Test Case #.	Test Scenario	Pre-requisite	Test Steps	Expected Result	Actual Result	Status
1	Ingest dataset in S3	Login Credentials/pem file	Step 1 - SCP to Leeds EMR Step 2 - Copy the project csv files	File should be copied successfully to the S3 bucket	File copied successfully	Pass
2	Read csv using PySpark and Save as Parquet	Access to S3 bucket	NA	Records should be divied and saved as parquet format inside the S3 bucket	Parquet file saved successfully	Pass
3	Read parquet file display df	Access to S3 bucket	NA	PySpark code should read the parquet files and display in the grid	Records displayed on notebook	Pass
4	Test Train Split(For Random Forest)		NA	Train and Test df records counts should be 80:20	Record count for both the sets show 80:20 ratio	Pass
5	Random Forest Classifier		NA	Model runs successfully	Model runs successfully	Pass
6	Confusion Matrix and Model Scores		NA	Model matrix should show TP, TN, FP and FN and accuracy should be ~90% with a reasonable precision	Model matrix shows all the values	Pass
7	Analyze and Visualize data		import pkgs - access data - analze/visualize	Discriptive statistics about the data with visualizations	Discriptive statistics about the data with visualizations	Pass