Chethan U Mahindrakar – SEC01 (NUID 002646783)

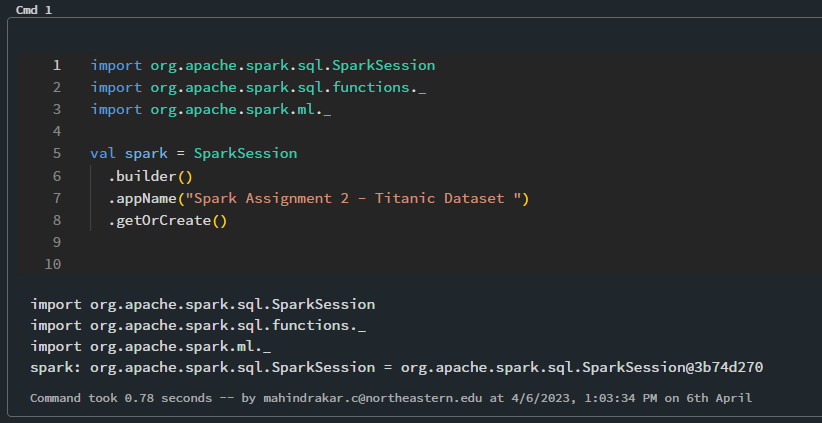
Big Data System Engineering with Scala  
Spring 2023   
Assignment No. Spark 2(Titanic Dataset)



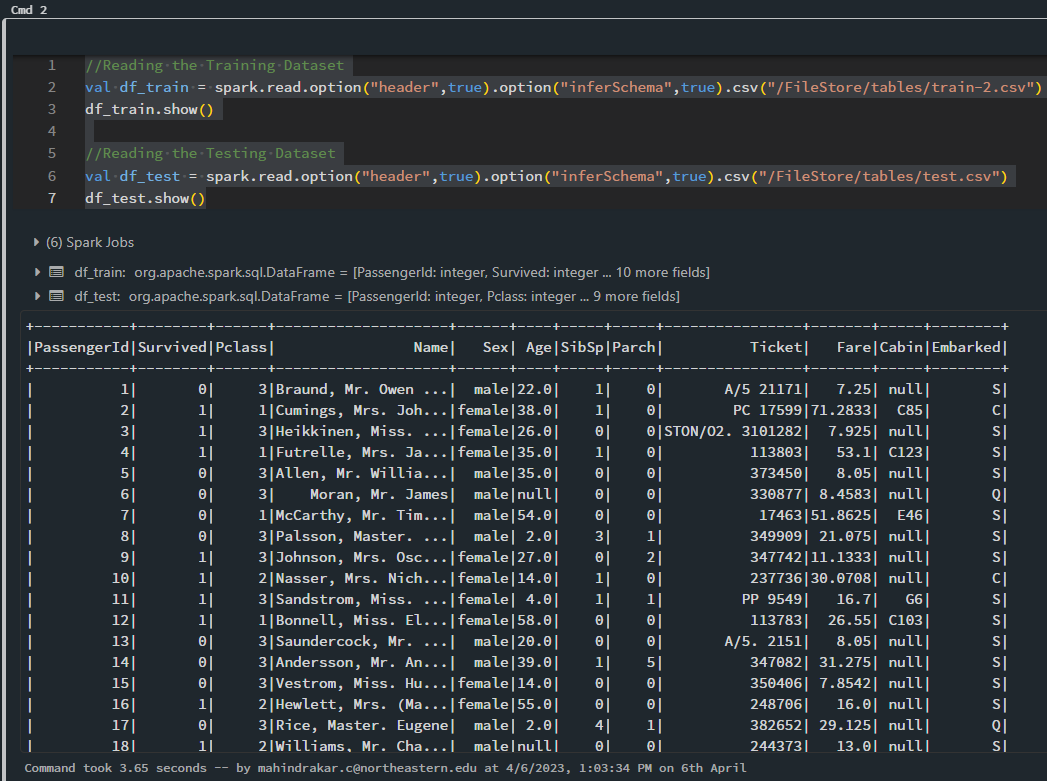
**Code and Analysis:**

(Most of the analysis is included within the comments of the code)

