DSA 610, Homework #2, 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 key features of a relational database? How do they ensure data integrity? What are their limitations?

2. What are the different types of NoSQL databases? How do they differ from relational databases? What are the challenges of working with NoSQL databases?

3. Can you give examples of real-world applications where different database types would be appropriate?

4. What are the main components of a conceptual model? How do you represent these components in a diagram?

5. What are some common data structures used in data analysis? When would you choose one data structure over another for a specific analysis task?

Part II:
Instructions: Use the attached dataset (**beer_preference_data_hw1.xlsx**) to complete the following tasks in Python. Report your answers to the questions on this homework sheet. Include your Jupyter notebook file along with your homework submission, saved as a PDF. Make sure that any graphs you create are quality graphs with a legend (as needed), axis titles, a descriptive title, appropriate ranges (bar graphs start at 0, etc.). They should be able to stand on their own. You may need to relabel some elements (such as replacing 0s and 1s with string labels).

Use Python to answer the following:

1. Create a bar graph for Gender and another for marital status with different plotting packages. What are the advantages and disadvantages of the packages you chose for this graph type.

2. Make a pie chart of Beer Preference. Which package did you use for this and why? Why do you think data analysts might not like pie charts in general?

3. Create a stacked bar graph of gender and beer preference. Create a cluster bar graph of the same data. What different stories do they tell?

4. What kind of datatypes are used in this dataset? Can you determine what kind of data you are working with solely from the datatype imports? Why or why not? Give examples.

5. Create a bar graph of average salary by marital status.

6. Create a bar graph of age by beer preference.