hdfs dfs -put crime\_incidents\_2013.csv

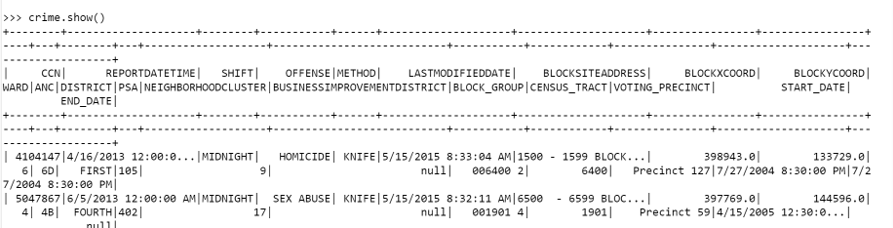
hdfs dfs -ls

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

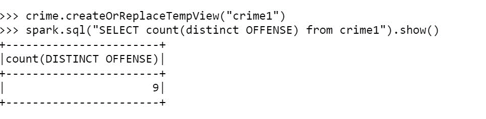
crime.createOrReplaceTempView("crime1")

crime.printSchema()

crime.show()



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





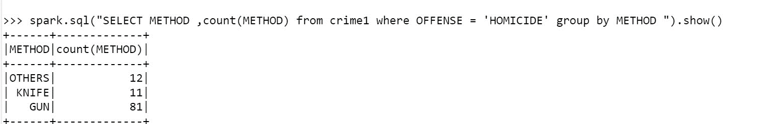
spark.sql("SELECT OFFENSE ,count(CCN) from crime1 group by OFFENSE").show()

A screenshot of a computer

Description automatically generated



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





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

A close-up of a white background

Description automatically generated