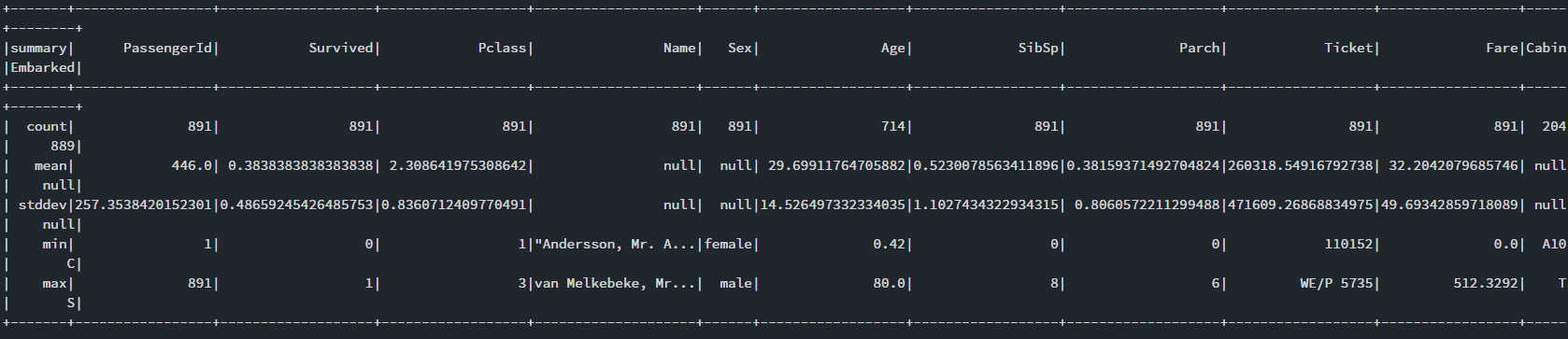
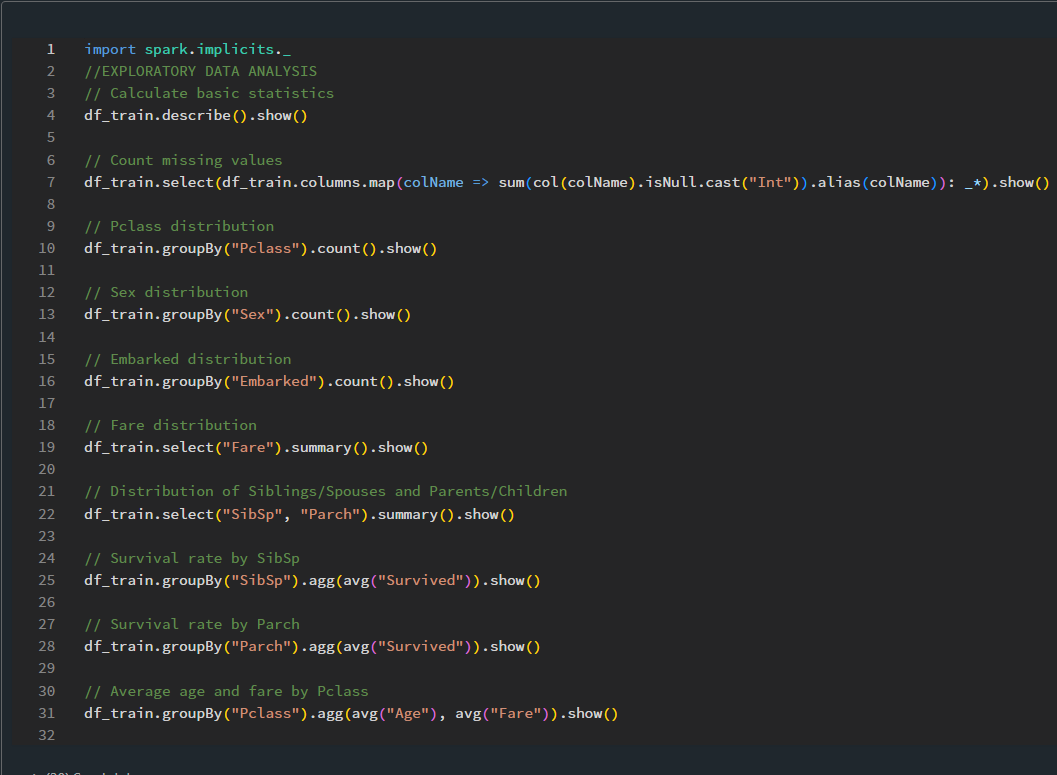
Importing the necessary Libraries and creating the spark session

