Magic Number: 45679

```
Ruchi Awasthi@DESKTOP-16T7K2C MINGW64 ~
$ ssh azureSandbox
ruchisharma26@104.43.136.236's password:
Permission denied, please try again.
ruchisharma26@104.43.136.236's password:
System is booting up. See pam_nologin(8)
Connection closed by 104.43.136.236 port 22

Ruchi Awasthi@DESKTOP-16T7K2C MINGW64 ~
$ ssh azureSandbox
ruchisharma26@104.43.136.236's password:
Last failed login: Thu Feb 21 07:47:18 UTC 2019 from 207.237.207.206 on ssh:nott
y
There were 5 failed login attempts since the last successful login.
Last login: Sun Feb 17 08:08:21 2019 from 207.237.207.206
[ruchisharma26@sandbox-host ~]$ ssh -P 2222 maria_dev@localhost
ssh: connect to host 2222 port 22: Invalid argument
[ruchisharma26@sandbox-host ~]$ ssh -p 2222 maria_dev@localhost
maria_dev@localhost's password:
Last login: Sun Feb 17 08:08:53 2019 from 172.17.0.1
[maria_dev@sandbox-hdp ~]$ java TestDataGen
Magic Number = 45679
[maria_dev@sandbox-hdp ~]$
```

Exercise 1) Magic Number: 45679

```
[maria_dev@sandbox-hdp ~]$ pyspark
SPARK_MAJOR_VERSION is set to 2, using Spark2
Python 2.6.6 (r266:84292, Aug 18 2016, 15:13:37)
[GCC 4.4.7 20120313 (Red Hat 4.4.7-17)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLeve
l(newLevel).
19/02/22 19:54:08 WARN Utils: Service 'SparkUI' could not bind on port 4040. Att
empting port 4041.
/usr/hdp/current/spark2-client/python/pyspark/context.py:205: UserWarning: Suppo
rt for Python 2.6 is deprecated as of Spark 2.0.0
  warnings.warn("Support for Python 2.6 is deprecated as of Spark 2.0.0")
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 2.2.0.2.6.4.0-91
      /_/

Using Python version 2.6.6 (r266:84292, Aug 18 2016 15:13:37)
SparkSession available as 'spark'.
>>> sc = SparkContext.getOrCreate()
>>> ex1RDD=sc.textFile('/user/maria_dev/foodratings31602.txt')
ex1RDD.take(5)Exercise 2)>>> ex1RDD=sc.textFile('/user/maria_dev/foodratings3160
>>> sc = SparkContext.getOrCreate()
>>> ex1RDD=sc.textFile('/user/maria_dev/foodratings45679.txt')
>>> ex1RDD.take(5)
[u'Jill,9,16,6,9,4', u'Joy,9,9,25,22,5', u'Mel,43,16,10,29,4', u'Joy,46,15,44,33,5', u'Jill,37,15,46,17,4']
>>>
```

