**SEMESTER - II 2023-2024**

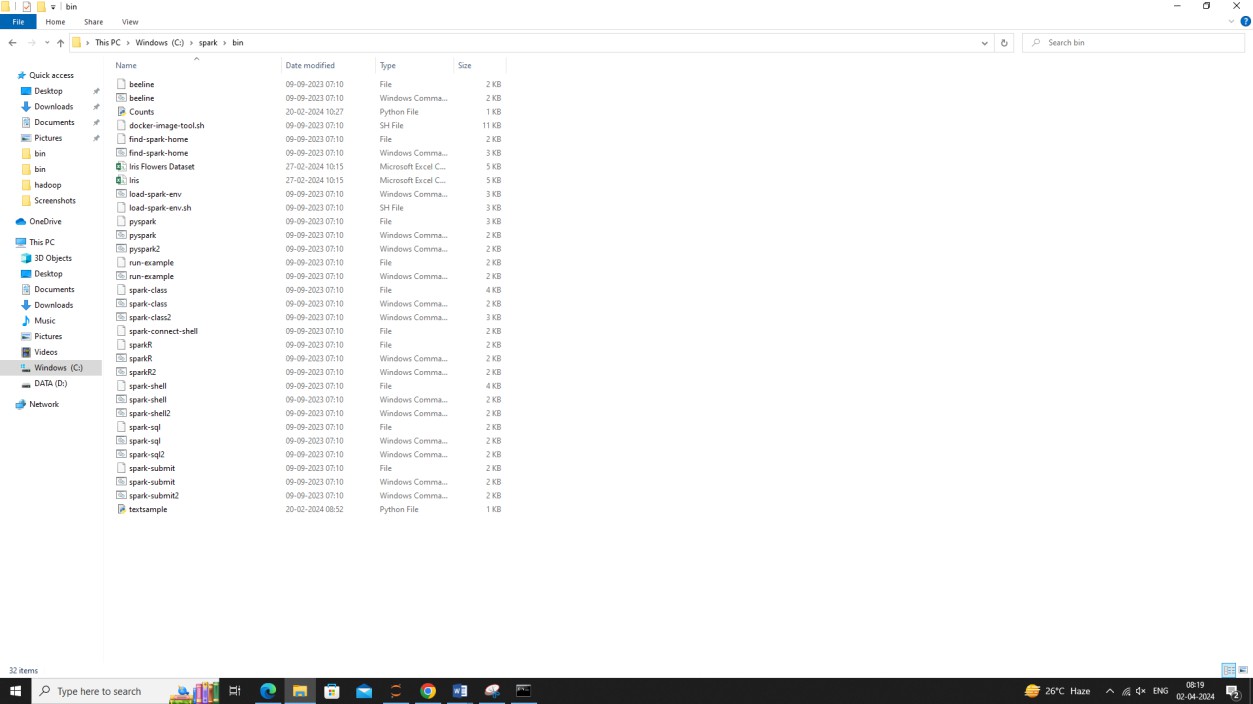
**BIG DATA TECHNOLOGIES**

**PRACTICAL DOCUMENTATION**

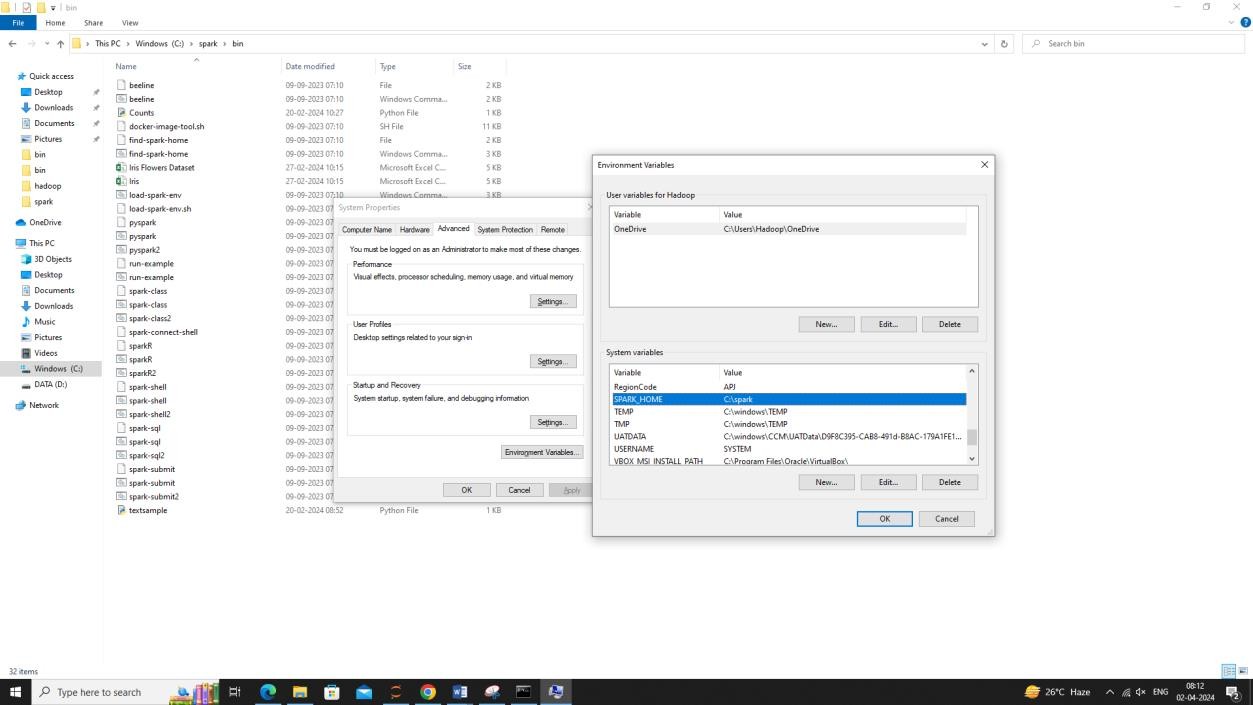
**LIST OF PRACTICALS**

|  |  |
| --- | --- |
| **1** | **Spark Installation** |
| **2** | **Demonstrate RDD operation in Apache Spark** |
| **3** | **Implement wordcount using Spark RDD** |
| **4** | **Demonstrate use of Spark DataFrames** |
| **5** | **Demonstrate use of Spark SQL** |
| **6** | **Case Study on Spark SQL** |
| **7** | **Regression using Spark ML** |
| **8** | **Classification using Spark ML** |
| **9** | **Demonstrate use of Evaluation metrics while build machine learning models in Spark ML** |
| **10** | **Demonstrate use of pipelines while building machine learning models in Spark ML** |

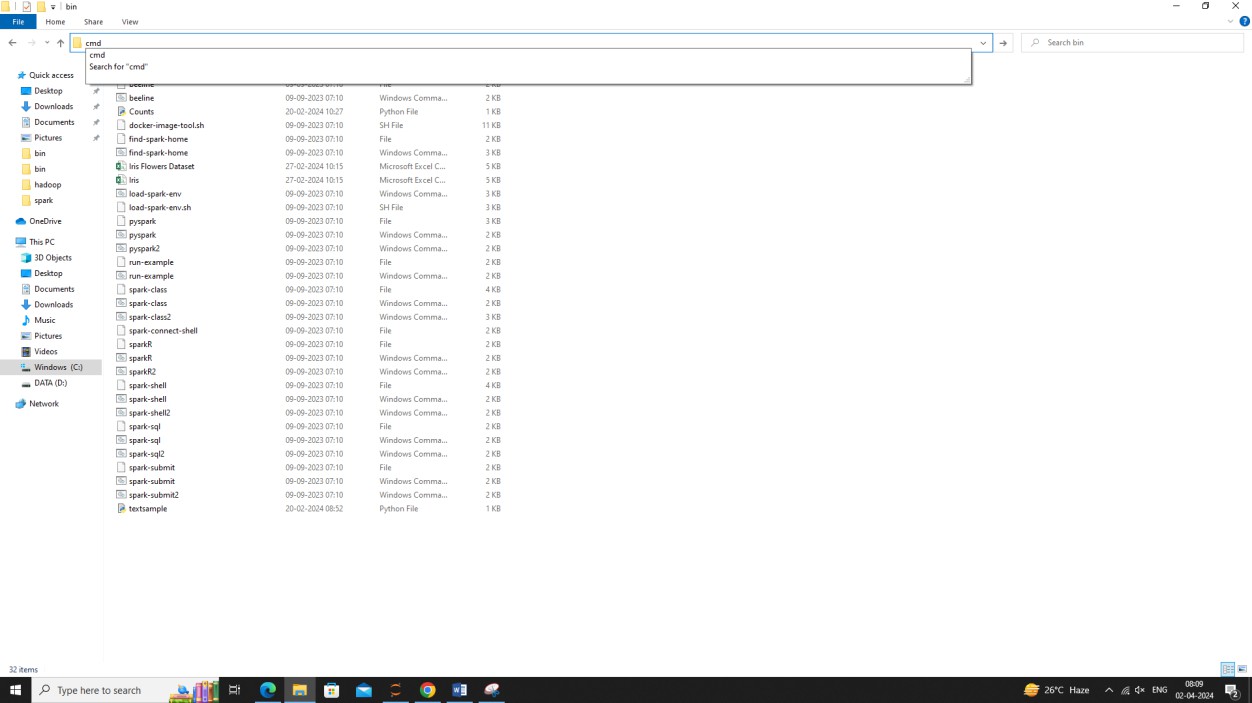
**PRACTICAL NO 1 :Spark Installation**

Step 1: Download Apache Spark and Hadoop

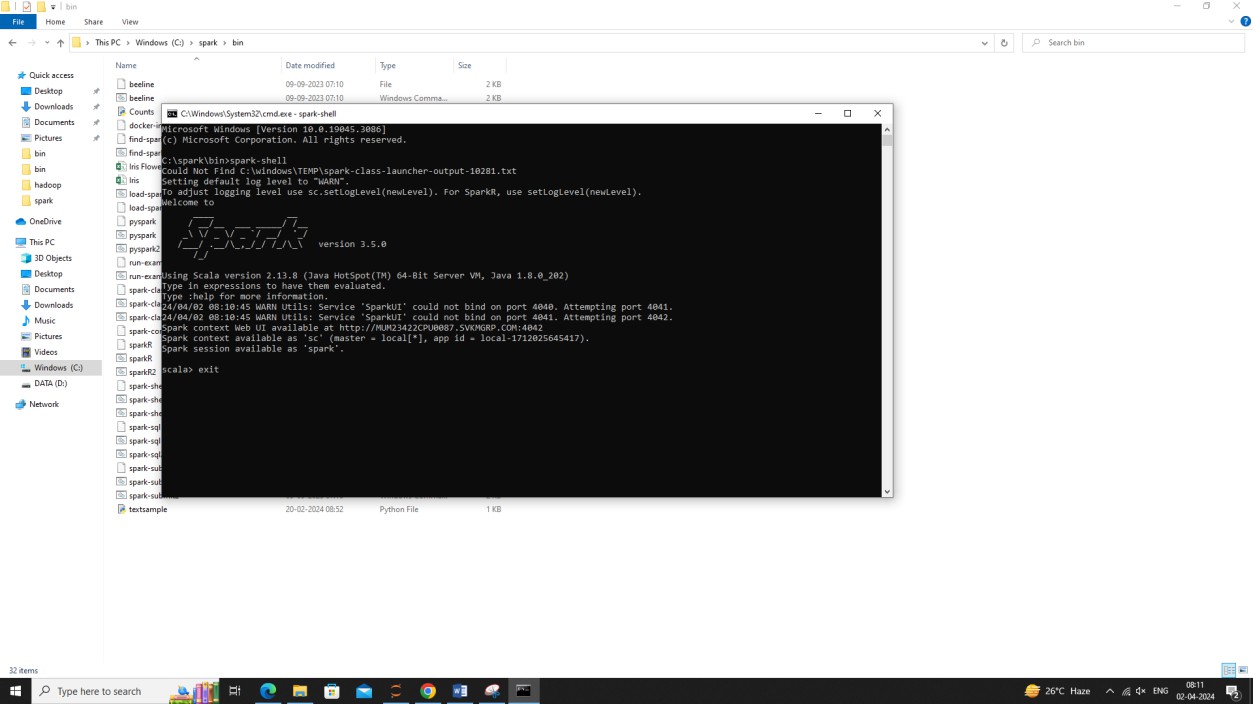
Step 2: Set\_up Spark in environment to create a path