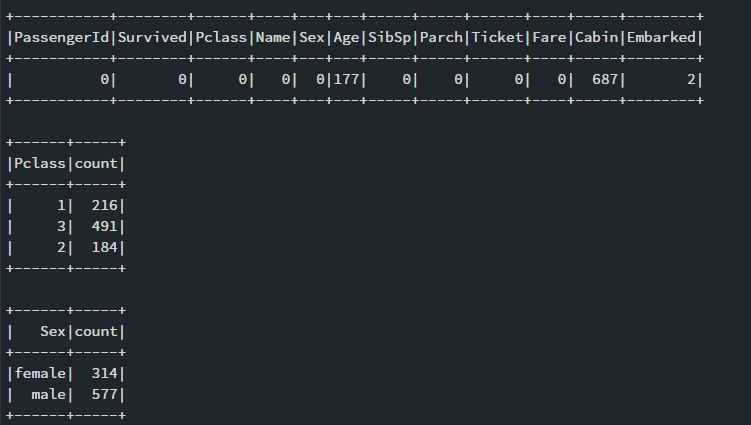


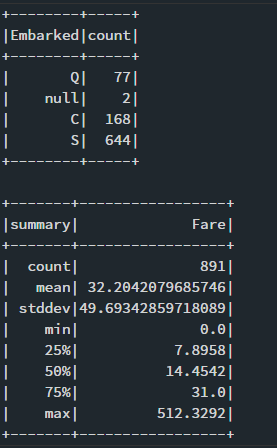
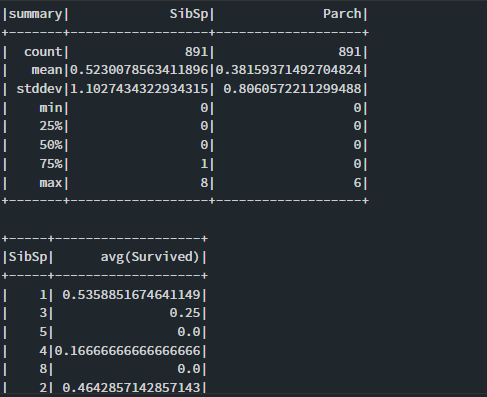
Reading the dataset using dataframes

