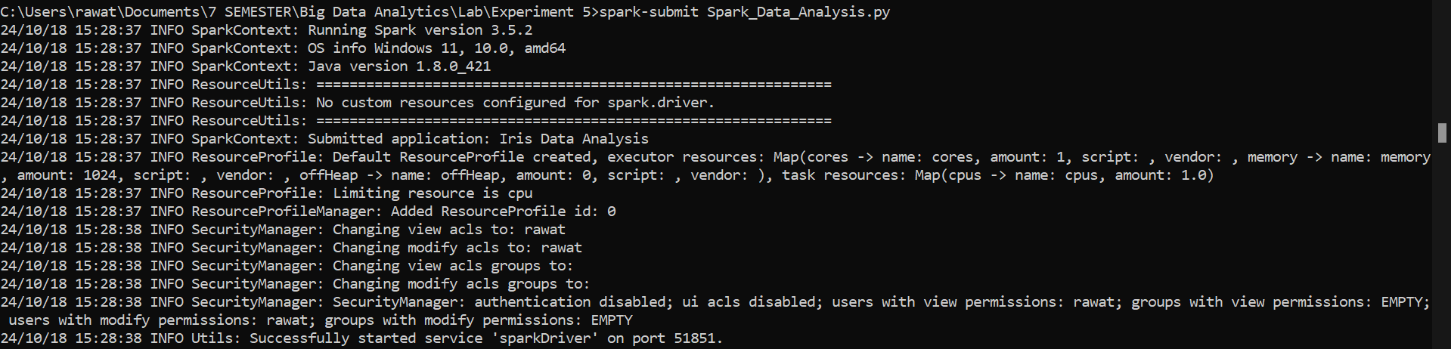
***Experiment 5***

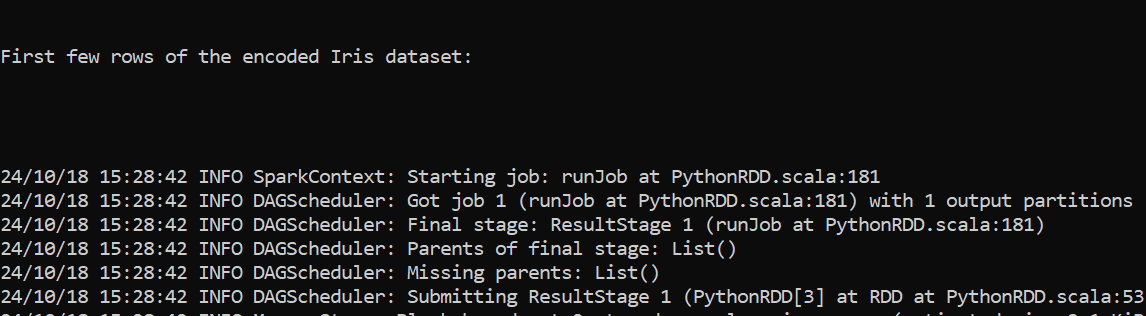
**Aim:** Data Analysis with Spark: Use Apache Spark to analyze large datasets by loading them into Spark RDDs (Resilient Distributed Datasets) and performing operations such as filtering, mapping, and aggregation.

***Directory:***

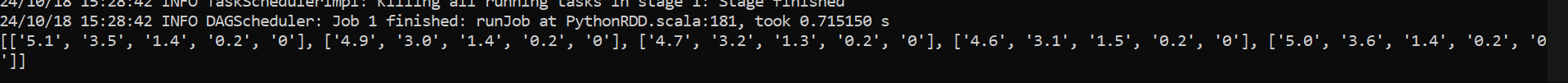


***Fig 1: Running Spark in Directory***

***Displaying Dataset:***

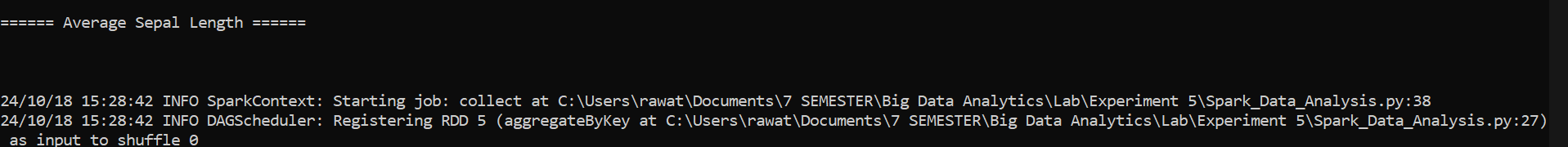
******

***Fig 2: Displaying First Rows Messages***

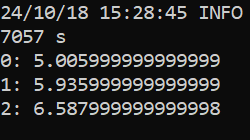
******

***Fig 3: Displaying First Rows***

***Filtering and Aggregation in Dataset:***

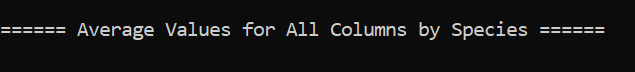
******

***Fig 4: Displaying Aggregation Messages***

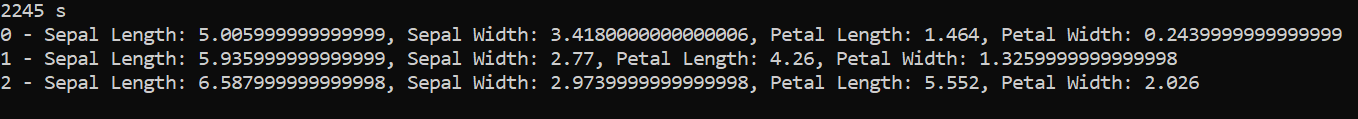
******

***Fig 5: Aggregation by Sepal Length***

***Mapping Species and Aggregation in Dataset:***

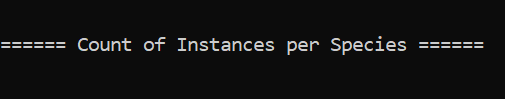
******

***Fig 6: Average Values for Species Column Message***

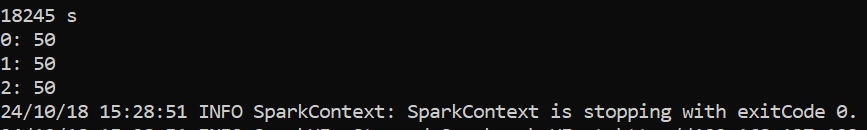
******

***Fig 7: Average Values for Species Column***

***Counting Occurrence of Species in Dataset:***

******

***Fig 8: Count of Column Message***

******

***Fig 9: Count of Column Unique Values in Species***