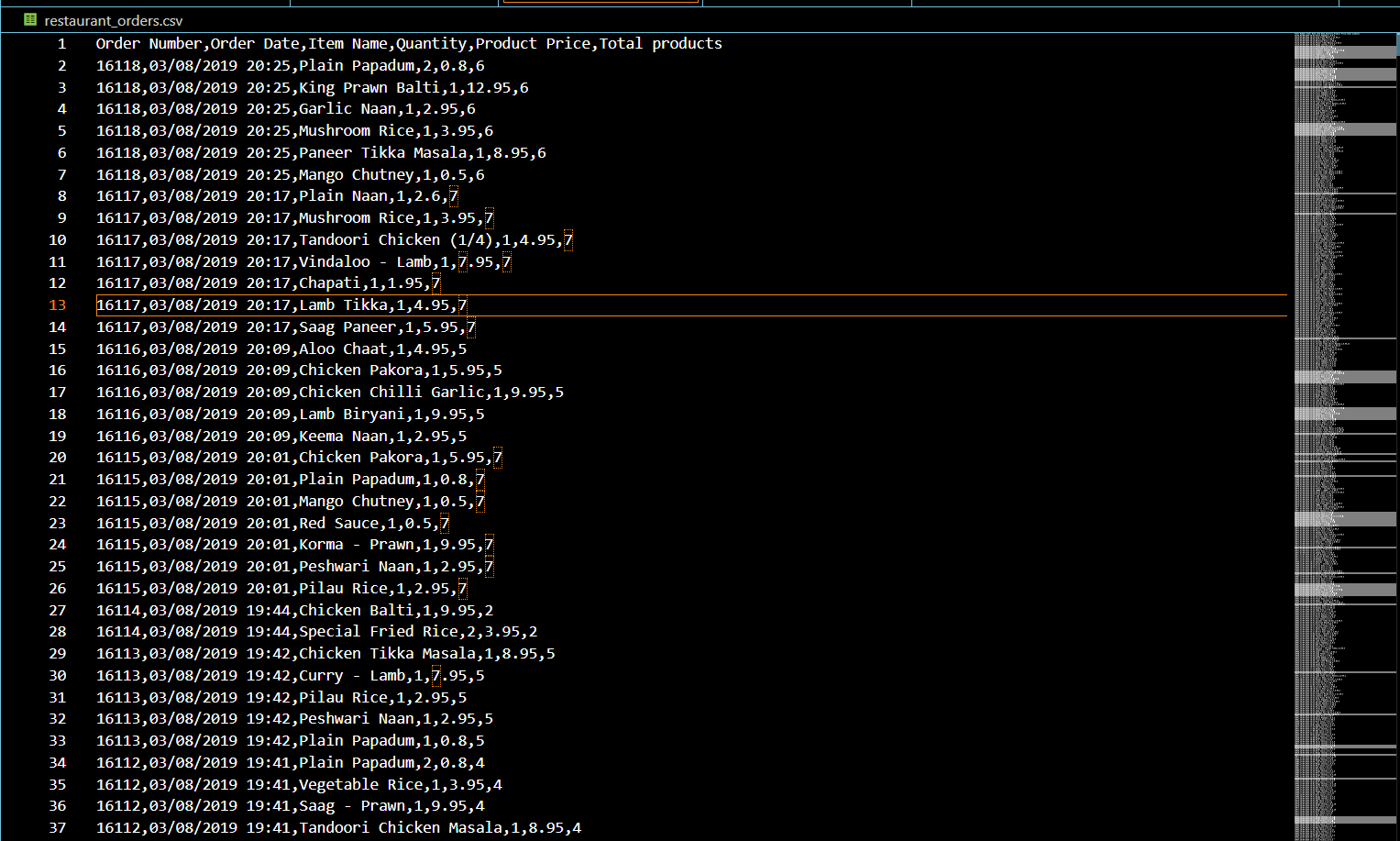
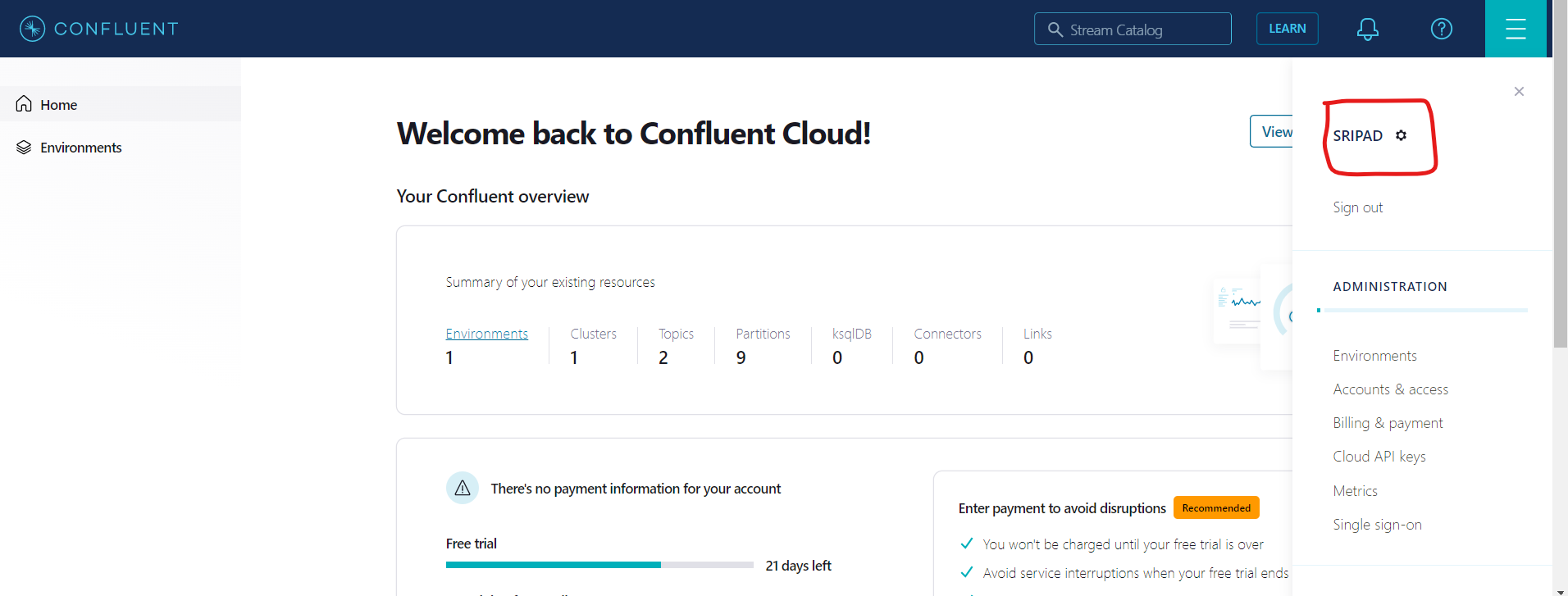
**Confluent Kafka Assignment 1**