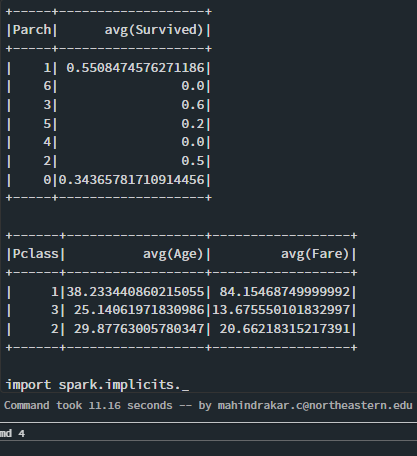


EDA (Analysis performed is included in the comments

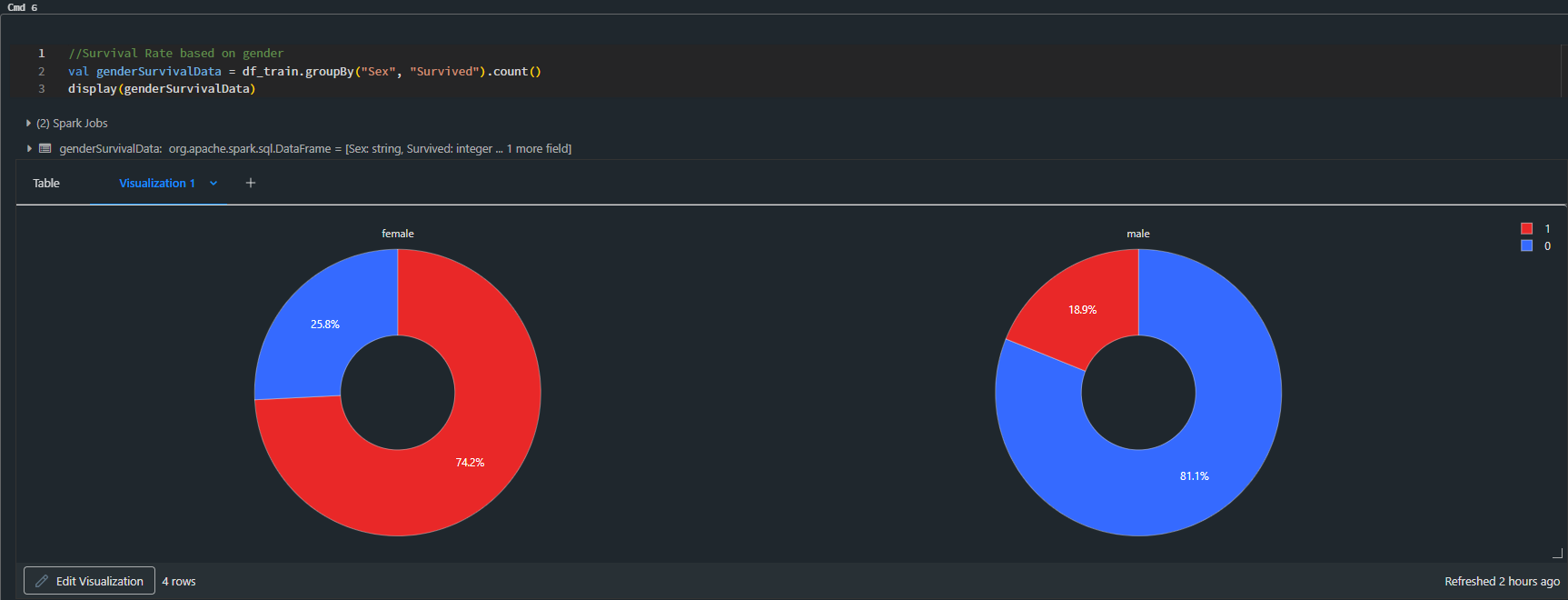
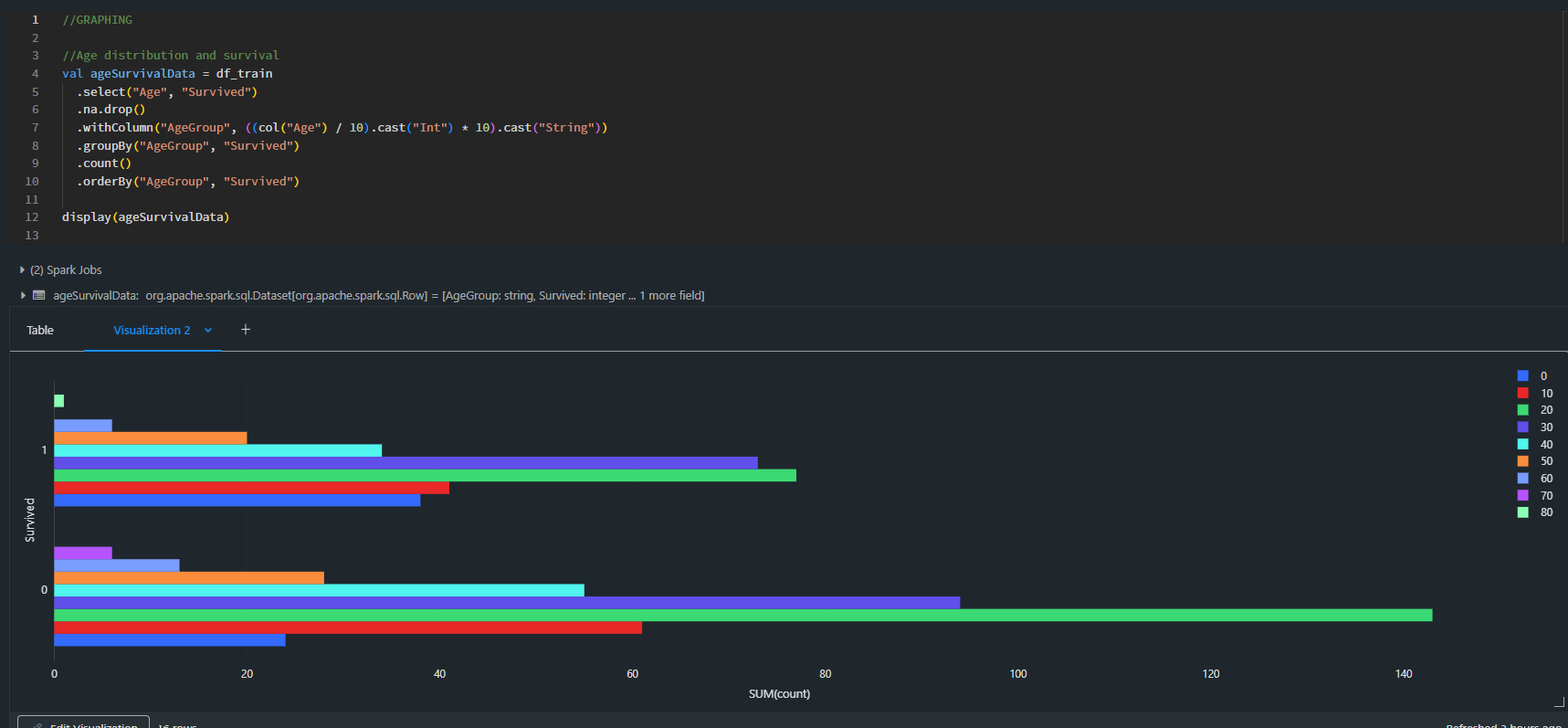




