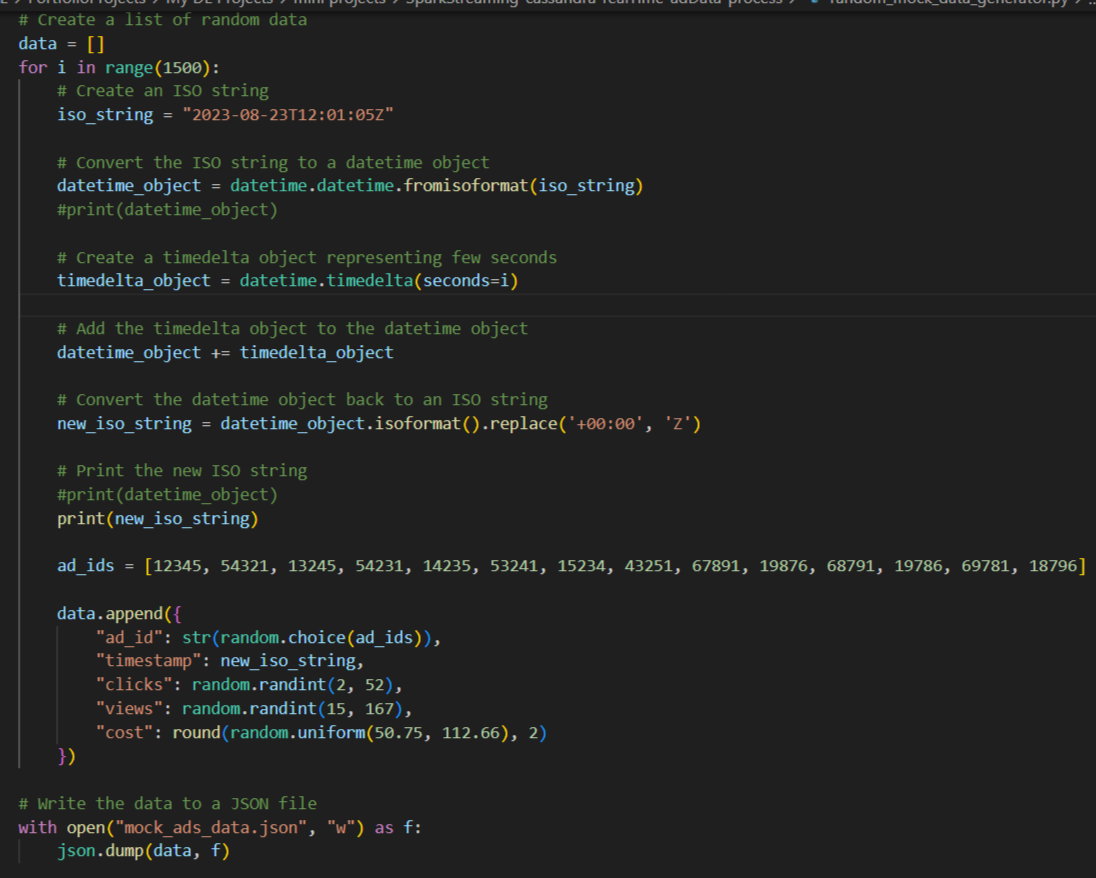
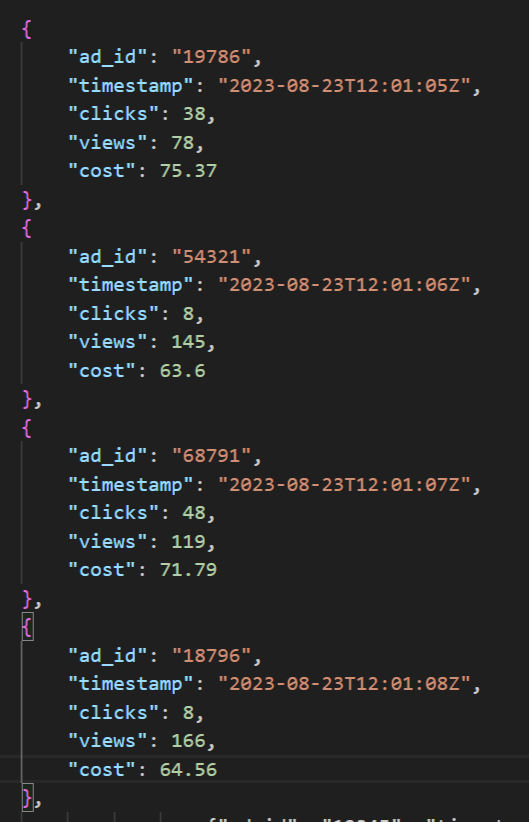
Real-Time Ads Data Processing Into Cassandra

1. This project was about processing real-time advertisement data using Spark Streaming to gain business insights and store the aggregated data into Cassandra.
2. I set up confluent kafka in local system and created topic named as ads\_data. Wrote a python script to generate ads data in json format using 'random\_mock\_data\_generator.py' file