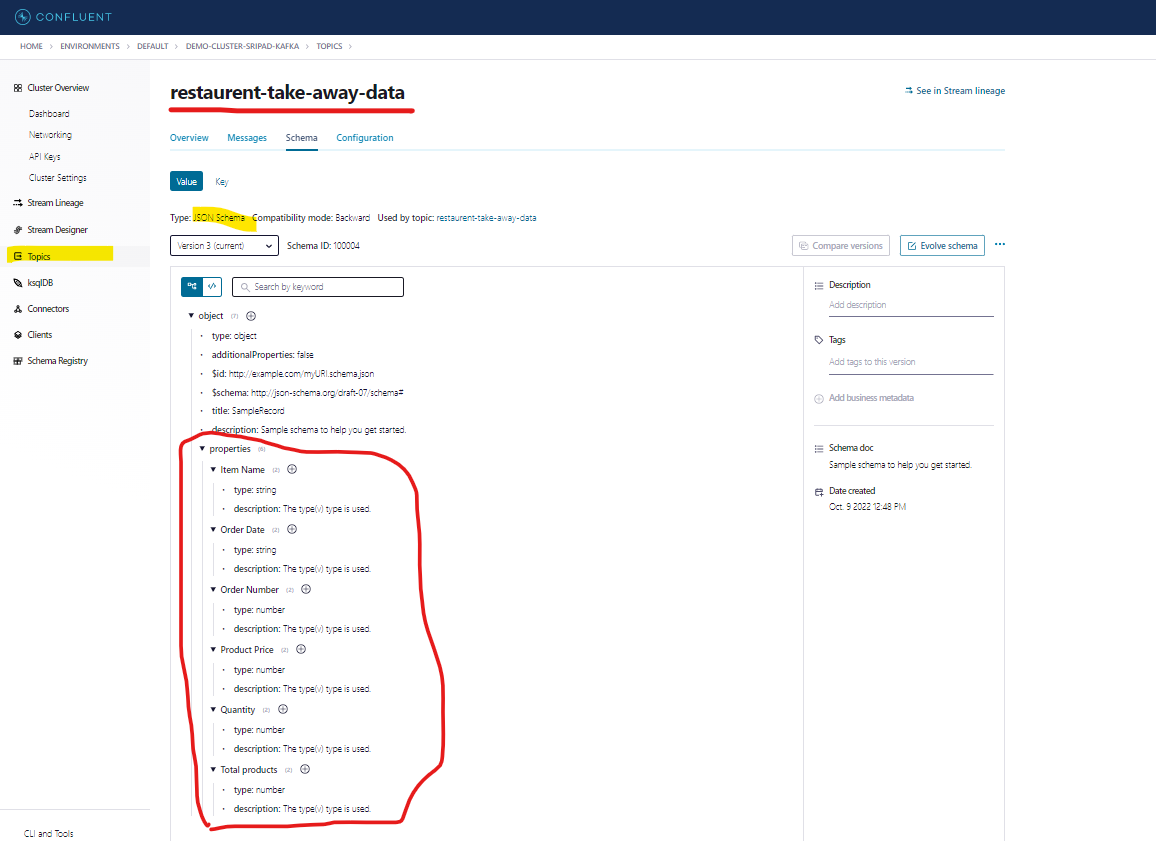
1. Download the Restaurant Order Dataset.