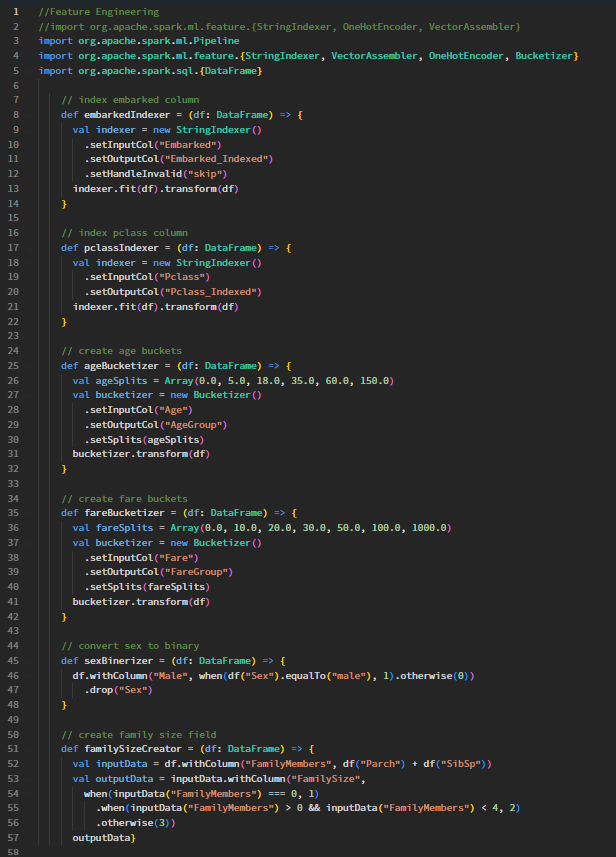
 

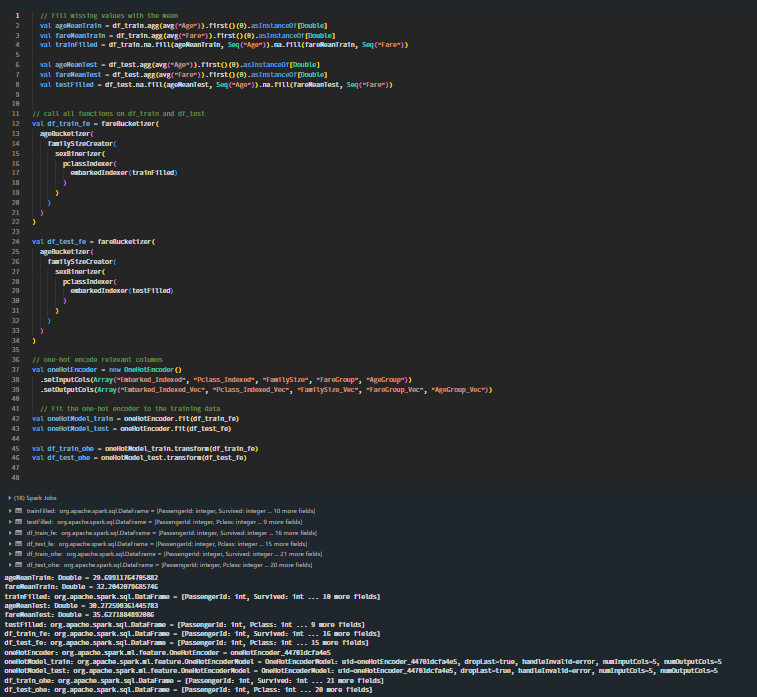




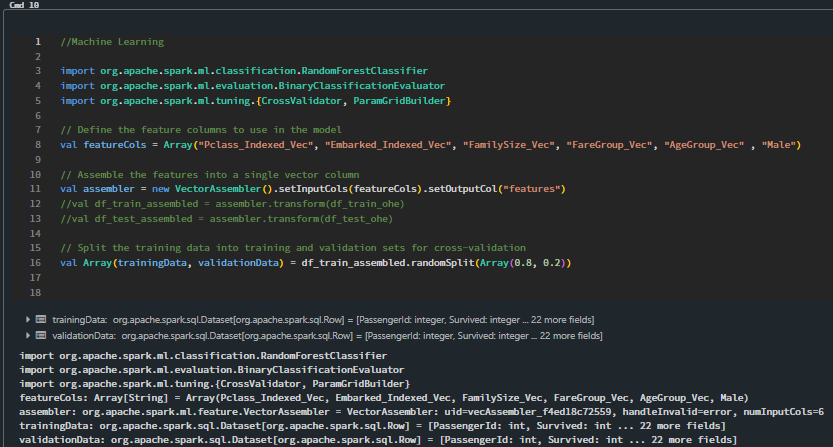


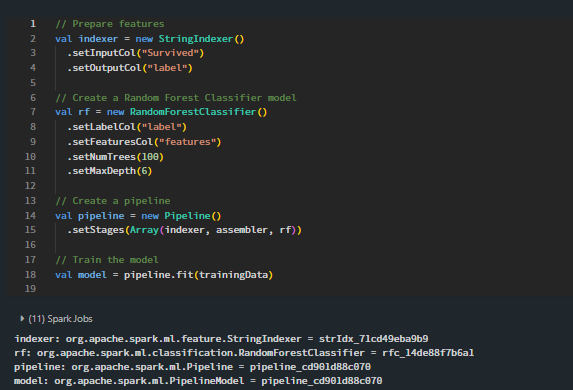
Feature Engineering

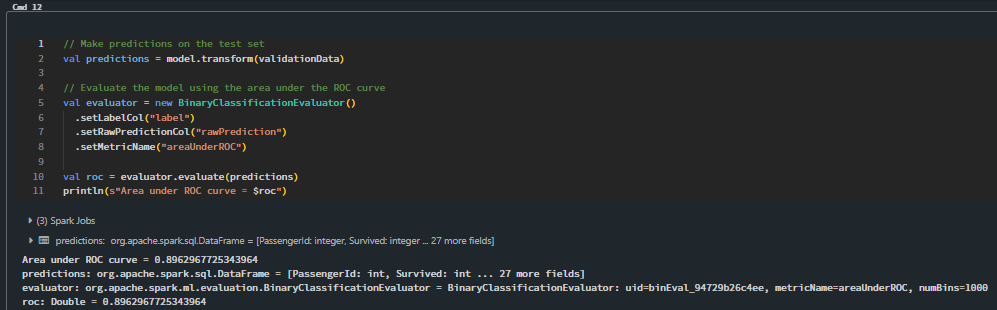




