University of Science and Technology
in Zewail City

جـــامعة العلـوم والتكنـولوجيـا
بمديـــنة زويل

Data Acquisition - DSAI 103
Spring 2024
Assignment 1

Once upon a time in a small town, there was a botanist named “Ann”. Ann was known for her exceptional skills in analyzing and studying various plant species. One day, she received an intriguing task from her mentor - to analyze the **Iris dataset** that was attached to the assignment.

Excited about the challenge, Ann began her analysis by loading the dataset into her preferred data analysis libraries. She imported the necessary libraries, such as pandas and matplotlib, to aid her in her analysis.

As she dove deeper into the dataset, Ann's first task was to check if the data of each flower was complete. She meticulously examined each row and column, looking for any missing or empty cells.

Following the guidelines provided, Ann came up with a plan to handle the missing data. If there was only one empty cell in a specific flower's data, she decided to fill it with the average value of that particular column. This ensured that the dataset remained as accurate as possible. However, if she found more than one empty cell for a flower, she made the difficult decision to remove that particular flower from the dataset.

As Ann continued to analyze the dataset, she stumbled upon a peculiar observation. In one of the columns, she noticed that there were more than 15 empty cells. Realizing that these records were unnecessary for her analysis, Ann promptly deleted them from the dataset. This ensured that her analysis would be based on accurate and meaningful data.

To further enhance her analysis, A decided to remove any repeated records from the dataset. By doing so, she eliminated any redundancy and maintained the integrity of her analysis.

Feeling satisfied with her data cleaning and preprocessing efforts, Ann moved on to visualize her findings. She plotted a pie graph to showcase the number of different flower types present in the dataset. The vibrant colors of the pie graph highlighted the diversity among the various **species** of Iris flowers.

In addition to the pie graph, A plotted a bar graph to illustrate the relationship between "**sepal length**" and **"sepal width**" across all the flowers. This visualization allowed her to observe any patterns or trends that might exist between these two variables.
As A completed her analysis and visualizations, she felt a sense of accomplishment. Her dedication to detail and her proficient use of the libraries she had learned allowed her to uncover valuable insights from the Iris dataset. She eagerly shared her findings with her mentor, who praised her exceptional work.