

Group09

a. Representative Screenshots of the console output when you execute the program.

```
cc@group09-1: /usr/local/spark-1.6.1-bin-hadoop1$ ./bin/spark-submit --master spark://10.56.3.221:7077 /home/cc/assign3.py
21/08/05 04:40:26 INFO spark.SparkContext: Running Spark version 1.6.1
21/08/05 04:40:26 INFO spark.SecurityManager: Changing view acls to: cc
21/08/05 04:40:26 INFO spark.SecurityManager: Changing modify acls to: cc
21/08/05 04:40:26 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(cc); users with modify permissions: Set(cc)
21/08/05 04:40:30 WARN internal.ThreadLocalRandom: Failed to generate a seed from SecureRandom within 3 seconds. Not enough entropy?
21/08/05 04:40:30 INFO util.Utils: Successfully started service 'sparkDriver' on port 43858.
21/08/05 04:40:30 INFO slf4j.Slf4jLogger: Slf4jLogger started
21/08/05 04:40:30 INFO Remoting: Starting remoting
21/08/05 04:40:30 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://sparkDriverActorSystem@10.56.3.221:42407]
21/08/05 04:40:30 INFO util.Utils: Successfully started service 'sparkDriverActorSystem' on port 42407.
21/08/05 04:40:30 INFO spark.SparkEnv: Registering MapOutputTracker
21/08/05 04:40:31 INFO spark.SparkEnv: Registering BlockManagerMaster
21/08/05 04:40:31 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-0edecb0b-96ff-4b3d-a587-21b7a32ae43f
21/08/05 04:40:31 INFO storage.MemoryStore: MemoryStore started with capacity 511.1 MB
21/08/05 04:40:31 INFO spark.SparkEnv: Registering OutputCommitCoordinator
21/08/05 04:40:31 INFO server.Server: jetty-8.y.z-SNAPSHOT
21/08/05 04:40:31 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
21/08/05 04:40:31 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.
21/08/05 04:40:31 INFO ui.SparkUI: Started SparkUI at http://10.56.3.221:4040
21/08/05 04:40:31 INFO spark.HttpFileServer: HTTP File server directory is /tmp/spark-91247bc6-5a5d-4820-9416-31ec1d7bb2f8/httpd-3f68ac45-aa19-483f-9208-749aef829bcc
21/08/05 04:40:31 INFO server.Server: jetty-8.y.z-SNAPSHOT
21/08/05 04:40:31 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:45750
21/08/05 04:40:31 INFO util.Utils: Successfully started service 'HTTP file server' on port 45750.
21/08/05 04:40:31 INFO util.Utils: Copying /home/cc/assign3.py to /tmp/spark-91247bc6-5a5d-4820-9416-31ec1d7bb2f8/userFiles-31441b90-783a-4626-9a72-3811abbc3add/assign3.py
21/08/05 04:40:31 INFO spark.SparkContext: Added file file:/home/cc/assign3.py at http://10.56.3.221:45750/files/assign3.py with timestamp 1628138431538
21/08/05 04:40:31 INFO client.AppClient$ClientEndpoint: Connecting to master spark://10.56.3.221:7077...
21/08/05 04:40:31 INFO cluster.SparkDeploySchedulerBackend: Connected to Spark cluster with app ID app-20210805044031-0083
21/08/05 04:40:31 INFO client.AppClient$ClientEndpoint: Executor added: app-20210805044031-0083/0 on worker-20210723002628-10.56.0.148:42246 (10.56.0.148:42246) with 2 cores
21/08/05 04:40:31 INFO cluster.SparkDeploySchedulerBackend: Granted executor ID app-20210805044031-0083/0 on hostPort 10.56.0.148:42246 with 2 cores, 1024.0 MB RAM
21/08/05 04:40:31 INFO client.AppClient$ClientEndpoint: Executor added: app-20210805044031-0083/1 on worker-20210723002629-10.56.3.221:39073 (10.56.3.221:39073) with 2 cores
21/08/05 04:40:31 INFO cluster.SparkDeploySchedulerBackend: Granted executor ID app-20210805044031-0083/1 on hostPort 10.56.3.221:39073 with 2 cores, 1024.0 MB RAM
21/08/05 04:40:31 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 42359.
21/08/05 04:40:31 INFO netty.NettyBlockTransferService: Server created on 42359
21/08/05 04:40:31 INFO storage.BlockManagerMaster: Trying to register BlockManager
21/08/05 04:40:31 INFO storage.BlockManagerMasterEndpoint: Registering block manager 10.56.3.221:42359 with 511.1 MB RAM, BlockManagerId(driver, 10.56.3.221, 42359)
21/08/05 04:40:31 INFO storage.BlockManagerMaster: Registered BlockManager
21/08/05 04:40:31 INFO client.AppClient$ClientEndpoint: Executor updated: app-20210805044031-0083/1 is now RUNNING
21/08/05 04:40:31 INFO client.AppClient$ClientEndpoint: Executor updated: app-20210805044031-0083/0 is now RUNNING
21/08/05 04:40:32 INFO cluster.SparkDeploySchedulerBackend: SchedulerBackend is ready for scheduling beginning after reached minRegisteredResourcesRatio: 0.0
1.) What are the top 3 crime types that were reported in the month of July?
  - PETIT LARCENY 7086
  - HARRASSMENT 2 5738
  - ASSAULT 3 & RELATED OFFENSES 4804
2.) How many crimes of type DANGEROUS WEAPONS were reported in the month of July?
  - DANGEROUS WEAPONS 701
cc@group09-1: /usr/local/spark-1.6.1-bin-hadoop1$
```

b. Output of your Spark program.

- What are the top 3 crime types that were reported in the month of July?**
 - PETIT LARCENY 7086
 - HARRASSMENT 2 5738
 - ASSAULT 3 & RELATED OFFENSES 4804
- How many crimes of type DANGEROUS WEAPONS were reported in the month of July?**
 - DANGEROUS WEAPONS 701

c. Describe the contribution of each group member

Sebastian Klemkosky: Completed the program and got the desired output

Nam Huynh: Removed extraneous pieces of code, reformatted the code and comments for readability.

Charley Mills: checked code for errors, formatted comments

Josue Chio: Fixed up and formatted output document for visualization and readability and formatted code output.