

# Case Instructions

## **Objectives of this case:**

1. You will use various tools to cleanup and manipulate data for the purpose of doing analytic work.
2. You will leverage your past coursework to explore the data and complete analysis that addresses the core questions asked in the case.
3. You will focus on telling a compelling story that utilizes your analytic insight to make specific recommendations regarding the questions in the case.

You may work in teams of up to five individuals. You may self-select into teams. At the end of the case, you will submit a management presentation that summarizes your findings.

## **Task:**

You have been hired by a large educational institution to examine trends in the analytic profession. The institution's goal is to effectively prepare and counsel students interested in analytic professions. Particularly, the company would like to build a tool that allows students to enter their qualifications and get a projected salary.

Your presentation should address the following questions:

1. What are the most significant factors driving Data Scientist salaries?
2. What are the most prevalent tools (software) and techniques (methods) being applied by Data Scientist today?
3. What tools and techniques are emerging in the field of Data Science?
4. How and where should aspiring data scientist invest their time and energy in order to prepare for the current and future Data Science environment?
5. Is formal education important to success as a Data Scientist?
6. How does the return on formal education compare to other types of learning?
7. How should educational institutions think about the role of formal education in the world of data science? Are there any specific recommendations in terms of methodologies that institutions of formal education should employ in the training of data scientists?

8. In addition, you need to create a tool (only a model...does not need to be a production piece of software) that students and young professionals can use to input their information and get a predicted income (you also need to validate that the model works). Make sure to include the model results and performance in your presentation.

### **Data Available:**

Kaggle conducted an international survey with over 25,000 responders. The survey was intended to capture the state of data science in 2021. You have three files:

1. Kaggle\_survey\_2022\_responses.csv : This file contains all answers to multiple choice questions from responders.
2. Kaggle\_survey\_2022\_answer\_choices.pdf: This file is a data dictionary with the actual questions asked in each column of the survey.
3. Kaggle\_survey\_2022\_methodology

### **Assignment:**

Your assignment is to develop and submit a management presentation that addresses the questions outlined. The management presentation should answer the questions asked above and cover the following (but does not have to be strictly arranged in these sections):

1. Executive Summary
2. Problem Statement
3. Data and Analysis
4. Findings/Recommendations
5. Limitations and Next Steps

The management presentation does not have to be long, but it should answer all questions, give management confidence in your findings, and leave them with an appropriate level of certainty in the results. If you prefer, you may record a video presentation (using Zoom or other video tools) of your presentation and include it with your final submission (you can save video in Google drive and provide me a link)

### **Grading:**

45% will be quality of analytics

45% will be quality of presentation (communication of findings/results)

10% will be your individual contribution (measured via survey)

**Deliverables:**

1. Management Presentation (should be clean and well structured) format can be PowerPoint or Google Slides. You need to turn in the presentation with all team members listed.