BankMarketing Campaign Data Analysis with pyspark SQL

Create Pyspark Application bank marketing-analysis.py

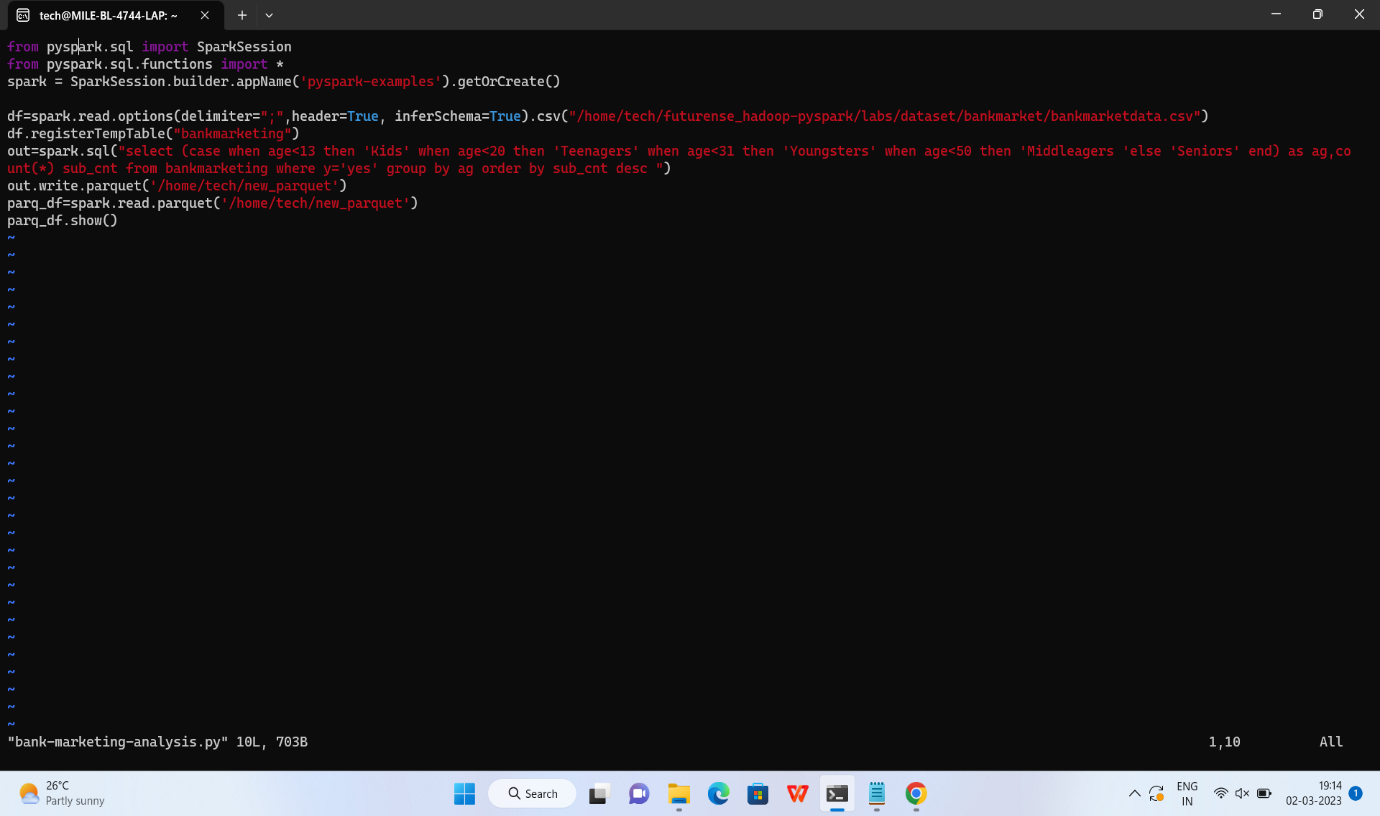
a) Load BankMarketing Campagin data from csv