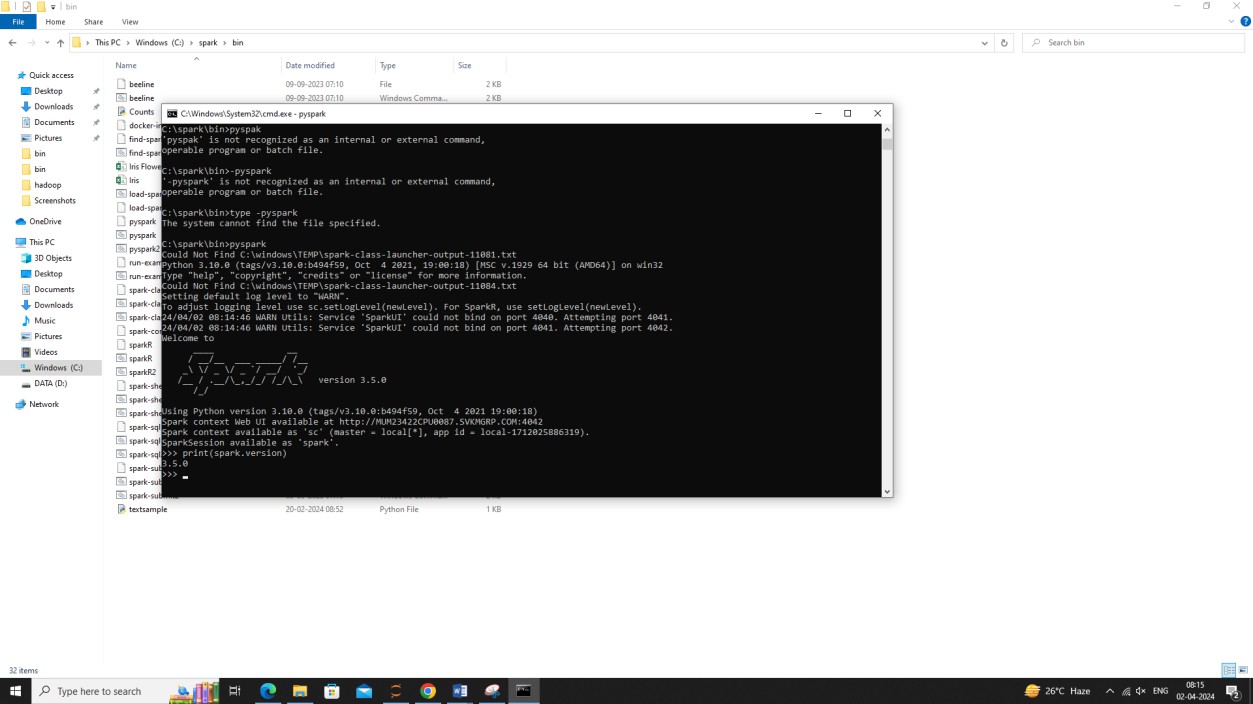
Step 3: Run cmd in bin

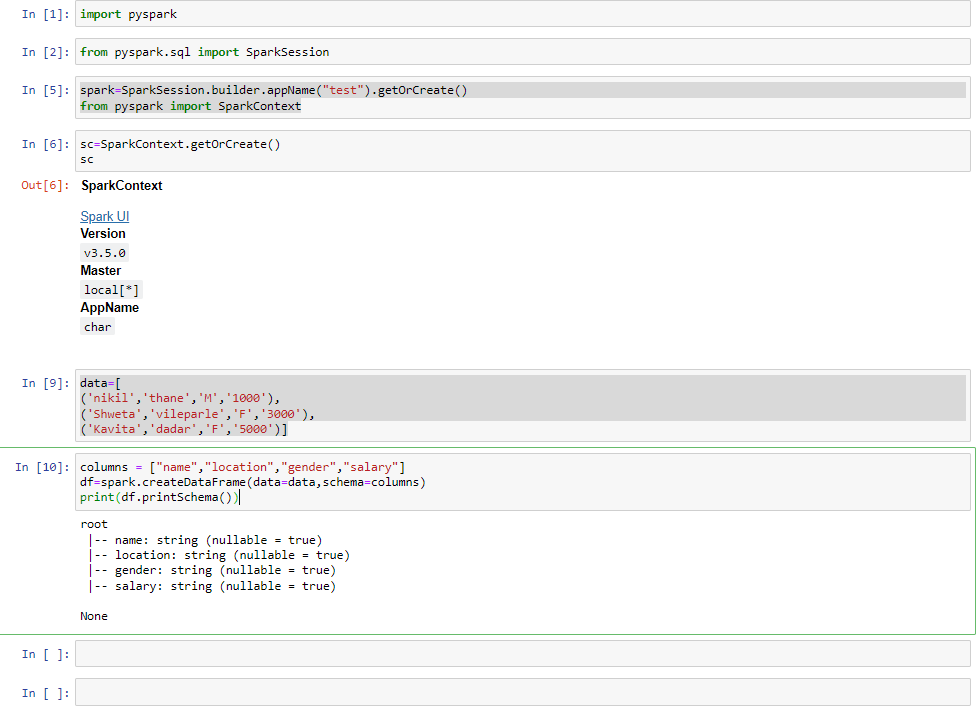


Step 4: type spark-shell on cmd to see if spark is installed

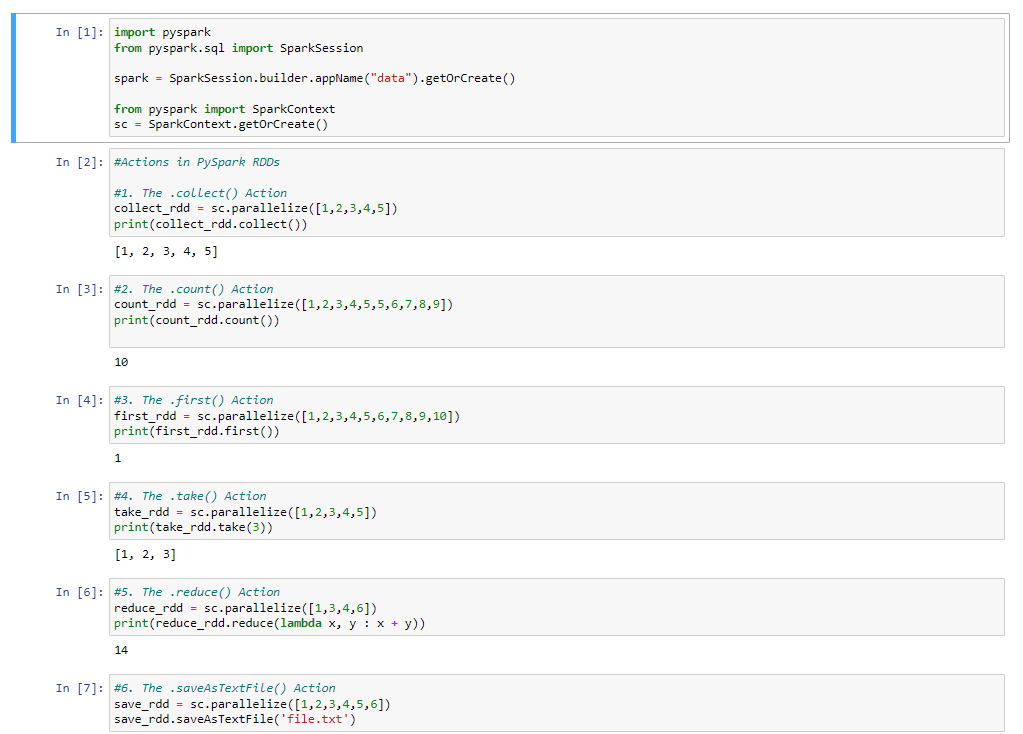


Step 5: Type pyspark and print pyspark version





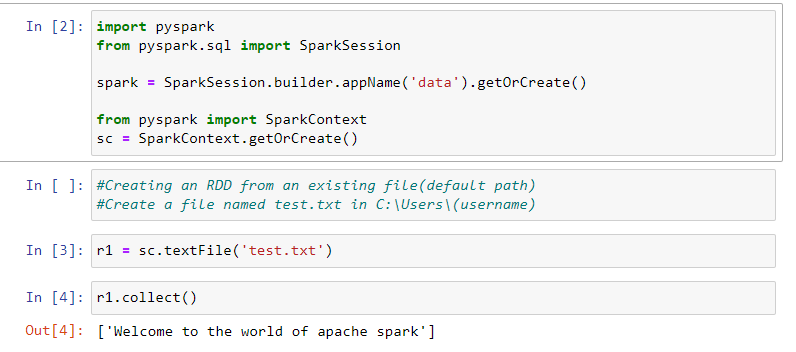
**PRACTICAL 2: Demonstrate RDD operation in Apache Spark**

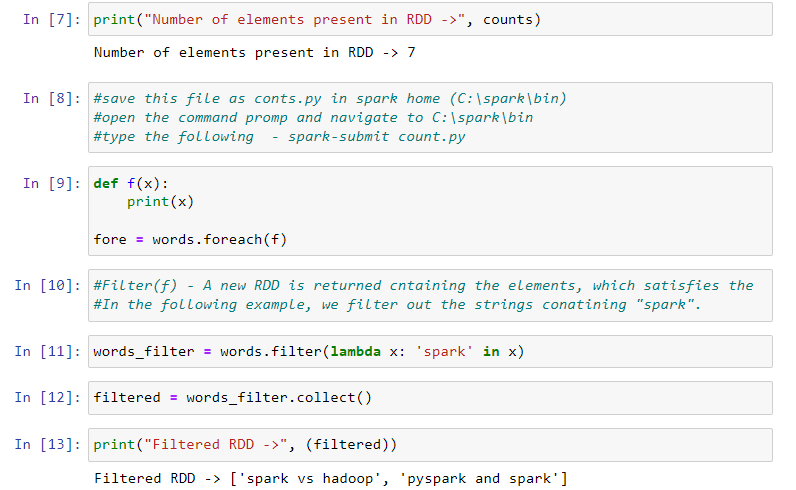
****

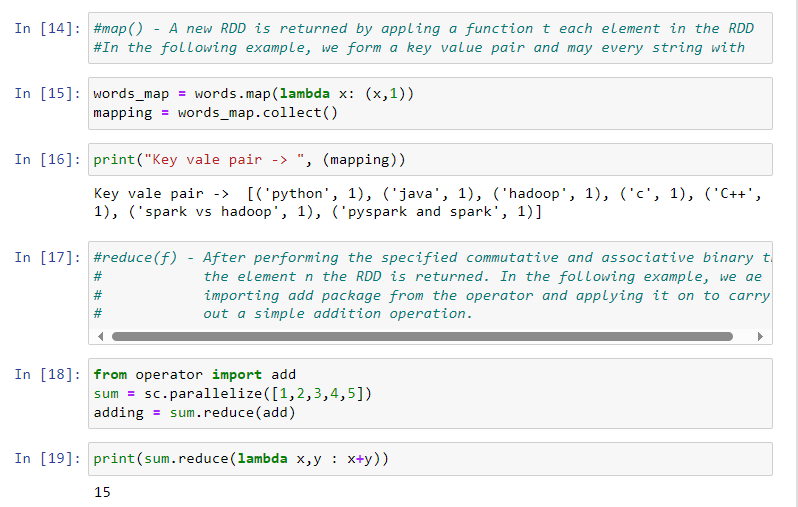
****

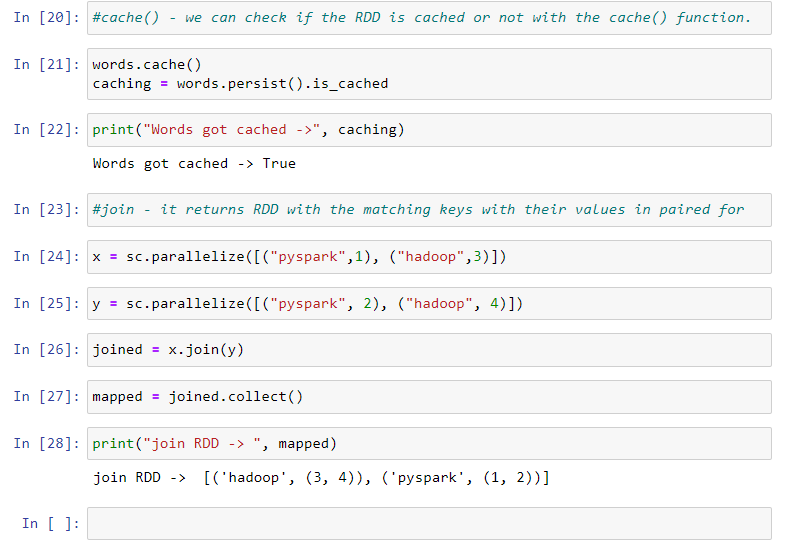
****

**PRACTICAL 3: Implement word count using spark RDD**

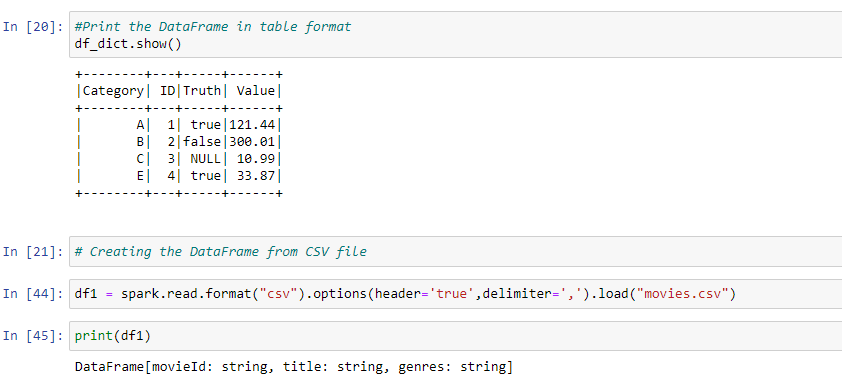
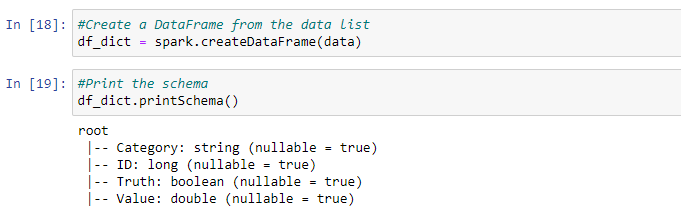
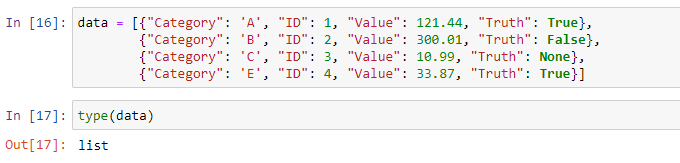
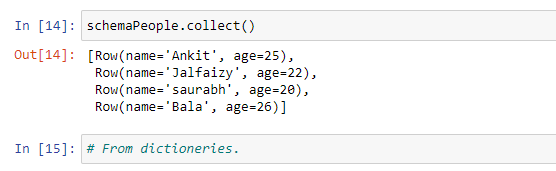
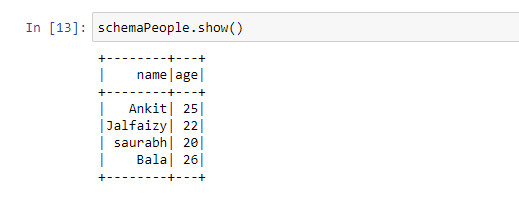
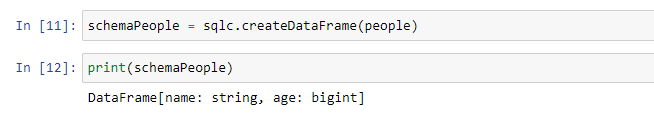
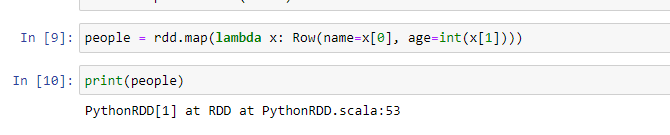
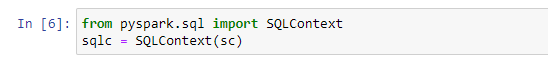
****

