Instructions

- This homework assignment is worth 41 points.
- Please submit a **.ipynb** file to Blackboard.
- **Please strive for clarity and organization.**
- **Due Date: September 15, 2023 by 11:59 pm.**

Exercise 1

(3 points) Which of the following is not a part of data analytics process?

(a) business problem

(b) data preparation

(c) model evaluation

(d) software development

(e) model deployment

Exercise 2

(4 points) What is predictive modeling?

Exercise 3

(4 points) What is the difference between supervised and unsupervised learning?

Exercise 4

(4 points) What is the difference between diagnostic and predictive analytics?

Exercise 5

(4 points) Ice cream sales and high crime rates correlate very well during the summer season. One could argue that ice cream sales could be used in a model that predicts crime rate in the summer. This is an example of bad analytics. Explain why.

Exercise 6

The following table list a dataset from the credit scoring domain.

ID	Occupation	Age	Loan-Salary Ratio
1	industrial	39	3.40
2	industrial	22	4.02
3	professional	30	2.70
4	professional	27	3.32
5	professional	40	2.04
6	professional	50	6.95
7	industrial	27	3.00
8	industrial	33	2.60
9	industrial	30	4.50
10	professional	45	2.78

```
Model 1
-----------
if Loan-Salary Ratio > 3.00 then Outcome = default
else Outcome = repay

Model 2
-----------
if Age = 50 then Outcome = default
else if Age = 39 then Outcome = default
else if Age = 30 and Occupation = industrial then Outcome = default
else if Age = 27 and Occupation = professional then Outcome = default
else Outcome = repay
```

(a) (5 points) Create the Outcome variable for the ten considered IDs using Model 1.

(b) (5 points) Create the Outcome variable for the ten considered IDs using Model 2.

Exercise 7

Consider the `StudentsPerformance.csv` data file. This data file contains information on students performance on different subjects. **In Python**, answer the following:

(a) (4 points) Using the pandas library, read the `StudentsPerformance.csv` file and create a data-frame called `students`.

(b) (4 points) Using the appropriate commands, report the number of female and male students.

(c) (4 points) Using the appropriate commands, report the average math score based on gender.