### Dataset Fundamental Analysis Activity

### Data Science Capstone Project

|  |  |
| --- | --- |
|  |  |

**Instructions**

1. Read the context of the activity and its data dictionary
2. Briefly explore (10-15 minutes) the dataset available on MyCourses (Week 2 - Dataset)
3. You can look for information on the internet but should not communicate with your classmates. If you don’t know an answer or how to perform the analysis skip the question.
4. Answer the questions published on MyCourses (Week 2 -Dataset fundamental analysis activity).
5. Always use 1 as the random seed value.
6. Submit your Tableau, Alteryx or Python workflow(s) on MyCourses (Week 2 - Workflows).

**Context**

You have been hired as a data analysis expert at a small high school. The principal has recently become interested in the capabilities of data analysis to understand more about the students, and has launched a project to design a new strategy to improve the student success rates of the school. This new strategy will be based on the analysis you perform and the insights you provide.

She talked to the information systems department that handles all the data, and they mentioned that there is a small data dictionary and a dataset available.

One main interest is to be able to predict in an early fashion what grade a student will get at the end of the course, based not only on their academic grades.

**Data dictionary**

|  |  |  |
| --- | --- | --- |
| **Feature** | **Description** | **Type of data** |
| student\_id | Student identifier | (numeric: from 0 to 649) |
| sex | student’s sex | (binary: female or male) |
| age | student’s age | (numeric: from 15 to 22) |
| failures | number of past class failures | (numeric: n if 1 ≤ n < 3, else 4) |
| Walc | weekend alcohol consumption | (numeric: from 1 – very low to 5 – very high) |
| Dalc | workday alcohol consumption | (numeric: from 1 – very low to 5 – very high) |
| G1 | first period grade | (numeric: from 0 to 20) |
| G2 | second period grade | (numeric: from 0 to 20) |
| G3 | final grade | (numeric: from 0 to 20) |