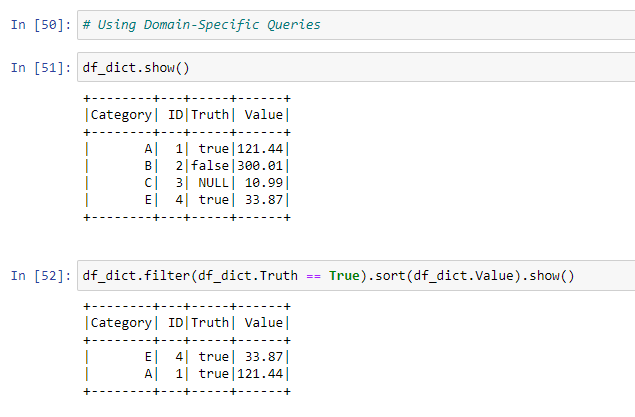
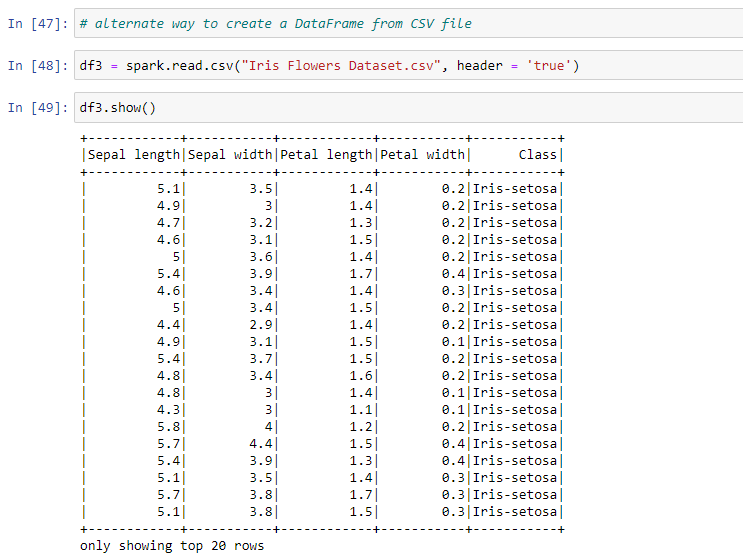
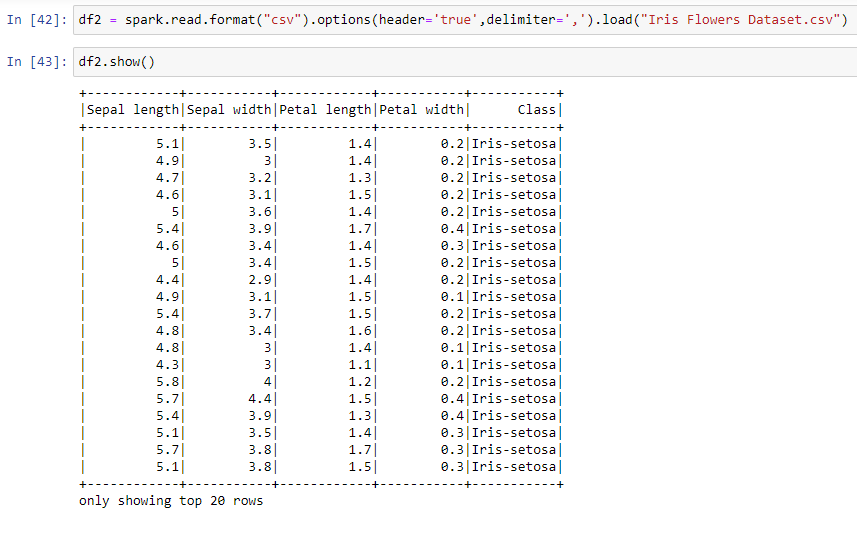
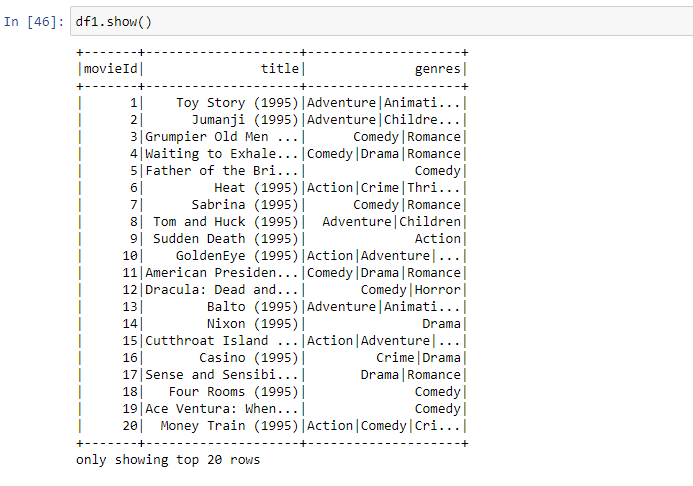
****

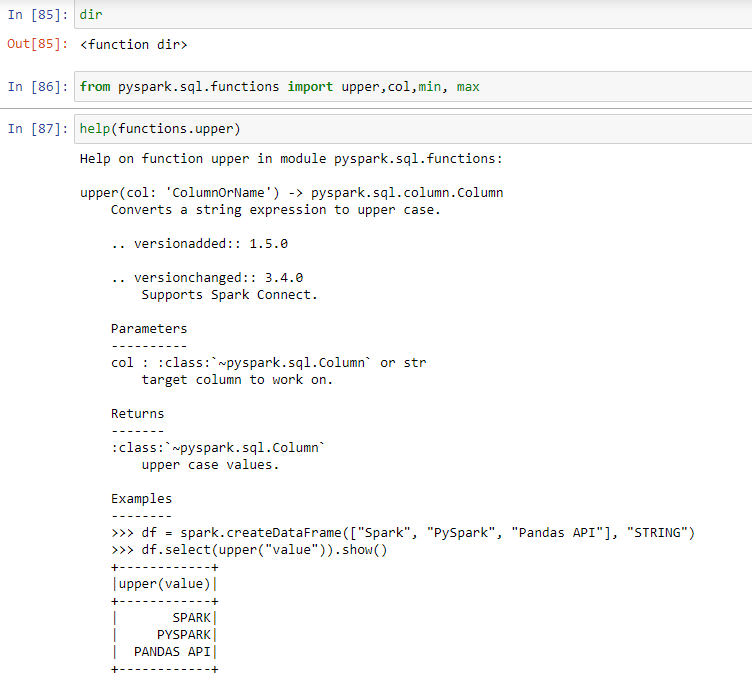
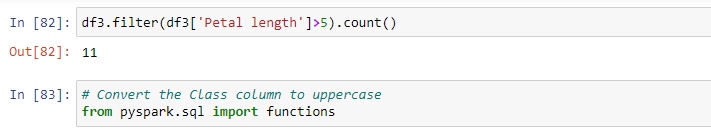
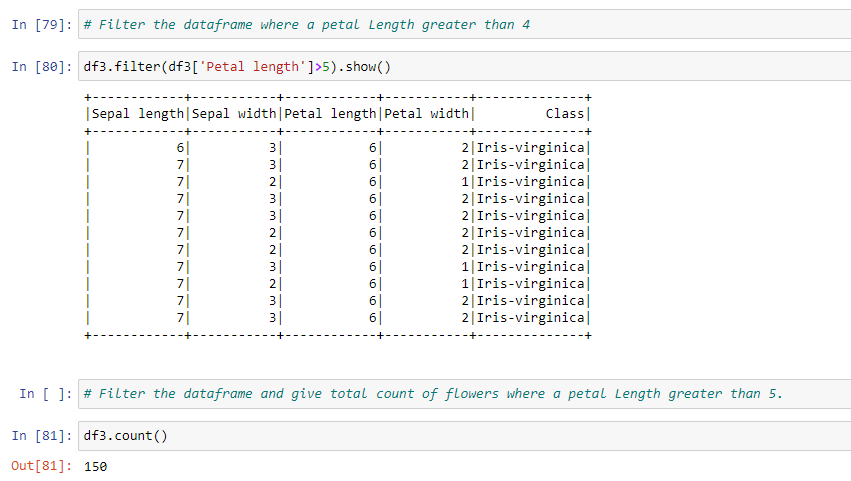
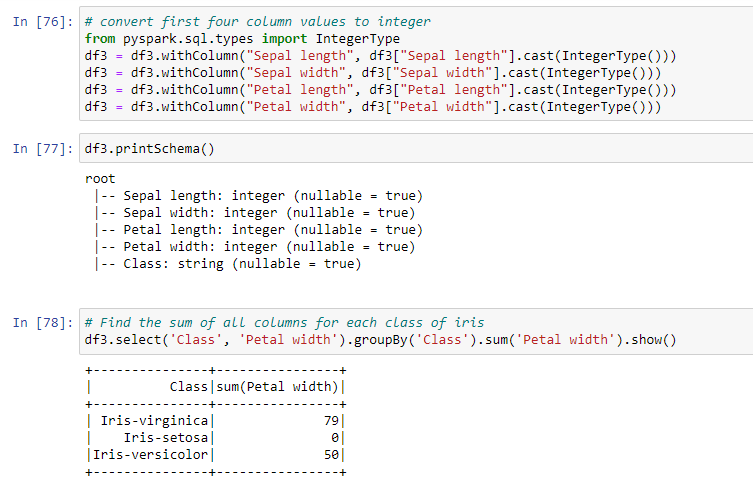
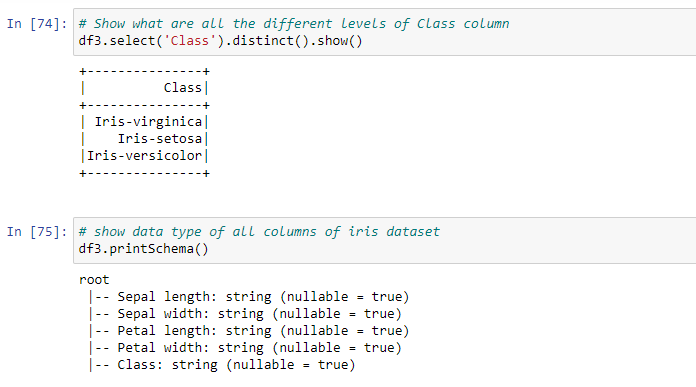
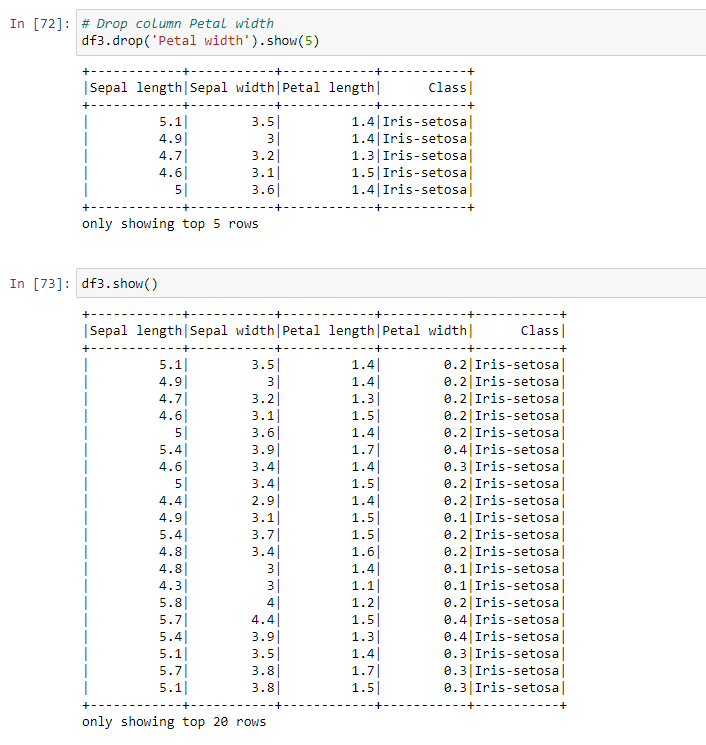
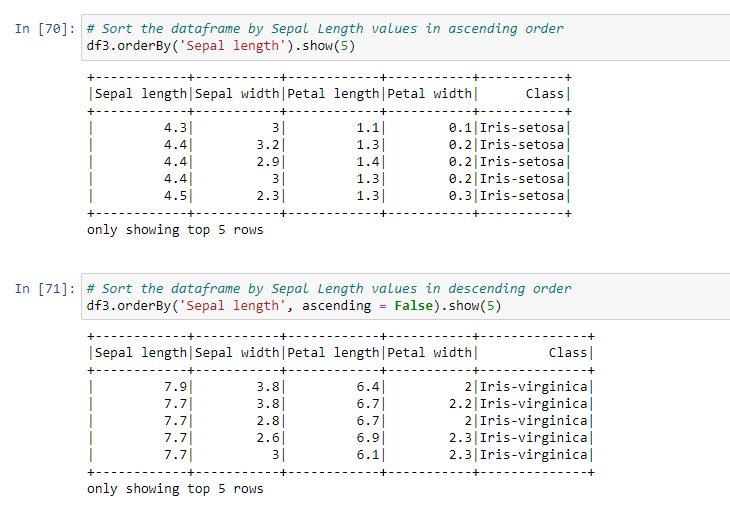
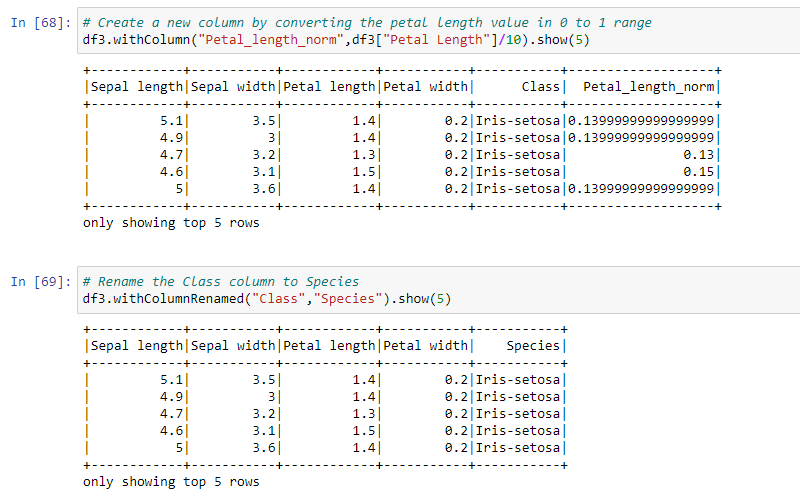
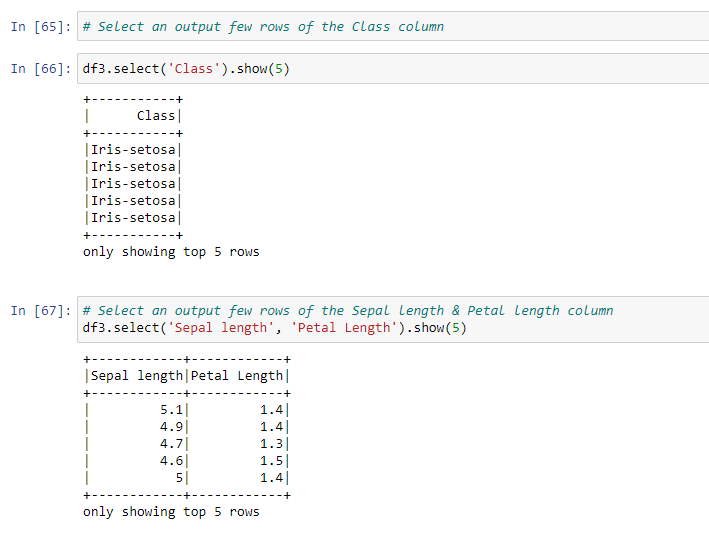
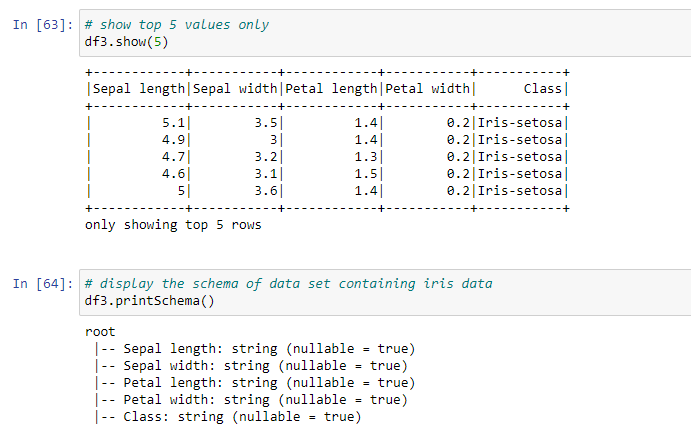
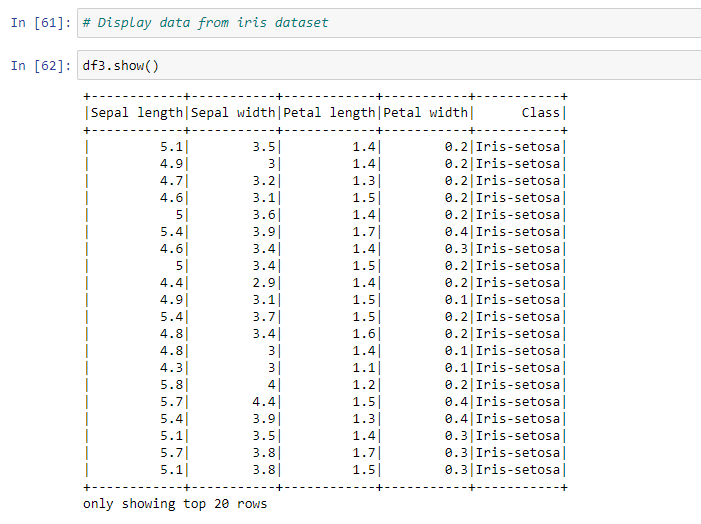
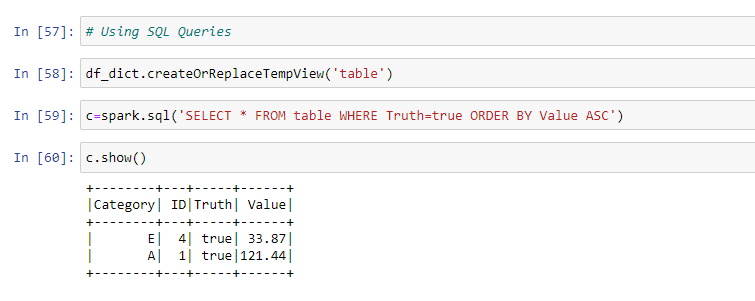
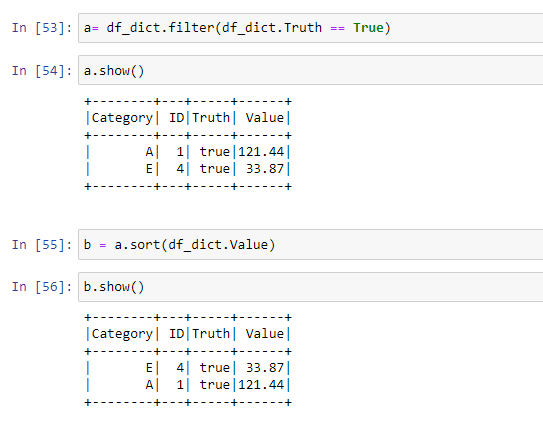
****