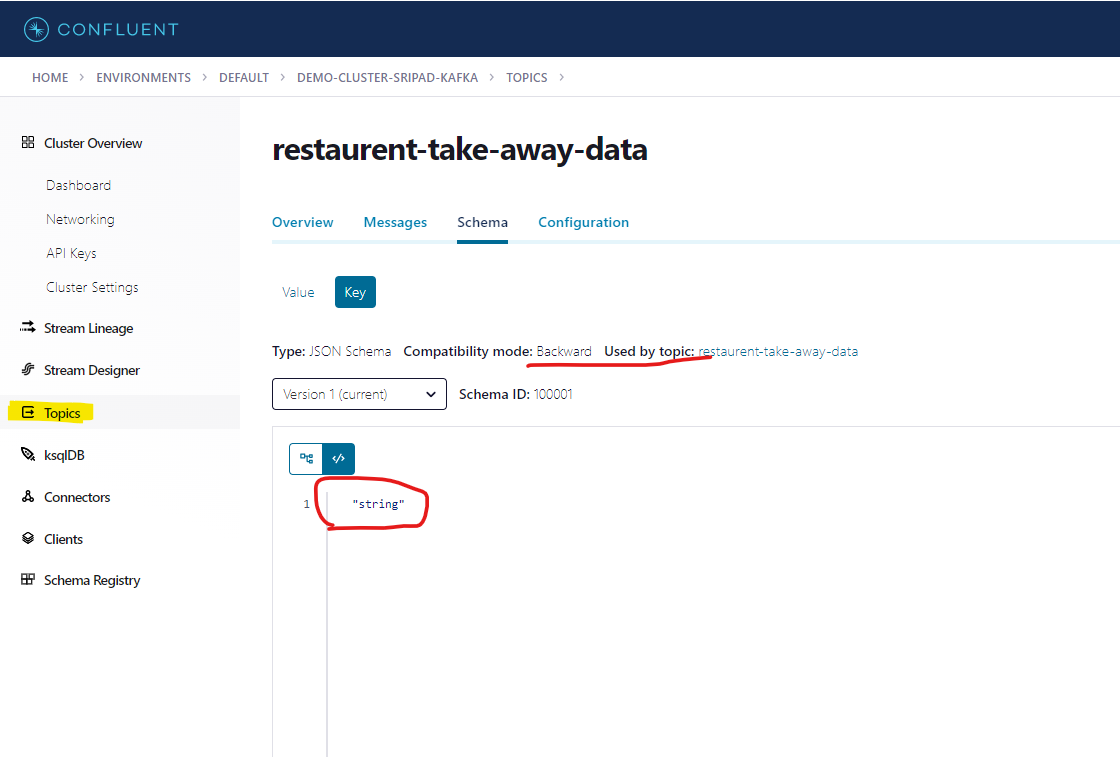
1. Setup Confluent Kafka Account



1. Setup key (string) & value (json) schema in the confluent schema registry
2. **Value:**



1. **Key:**



1. 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.



Code is attached in the github repo.

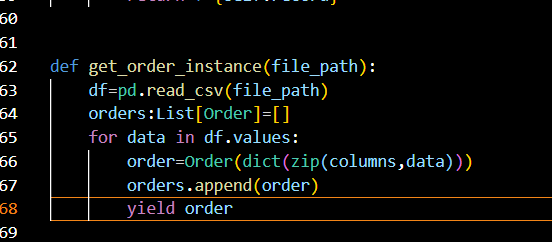
Schema is not hardcoded as can been seen in this below snippet.It is being automatically pulled from following methods:

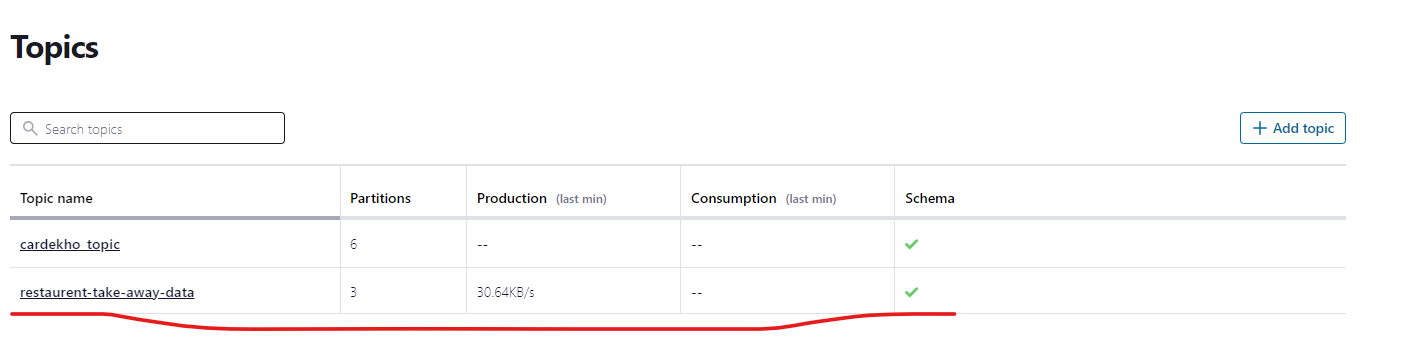
1. get\_subjects() ->returns list of subjects.
2. get\_latest\_version() -> Parameter: subject(Schema name),returns: schemaversion
3. get\_schema() -> Parameter : Schema \_id , returns: Schema\_str

