

```
at org.apache.spark.rdd.RDD.count(RDD.scala:1274)
at org.apache.spark.rdd.RDD.$anonfun$takeSample$1(RDD.scala:626)
at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:151)
at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:112)
at org.apache.spark.rdd.RDD.withScope(RDD.scala:406)
at org.apache.spark.rdd.RDD.takeSample(RDD.scala:615)
at org.apache.spark.mllib.clustering.KMeans.initKMeansParallel(KMeans.scala:384)
at org.apache.spark.mllib.clustering.KMeans.runAlgorithmWithWeight(KMeans.scala:257)
at org.apache.spark.mllib.clustering.KMeans.runWithWeight(KMeans.scala:231)
at org.apache.spark.ml.clustering.KMeans.$anonfun$fit$1(KMeans.scala:354)
at org.apache.spark.ml.util.Instrumentation$.anonfun$instrumented$1(Instrumentation.scala:191)
at scala.util.Try$.apply(Try.scala:213)
at org.apache.spark.ml.util.Instrumentation$.instrumented(Instrumentation.scala:191)
at org.apache.spark.ml.clustering.KMeans.fit(KMeans.scala:329)
at org.apache.spark.ml.clustering.KMeans.fit(KMeans.scala:272)
at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.base/java.lang.reflect.Method.invoke(Method.java:566)
at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)
at py4j.Gateway.invoke(Gateway.java:282)
at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)
at py4j.commands.CallCommand.execute(CallCommand.java:79)
at py4j.ClientServerConnection.waitForCommands(ClientServerConnection.java:182)
at py4j.ClientServerConnection.run(ClientServerConnection.java:106)
at java.base/java.lang.Thread.run(Thread.java:829)
Caused by: java.lang.OutOfMemoryError: Java heap space
at java.base/java.lang.reflect.Array.newInstance(Array.java:78)
at java.base/java.io.ObjectInputStream.readArray(ObjectInputStream.java:2098)
at java.base/java.io.ObjectInputStream.readObject0(ObjectInputStream.java:1675)
at java.base/java.io.ObjectInputStream.readArray(ObjectInputStream.java:2134)
at java.base/java.io.ObjectInputStream.readObject0(ObjectInputStream.java:1675)
at java.base/java.io.ObjectInputStream.defaultReadFields(ObjectInputStream.java:2496)
at java.base/java.io.ObjectInputStream.readSerialData(ObjectInputStream.java:2390)
at java.base/java.io.ObjectInputStream.readOrdinaryObject(ObjectInputStream.java:2228)
at java.base/java.io.ObjectInputStream.readObject0(ObjectInputStream.java:1687)
at java.base/java.io.ObjectInputStream.readObject(ObjectInputStream.java:489)
at java.base/java.io.ObjectInputStream.readObject(ObjectInputStream.java:447)
at org.apache.spark.serializer.JavaDeserializationStream.readObject(JavaSerializer.scala:87)
at org.apache.spark.serializer.DeserializationStream$$anon$1.getNext(Serializer.scala:168)
at org.apache.spark.util.NextIterator.hasNext(NextIterator.scala:73)
at org.apache.spark.storage.memory.PartiallyUnrolledIterator.hasNext(MemoryStore.scala:772)
at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterator.scala:31)
at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterator.scala:37)
at scala.collection.Iterator$$anon$12.hasNext(Iterator.scala:513)
at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
at org.apache.spark.sql.catalyst.expressions.GeneratedClass$SpecificColumnarIterator.hasNext(Unknown Source)
at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedIteratorForCodegenStage1.processNext(Unknown Source)
at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(BufferedRowIterator.java:43)
at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNext(WholeStageCodegenExec.scala:760)
at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
at org.apache.spark.storage.memory.MemoryStore.putIterator(MemoryStore.scala:223)
at org.apache.spark.storage.memory.MemoryStore.putIteratorAsValues(MemoryStore.scala:302)
at org.apache.spark.storage.BlockManager.$anonfun$doPutIterator$1(BlockManager.scala:1508)
at org.apache.spark.storage.BlockManager$$Lambda$2244/0x0000000840fa6040.apply(Unknown Source)
at org.apache.spark.storage.BlockManager.org$apache$spark$storage$BlockManager$$doPut(BlockManager.scala:1435)

24/12/09 21:18:35 INFO MemoryStore: MemoryStore cleared
24/12/09 21:18:35 INFO BlockManager: BlockManager stopped
24/12/09 21:18:35 INFO BlockManagerMaster: BlockManagerMaster stopped
24/12/09 21:18:35 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
24/12/09 21:18:35 INFO SparkContext: Successfully stopped SparkContext
24/12/09 21:18:35 INFO ShutdownHookManager: Shutdown hook called
24/12/09 21:18:35 INFO ShutdownHookManager: Deleting directory /tmp/spark-0f2a2396-794b-400f-b8f7-a5e37c3e0a83
24/12/09 21:18:35 INFO ShutdownHookManager: Deleting directory /tmp/spark-20e06261-6d7d-4499-833b-32216b97f7f5
24/12/09 21:18:35 INFO ShutdownHookManager: Deleting directory /tmp/spark-20e06261-6d7d-4499-833b-32216b97f7f5/pyspark-d07dd1c7-afd6-4ef8-b11d-a52f9ce9f5a8
Execution time: 424 seconds
[tkk5297@p-bc-5014 MiniProj3]$
```