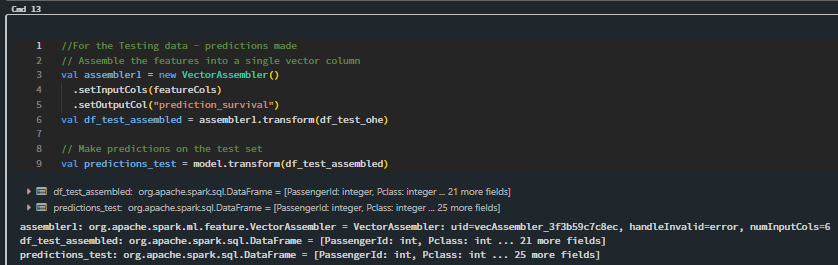
The ML model

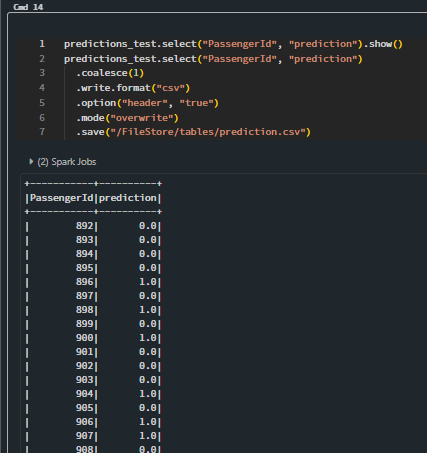






Accuracy of about 89% on the validation set.





Github Link - <https://github.com/chethanmahindrakar/Big-Data-Engineering-Systems---CSYE7200>