**Crime Data Analysis**

Dataset: crime\_incidents\_2013.csv

hdfs dfs -put crime\_incidents\_2013.csv

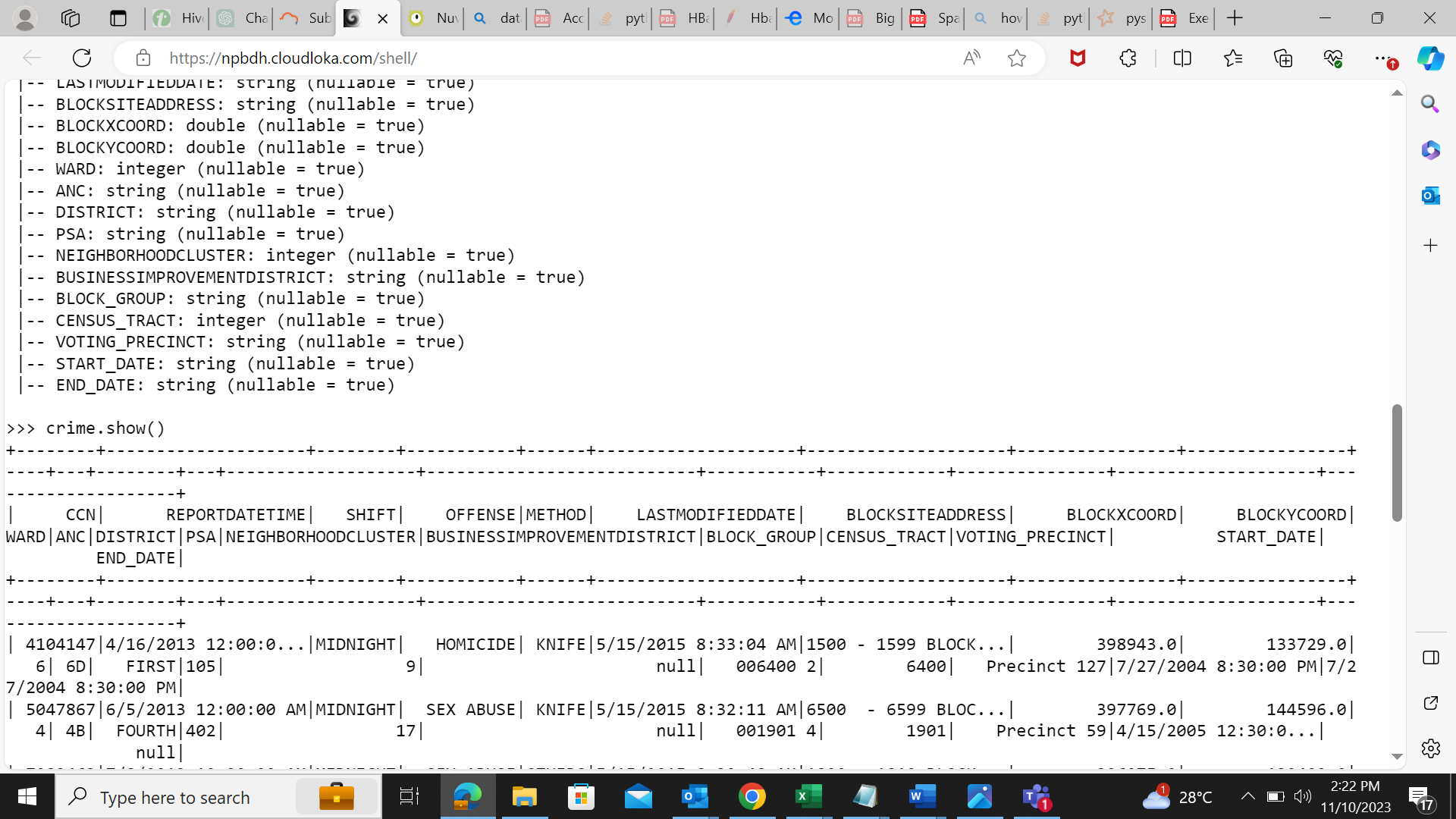
hdfs dfs -ls

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

crime.createOrReplaceTempView("crime1")

crime.printSchema()

crime.show()



