St Lawrence College, School of Business

Graduate Certificate Program, Business Analytics

**ADMN5003: Data Acquisition, Analysis and Modelling Techniques (Fall 2022)**

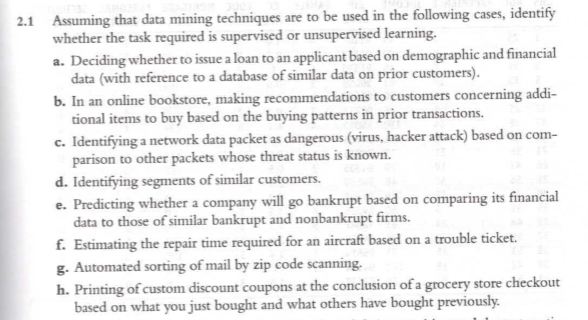
Assignment 1 – Due on Friday, September 23rd, 2022 @ 2:00pm

12 Marks

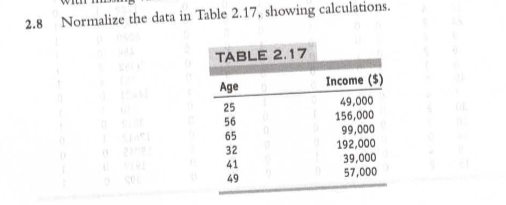
* Open book assignment.
* You need to submit your assignment to Blackboard in a group of 2 students.
* Submit a Word file with the file name in this format: Assign1\_First name1\_First name2.
* Late submission penalty is 10% per day.

**Data Mining for Business Analytics text (The Green Book)**

1. Problems 2.1 a, b, c, d, e, f, g, h (4 marks, 0.5 mark each)



1. Supervised learning
2. Unsupervised learning
3. Supervised learning
4. Unsupervised learning
5. Supervised learning
6. Supervised learning
7. Supervised learning
8. Unsupervised learning
9. Problem 2.8 (show calculation steps for each column, Age and Income) (4 marks)



Text

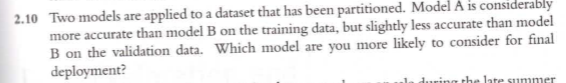
Description automatically generated

Normalization

A picture containing text, scoreboard

Description automatically generated

1. Problem 2.10 (1 mark)



Model B

**Data Science and Big Data Analytics text (The Black Book)**

Text

Description automatically generated

1. Chapter 3 Problem 1 (0.5 mark)
   1. 3 levels
2. Chapter 3 Problem 2 (0.5 mark)
   1. A picture containing text, scoreboard

      Description automatically generated
   2. A screen shot of numbers

      Description automatically generated with low confidence
3. Chapter 3 Problem 4 (0.5 mark)
   1. Install.packages()
   2. The name of the package is enclosed in the parenthesis with quotation marks (e.g., install.packages(“psych”))
4. Chapter 3 Problem 5 (0.5 mark)
   1. factor
5. Short answer: State the difference between a matrix and a data frame in R (0.5 mark).

* Matrix - All columns in a matrix must have the same data type (numeric, character, etc.) and the same length.
* Dataframe - A data frame is more general than a matrix, in that different columns can have different data types (numeric, character, factor, etc).

1. Short answer: List the basic data types in R (0.5 mark).
   1. Numeric
   2. Character
   3. Logical