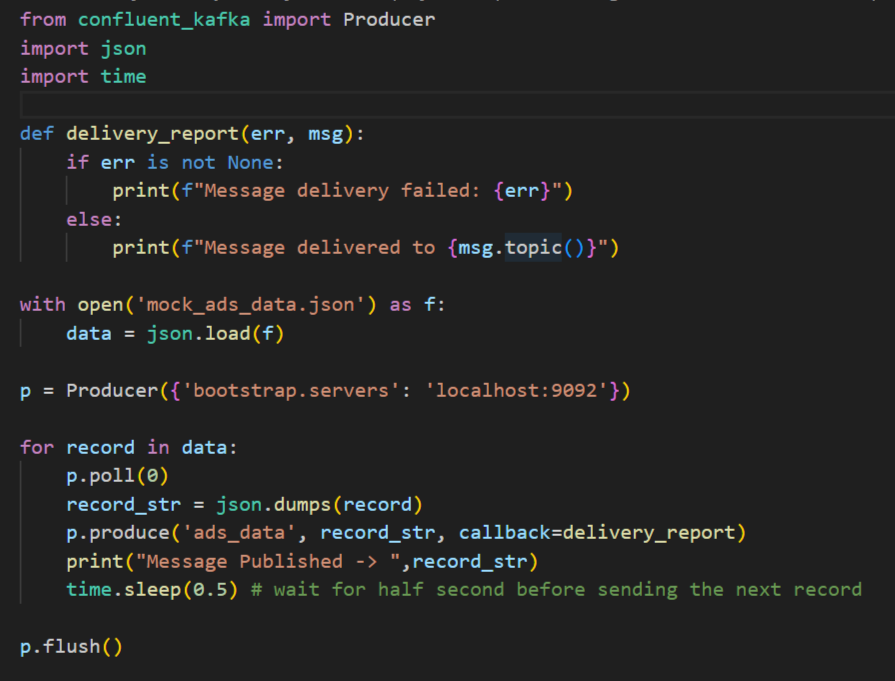
random\_mock\_data\_genrator.py:



mock\_ads\_data.json:



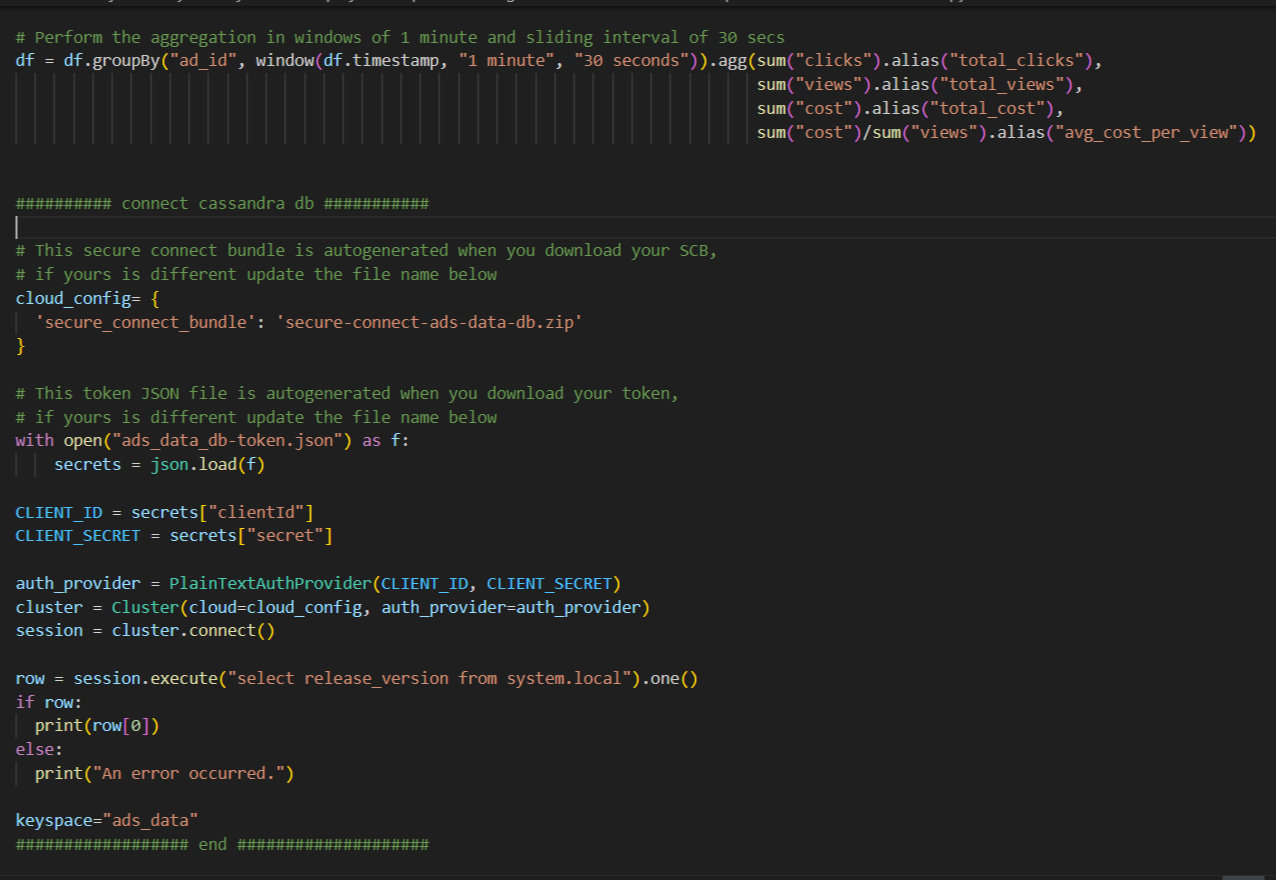
1. Created a producer script which was publishing those json ads data into kafka topic.



1. Next I set up a spark streaming application where I used the kafka connector to read data from ads\_data topic. Then parsed the json data into proper structure using my defined schema.



1. Further transformed data by performed windowed based aggregation and included cassandra connection script into spark stream app so that I can write aggregated data into cassandra table.



1. But writing data with the logic was like if entry already existed with 'ad\_id' then updated that record or else inserting new entry in table.

