Quiz 4

MSAN 694, Diane Woodbridge

Description

Use the given *file_name*, *app_name and id* in “user_definition.py” to complete the python code “quiz4.py”.

```
app_name = "sparksql_basic"
file_name = "italianVotes.csv"
id = 2675
```

Using the input file, generate a DataFrame with the format of

id (integer, not-nullable) postId(integer, not-nullable) voteType(integer, nullable) time(timestamp, nullable)	input example.. italianVotes.csv id,postId, voteType, time 2657~135~2~2013-11-22 00:00:00.0 2658~142~2~2013-11-22 00:00:00.0 2659~142~1~2013-11-22 00:00:00.0 2660~140~2~2013-11-22 00:00:00.0 2661~140~1~2013-11-22 00:00:00.0 2662~1354~2~2013-11-22 00:00:00.0 2663~1356~2~2013-11-22 00:00:00.0 2664~1353~2~2013-11-22 00:00:00.0 2665~1351~2~2013-11-22 00:00:00.0 2667~1357~2~2013-11-22 00:00:00.0 2668~1357~2~2013-11-22 00:00:00.0 2669~1351~2~2013-11-22 00:00:00.0 2670~1351~1~2013-11-22 00:00:00.0

“quiz4.py” should **print** the schema of your DataFrame (orders of the columns/rows do not matter.) and **print** postId of the given id.

Submit the quiz4.py file (**ONLY**)- the name of your file should be **quiz4_LastName_Firstname.py** on the link below.

If you run spark-submit quiz4_Woodbridge_Diane.py > out, the output should look like as below.

output

```
root
 |-- id: integer (nullable = false)
 |-- postId: integer (nullable = false)
 |-- time: timestamp (nullable = true)
 |-- voteType: integer (nullable = true)

+------+
|postId|
+------+
|  1340|
+------+
```