**Crime Data Analysis**

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

**Solution**:

from pyspark.sql import SparkSession

from pyspark.sql.functions import col

spark = SparkSession.builder.appName("CrimeAnalysis").getOrCreate()

file\_path = "dbfs:/FileStore/tables/crime\_incidents\_2013.csv"

crime\_df = spark.read.csv(file\_path, header=True, inferSchema=True)

crime\_df.display()

offenses\_count = crime\_df.filter(col("END\_DATE") == 2013).select("OFFENSE").distinct().count()

print(f"Number of offenses in 2013: {offenses\_count}")

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

**Solution:**

crimes\_per\_offense = crime\_df.groupBy("OFFENSE").count()

crimes\_per\_offense.show()

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

**Solution:**

homicide\_methods\_count = crime\_df.filter(col("OFFENSE") == "HOMICIDE").groupBy("Method").count()

homicide\_methods\_count.show()

**10. Which shift had the maximum crimes?**

**Solution:**

max\_crimes\_shift = crime\_df.groupBy("Shift").count().orderBy(col("count").desc()).first()["Shift"]

print(f"The shift with the maximum crimes is: {max\_crimes\_shift}")