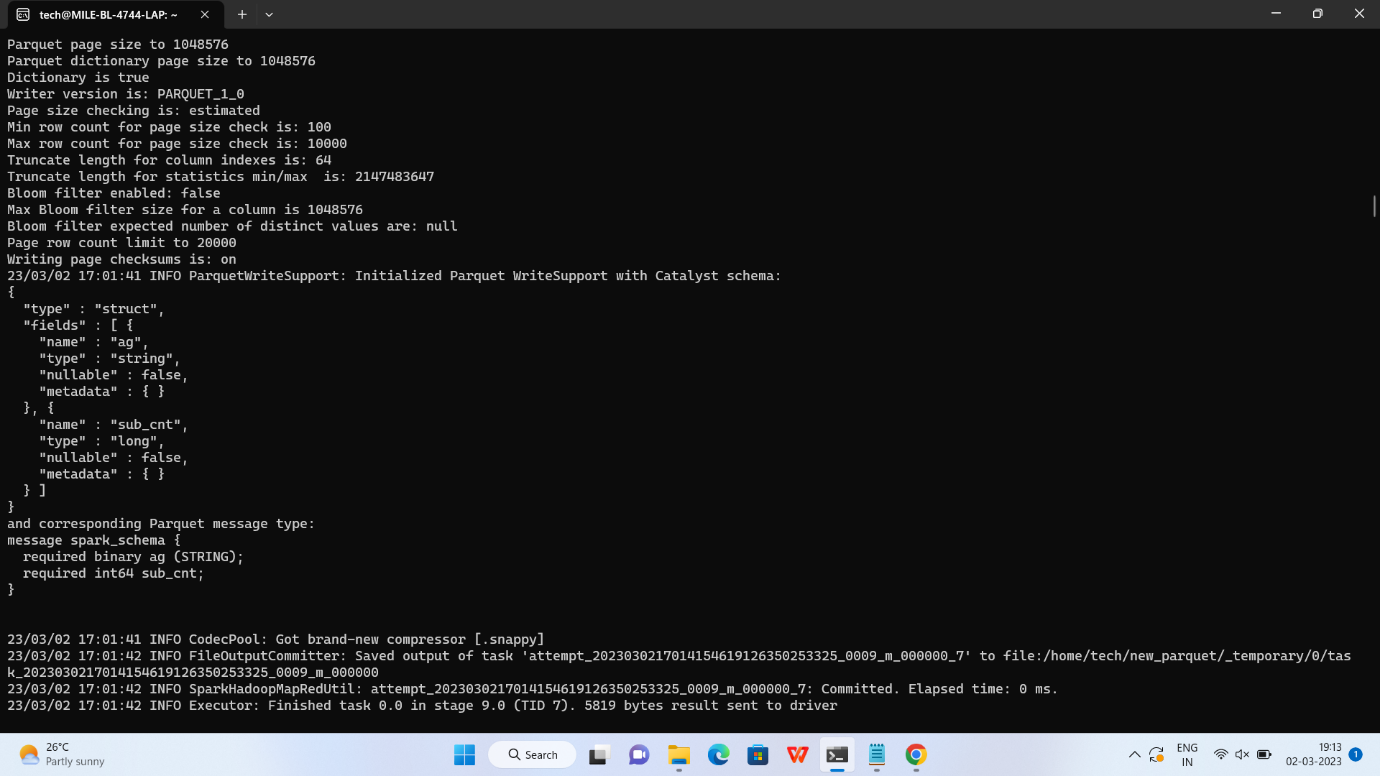
b) Get AgeGroup wise subscripitionCount

c) Write the output in