**Assignment 5: Spark RDD**

================================

**Note: Dataset are named as per the problem number.**

1. Given an Input data set with name, age and city

if age > 18 add a new column that’s populated with ‘Y’ else ‘N’

1. Given the input file where columns are stationId, timeOfTheReading, readingType, temperatureRecorded, and few other columns...

We need to find the minimum temperature for each station id.

1. Given the input data, find the top 10 words searched.
2. Find the top 10 customers who spend the maximum amount on shopping.

File Details

============

CustomerId, ProductId, Amount

Map the above column name in the data respectively.

1. Given a data set, write a spark code to answer below questions.

* How many times do movies rated 5 stars?
* How many times do movies rated 4 stars?
* How many times do movies rated 3 stars?
* How many times do movies rated 2 start?
* How many times do movies rated 1 start?

Data Set Columns Respectively

==========================

userid, movieid, rating, timestamp