Yuhan Zhang– SEC01 (NUID 002846159)

Big Data System Engineering with Scala
Spring 2024
Assignment No. 1

- GitHub Repo URL –

https://github.com/LelouchCcCC/CSYE7200/tree/Spring2024/a1

- List of Tasks Implemented

Implemented average ticket fare for every class

Implemented the average survival percentage

Found 2 Rose

Found 23 Jack

Split the age group

- Code

SparkAssignment1.scala

- Unit tests / Results

1.

```
+------+------------------+
|Pclass|      average_fare|
+------+------------------+
|     1| 84.15468749999992|
|     3|13.675550101832997|
|     2| 20.66218315217391|
+------+------------------+
```

2.

```
+------+------------------+
|Pclass|     survival_rate|
+------+------------------+
|     1| 0.6296296296296297|
|     3|0.24236252545824846|
|     2|0.47282608695652173|
+------+------------------+
```

the max survivalrate class is 1

3.

there is 2 person might be Rose

4.

there is 23 person might be Jack

5.

```
+--------+------------------+------------------+
|AgeGroup|      average_fare|     survival_rate|
+--------+------------------+------------------+
|   11-20|29.529531304347838| 0.3826086956521739|
| Unknown|22.158566666666673| 0.2937853107344633|
|   21-30|28.306718695652194| 0.3652173913043478|
|   71-80|25.936680000000003|                0.2|
|   01-10|30.434439062500008|            0.59375|
|   41-50| 41.16318139534884|0.38372093023255816|
|   31-40| 42.49610000000002|0.44516129032258067|
|   51-60| 44.77480238095238|0.40476190476190477|
|   61-70| 45.91078235294117|0.23529411764705882|
+--------+------------------+------------------+
```
The young and old paid less fare than the middle-aged
Age Group with highest survival rate: 01-10