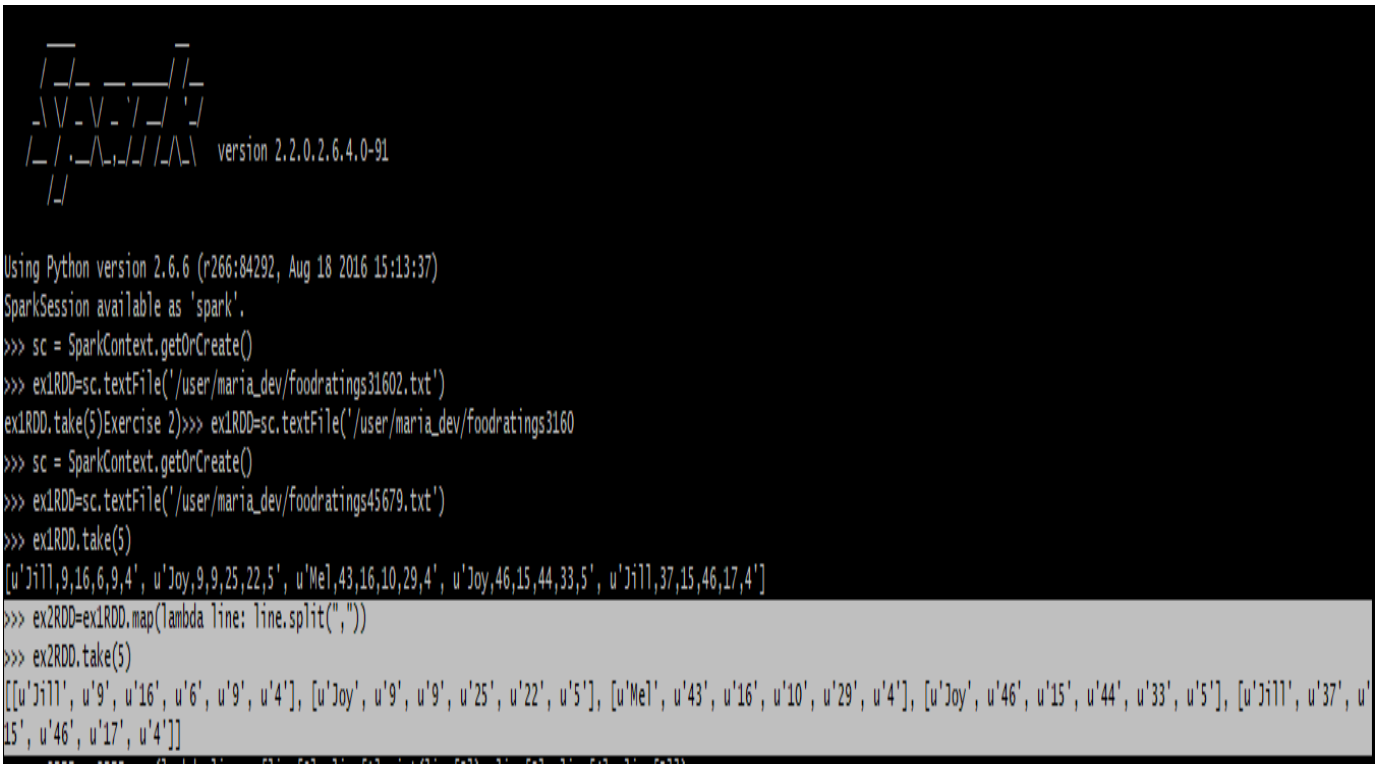
Command:

sc = SparkContext.getOrCreate()

ex1RDD=sc.textFile('/user/maria_dev/foodratings45679.txt')

ex1RDD.take(5)

Exercise 2)

