Project Requirement

Step 1:- Stream of text files are coming (Here you will have a text file) in your local taken from Kaggle

[Fraud Detection Dataset (kaggle.com)](https://www.kaggle.com/datasets/ranjeetshrivastav/fraud-detection-dataset)

Step2:- Set up kafka cluster using docker with 3 brokers, kafka-ui and schema registry

Step3:- Test Step 2

Write a simple scala script to send a record to kafka topic

Step4:- Write a producer code which takes data from local filepath created in Step1 and sends it to Kafka Topic

Step5:- Create schema of the data being produced in avro format, use lcoalhost 8081 to register the schema and send the data to Kafka topic

Step6:- Test by sending a single record which is not as per the schema to ensure it fails if data is not as per pre decided schema

Step7:- Write a consumer code to take data as per the schema, add a filter condition and then write it to s3

Step8:- Setup Snowflake, a table as per the consumer schema, a snowpipe

Step9:- Check by adding data in s3 and data should directly be getting populated in table through snowpipe