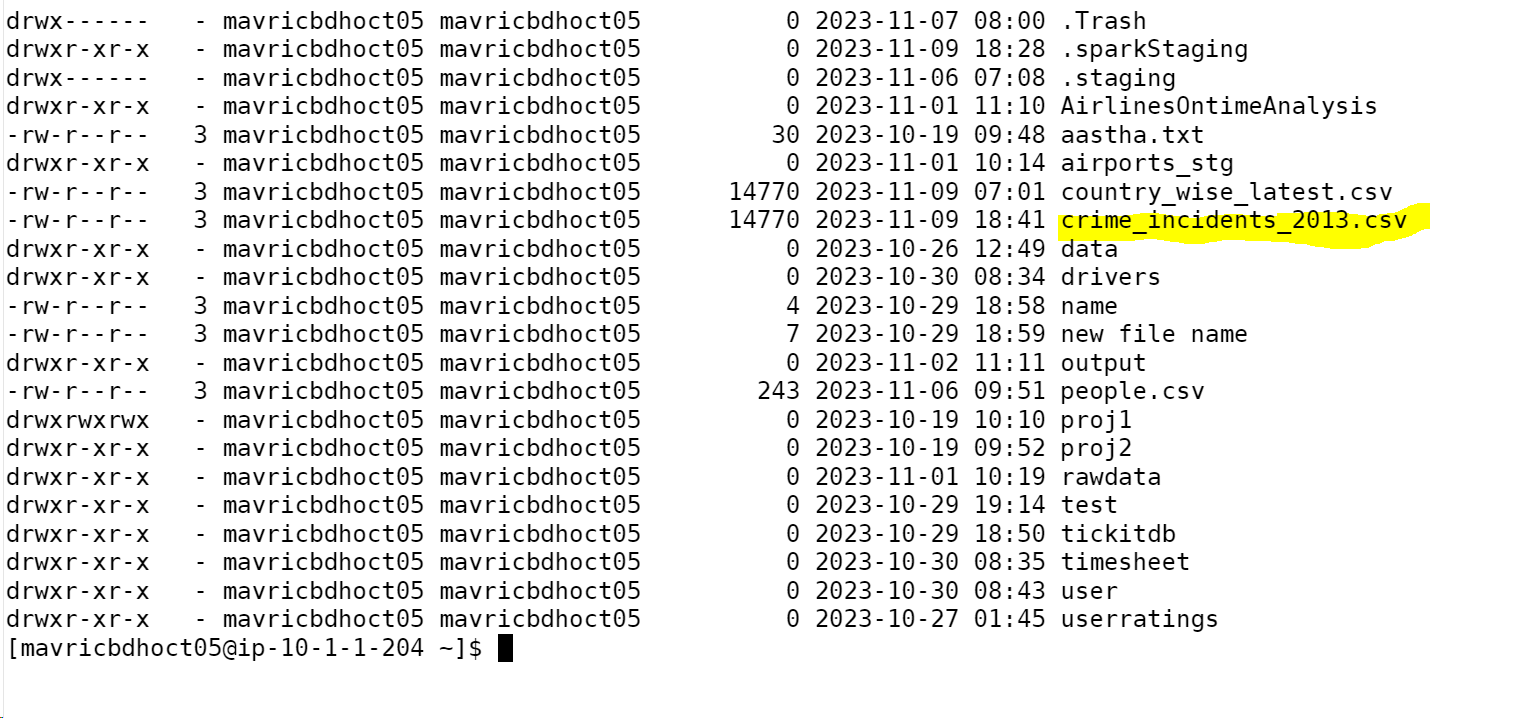
**Task 2 crime Analysis:**

Crime Data Analysis

Dataset: crime\_incidents\_2013.csv

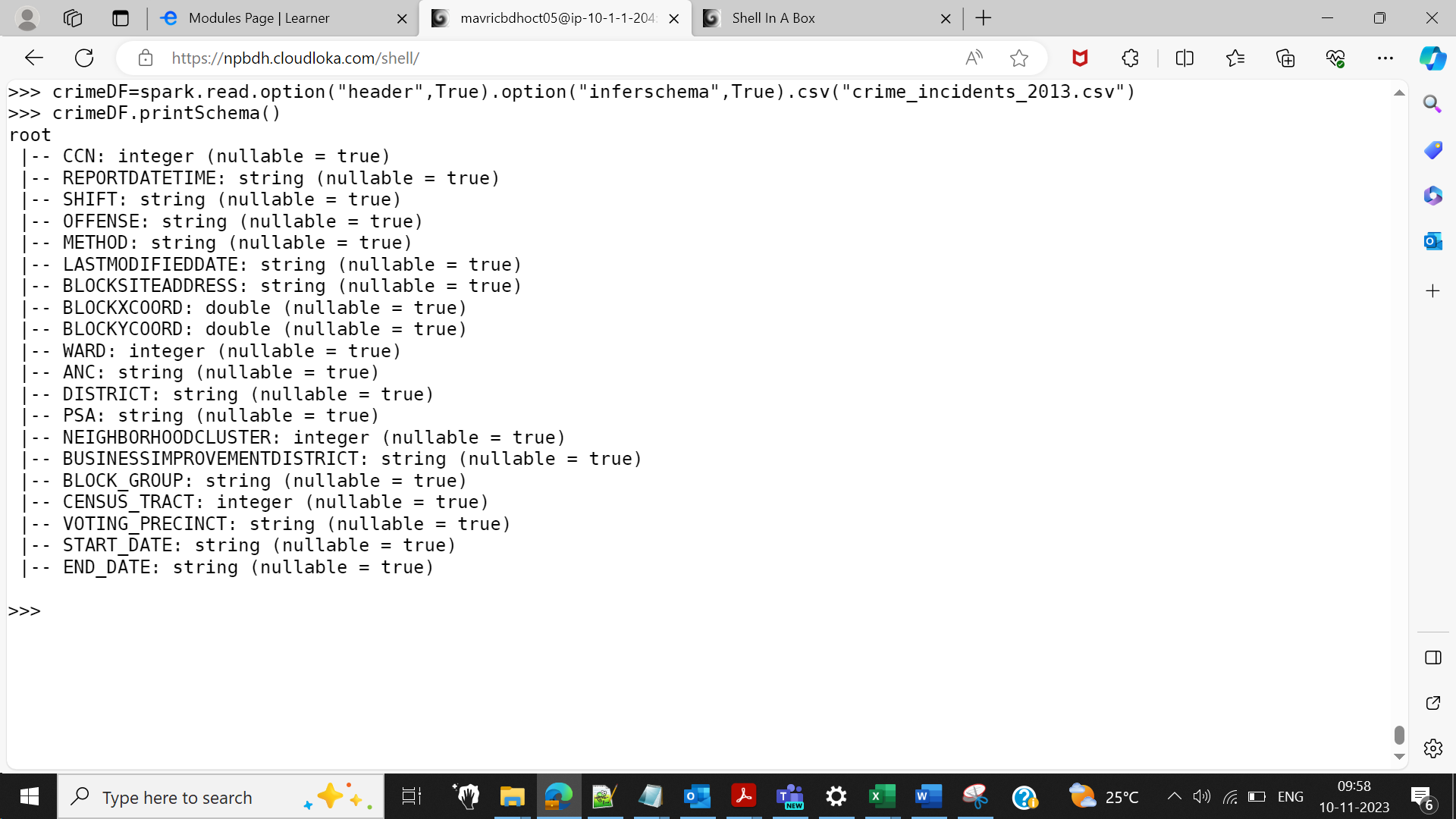
hdfs dfs -put crime\_incidents\_2013.csv

hdfs dfs -ls



crimeDF=spark.read.option("header",True).option("inferschema",True).csv("crime\_incidents\_2013.csv")

crimeDF.printSchema()



crimeDF.show()

A screenshot of a computer

Description automatically generated

crimeDF.createOrReplaceTempView("crime")

7. How many offenses were there in 2013?

spark.sql("SELECT count(distinct OFFENSE) from crime").show()

>>> spark.sql("SELECT count(distinct OFFENSE) from crime").show()