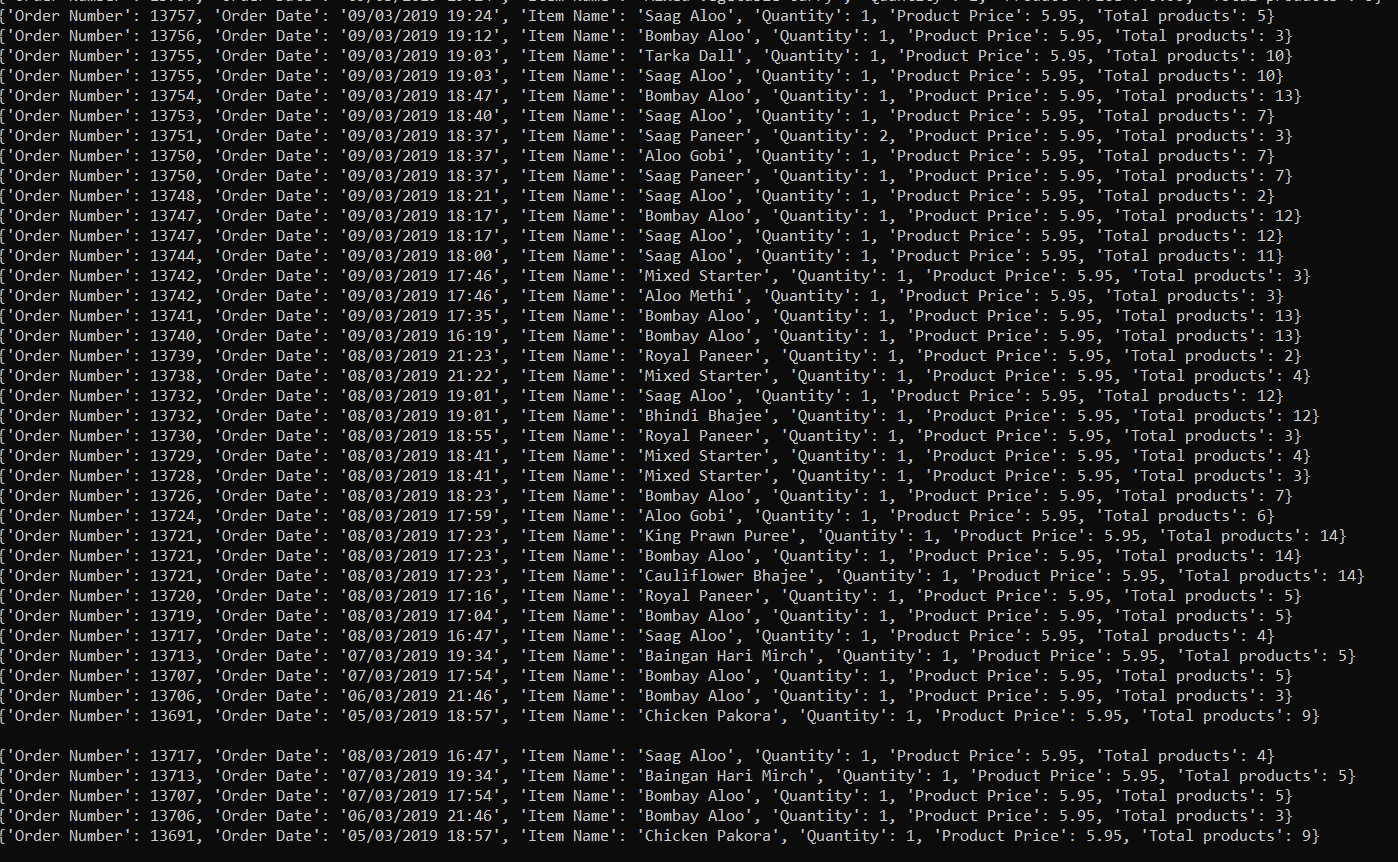
Ref:https://docs.confluent.io/5.5.1/clients/confluent-kafka-python/\_modules/confluent\_kafka/schema\_registry/schema\_registry\_client.html#SchemaRegistryClient

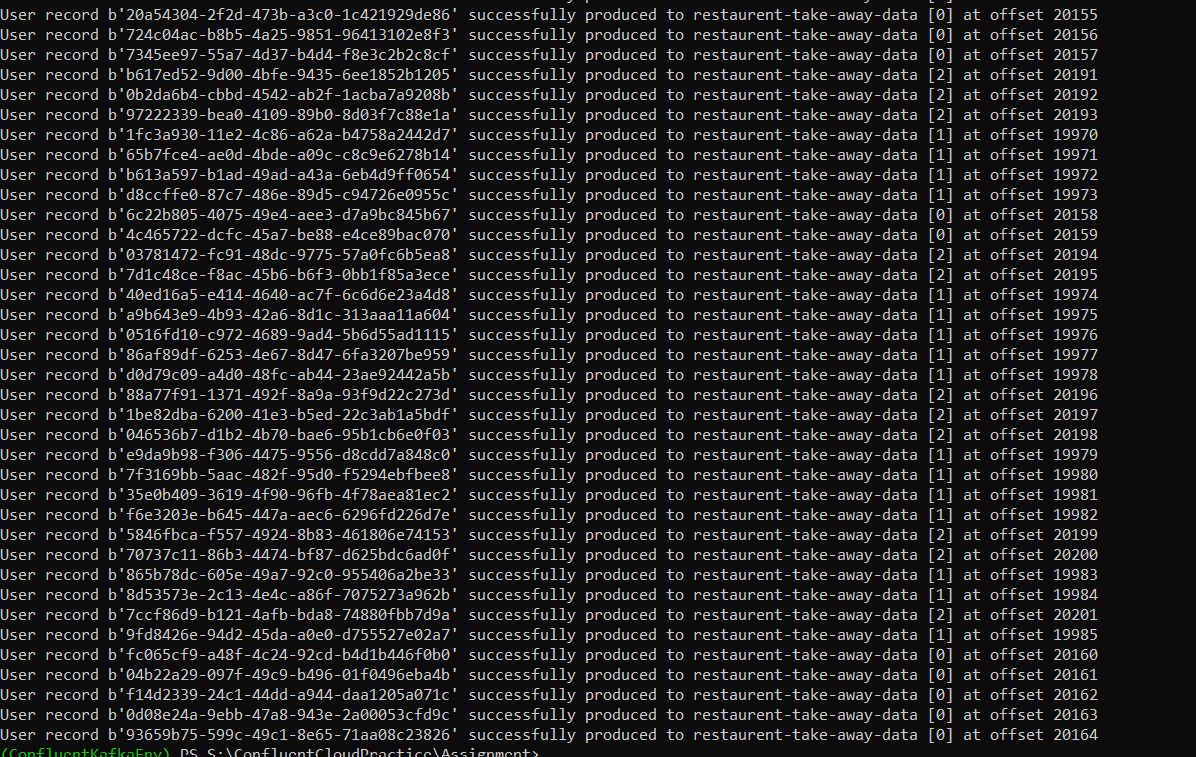


1. From producer code, publish data in Kafka Topic one by one and use dynamic key while publishing the records into the Kafka Topic







