Deepika Balasubramanian – SEC01 (NUID 002194564)

Big Data System Engineering with Scala  
Fall 2022   
Assignment No. 7



**-List of Tasks Implemented**

* **Case class MovieRatingAnalyzer containing method to process dataframe**
* **Object MovieDatabaseAnalyzer**

**Code:**

**Code for** Case class MovieRatingAnalyzer:

case class MovieRatingAnalyzer()  
{  
 val *spark*: SparkSession = SparkSession  
 .*builder*()  
 .appName("WordCount")  
 .master("local[\*]")  
 .getOrCreate()  
  
 *spark*.sparkContext.setLogLevel("ERROR") // We want to ignore all of the INFO and WARN messages.  
  
 val *path* = "spark-csv/src/main/resources/movie\_metadata.csv"  
 val *df* = *spark*.read.option("header", "true").csv(*path*)  
  
 def test(sampleDF: DataFrame): DataFrame = {  
 sampleDF.select(*mean*(*df*("imdb\_score")))  
 sampleDF.select(*stddev*(*df*("imdb\_score")))  
 }  
}

**Code for MovieDatabaseAnalyzer:**

object MovieDatabaseAnalyzer extends App {  
 MovieRatingAnalyzer()  
}

**Screenshots:**

**Unit Tests Screenshot:**

**Passed all test cases**

Text

Description automatically generated

Link to repo: <https://github.com/Deepikabala14/CSYE7200>