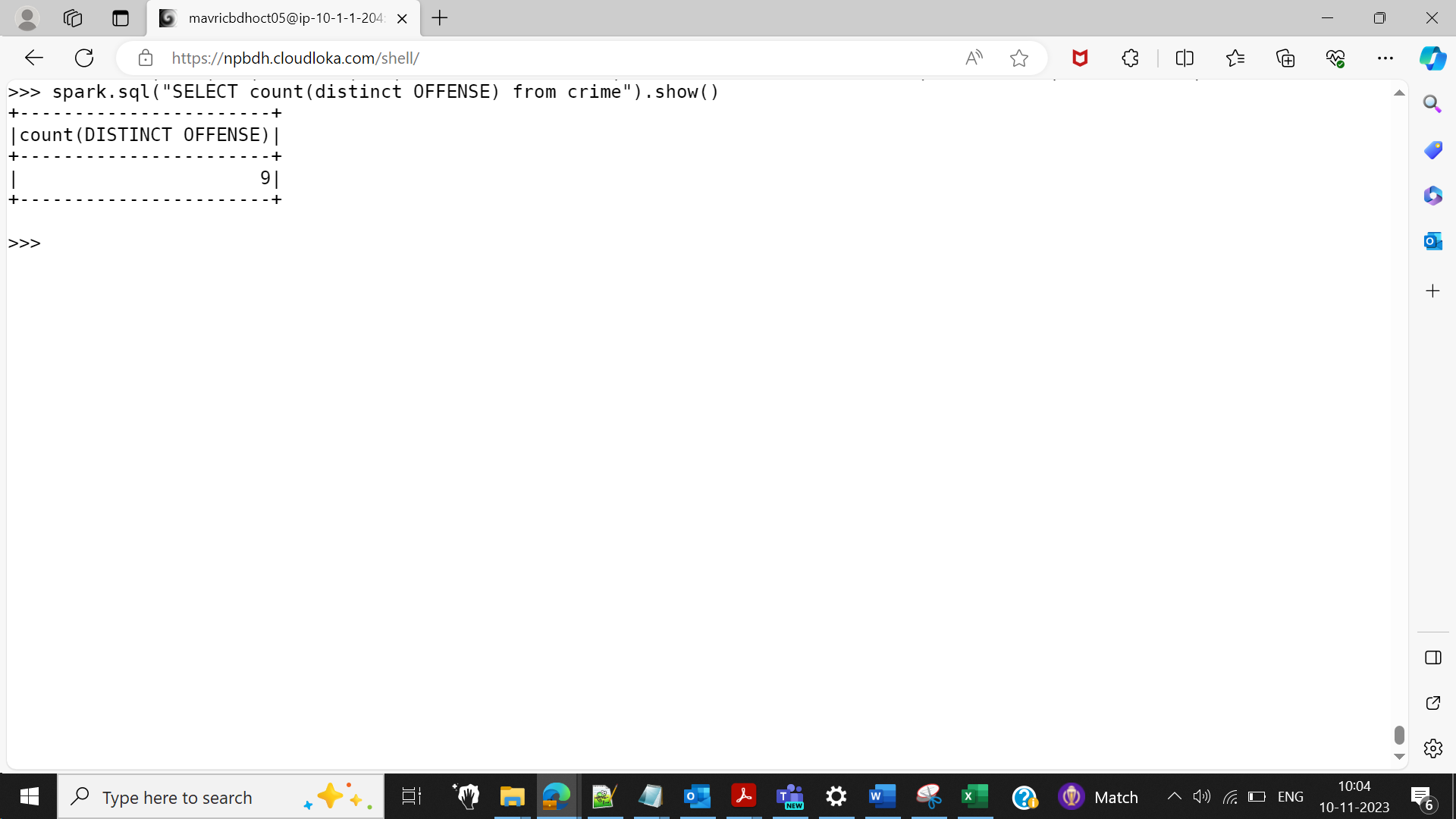
+-----------------------+

|count(DISTINCT OFFENSE)|

+-----------------------+

| 9|

+-----------------------+



8.How many crimes were committed in each offense?

spark.sql("SELECT OFFENSE ,count(CCN) from crime group by OFFENSE").show()

>>> spark.sql("SELECT OFFENSE ,count(CCN) from crime group by OFFENSE").show()