parquet file format

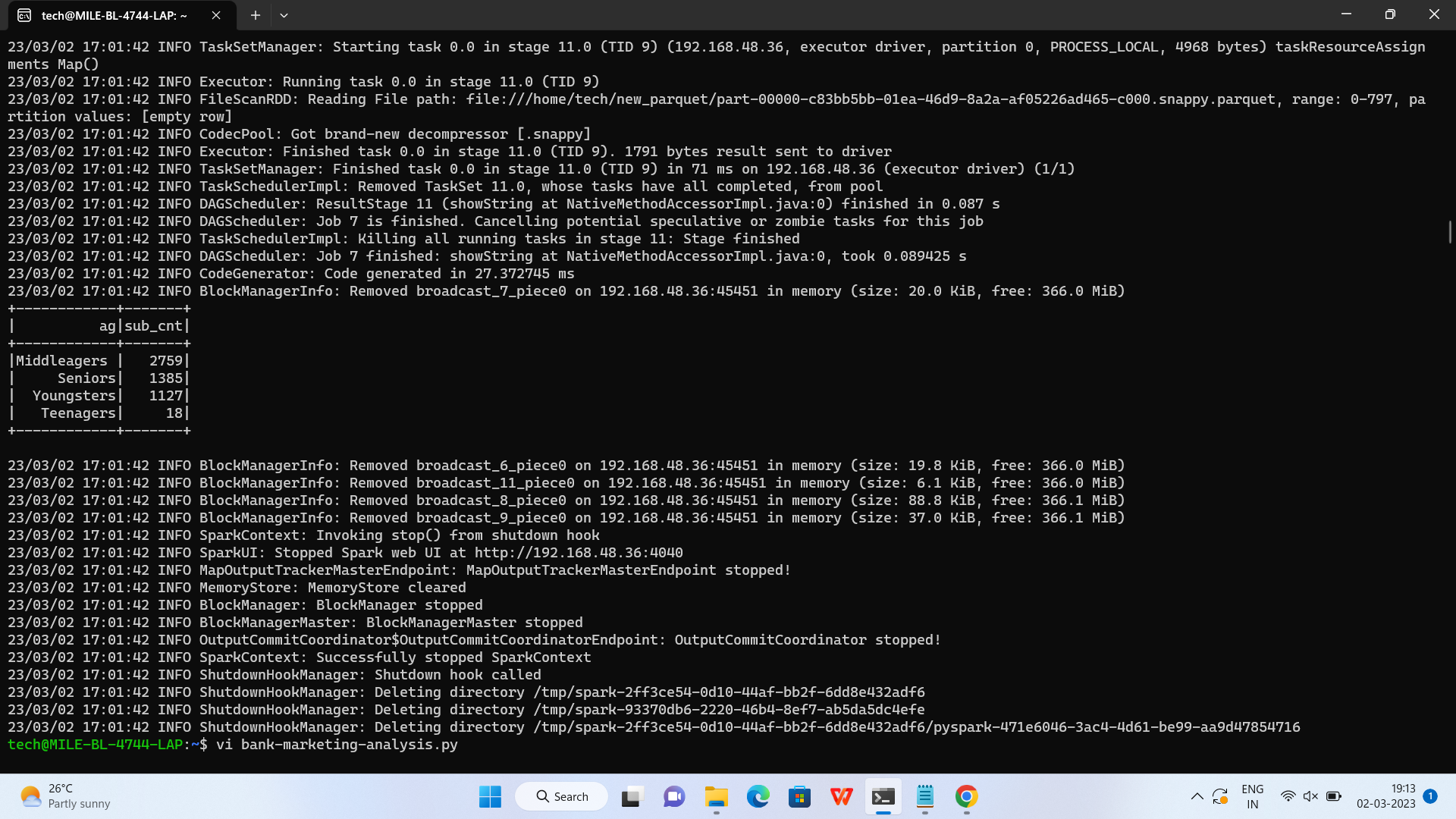
d) Load the data from parquest file written above

