Use the EH\_RESP2015 dataset

1. Clean the data

Make the following corrections to the ERBMI, ERINCOME, EUEXFREQ, EUFASTFDFRQ data:

A) Eliminate all rows with missing data.

B) Eliminate all rows with outliers (with a z-score above 2 or below -2).

C) Add a binary OVERWEIGHT indicator column with a value of 1 if ERBMI > 25.0; o/w 0.

2. Numerical Prediction Problem

Predict ERBMI (Body Mass Index) based on ERINCOME, EUEXFREQ, EUFASTFDFRQ

A) Using KNN

B) Multiple Regression

3. Classification Problem

Predict the OVERWEIGHT based on ERINCOME, EUEXFREQ, EUFASTFDFRQ.

A) Using KNN

B) Using Logistic Regression

C) Using Naive Bayes

Then evaluate #2 and #3 separately and submit ONLY your conclusions and any necessary charts or data to justify your findings. Write it like an email you'd write to a supervisor or an investor in a business context. Focus on the findings and provide justifications for them. Do not submit Excel files or Google Spreadsheet links or Python code or photos of you with an abacus or anything else that you wouldn't send in a business email. You can include some brief end notes if you feel you absolutely must (eg "I also tried method X but it's accuracy was 34% as opposed to the 89% of the method Y that I adopted.")

You could use this link to generate the models

https://www.wolframcloud.com/obj/philip/ai