Spark Queries

To connect spark => spark-shell
To load the data => val df = spark.read.format("csv").option("inferSchema", "true").option("header","true").load("s3://bigdataassignm/DelayedFlights-updated.csv")

df.createOrReplaceTempView("delay_flights");

spark.time(spark.sql("SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights WHERE Year>=2003 AND Year<=2010 GROUP BY Year ORDER BY Year").show())

spark.time(spark.sql("SELECT Year, avg((NASDelay /ArrDelay)*100) from delay_flights WHERE Year>=2003 AND Year<=2010 GROUP BY Year ORDER BY Year").show())

spark.time(spark.sql("SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_flights WHERE Year>=2003 AND Year<=2010 GROUP BY Year ORDER BY Year").show())

spark.time(spark.sql("SELECT Year, avg((SecurityDelay /ArrDelay)*100) from delay_flights WHERE Year>=2003 AND Year<=2010 GROUP BY Year ORDER BY Year").show())

spark.time(spark.sql("SELECT Year, avg((LateAircraftDelay /ArrDelay)*100) from delay_flights WHERE Year>=2003 AND Year<=2010 GROUP BY Year ORDER BY Year").show())