1. Lab Task:

1. Complete the **Kafka Producer** in the file `LT1-Producer.ipynb`. Here you are required to address the following conditions:
    - Read the data frome the csv file `clickstream.csv` continiously.
    - Add the current timestamp to the data being sent
    - Publish **5-10** number of rows randomly every **5 seconds** to a topic named:**clickstream**.

In [ ]:

```
# import statements
from time import sleep
from json import dumps
from kafka import KafkaProducer
import random
import datetime as dt
import csv


def read_csv(fileName):
    '''Read the CSV file clickstream.csv'''
    '''you can use csv.DictReader'''
    list = []
    with open(fileName, 'rt') as f:
        reader = csv.DictReader(f)
        for row in reader:
            list.append(row)
    return list

def publish_message(producer_instance, topic_name, data):
    try:
        producer_instance.send(topic_name, data)
        print('Message published successfully. Data: ' + str(data))
    except Exception as ex:
        print('Exception in publishing message.')
        print(str(ex))

def connect_kafka_producer():
    _producer = None
    try:
        _producer = KafkaProducer(bootstrap_servers=['localhost:9092'],
                                  value_serializer=lambda x:
dumps(x).encode('ascii'),
                                  api_version=(0, 10))
    except Exception as ex:
        print('Exception while connecting Kafka.')
        print(str(ex))
    finally:
        return _producer

if __name__ == '__main__':

    topic = 'clickstream'
    cRows = read_csv('clickstream.csv')

    print('Publishing records..')
```

```
    producer = connect_kafka_producer()

    #WRITE THE CODE HERE
    #A while True loop to infinitely loop through records
    #A random function to get random number of rows between 5-10 use
random.randint(5,10)
    #Create the data object to publish with the selected rows, also include the
current timestamp as ts
    #Sleep for 5 seconds
```