e) show the data





spark-submit bank-marketing-analysis.py



f) Filter AgeGroup with SubcripitionCount>2000 and Write into Avro file format

g)Load the data from parquet file written above

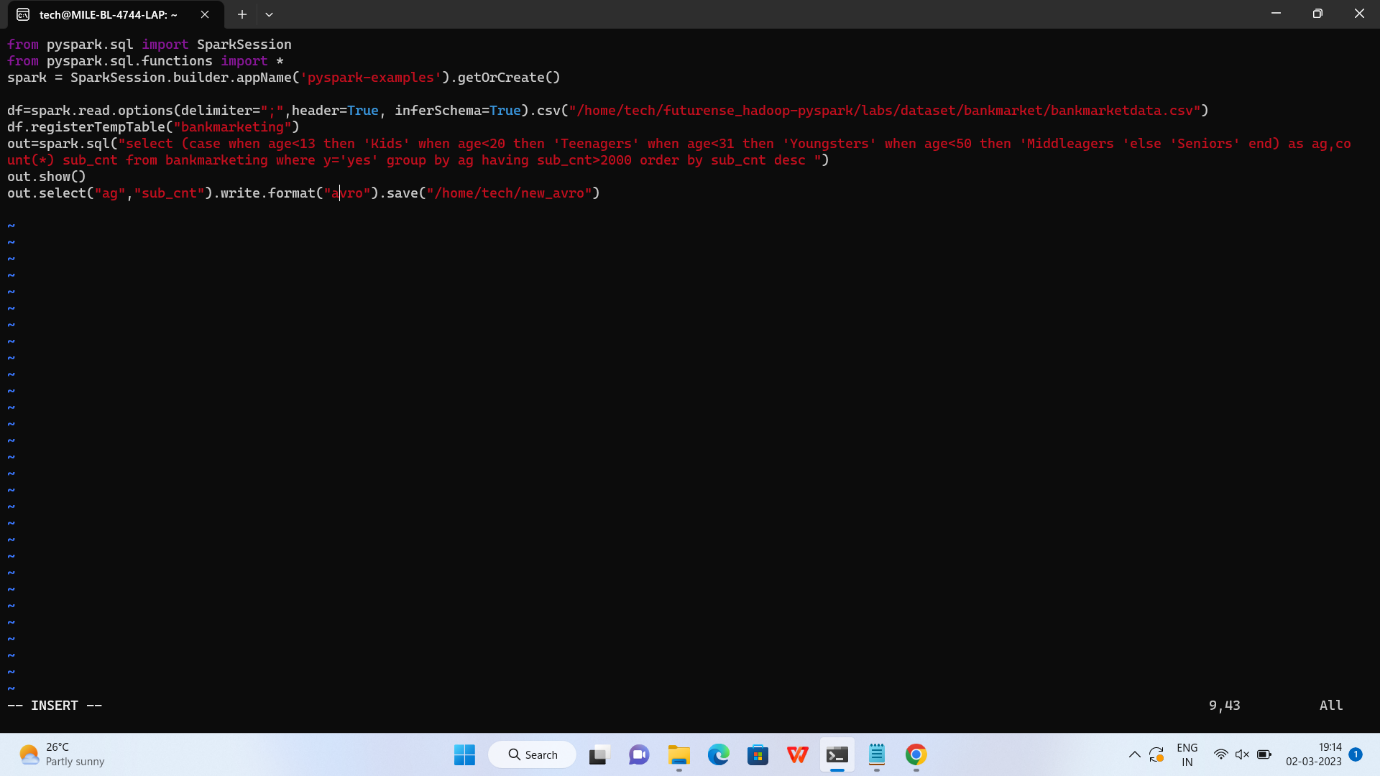
h) show the data

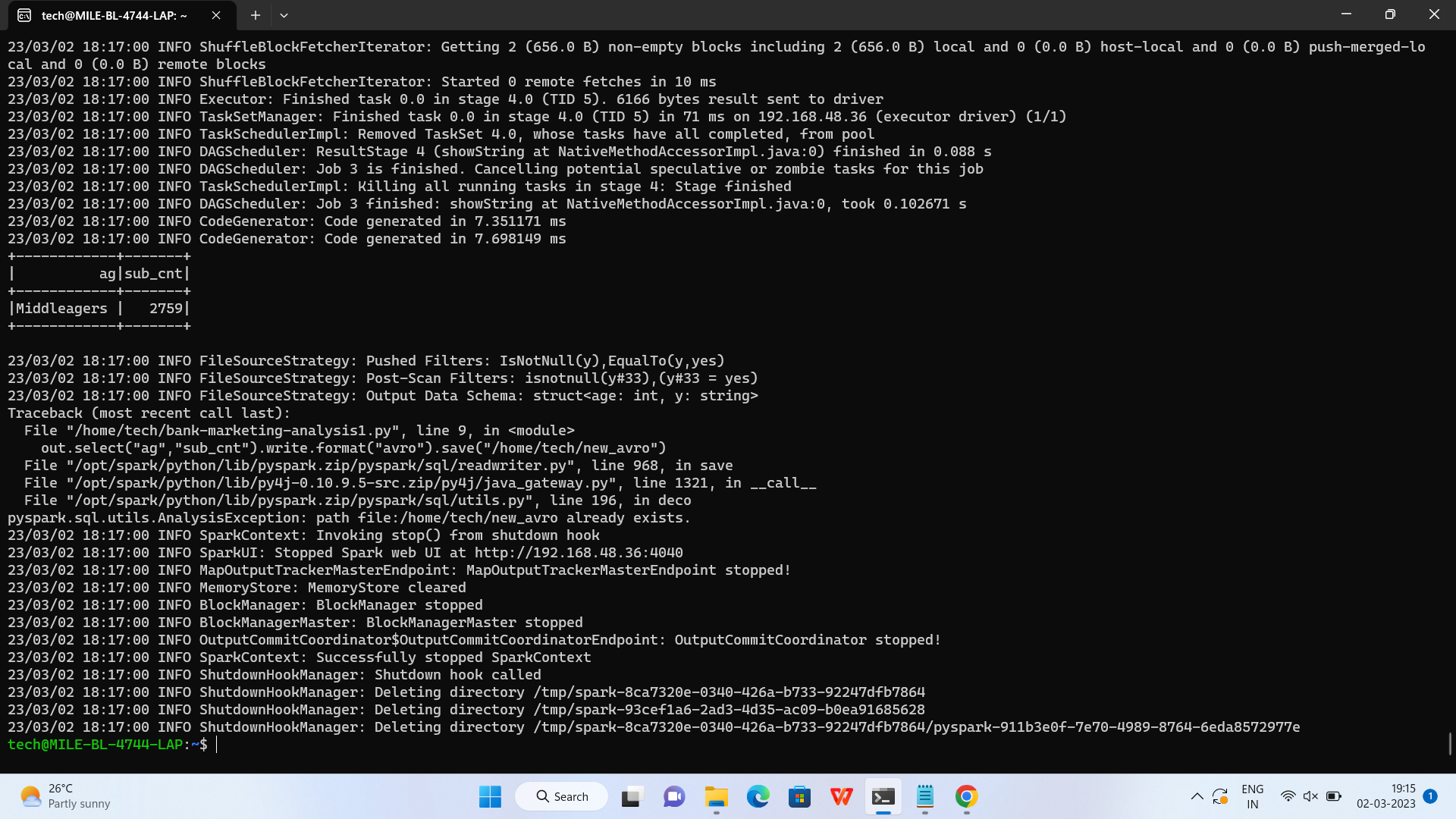
1] spark-submit bank-marketing-analysis.py => Runs in local mode

3] schedule to PysparkApplication to run every N mins

tech@MILE-BL-4744-LAP:~$ vi bank-marketing-analysis1.py

tech@MILE-BL-4744-LAP:~$ spark-submit --packages org.apache.spark:spark-avro\_2.12:3.3.2 bank-marketing-analysis1.py





2] Runs in cluster mode