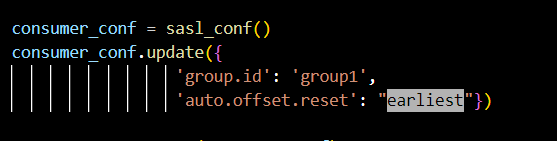


1. Write kafka consumer code and create two copies of same consumer code and save it with different names (kafka\_consumer\_1.py & kafka\_consumer\_2.py), again make sure lates schema version and schema\_str is not hardcoded in the consumer code, read it automatically from the schema registry to desrialize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

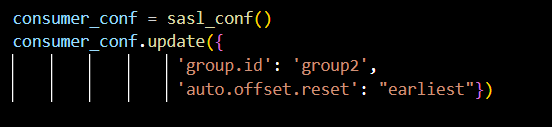
Now test two scenarios with your consumer code:

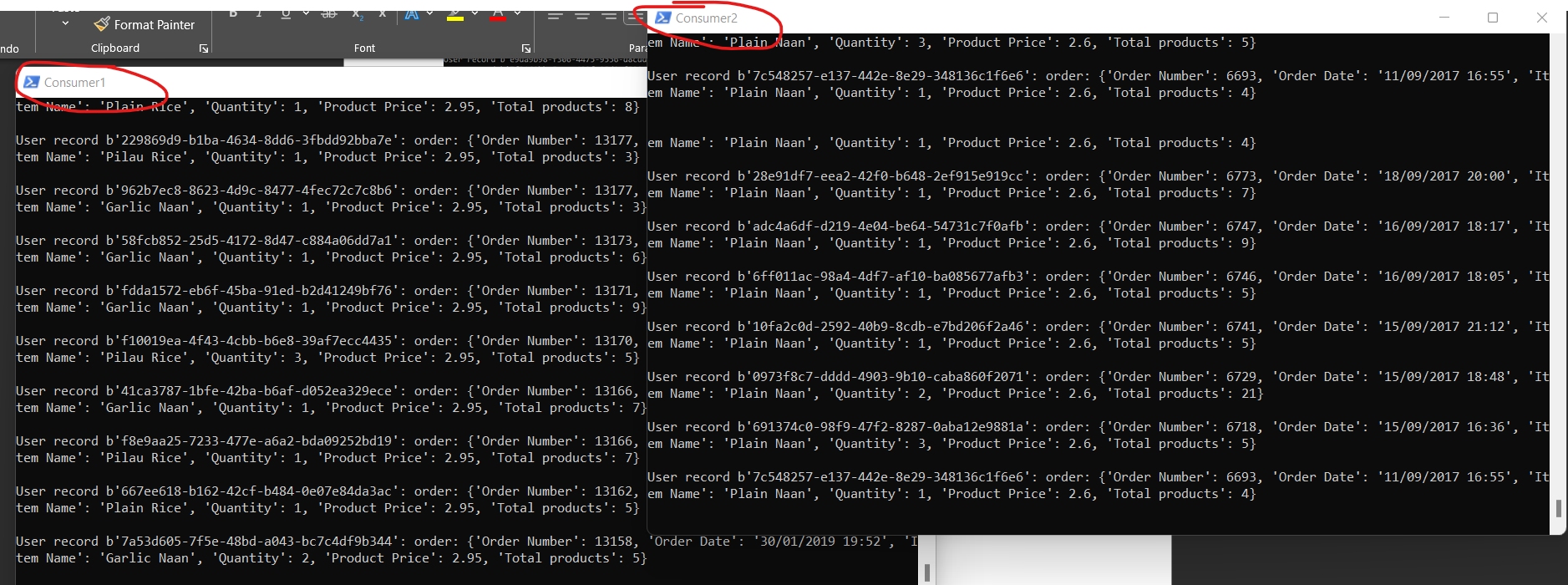
a.) Use "group.id" property in consumer config for both consumers and mention different group\_ids in kafka\_consumer\_1.py & kafka\_consumer\_2.py,apply "earliest" offset property in both consumers and run these two consumers from two different terminals. Calculate how many records each consumer consumed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

