DSA 610, Homework #1, Spring 2025

Part I:
Instructions: Answer each discussion question in your own words. You may use posted resources or other online resources to answer the questions (**cite your sources**). Thoroughly explain your responses with a minimum of one paragraph (3-5 sentences) in length. Be thoughtful. We will discuss the answers in class.

1. What are the biggest challenges organizations face in managing data effectively throughout its lifecycle? Give some specific examples.

2. How does the increasing volume and variety of data impact the data management lifecycle?

3. How does the data analysis lifecycle differ from the broader data management lifecycle?

4. Why have a "lifecycle" at all, for either data management or data analysis? What purpose does it serve? What are the drawbacks of not being aware of or ignoring these lifecycles?

5. Discuss the ethical implications of using data analysis to target advertising.

Part II:
Instructions: Use the attached dataset (**beer_preference_data_hw1.xlsx**) to complete the following tasks in Excel. Report your answers to the questions on this homework sheet. Include your Excel (or Sheets) file along with your homework submission.

Use Excel to answer the following:

1. Create a table of counts for the Gender and Martial Status Variables. Then create a two-way table that includes both variables. Create graphs of your tables.

2. When making a graph of the two-way table, you can do a cluster graph (side-by-side), or a stacked graph (counts or percentages). Make each type and explain how each graph is different and what story it tells about the data (you need one explanation per graph type).

3. Describe the distribution of Salary in the dataset. Give some basic statistical information and construct a histogram of the data.

4. Create a comparative boxplot of Age by Beer Preference. Does there appear to be a relationship? Describe it.

5. Does beer preference vary by gender? By marital status? How do you know?

6. Are there any outliers in the age or income variables? How could you tell?

7. Describe any limitations to the dataset.

8. Could the information from this dataset be used in a way that is discriminatory or unfair?