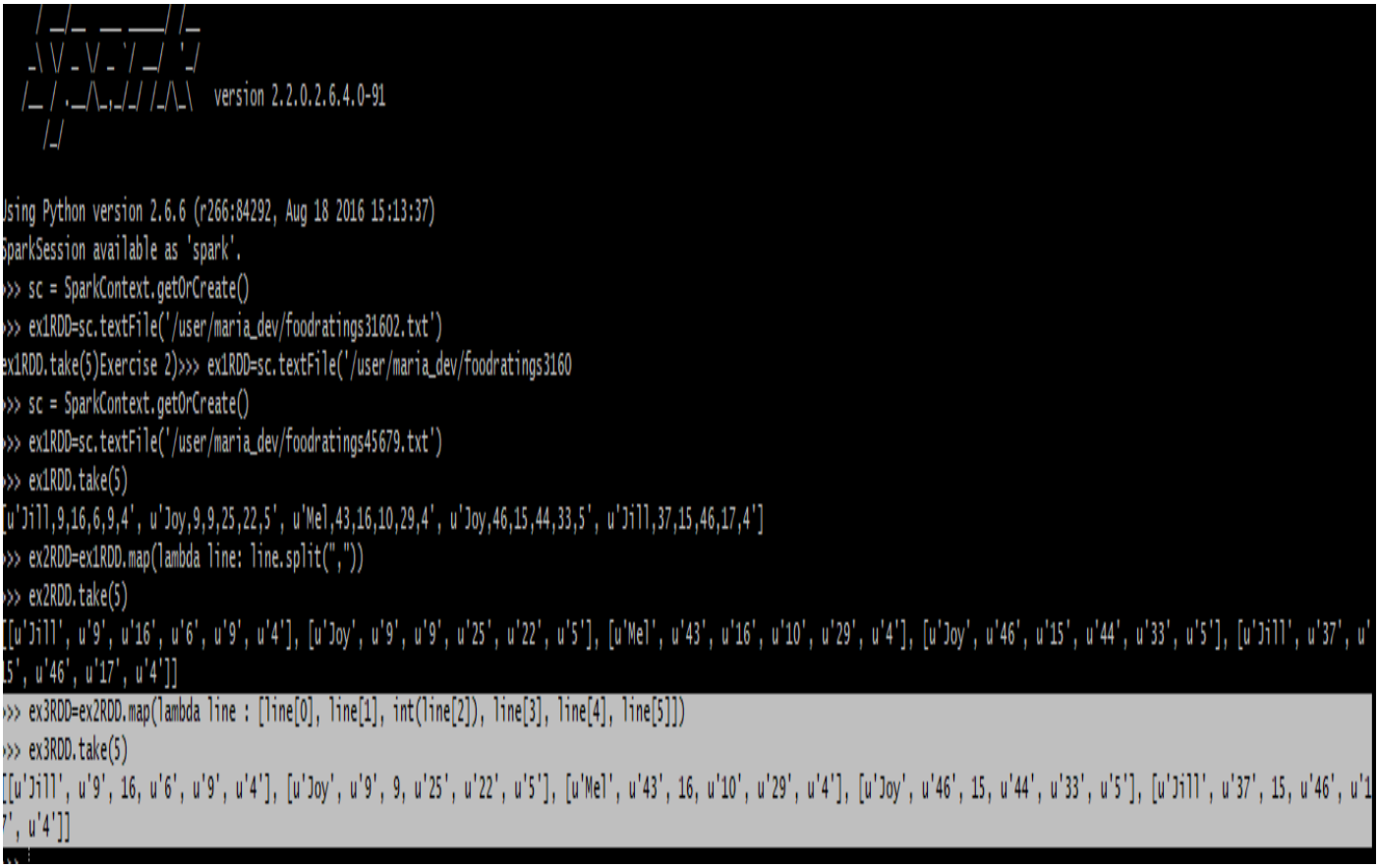


Command:

ex2RDD=ex1RDD.map(lambda line: line.split(","))

ex2RDD.take(5)

Exercise 3)

