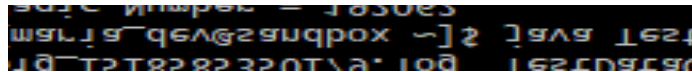


CSP 595 - Assignment 7

Name: Jagruti Vichare
Email ID: jvichare@hawk.iit.edu
CWID: A20378092

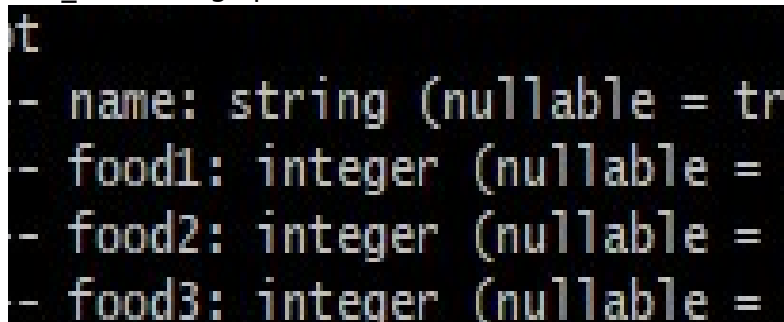
Exercise 1

Step A

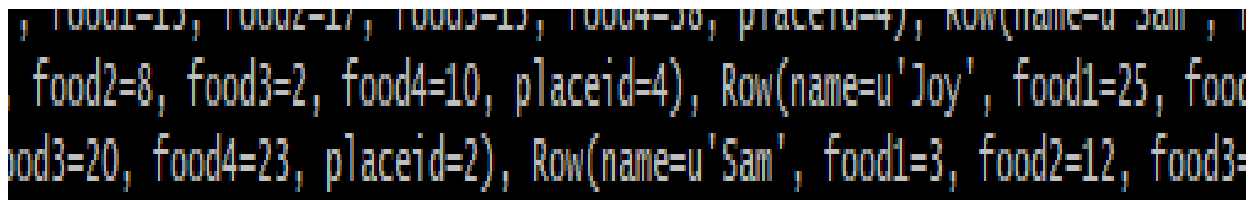


Step B

```
from pyspark.sql.types import *
foodratingstruct = StructType(
    [
        StructField("name", StringType(), True),
        StructField("food1", IntegerType(), True),
        StructField("food2", IntegerType(), True),
        StructField("food3", IntegerType(), True),
        StructField("food4", IntegerType(), True),
        StructField("placeid", IntegerType(), True)
    ]
)
ex1_foodratings=spark.read.schema(foodratingstruct).csv('/user/maria_dev/foodr
atings192062.txt')
ex1_foodratings.printSchema()
```

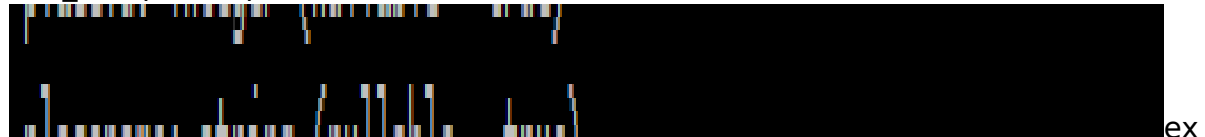


```
ex1_foodratings.head(5)
```

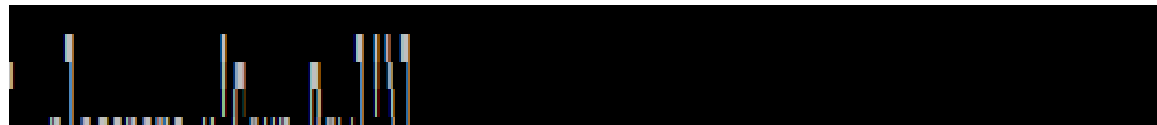


Exercise 2

```
from pyspark.sql.types import *
foodplacestruct = StructType(
    [
        StructField("placeid", IntegerType(), True),
        StructField("placename", StringType(), True)
    ]
)
ex2_foodplaces =
spark.read.schema(foodplacestruct).csv('/user/maria_dev/foodplaces192062.txt')
ex2_foodplaces.printSchema()
```



```
ex2_foodplaces.head(5)
```