****

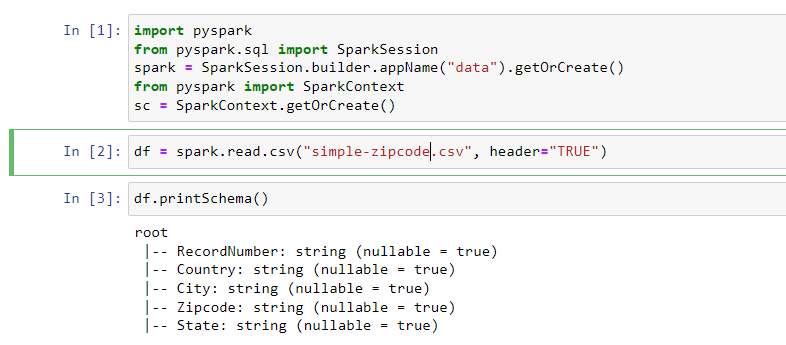
**PRACTICAL 4: Demonstrate use of Spark DataFrames.**

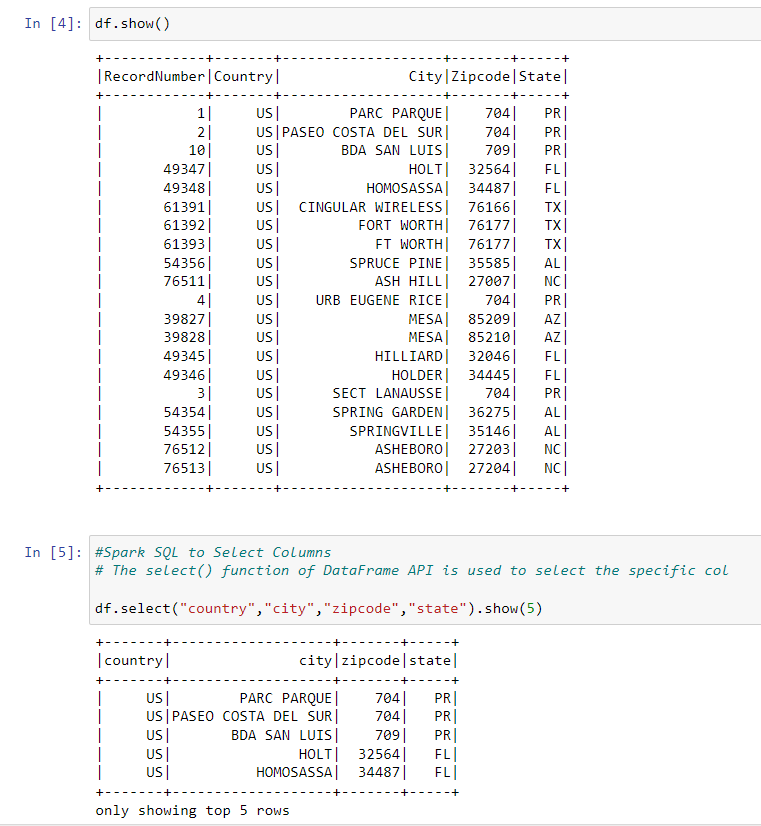
****

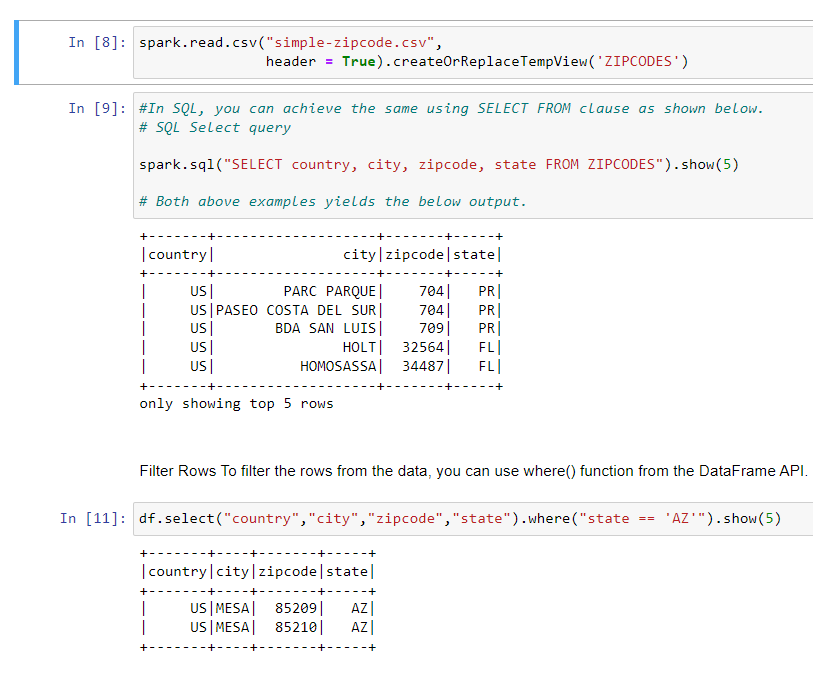
****

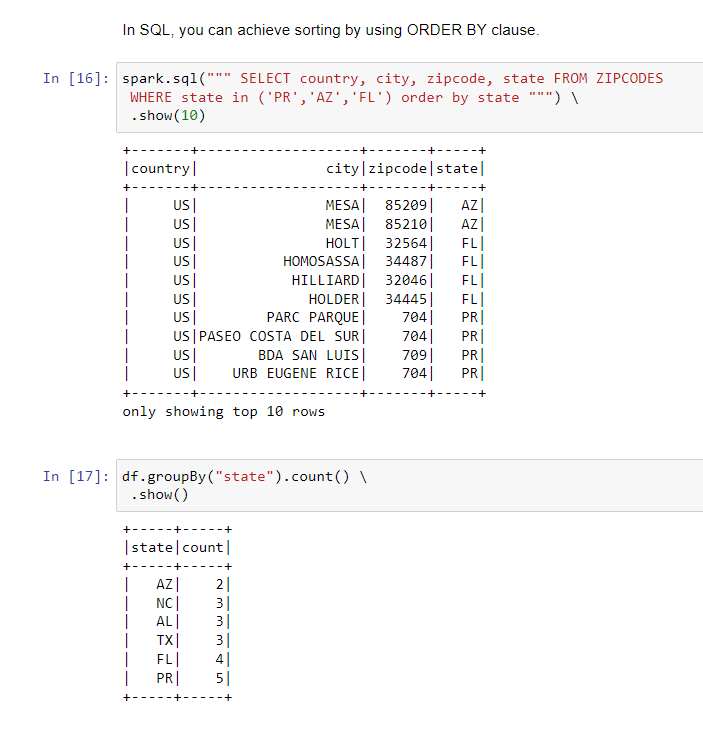
****

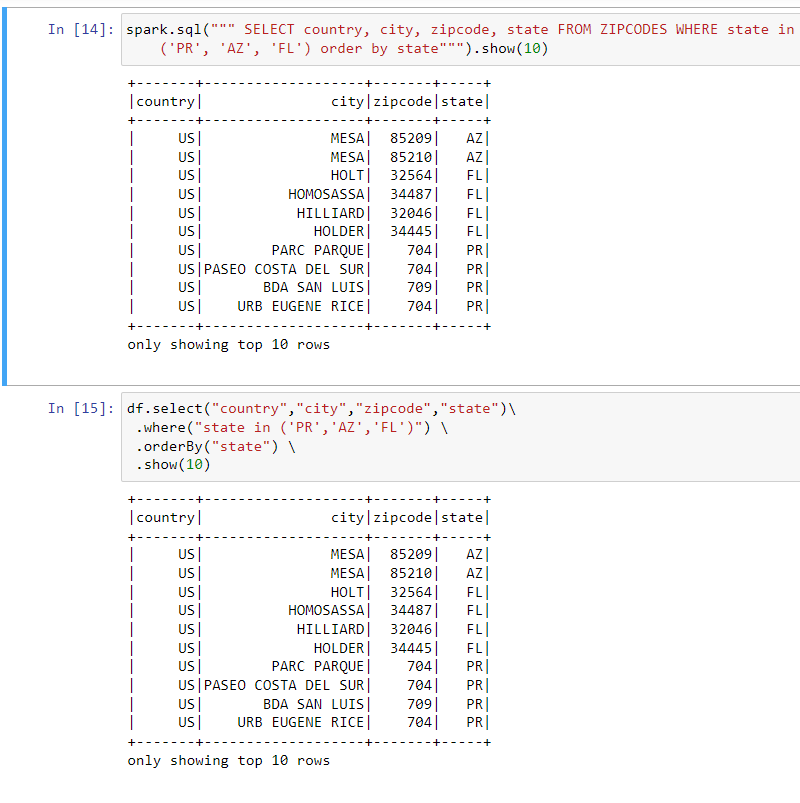
**PRACTICAL 5: Demonstrate use of Spark DataFrames.**

