

```
sreeti@sreeti-VirtualBox: ~
21/05/23 23:37:28 WARN Utils: Your hostname, sreeti-VirtualBox resolves to a lo
opback address: 127.0.1.1; using 10.0.2.15 instead (on interface enp0s3)
21/05/23 23:37:28 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another
address
21/05/23 23:37:35 WARN NativeCodeLoader: Unable to load native-hadoop library f
or your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLev
el(newLevel).
Welcome to

  ____      _
 / ___|    / \
| |  | |  / _ \
| |  | | / ___ \
| |  | |/_/   \_\
| |  | |
|_|  |_|

 version 3.1.1

Using Python version 3.9.4 (default, Apr  4 2021 19:38:44)
Spark context Web UI available at http://10.0.2.15:4040
Spark context available as 'sc' (master = local[*,] app id = local-162182745991
5).
SparkSession available as 'spark'.
>>> def inside(p):
... x, y = random.random(), random.random()
    File "<stdin>", line 2
        x, y = random.random(), random.random()
        ^
IndentationError: expected an indented block
>>>
```

```
>>> sparkContext.textFile("hdfs://...").flatMap(line => line.split(" ")).map(word
rd => (word, 1)).reduceByKey(_ + _).saveAsTextFile("hdfs://...")
    File "<stdin>", line 1
        sparkContext.textFile("hdfs://...").flatMap(line => line.split(" ")).map(word
rd => (word, 1)).reduceByKey(_ + _).saveAsTextFile("hdfs://...")
                                   ^
SyntaxError: expression cannot contain assignment, perhaps you meant "=="?
```

```
>>> df=spark.read.json("examples/src/main/resources/people.json")j
    File "<stdin>", line 1
        df=spark.read.json("examples/src/main/resources/people.json")j
                                   ^
SyntaxError: invalid syntax
>>> df=spark.read.json("examples/src/main/resources/people.json")
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
    File "/home/sreeti/spark-3.1.1-bin-hadoop2.7/python/pyspark/sql/readwriter.py", line 372, in json
    return self._df(self._jreader.json(self._spark._sc._jvm.PythonUtils.toSeq(p
ath)))
    File "/home/sreeti/spark-3.1.1-bin-hadoop2.7/python/lib/py4j-0.10.9-src.zip/p
y4j/java_gateway.py", line 1304, in __call__
    File "/home/sreeti/spark-3.1.1-bin-hadoop2.7/python/pyspark/sql/utils.py", li
ne 117, in deco
    raise converted from None
pyspark.sql.utils.AnalysisException: Path does not exist: file:/home/sreeti/exa
mples/src/main/resources/people.json
>>> text_file = sc.textFile("hdfs://...")
>>> counts = text_file.flatMap(lambda line: line.split(" ")).map(lambda word: (
```