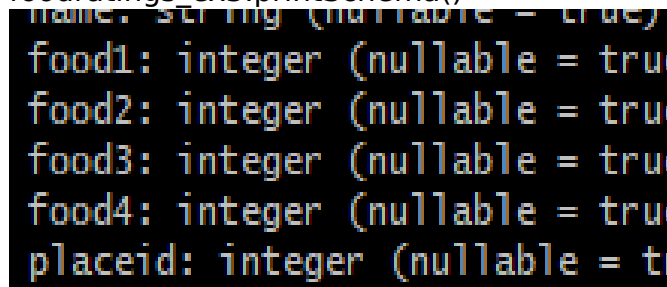
Exercise 3

Step A

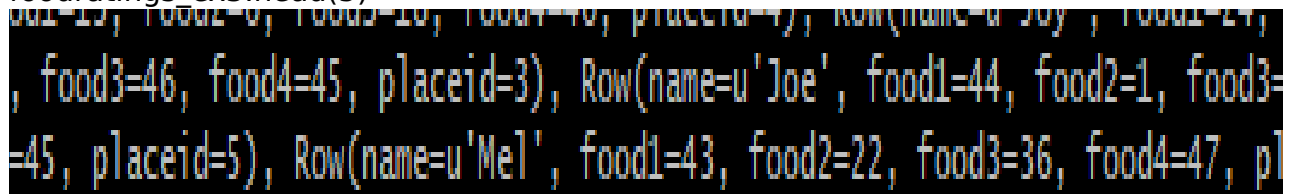
```
from pyspark.sql.types import *
ex1_foodratings.createOrReplaceTempView("foodratingsT")
ex2_foodplaces.createOrReplaceTempView("foodplacesT")
```

Step B

```
foodratings_ex3 = spark.sql("SELECT * FROM foodratingsT WHERE food2 < 25
AND food4 > 40")
foodratings_ex3.printSchema()
```



```
foodratings_ex3.head(5)
```



Step C

```
foodplaces_ex3 = spark.sql("SELECT * FROM foodplacesT WHERE placeid > 3")
foodplaces_ex3.printSchema()
```

```
placeid: integer (nullable = true)
placename: string (nullable = true)
```

```
foodplaces_ex3.head(5)
```

```
(placeid=4, placename="Kombi", placeid=5, placename="Zombi", placeid=6, placename="Zombi", placeid=7, placename="Zombi", placeid=8, placename="Zombi", placeid=9, placename="Zombi")
```

Exercise 4

```
foodratings_ex0 = ex1_foodratings.filter(ex1_foodratings.name=="Mel")
foodratings_ex4 = foodratings_ex0.filter(ex1_foodratings.food3<25)
foodratings_ex4.printSchema()
foodratings_ex4.head(5)
```

```
name: string (nullable = true)
food1: integer (nullable = true)
food2: integer (nullable = true)
food3: integer (nullable = true)
food4: integer (nullable = true)
placeid: integer (nullable = true)

Foodratings_ex4.head(5)
(name=u'Mel', food1=13, food2=5, food3=6, food4=49, placeid=3), Row(name=u'Mel', food1=43, food2=18, food3=8, food4=18, placeid=3), Row(name=u'Mel', food1=36, food2=38, food3=22, food4=11, placeid=3), Row(name=u'Mel', food1=13, food2=45, food3=7, food4=37, placeid=4), Row(name=u'Mel', food1=26, food2=17, food3=16, food4=33, placeid=3)
```

Exercise 5

```
foodratings_ex5 = ex1_foodratings.select('name','placeid')
foodratings_ex5.printSchema()
foodratings_ex5.head(5)
```

```
name: string (nullable = true)
placeid: integer (nullable = true)

foodratings_ex5.head(5)
(name=u'Joe', placeid=4), Row(name=u'Joe', placeid=4), Row(name=u'Sam', placeid=4), Row(name=u'Joe', placeid=3), Row(name=u'Sam', placeid=3)
```

Exercise 6

```
ex6=ex1_foodratings.join(ex2_foodplaces,ex1_foodratings.placeid==ex2_foodplaces.placeid,'inner')
ex6.printSchema()
ex6.head(5)
```

```
name: string (nullable = true)
food1: integer (nullable = true)
food2: integer (nullable = true)
food3: integer (nullable = true)
food4: integer (nullable = true)
placeid: integer (nullable = true)
placeid: integer (nullable = true)
placename: string (nullable = true)
```

```
ex6.head(5)
```

```
(name=u'Joe', food1=44, food2=10, food3=28, food4=8, placeid=4, placeid=4, placename=u"Jake's"
name=u'Joe', food1=13, food2=17, food3=15, food4=38, placeid=4, placeid=4, placename=u"Jake's"
name=u'Sam', food1=33, food2=8, food3=2, food4=10, placeid=4, placeid=4, placename=u"Jake's"),
me=u'Joy', food1=25, food2=43, food3=20, food4=23, placeid=2, placeid=2, placename=u'Atlantic'
name=u'Sam', food1=3, food2=12, food3=27, food4=19, placeid=3, placeid=3, placename=u'Food Tow
```