**Questions:**

1. Setup Confluent Kafka Account

2. Create one kafka topic named as "restaurent-take-away-data" with 3 partitions

3. Setup key (string) & value (json) schema in the confluent schema registry

4. Write a kafka producer program (python or any other language) to read data records from restaurent data csv file, make sure schema is not hardcoded in the producer code, read the latest version of schema and schema\_str from schema registry and use it for data serialization.

5. From producer code, publish data in Kafka Topic one by one and use dynamic key while publishing the records into the Kafka Topic

6. Write kafka consumer code and create two copies of same consumer code and save it with different names (kafka\_consumer\_1.py & kafka\_consumer\_2.py), again make sure lates schema version and schema\_str is not hardcoded in the consumer code, read it automatically from the schema registry to desrialize the data.

Now test two scenarios with your consumer code:

a.) Use "group.id" property in consumer config for both consumers and mention different group\_ids in kafka\_consumer\_1.py & kafka\_consumer\_2.py, apply "earliest" offset property in both consumers and run these two consumers from two different terminals. Calculate how many records each consumer consumed and printed on the terminal

b.) Use "group.id" property in consumer config for both consumers and mention same group\_ids in kafka\_consumer\_1.py & kafka\_consumer\_2.py, apply "earliest" offset property in both consumers and run these two consumers from two different terminals. Calculate how many records each consumer consumed and printed on the terminal

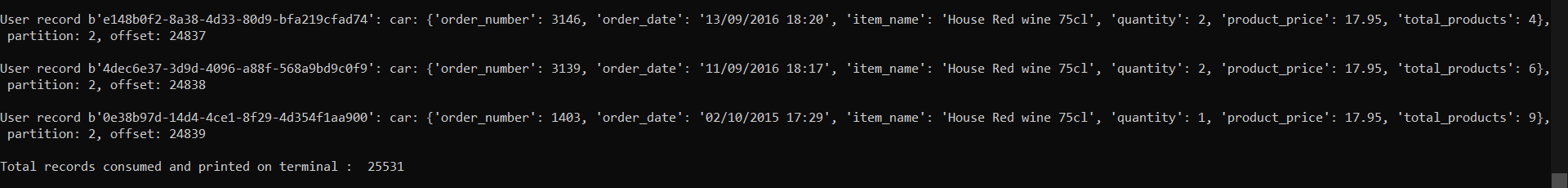
7. Once above questions are done, write another kafka consumer to read data from kafka topic and from the consumer code create one csv file "output.csv" and append consumed records output.csv file

**Solution:**

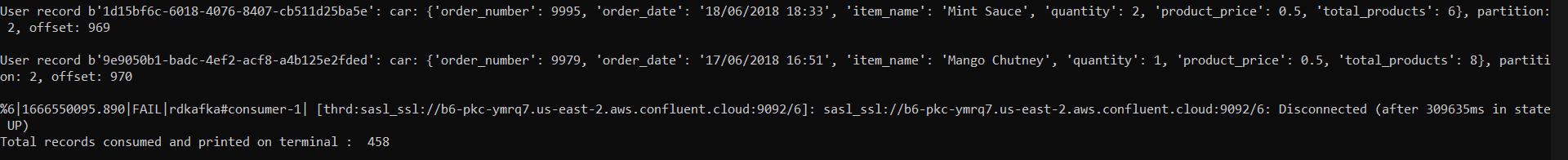
* Total records: 74818

**Note:** Codes were executed before and the offset is set to earlier. Hence the difference in records consumed

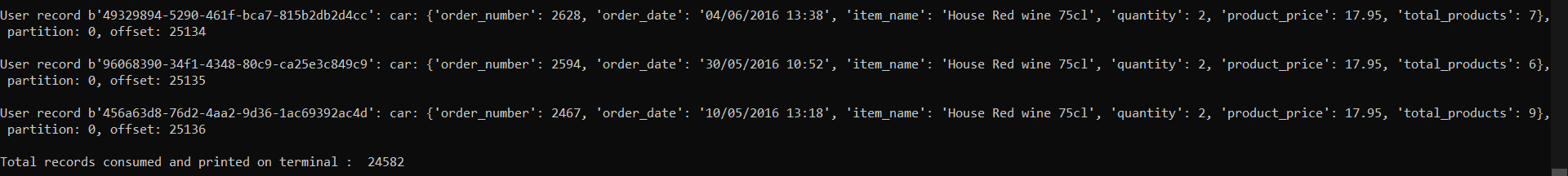
* Records consumed by consumer-1 with group\_id =1:



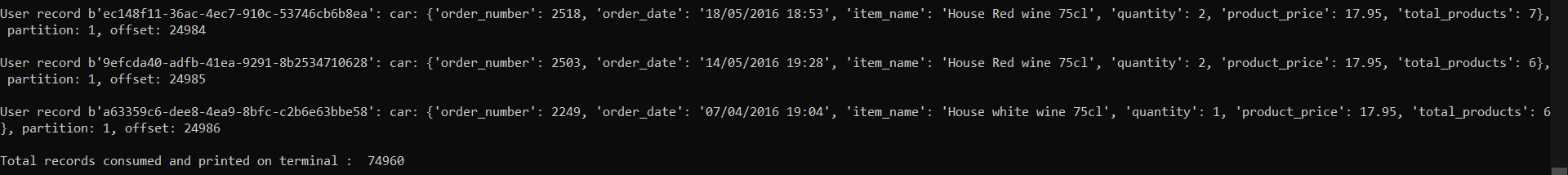
* Records consumed by consumer-2 with group\_id =1:



* Records consumed by consumer-1 with group\_id =1:



* Records consumed by consumer-2 with group\_id =2:



* Total records written into csv file (with group\_id set as 1):

