PySpark Interview

Interviewer:

You have a dataset containing customer orders, and you need to find the most frequently ordered item. How would you do this using PySpark ?

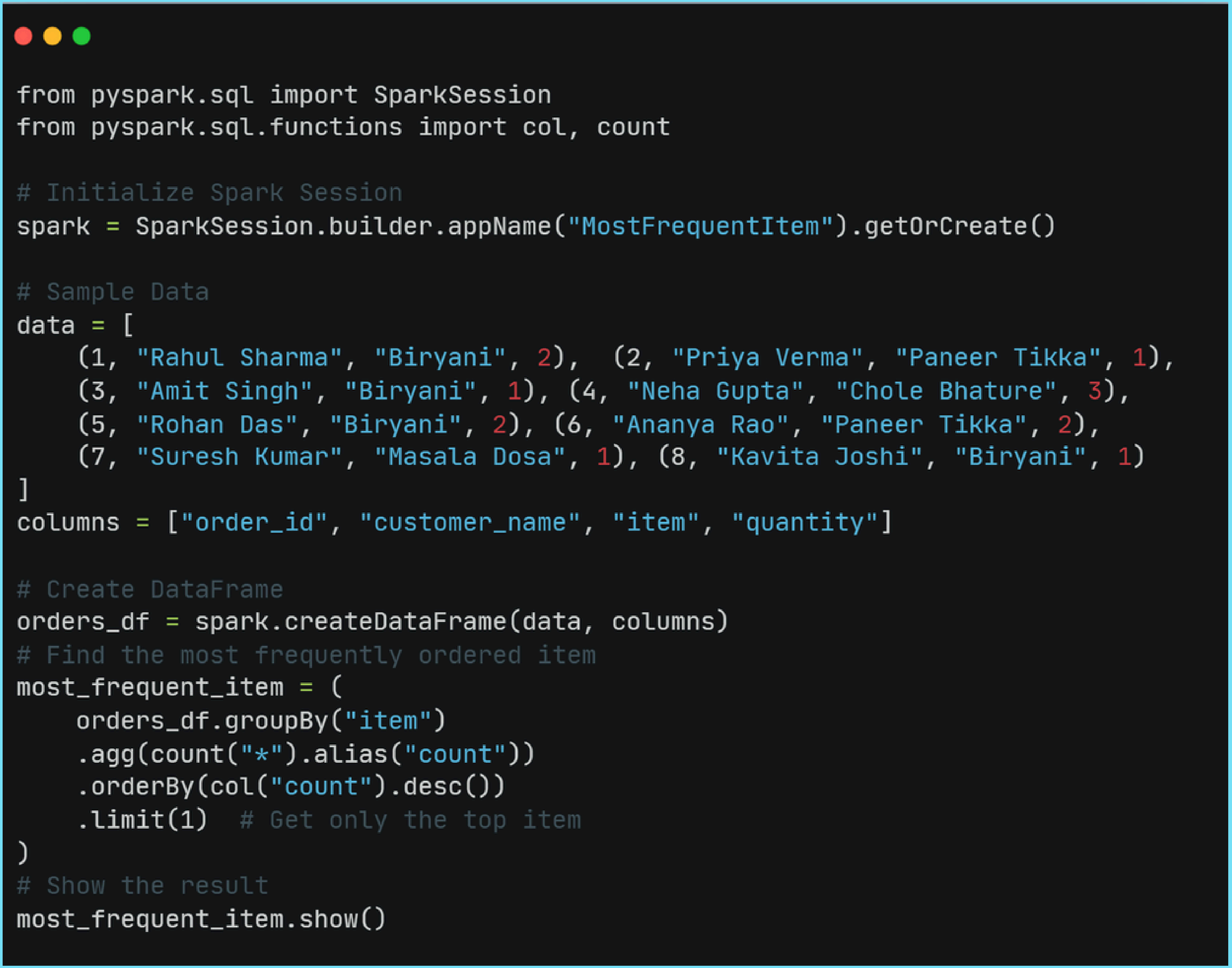
Sample Data :

```
data = [ (1, "Rahul Sharma", "Biryani", 2),
         (2, "Priya Verma", "Paneer Tikka", 1),
         (3, "Amit Singh", "Biryani", 1),
         (4, "Neha Gupta", "Chole Bhature", 3),
         (5, "Rohan Das", "Biryani", 2),
         (6, "Ananya Rao", "Paneer Tikka", 2),
         (7, "Suresh Kumar", "Masala Dosa", 1),
         (8, "Kavita Joshi", "Biryani", 1) ]

columns = ["order_id", "customer_name", "item", "quantity"]
```

https://www.seekhobigdata.com/

Solution :



```
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, count

# Initialize Spark Session
spark = SparkSession.builder.appName("MostFrequentItem").getOrCreate()

# Sample Data
data = [
    (1, "Rahul Sharma", "Biryani", 2),  (2, "Priya Verma", "Paneer Tikka", 1),
    (3, "Amit Singh", "Biryani", 1), (4, "Neha Gupta", "Chole Bhature", 3),
    (5, "Rohan Das", "Biryani", 2), (6, "Ananya Rao", "Paneer Tikka", 2),
    (7, "Suresh Kumar", "Masala Dosa", 1), (8, "Kavita Joshi", "Biryani", 1)
]
columns = ["order_id", "customer_name", "item", "quantity"]

# Create DataFrame
orders_df = spark.createDataFrame(data, columns)
# Find the most frequently ordered item
most_frequent_item = (
    orders_df.groupBy("item")
    .agg(count("*").alias("count"))
    .orderBy(col("count").desc())
    .limit(1)  # Get only the top item
)
# Show the result
most_frequent_item.show()
```

https://www.seekhobigdata.com/

Output:

```
+---------+----------+
|item     | count  |
+---------+----------+
|Biryani |   4      |
+---------+----------+
```

Explanation:

1. We group the data by item to count occurrences.
2. We sort in descending order to get the most frequent item.
3. We use .limit(1) to return only the top result.

https://www.seekhobigdata.com/

IF YOU FIND THIS HELPFUL LIKE AND SHARE

Karthik Kondpak

https://www.seekhobigdata.com/

☎ +91 99894 54737