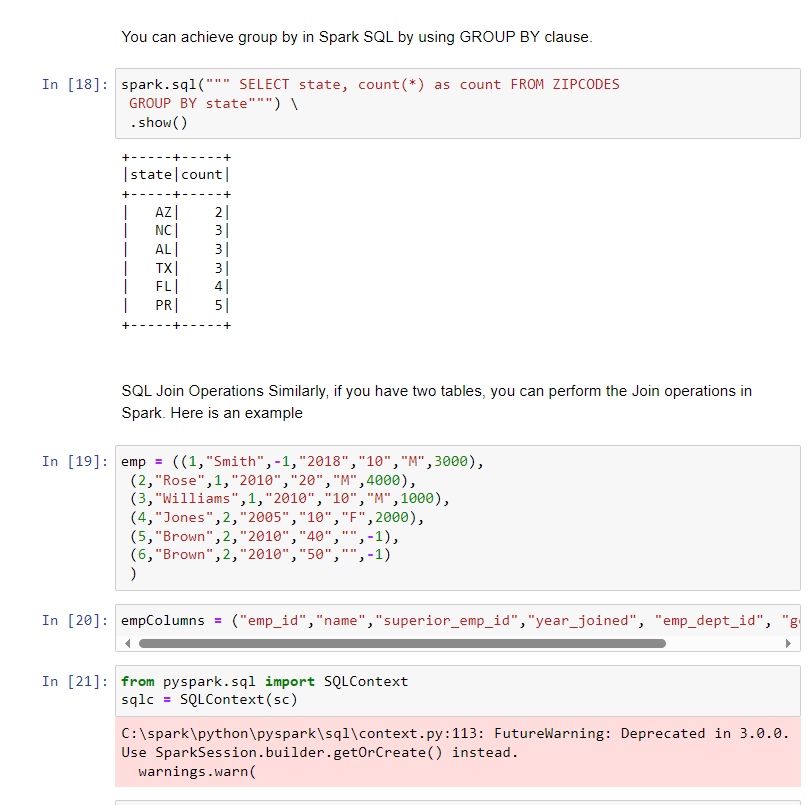
****

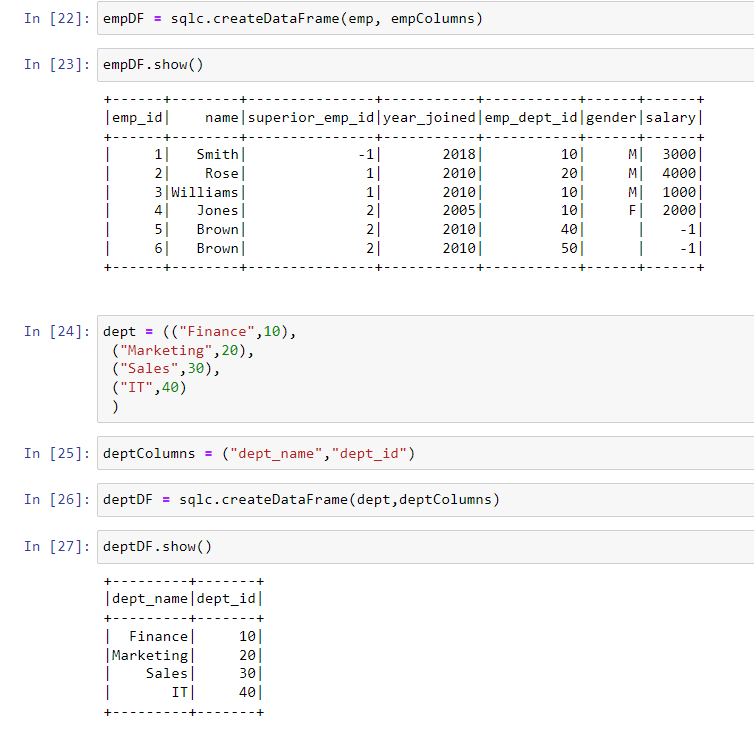
****

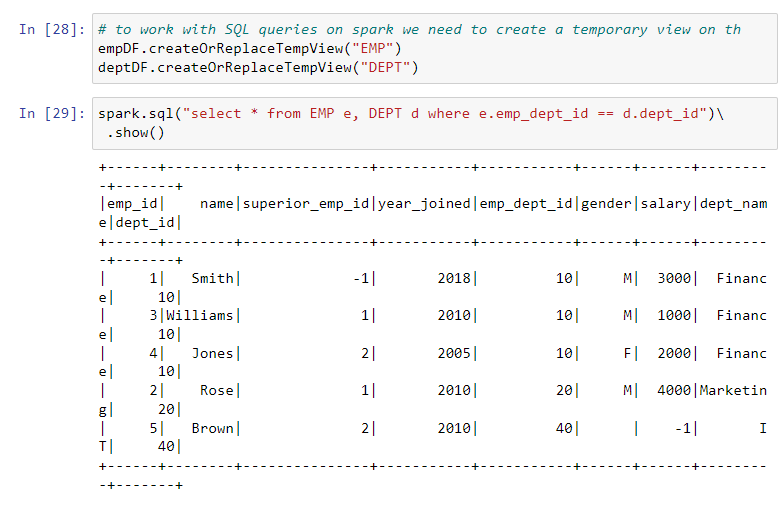
****

****

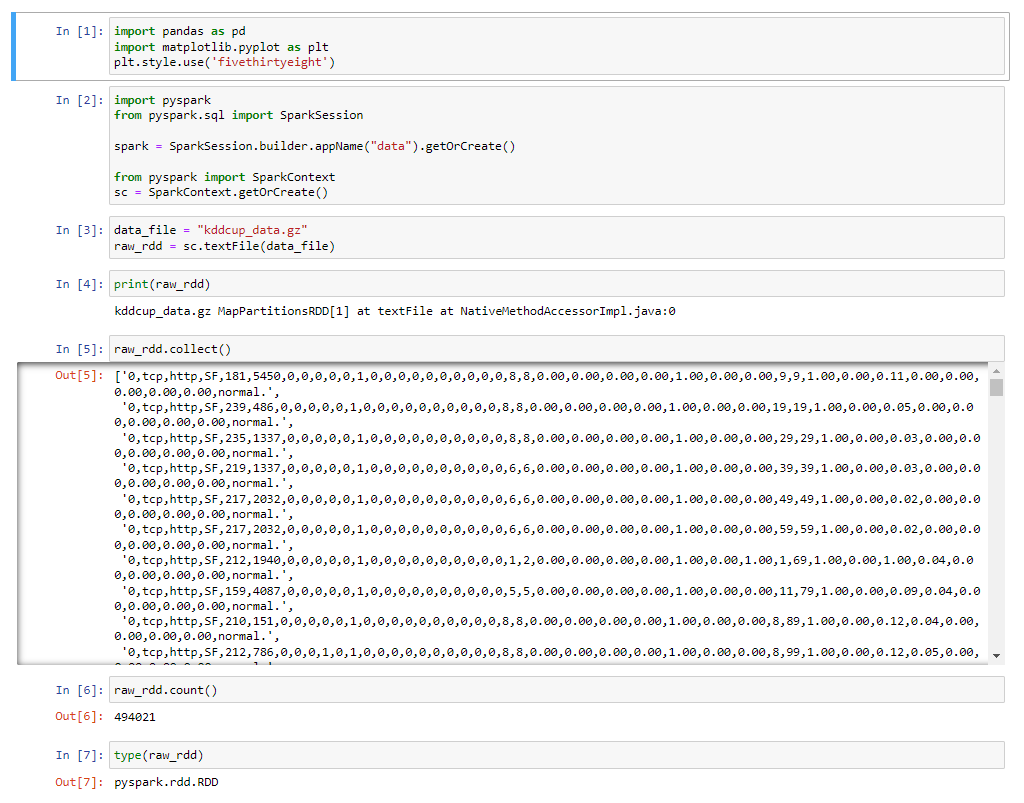
****

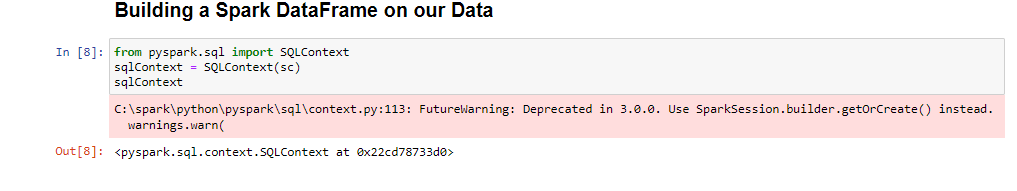
****

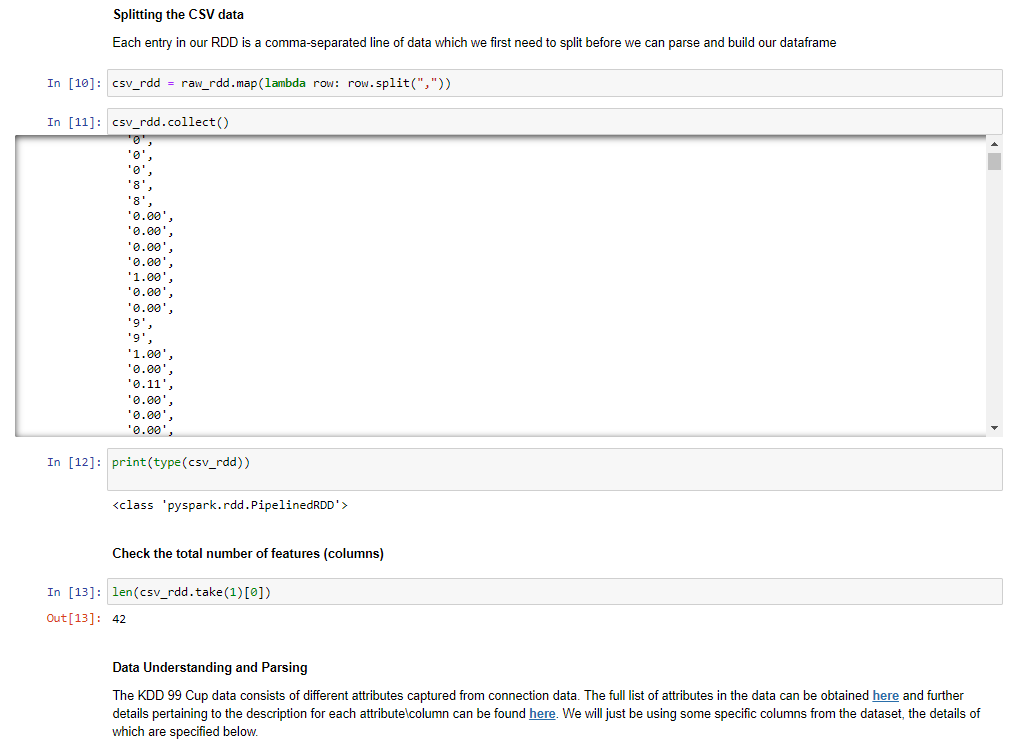
****

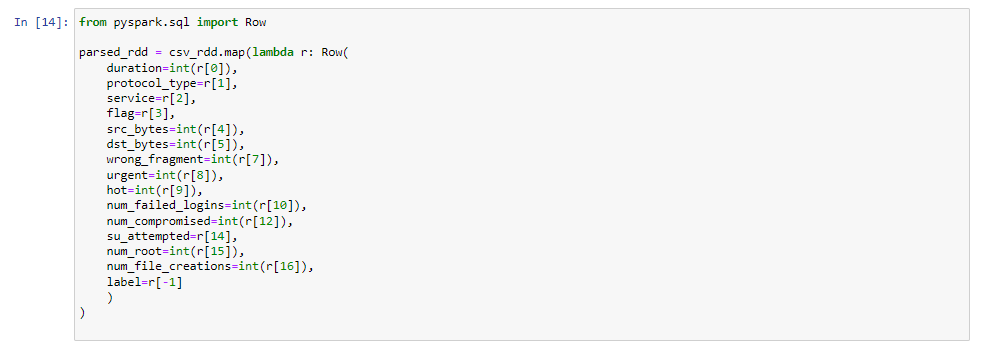
****

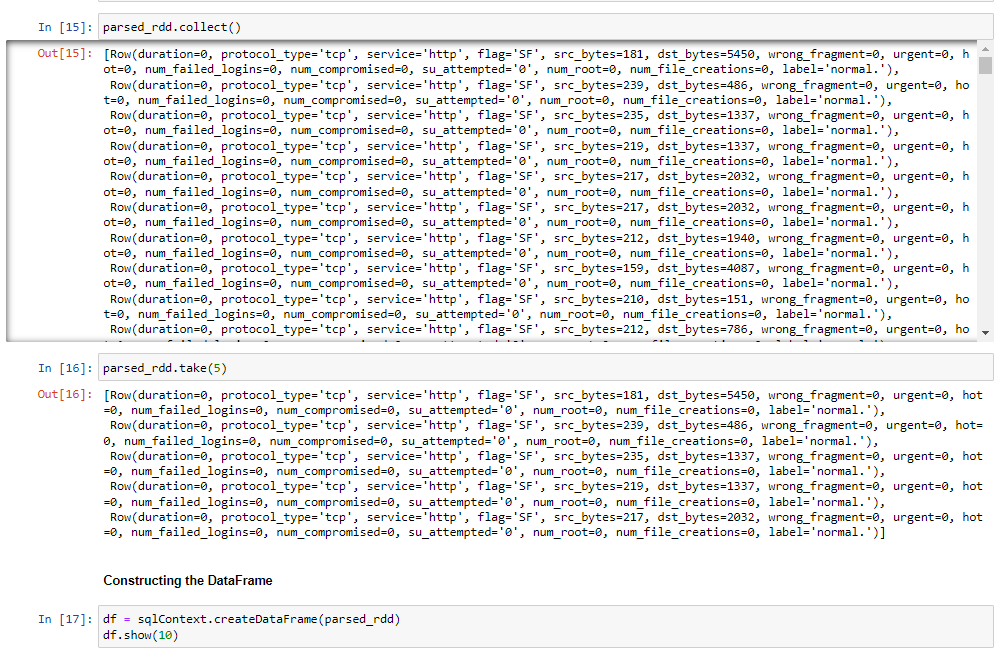
**PRACTICAL 6: Demonstrate use of Spark SQL**

