



## **Session 17**

### **Assignment 2 Questions**

**ACADGILD**

# *Session 17: Assignment 1*

## **Table of Contents**

- [1. Introduction](#)
- [2. Problem Statement](#)
- [3. Output](#)

## 1. Introduction

This assignment will help you to consolidate the concepts learnt in the session.

## 2. Problem Statement

Given a dataset of college students as a text file (name, subject, grade, marks) :

### Dataset

#### **Problem Statement 1:**

1. Read the text file, and create a tupled rdd.
2. Find the count of total number of rows present.
3. What is the distinct number of subjects present in the entire school
4. What is the count of the number of students in the school, whose name is Mathew and marks is 55

#### **Problem Statement 2:**

1. What is the count of students per grade in the school?
2. Find the average of each student (Note - Mathew is grade-1, is different from Mathew in some other grade!)
3. What is the average score of students in each subject across all grades?
4. What is the average score of students in each subject per grade?
5. For all students in grade-2, how many have average score greater than 50?

#### **Problem Statement 3:**

Are there any students in the college that satisfy the below criteria :

1. Average score per student\_name across all grades is same as average score per student\_name per grade

Hint - Use Intersection Property.

## 3. Output

Code and Output screenshot(s).