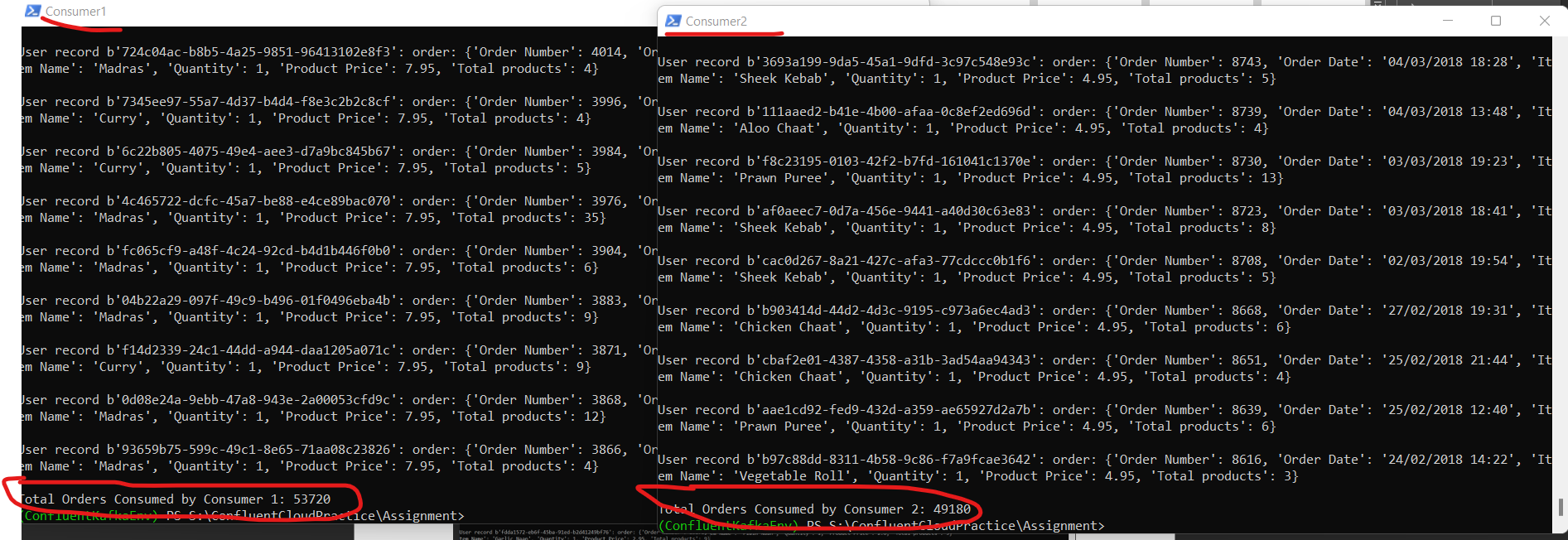
Kafka\_consumer\_1.py



Kafka\_consumer\_2.py





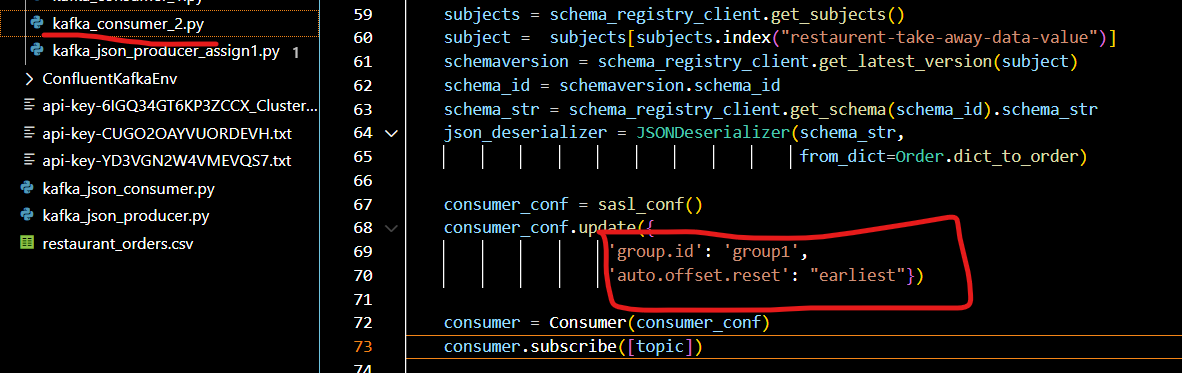


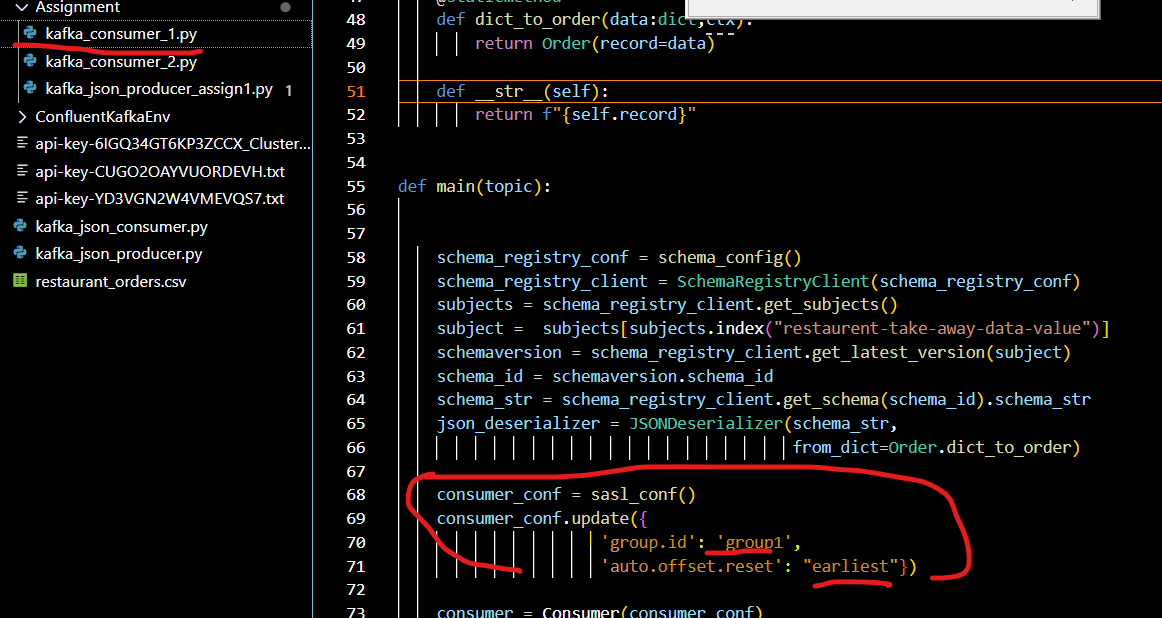
Total Orders Consumed by Consumer 1: 53720

Total Orders Consumed by Consumer 2: 49180

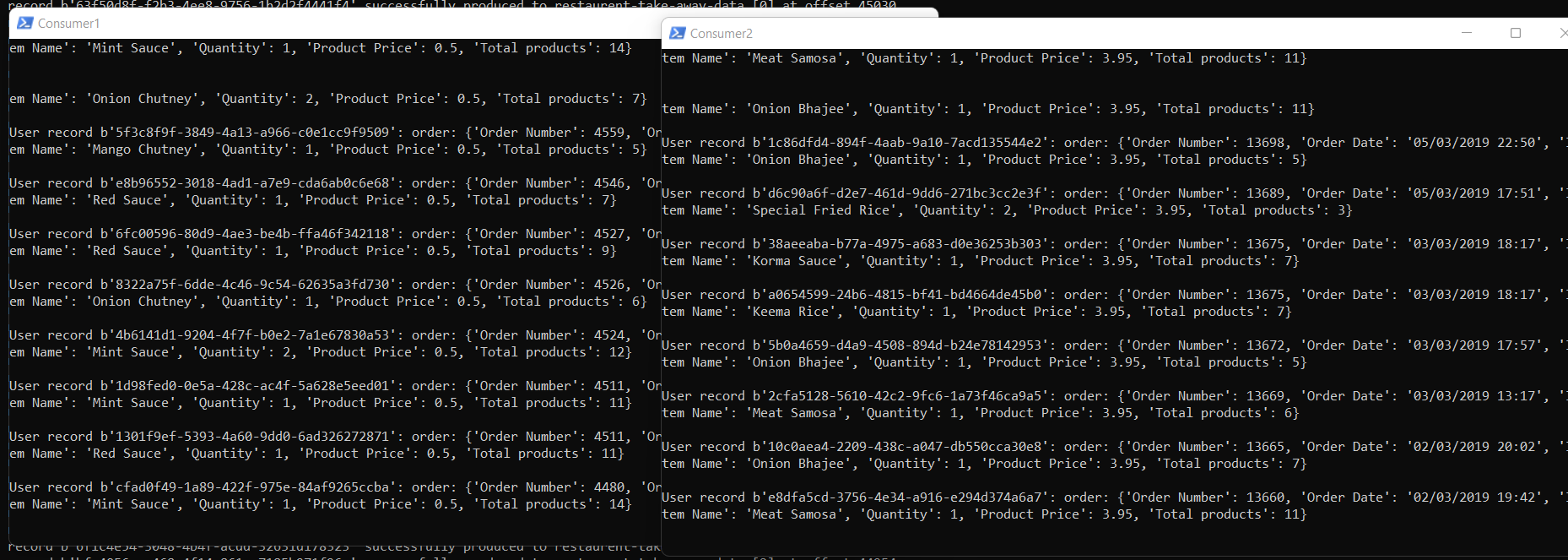
 

b.) Use "group.id" property in consumer config for both consumers and mention same group\_ids in kafka\_consumer\_1.py & kafka\_consumer\_2.py, apply "earliest" offset property in both consumers and run these two consumers from two different terminals. Calculate how many records each consumer consumed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



