**Crime Data Analysis**

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

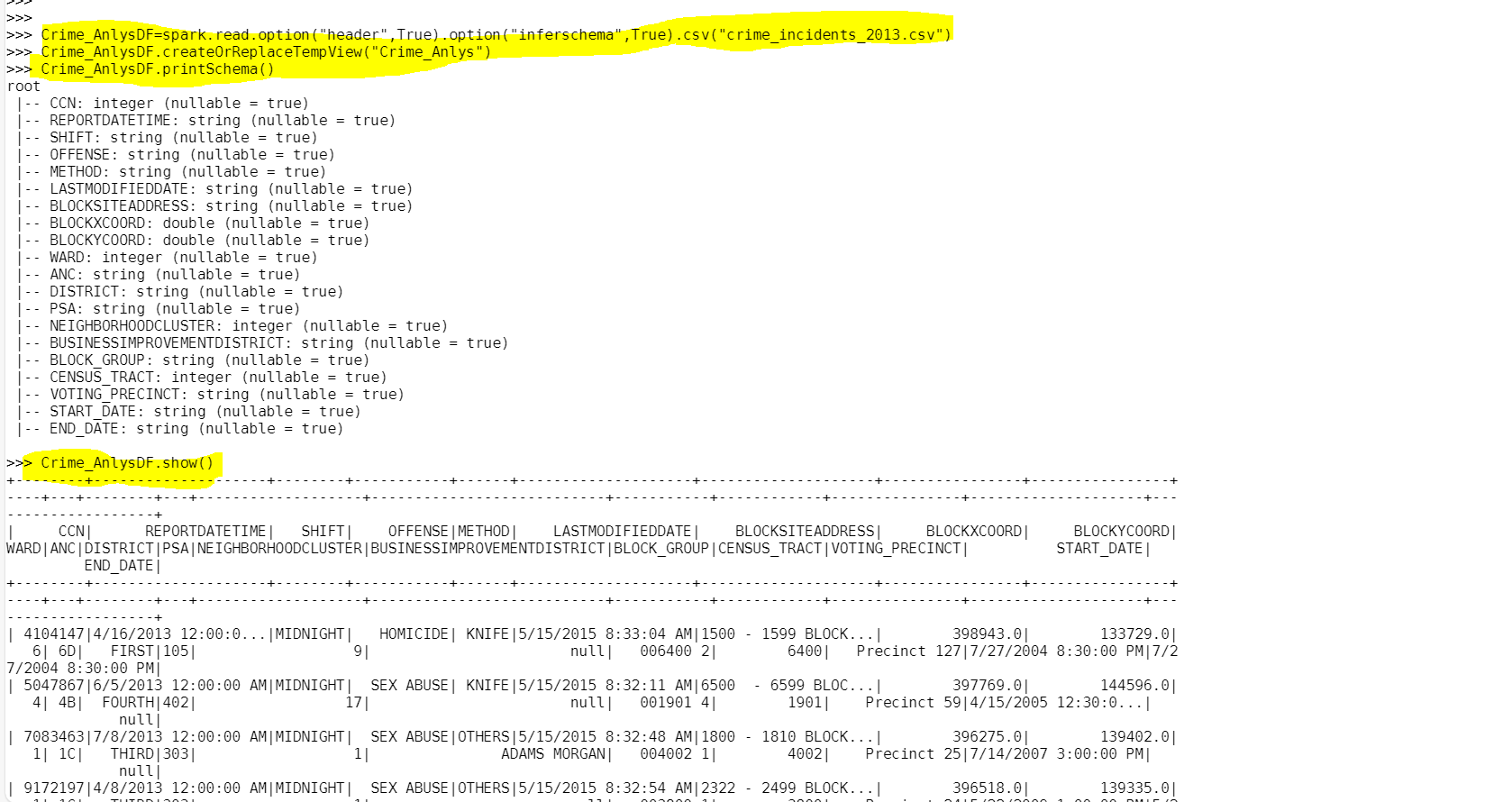
Step 1: Loading data into Dataframe

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

*>>> Crime\_AnlysDF.createOrReplaceTempView ("Crime\_Anlys")*

*>>> Crime\_AnlysDF.printSchema()*

*>>> Crime\_AnlysDF.show()*



1. How many offenses were there in 2013?

Distinct Offenses happened in 2013

*>>> spark.sql("SELECT count(distinct OFFENSE) from Crime\_Anlys").show()*

A close-up of a white background

Description automatically generated

No of Offenses happened in 2013

*>>> spark.sql("SELECT count(OFFENSE) from Crime\_Anlys").show()*

A screenshot of a computer

Description automatically generated

1. How many crimes were committed in each offense?

*>>> spark.sql("SELECT OFFENSE ,count(CCN) from Crime\_Anlys group by OFFENSE").show()*

A screenshot of a computer

Description automatically generated

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

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

A white screen with black text

Description automatically generated

1. Which shift had the maximum crimes?

Evening shift had the maximum crimes.

*>>> spark.sql("SELECT SHIFT ,max(CCN) from Crime\_Anlys group by SHIFT order by 2 desc").show()*

*>>> spark.sql("SELECT SHIFT ,max(CCN) from Crime\_Anlys group by SHIFT order by 2 desc limit 1").show()*

A screenshot of a computer

Description automatically generated