1. How many offenses were there in 2013?

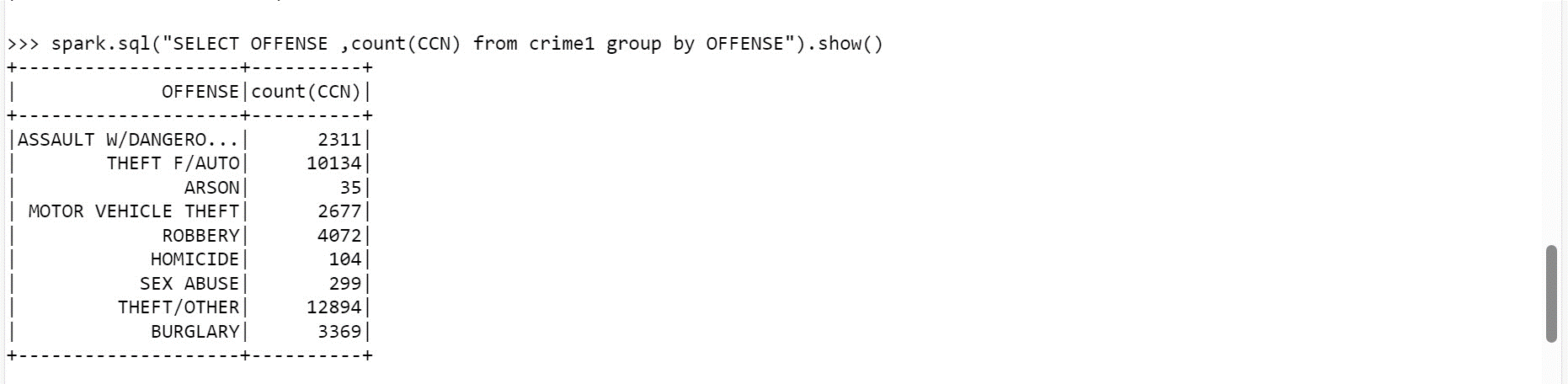
spark.sql("SELECT count(distinct OFFENSE) from crime1").show()

A close-up of a white background

Description automatically generated

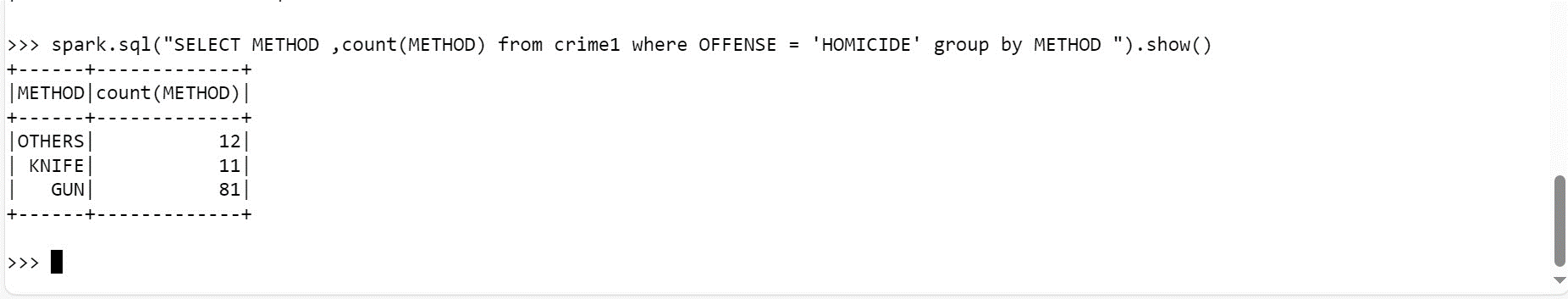
1. . How many crimes were committed in each offense?

spark.sql("SELECT OFFENSE ,count(CCN) from crime1 group by OFFENSE").show()



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

spark.sql("SELECT METHOD ,count(METHOD) from crime1 where OFFENSE = 'HOMICIDE' group by METHOD ").show()



1. Which shift had the maximum crimes?

spark.sql("SELECT SHIFT ,max(CCN) from crime1 group by SHIFT order by 2 desc").show()