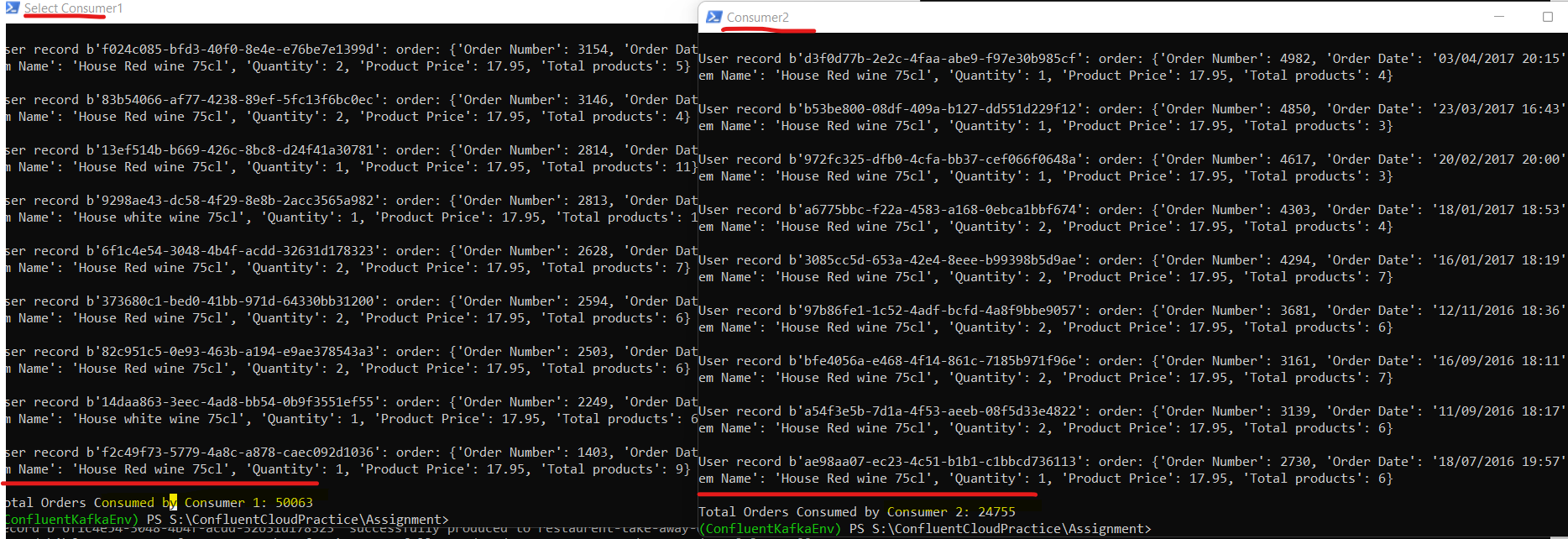


Both Consumer 1 and Consumer 2 same group id.



Consumer 1 consumed: 50063

Consumer 2 consumed: 24755

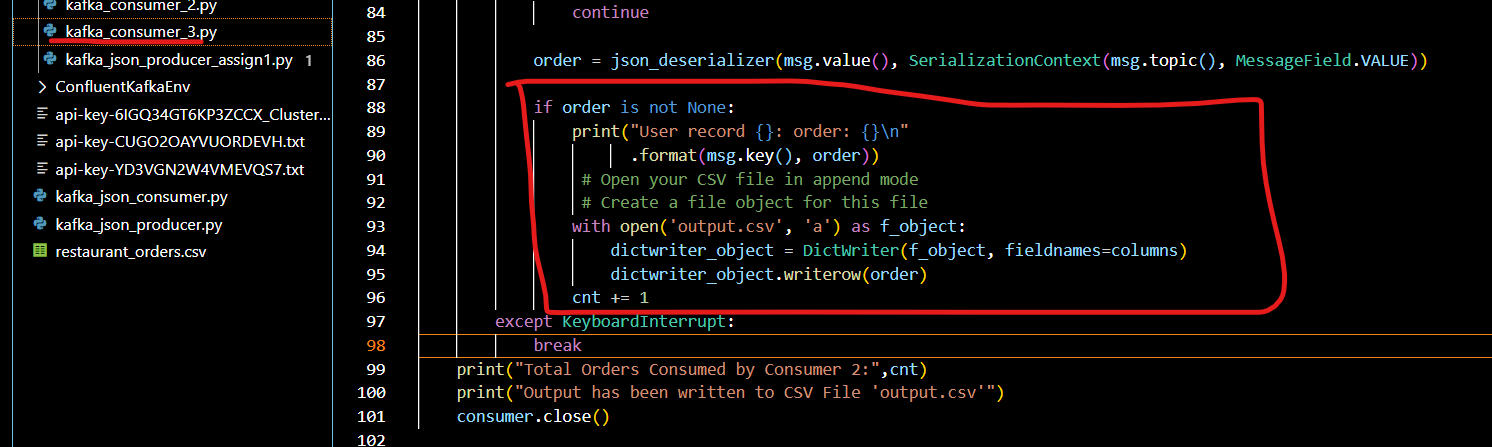


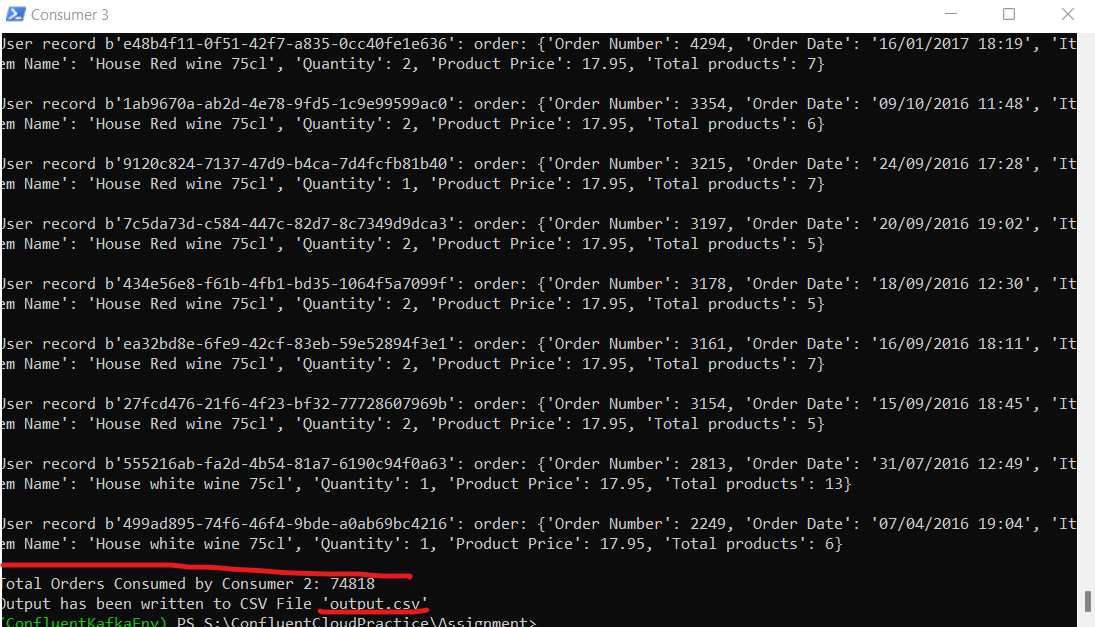




Embedded Code Snippet for Second Requirement two consumers with same group id.

1. 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.





Output Data Appended to created “Output.csv”.

