QTM2000   Name_____

Case Studies in Business Analytics
Professor Mathaisel
Fall 2021

Mid-Term Exam 1

Individual Exercise

Due: 11:59pm Thursday, October 7.

I pledge my honor that I have neither received nor provided unauthorized assistance during the completion of this work. Please Initial:_____

Note: This is an Individual Exercise. No group submission. In addition to the instructor, feel free to reach out to others for help, but the work must be your original contribution.

1 .	10 points
2.	20 points
3.	35 points
4 .	35 points
	_____
	100 points

Introduction to the Case

In this exercise, we will explore a dataset from "Campus Recruitment - Academic and Employability Factors Influencing Placement" provided by Ben Roshan in Business Analytics at Jain University in Bangalore. "This data set consists of placement data of students in our campus. It includes secondary and higher secondary school percentage and specialization. It also includes degree specialization, type, work experience, and salary offers to the placed students." – Ben Roshan (https://www.kaggle.com/benroshan/factors-affecting-campus-placement). To better understand the meaning of the school boards for education in India: https://www.jagranjosh.com/articles/different-school-boards-in-india-1525780883-1. The most common school board in India is the Central Board whose general goal is: "Always works with a vision to provide stress-free education to its students by adapting innovative teaching methods infused with psychological and instructive principles". Other boards have different goals and adopt different methods of teaching.

Source of the dataset:
https://www.kaggle.com/benroshan/factors-affecting-campus-placement

Data Dictionary:
This is the only data dictionary provided by the source.

- **sl_no** = Serial Number
- **gender** = Gender: Male='M', Female='F'
- **ssc_p** = Secondary Education percentage - 10th Grade
- **ssc_b** = Board of Education- Central/ Others
- **hsc_p** = Higher Secondary Education percentage - 12th Grade
- **hsc_b** = Board of Education - Central/ Others
- **hsc_s** = Specialization in Higher Secondary Education
- **degree_p** = Degree Percentage
- **degree_t** = Under Graduation(Degree type) - Field of degree education
- **workex** = Work Experience
- **etest_p** = Employability test percentage (conducted by college)
- **specialisation** = Post Graduation(MBA)- Specialization
- **mba_p** = MBA percentage
- **status** = Status of placement- Placed/Not placed
- **salary** = Salary offered by corporate to candidates

1. Data Description and Structure
Briefly and concisely describe the data: its source, its structure, and units for each attribute (if possible).

2. Data Visualization
Briefly and concisely tell a story of what you visualize from each of the plots generated by the R script. Use as few words as possible, but comprehensively convey your story. Address all of the visualizations presented by the script.

3. Linear Regression
Summarize the results of the linear regression model. Your analysis should include the following:

    a. A brief discussion of what the model(s) are trying to accomplish.
    b. A brief discussion of the visualizations.
    c. A description of the resulting model(s) with an explanation of each of the coefficients in the model(s) using the attribute information that is provided (above) in the data dictionary.
    d. Which variables are the most significant in their order of significance?
    e. An overall evaluation of the models, limiting the evaluation to: the t-values / p-values; the standard error; the R^2 Adjusted; and F-ratio tests. Be sure to explain these tests in words, and in the context of the case. It is not necessary to say whether the model is good or bad at explaining the target variable.

4. k-NN
Summarize the results of the k-NN model. Be sure to include the following:

    a. A brief discussion of what the model is trying to accomplish.
    b. A brief discussion of the visualizations.
    c. An evaluation of the model, including, but not limited to, the Classification Matrix.