DATA ANALYTICS AND VISUALIZATION COURSE

CAPSTONE PROJECT

By Sheriffdeen Hammed

Project Steps:

1. Problem Definition:

Objective: Analyze the organization’s patient database to discover trends and insights. The company is interested in identifying common health issues, treatment outcomes, and geographical distribution of diseases to better tailor its services.

1. Data Cleaning:
2. Changing the case of the Patients’ names

- First, I inserted a temporary column next to the right column that contains the name.

- In cell B2, I typed =PROPER(A2), then pressed Enter. I then filled down the formula in the column by double-clicking the small black square that appears in the lower-right corner of the cell.

-  I selected all the values in the new column (B) and I pressed CTRL+C to copy them to the Clipboard.

- I right-click cell A2, clicked Paste, and then clicked Values. Then, I deleted column (B).

1. Standardisation of “Gender” Column to fix inconsistent entries:

- Replaced variations of "Female" F

- Replaced variations of "Male" with M.

1. Remove Duplicates

- I highlighted the entire dataset.

- I went to Data → Remove Duplicates.

- I selected all columns to identify duplicates, then clicked OK.

1. Data Cleaning

Key Questions:

■ What are the most common diagnoses based on region?

A pivot table was created showing the most common diagnoses based on each hospital location.

■ What are the success rates of the various treatment plans?

I used the “COUNTIFS” formula to count the number of times the treatment result of each medication is “Normal”. I found out that Ibuprofen has the highest success rate.

■ What are the demographics (age, gender) of patients with specific health

conditions?

I calculated the average, maximum and minimum age of patients with each of the health condition. Then, I created a pivot table to show the number of male and female with specific health condition.