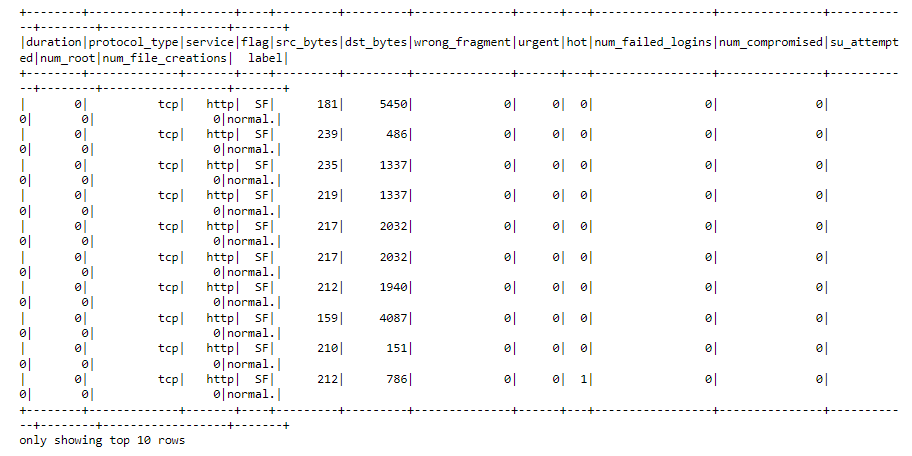
****

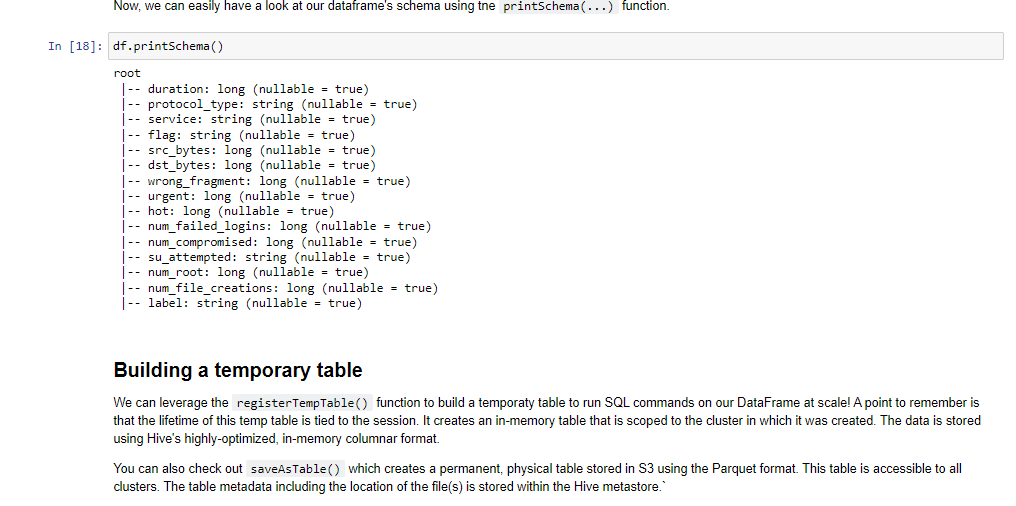
****

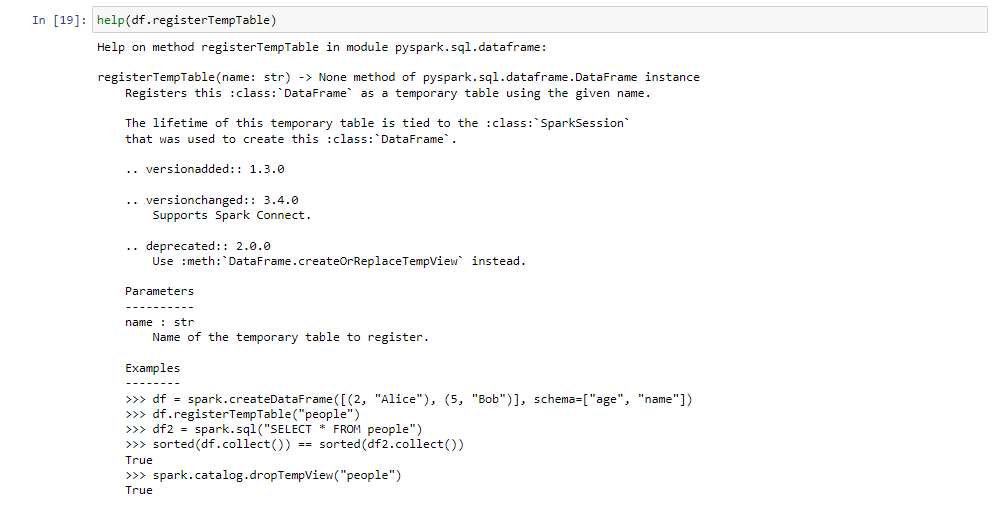
****

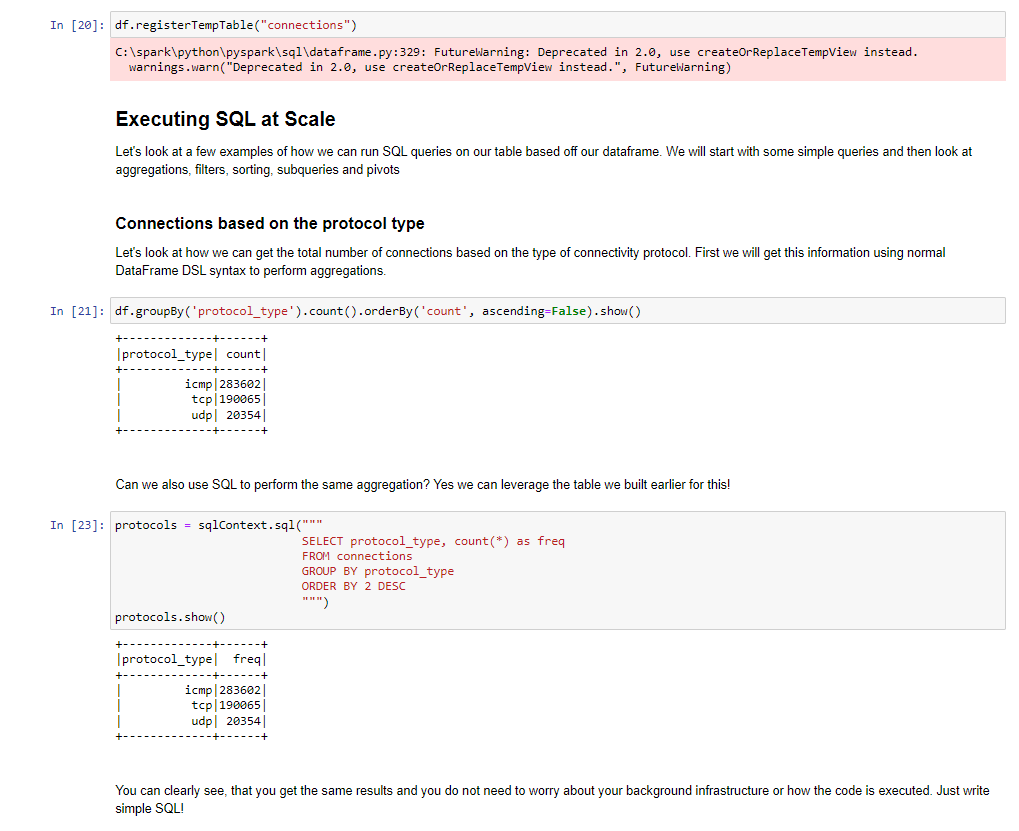
****

****

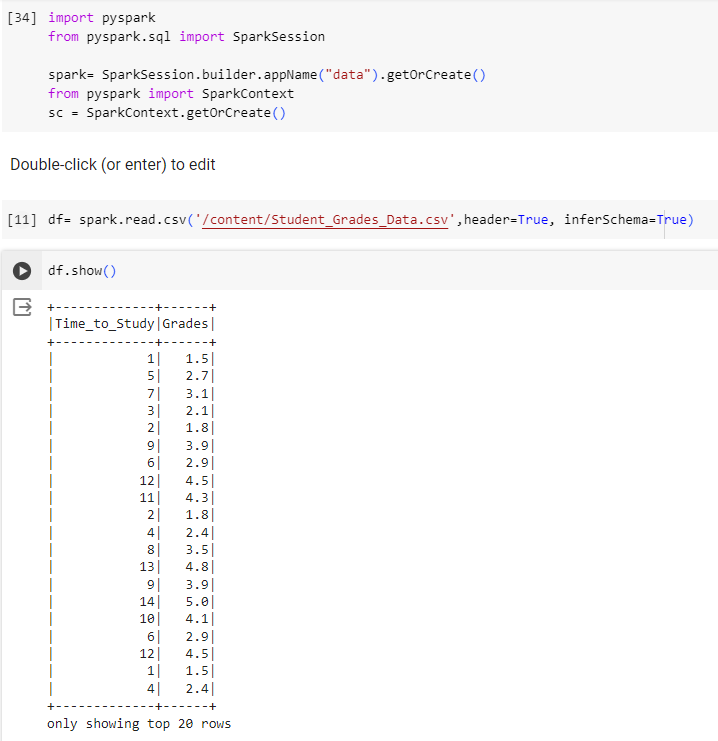
****

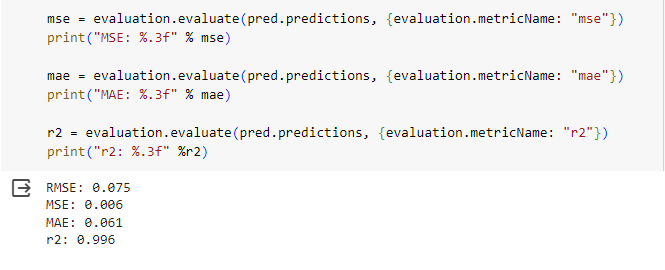
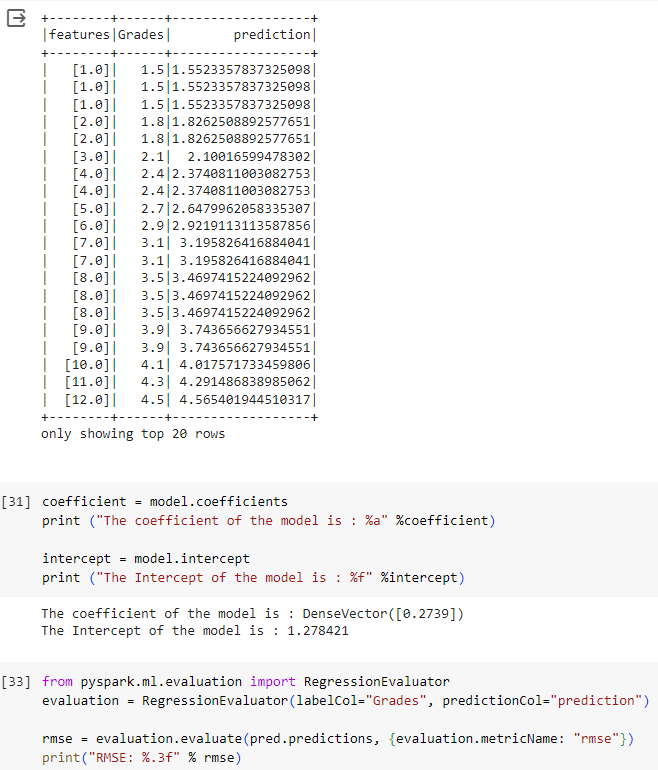
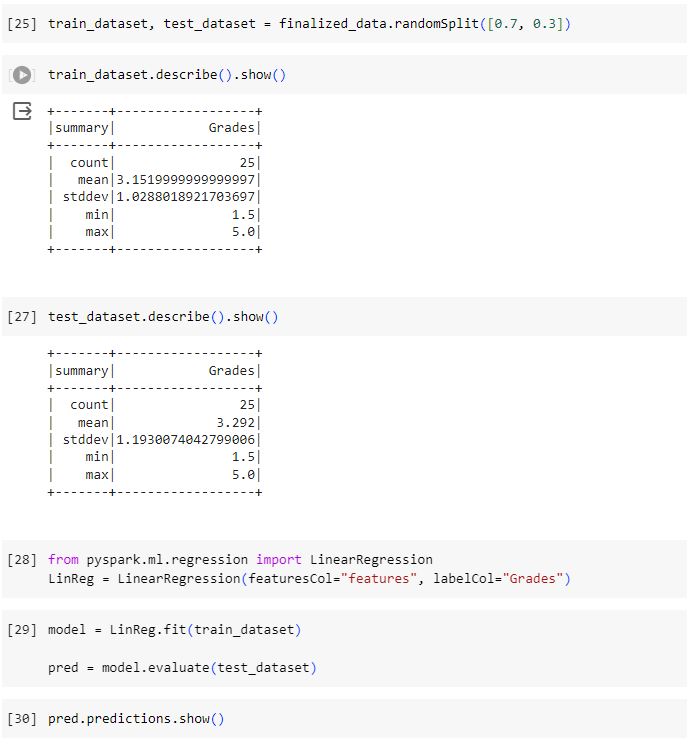
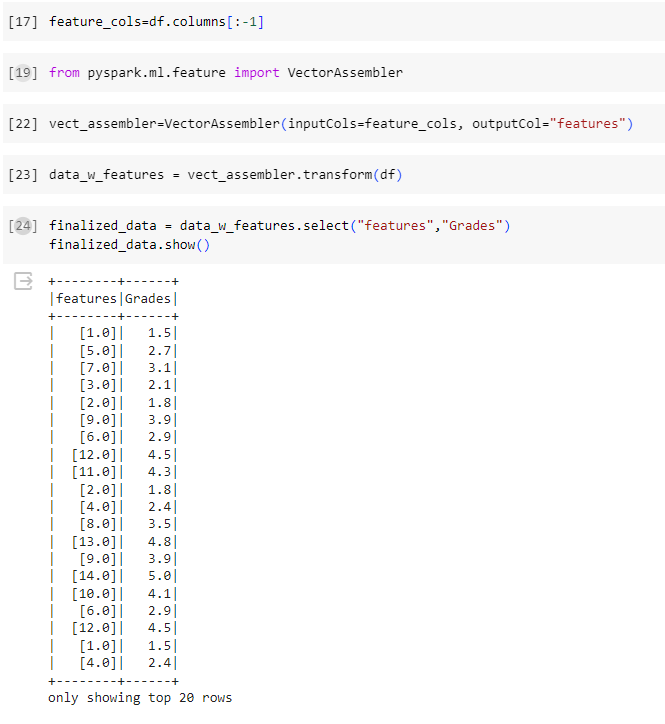
****

****

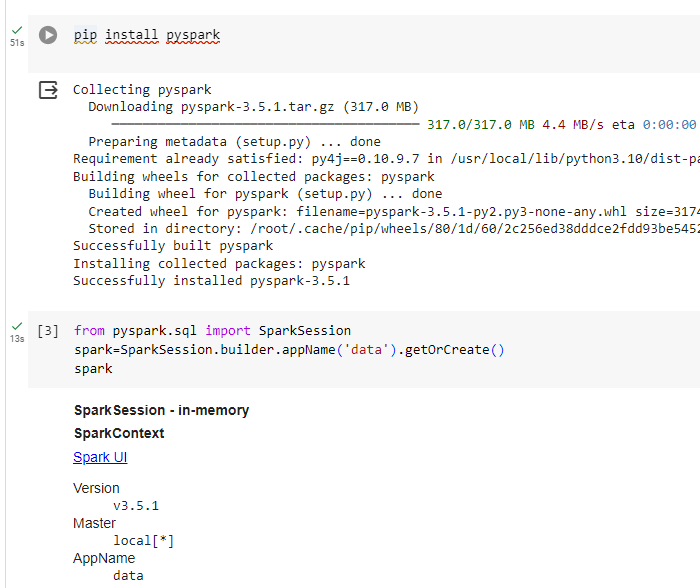
****

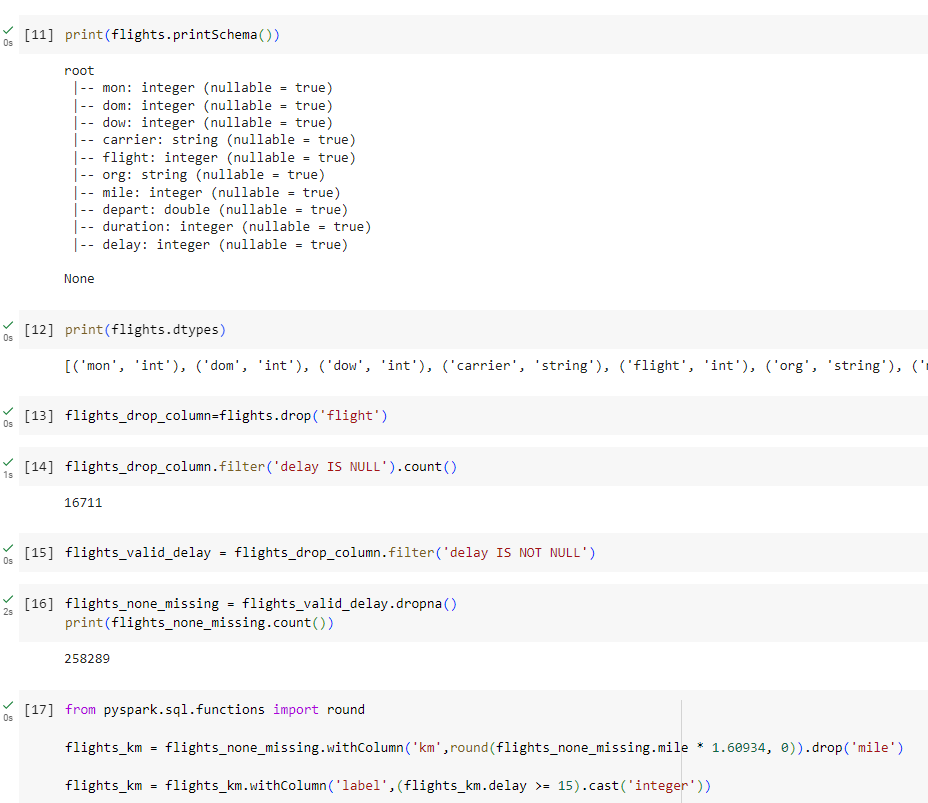
**PRACTICAL 7: Linear Regression using pyspark.**

