**Pyspark Assignment Task 2**

**Crime Data Analysis**

**Dataset: crime\_incidents\_2013.csv**

hdfs dfs -put crime\_incidents\_2013.csv

hdfs dfs -ls

A screenshot of a computer

Description automatically generated

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

crimeDF.createOrReplaceTempView("crime")

crimeDF.printSchema()

A screenshot of a computer

Description automatically generated

crimeDF.show()

A screenshot of a computer

Description automatically generated

1. **How many offenses were there in 2013?**

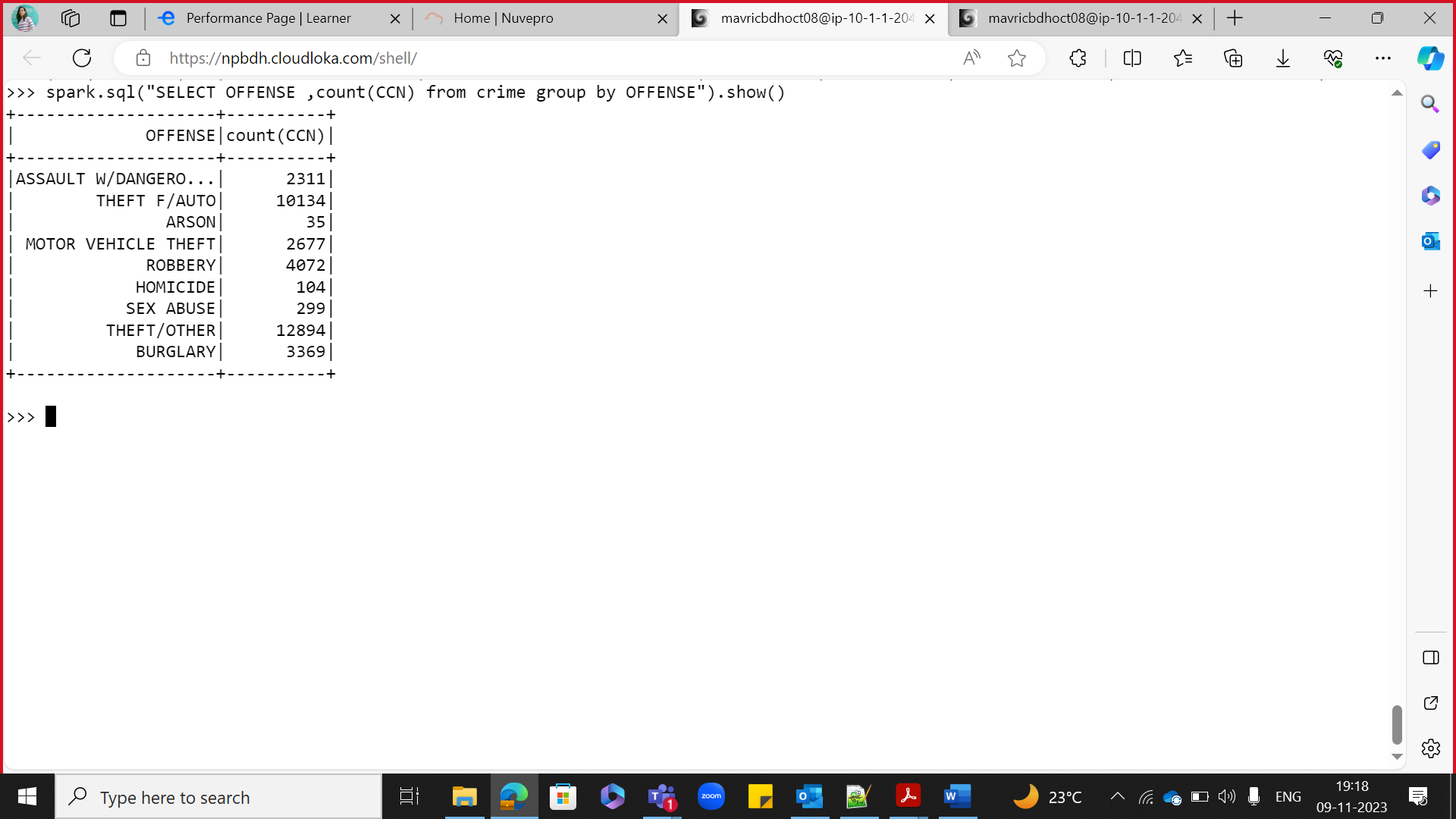
spark.sql("SELECT count(distinct OFFENSE) from crime").show()

A screenshot of a computer

Description automatically generated

**8.How many crimes were committed in each offense?**

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



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

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

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)

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated