UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

Artificial Intelligence

Assignment 2: Processing Datasets in Python

Name	ID
Sergio Orlando Alanís De La Rosa	2043932
Diego Castro Galindo	2109304
Sebastian Hernandez Renteria	2109140
Braulio Azael García Treviño	2043046
Miguel Angel Perez Luevano	2052579

Professor Name: Daniel Isaías Lopez Paez

Career: IB   **Group:** 003

Date: 5/9/25

Google Colab:

Conclusion:

First, it checks the most important columns (like glucose, blood pressure, insulin, etc.) to detect if they have weird values, such as zeros or empty spaces. These values are a problem because, in real life, it wouldn't make sense for someone to have "0" glucose.

Then, instead of deleting those records, the program fixes them by replacing them with the median of each column. This way, the information isn't lost, but it also doesn't keep incorrect values.

Next, the data is adjusted to the same scale (between 0 and 1). This is important so that all the variables have equal weight and can be used properly in further analysis or prediction models.

Finally, some histogram plots are created to compare how glucose values look before and after normalization. This helps confirm that the change was applied correctly.