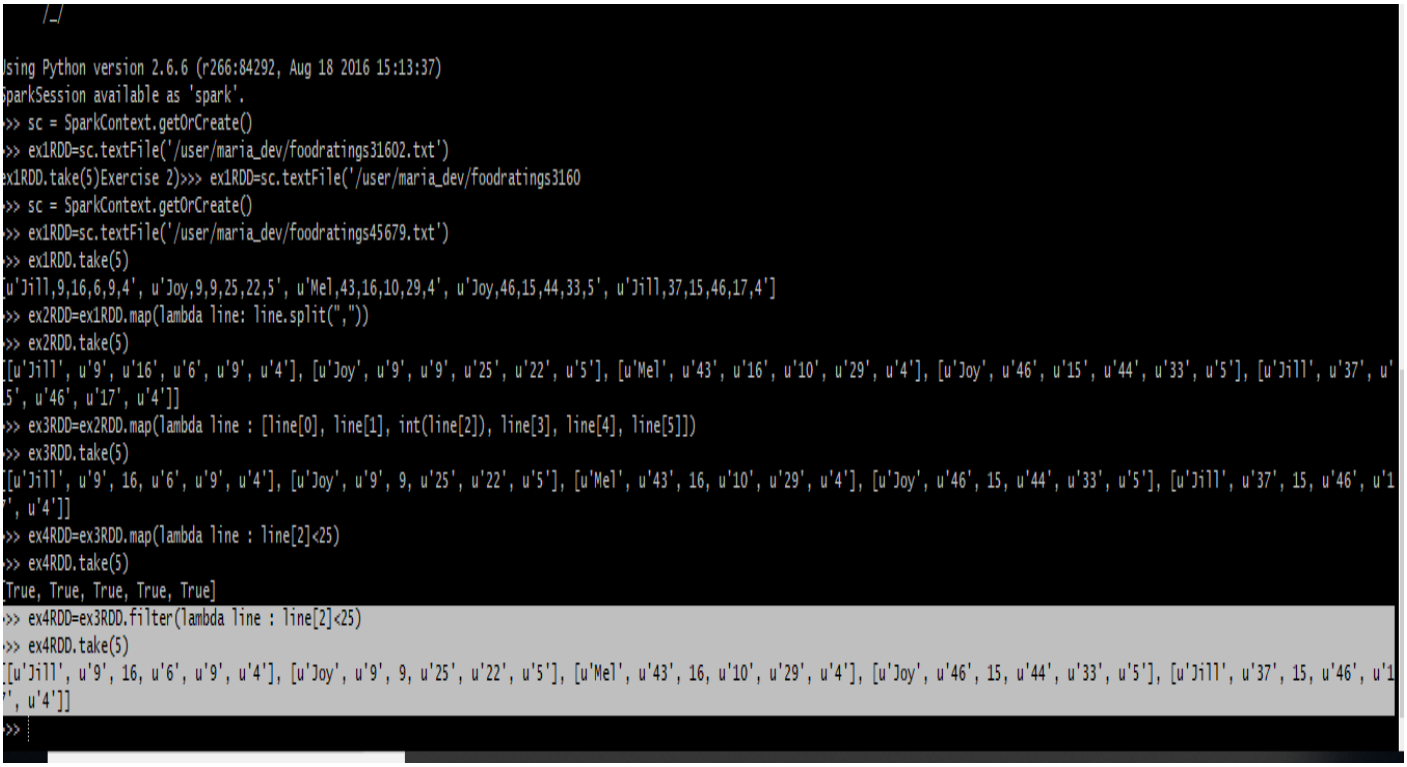


Command:

ex3RDD=ex2RDD.map(lambda line : [line[0], line[1], int(line[2]), line[3], line[4], line[5]])

ex3RDD.take(5)

Exercise 4)



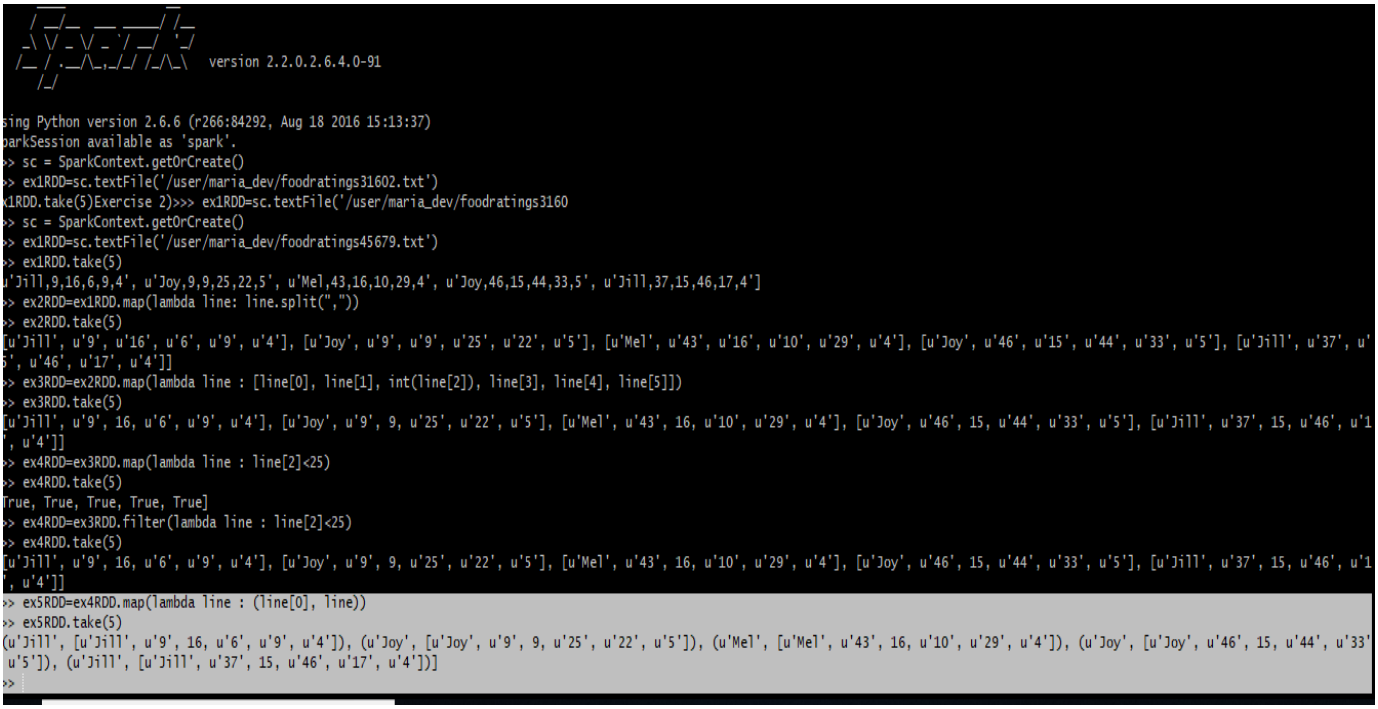
Command:

