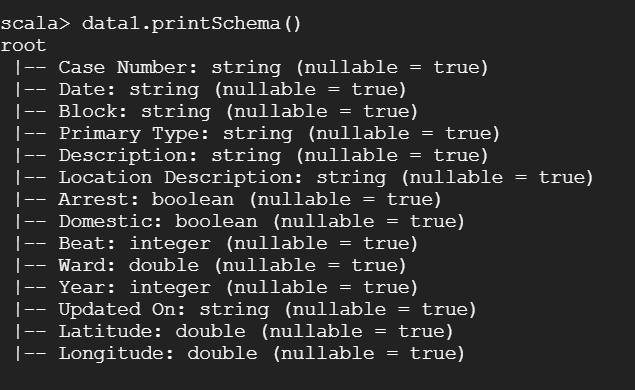
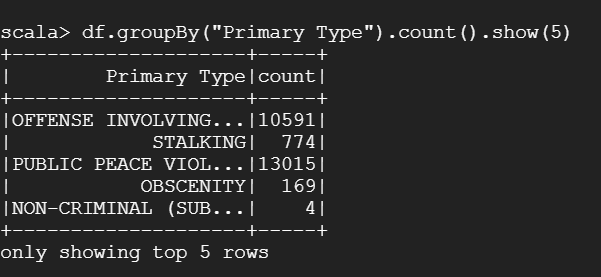
val data1 = spark.read.options(Map("inferSchema"->"true","delimiter"->",","header"->"true")).csv("hdfs://cluster-9eee-m/user/soumi\_mitra2/Crimes.csv")

data1.printSchema()

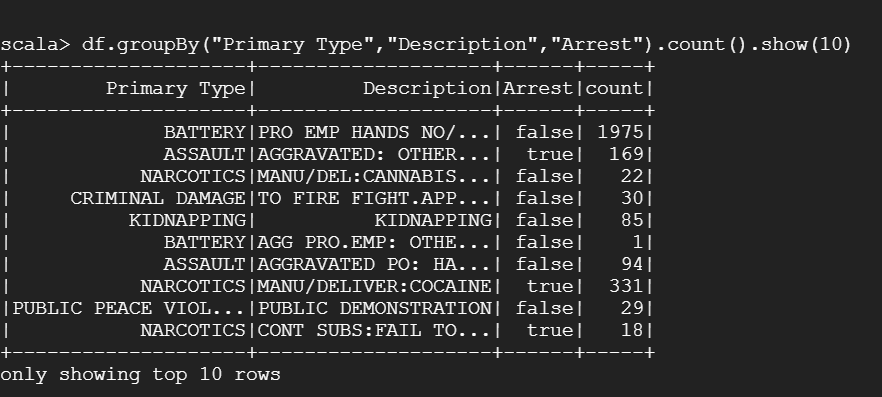


val df= data1.toDF("Case Number","Date","Block","Primary Type","Description","Location Description","Arrest","Domestic","Beat","Ward","Year","Updated On","Latitude","Longitude")

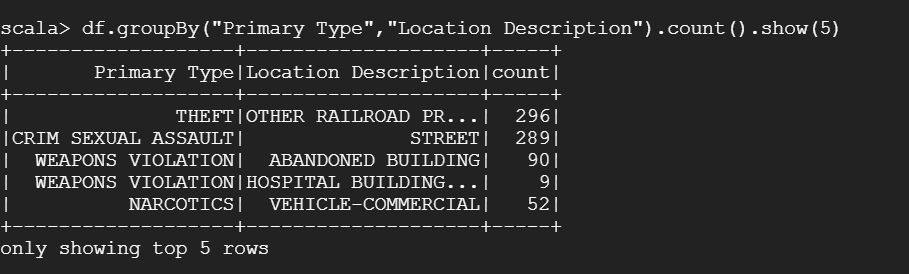
df.groupBy("Primary Type").count().show(5)



df.groupBy("Primary Type","Description","Arrest").count().show(10)



df.groupBy("Primary Type","Location Description").count().show(5)



df.groupBy("Description","Arrest").count().show(5)

