168|     109.28|     35.84|     455479.04| 149381.12|    306097.92|
|               Asia|          Bangladesh|       Clothes|       Online|          L| 1/13/2017|187310731|  3/1/2017|      8
263|     109.28|     35.84|     902980.64| 296145.92|    606834.72|
|Central America a...|            Honduras|      Household|      Offline|          H|  2/8/2017|522840487| 2/13/2017|      8
974|     668.27|    502.54|    5997654.98|4509793.96|   1487261.02|
|               Asia|            Mongolia|  Personal Care|      Offline|          C| 2/19/2014|832401311| 2/23/2014|      4
901|      81.73|     56.67|     400558.73| 277739.67|    122819.06|
|             Europe|            Bulgaria|       Clothes|       Online|          M| 4/23/2012|972292029|  6/3/2012|      1
673|     109.28|     35.84|     182825.44|  59960.32|    122865.12|
|               Asia|           Sri Lanka|      Cosmetics|      Offline|          M|11/19/2016|419123971|12/18/2016|      6
952|     437.20|    263.33|    3039414.40|1830670.16|   1208744.24|
| Sub-Saharan Africa|            Cameroon|      Beverages|      Offline|          C|  4/1/2015|519820964| 4/18/2015|      5
430|      47.45|     31.79|     257653.50| 172619.70|     85033.80|
|               Asia|          Turkmenistan|      Household|      Offline|          L|12/30/2010|441619336| 1/20/2011|      3
830|     668.27|    502.54|    2559474.10|1924728.20|    634745.90|
+--------------------+--------------------+--------------+-------------+-------------+----------+---------+----------+-------
---+----------+---------+-------------+----------+------------+
only showing top 20 rows
```

In [27]:
```
df.show(5)
```

```
+--------------------+--------------------+--------------+-------------+-------------+----------+---------+----------+--------
--+----------+---------+-------------+----------+------------+
|                 _c0|                 _c1|           _c2|          _c3|          _c4|       _c5|      _c6|      _c7|       _
c8|       _c9|      _c10|         _c11|       _c12|        _c13|
+--------------------+--------------------+--------------+-------------+-------------+----------+---------+----------+--------
--+----------+---------+-------------+----------+------------+
|              Region|             Country|     Item Type|Sales Channel|Order Priority|Order Date| Order ID|Ship Date|Units So
ld|Unit Price|Unit Cost|Total Revenue|Total Cost|Total Profit|
|Australia and Oce...|              Tuvalu|     Baby Food|      Offline|          H| 5/28/2010|669165933| 6/27/2010|      99
25|    255.28|   159.42|   2533654.00|1582243.50|   951410.50|
|Central America a...|             Grenada|        Cereal|       Online|          C| 8/22/2012|963881480| 9/15/2012|      28
```

(1) WhatsApp | Home Page - Select or create a | createrdd - Jupyter Notebook | Learning Apache Spark with Py

localhost:8888/notebooks/createrdd.ipynb#By-using-createDataFrame(-)-function

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code

```
|                Europe|                Russia|Office Supplies|     Offline|          C| 5/2/2014|541417157| 5/8/2014|      17
79|    651.21|    524.96|   1158502.59| 933903.84|   224598.75|
|  Sub-Saharan Africa|Sao Tome and Prin...|         Fruits|      Online|          C| 6/20/2014|514321792| 7/5/2014|      81
02|      9.33|      6.92|      75591.66|  56065.84|    19525.82|
+--------------------+--------------------+---------------+------------+-----------+----------+---------+---------+--------
--+----------+---------+-------------+----------+------------+
only showing top 5 rows
```

In [ ]:

In [4]:
```
from pyspark.sql import SparkSession

# Create a Spark session
spark = SparkSession.builder.appName("example").getOrCreate()

# Create an RDD using parallelize
rdd_data = [1, 2, 3, 4, 5]
rdd = spark.sparkContext.parallelize(rdd_data)

# Display RDD contents
print("RDD content:")
rdd.collect()
```

```
RDD content:
```

Out[4]: [1, 2, 3, 4, 5]

In [11]:
```
from pyspark.sql import SparkSession
spark = SparkSession \
.builder \
```


Type here to search | GBP... | ENG | 00:22 06-02-2024