ex4RDD=ex3RDD.map(lambda line : line[2]<25)

ex4RDD=ex3RDD.filter(lambda line : line[2]<25) //for assignment purpose

ex4RDD.take(5)

Exercise 5)

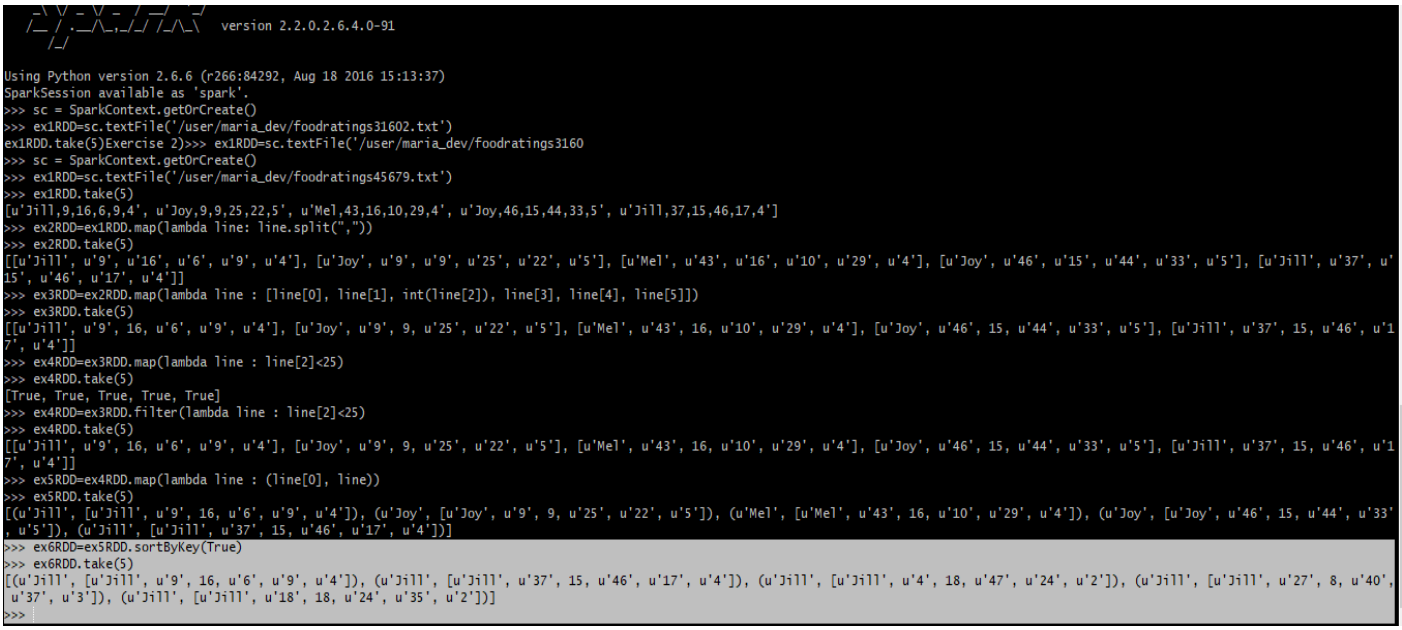


Command:

ex5RDD=ex4RDD.map(lambda line : (line[0], line))

ex5RDD.take(5)

Exercise 6)



ex6RDD=ex5RDD.sortByKey(True)

ex6RDD.take(5)