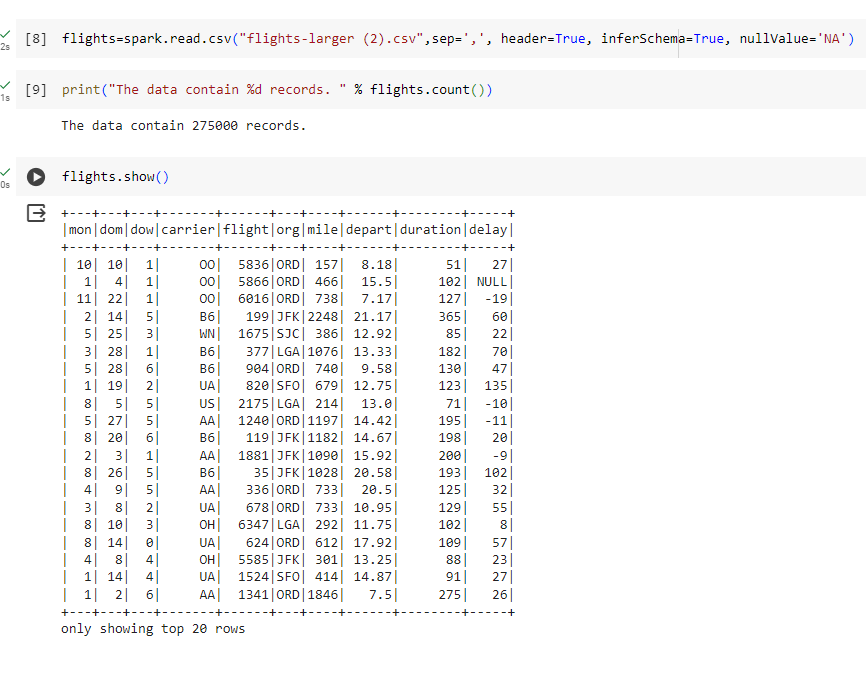
****

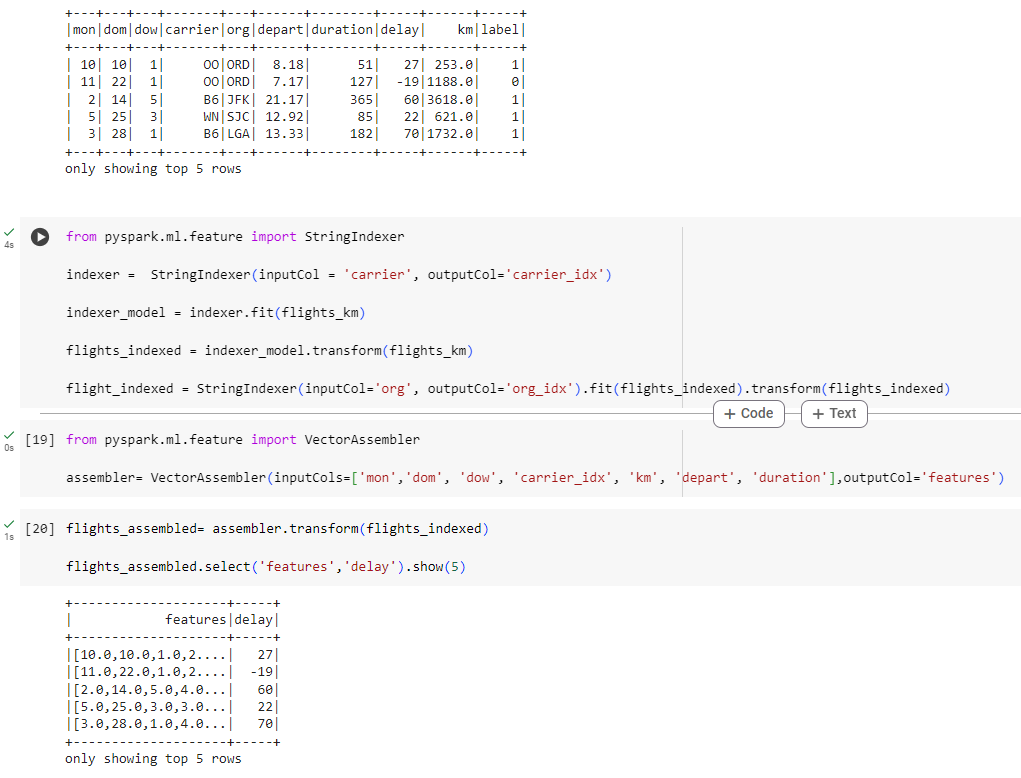
****

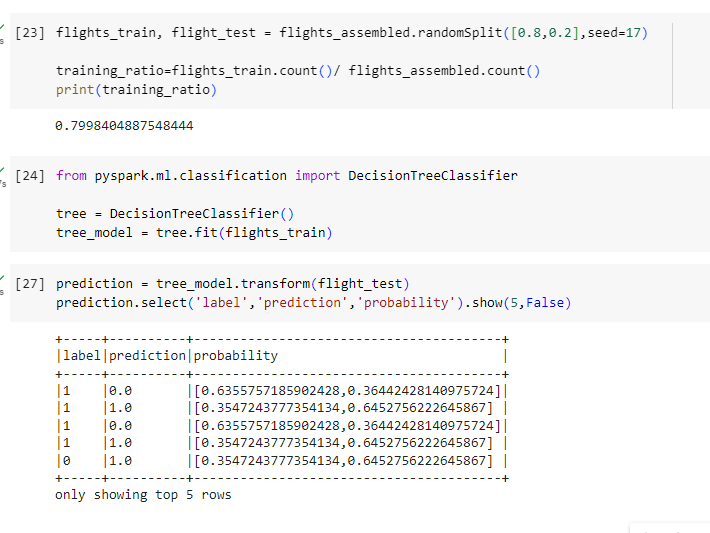
**Practical 8: Classification in Spark ML**

****

****

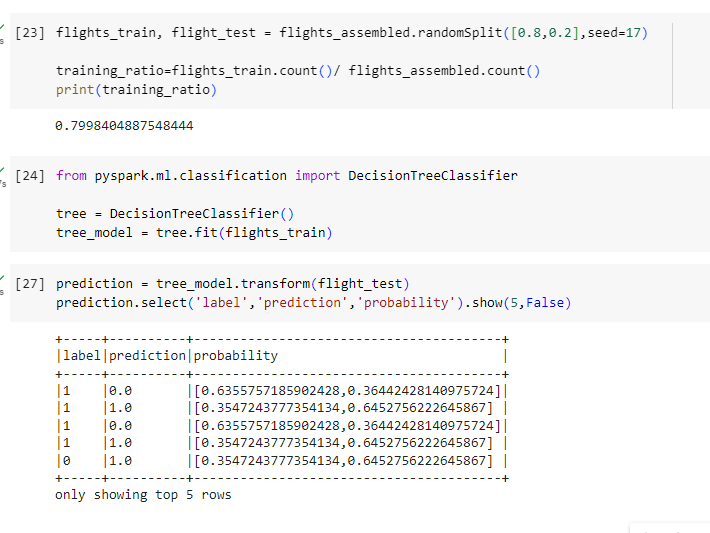
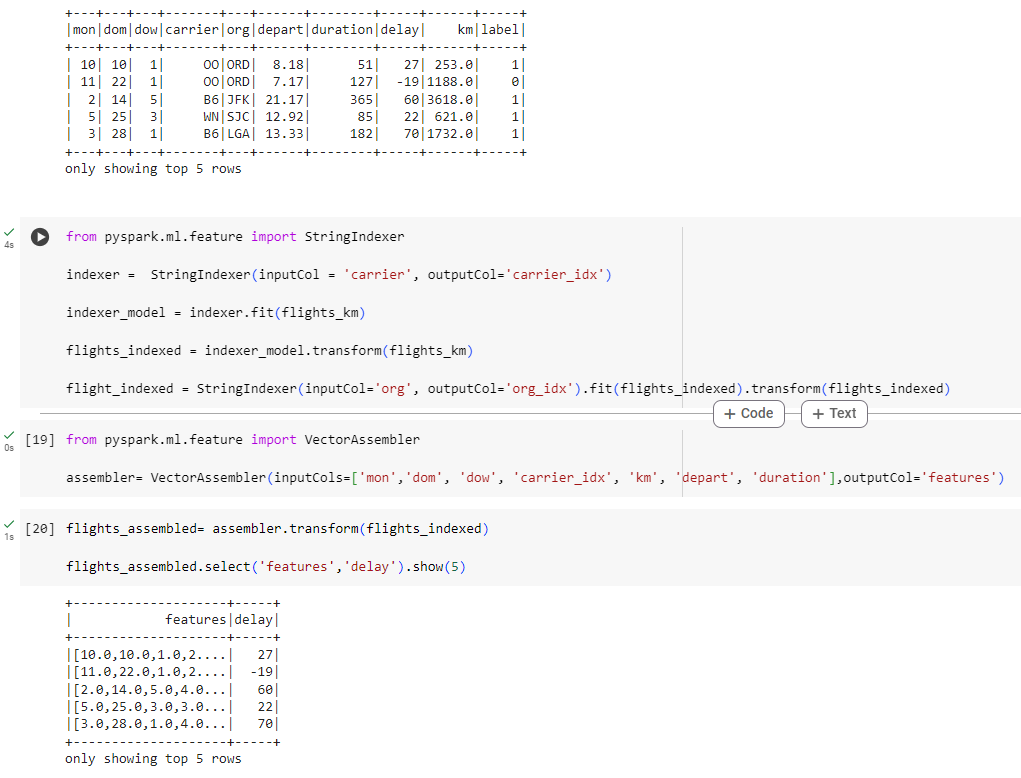
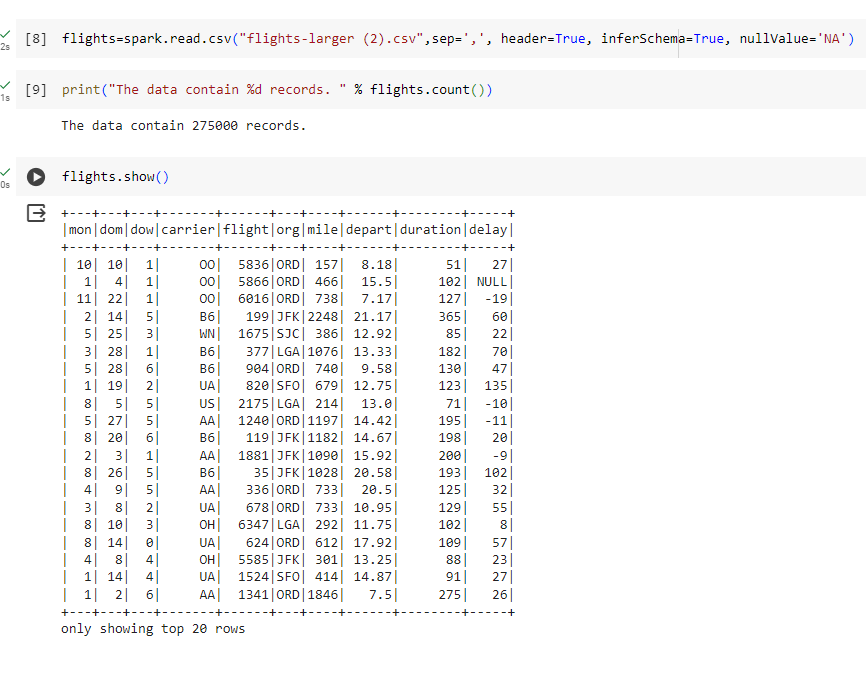
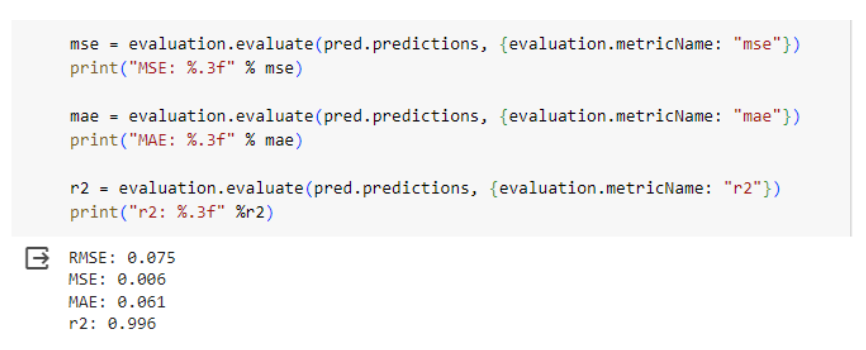
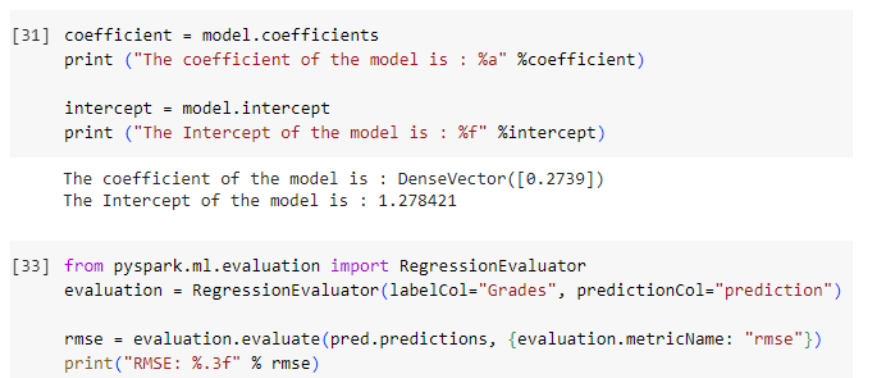
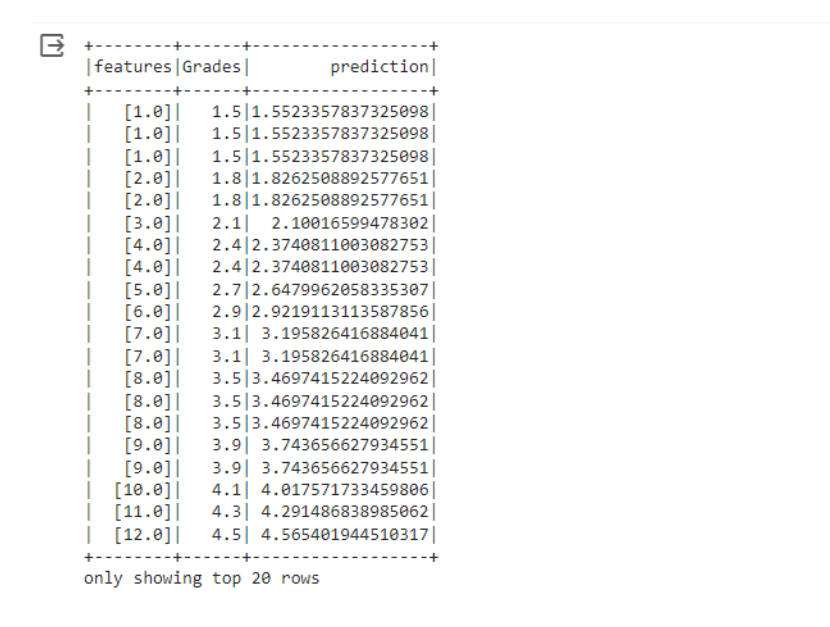
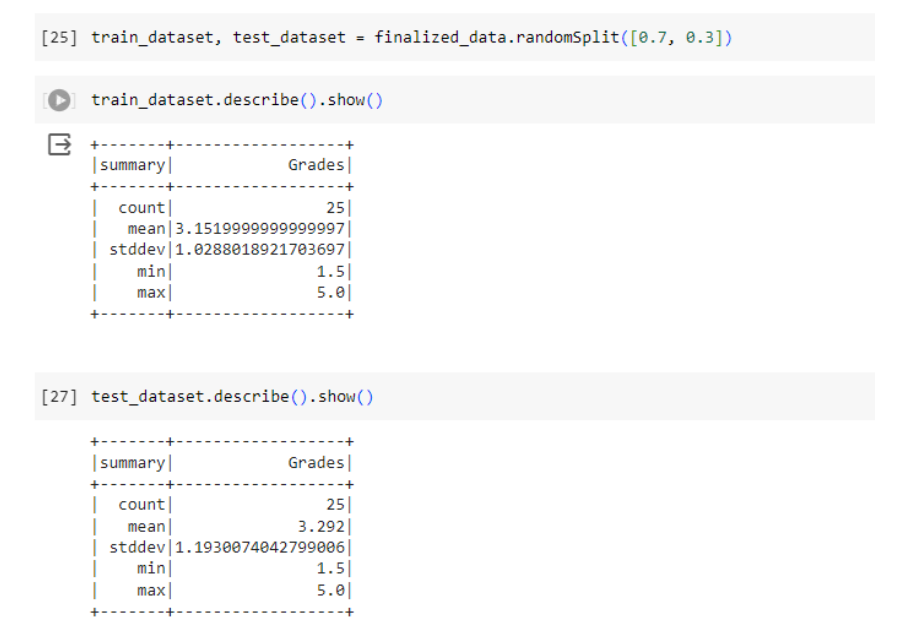
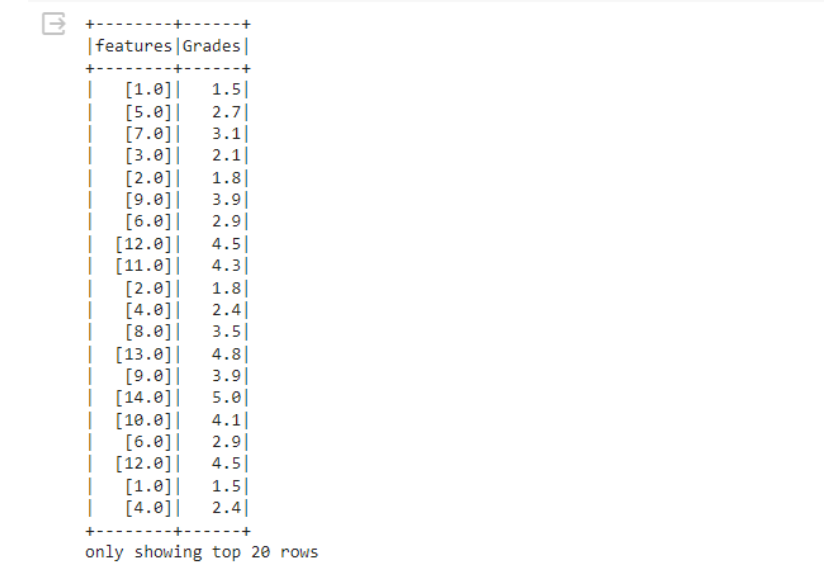
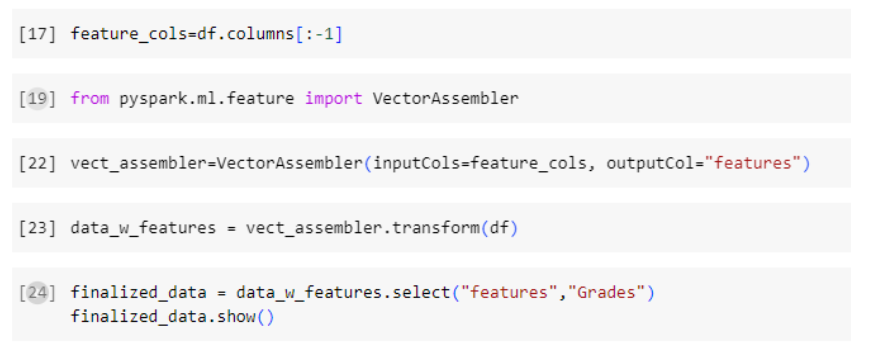
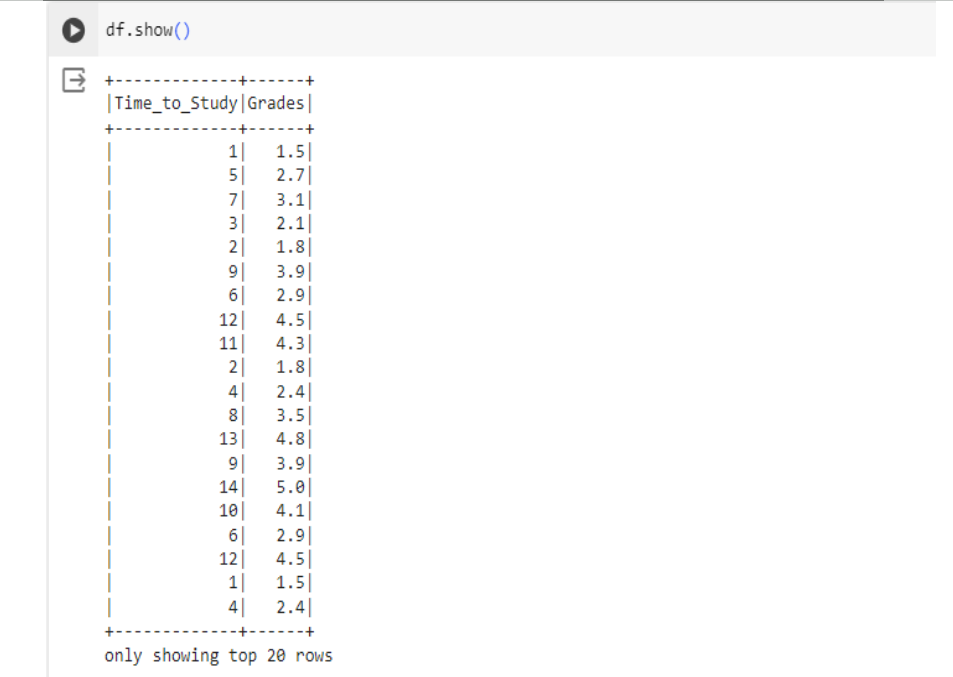
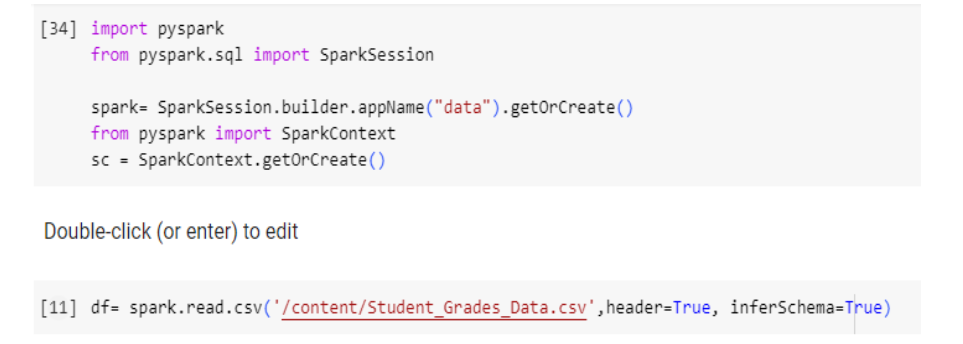
****