+--------------------+----------+

| OFFENSE|count(CCN)|

+--------------------+----------+

|ASSAULT W/DANGERO...| 2311|

| THEFT F/AUTO| 10134|

| ARSON| 35|

| MOTOR VEHICLE THEFT| 2677|

| ROBBERY| 4072|

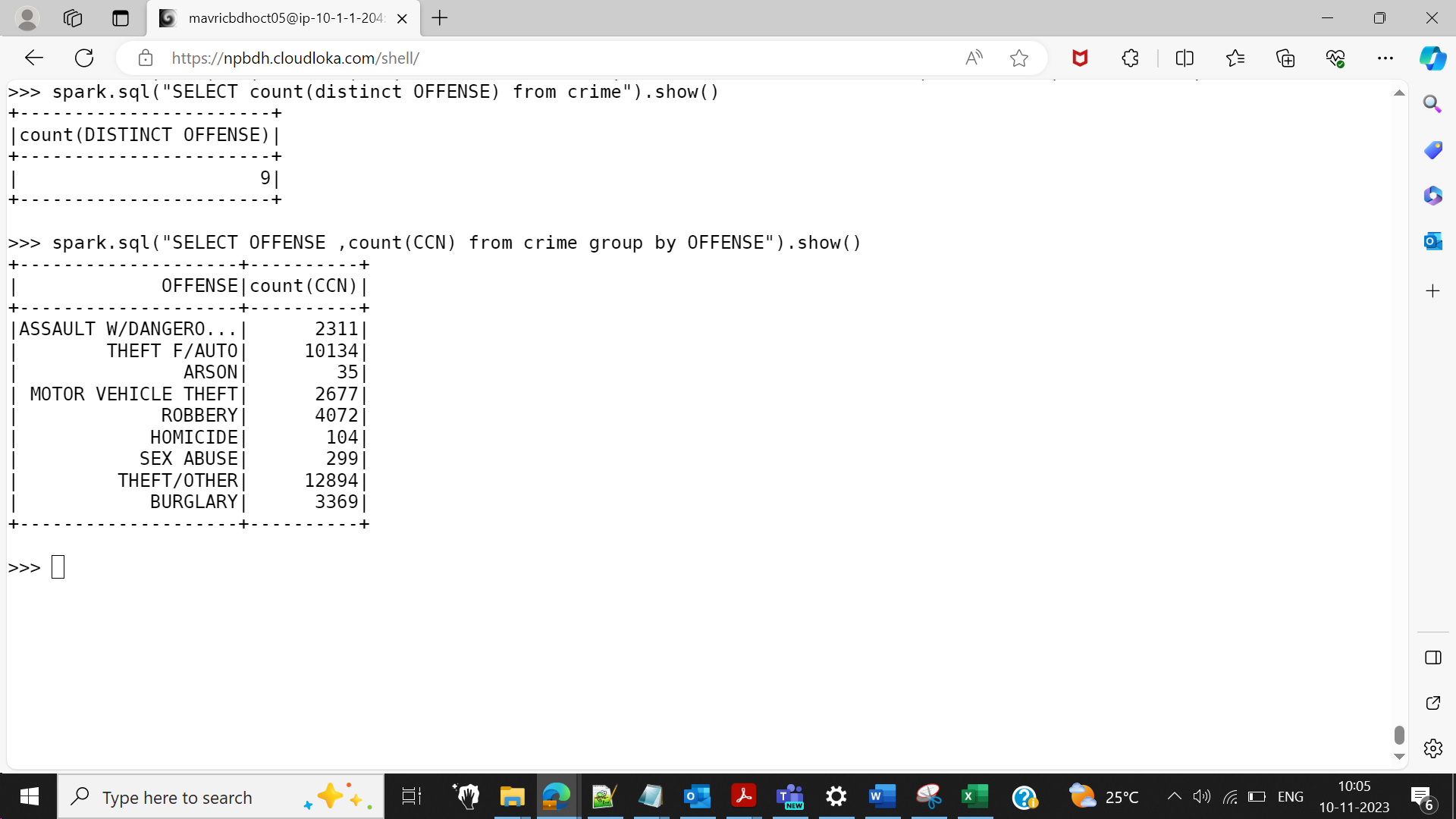
| HOMICIDE| 104|

| SEX ABUSE| 299|

| THEFT/OTHER| 12894|

| BURGLARY| 3369|

+--------------------+----------+



9.How many different methods were used (And their count) in offense

“Homicide”

spark.sql("SELECT METHOD ,count(METHOD) from crime where OFFENSE = 'HOMICIDE' group by METHOD ").show()

>>> spark.sql("SELECT METHOD ,count(METHOD) from crime where OFFENSE = 'HOMICIDE' group by METHOD ").show()

+------+-------------+

|METHOD|count(METHOD)|

+------+-------------+

|OTHERS| 12|

| KNIFE| 11|

| GUN| 81|

+------+-------------+

A screenshot of a computer

Description automatically generated

10.Which shift had the maximum crimes?

spark.sql("SELECT SHIFT ,max(CCN) from crime group by SHIFT order by 2 desc").show(1)

>>> spark.sql("SELECT SHIFT ,max(CCN) from crime group by SHIFT order by 2 desc").show(1)

+-------+--------+

| SHIFT|max(CCN)|

+-------+--------+

|EVENING|14400271|

+-------+--------+

only showing top 1 row

>>>