****

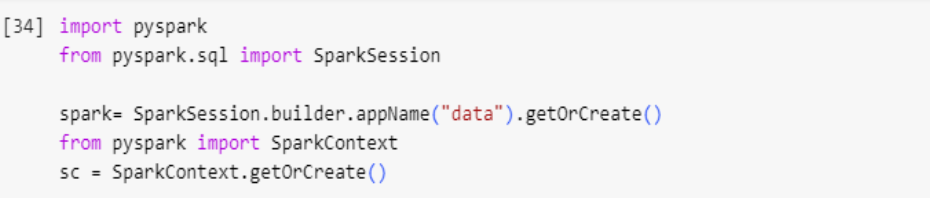
****

****

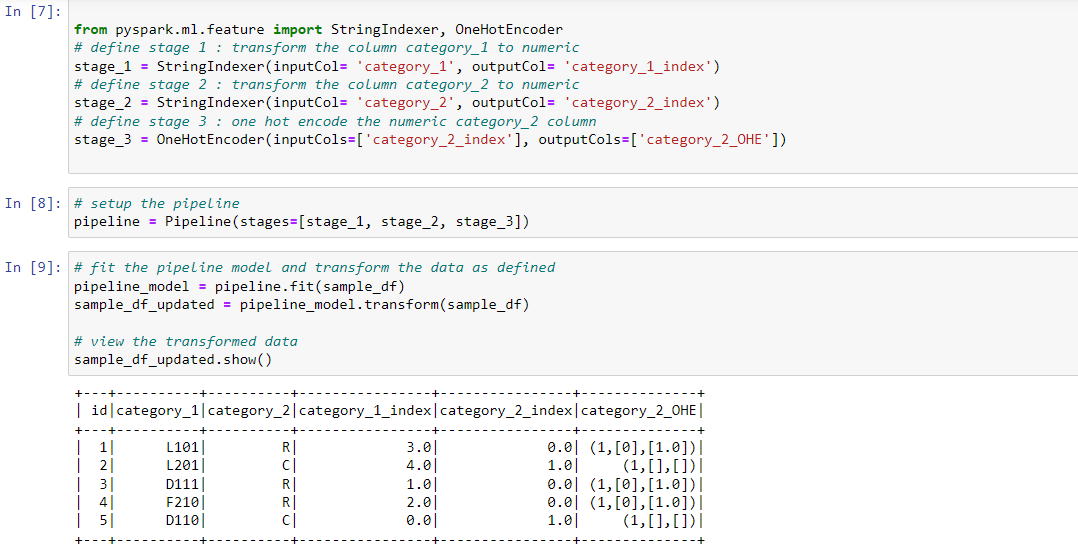
**PRACTICAL NO 9 : Demonstrate use of Evaluation metrics while build machine learning models in Spark ML.**

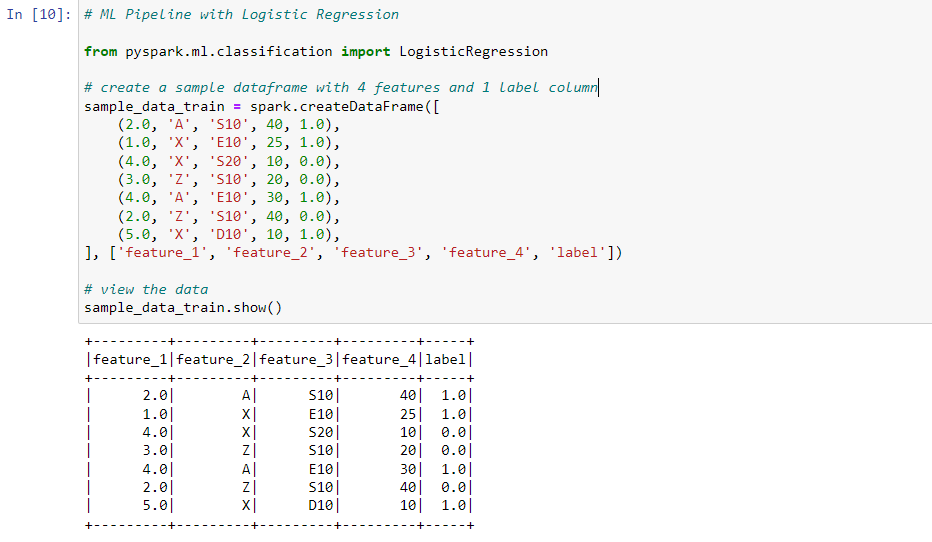
****

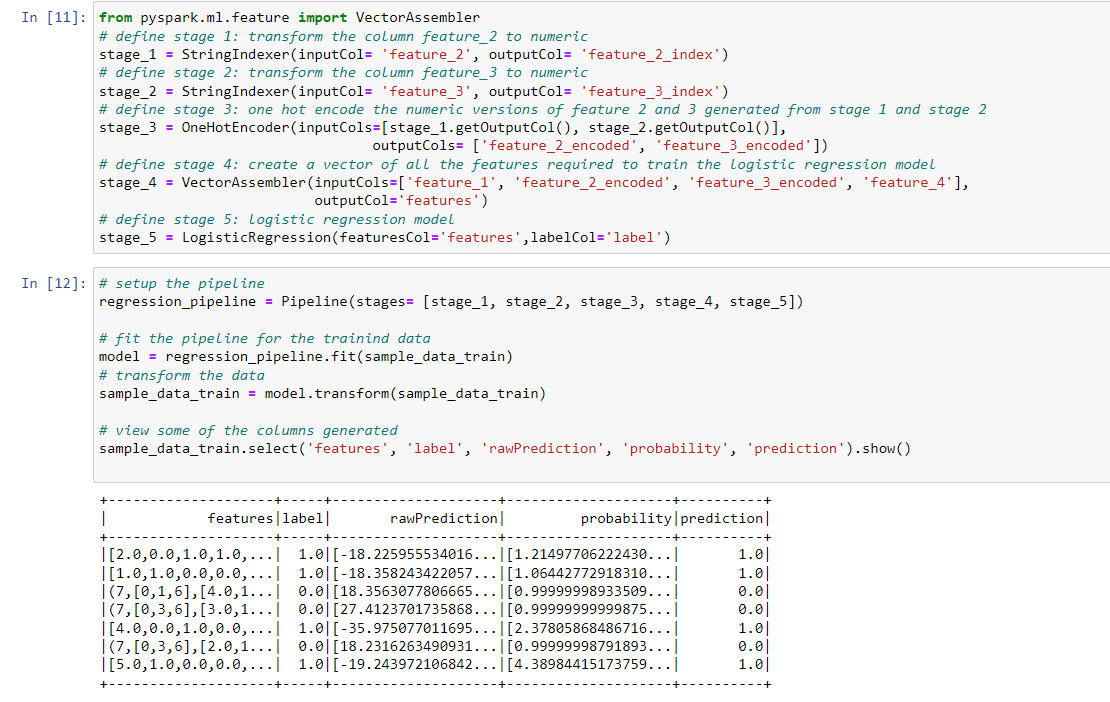
**PRACTICAL NO 10 : Demonstrate use of pipelines while building machine learning models in Spark ML**

****

****

****

****

****