**Example 1**

**Background Knowledge:**

This dataset is structured to analyze how eight variables—X1 (Relative Compactness), X2 (Surface Area), X3 (Wall Area), X4 (Roof Area), X5 (Overall Height), X6 (Orientation), X7 (Glazing Area), and X8 (Glazing Area Distribution)—affect Y1 (Energy Heating Load). This analysis is conducted using a linear regression model.

Please evaluate X1 to X8 as independent variables in relation to Y1, the dependent variable. Utilize a linear regression approach to investigate the following questions:

**Question Input:**

Among these factors, which one has the greatest impact on Y1?

**LLM output:**

Correlation with Y1:

X1 (Relative Compactness): 0.62

X2 (Surface Area): -0.66

X3 (Wall Area): 0.46

X4 (Roof Area): -0.86

X5 (Overall Height): 0.89

X6 (Orientation): -0.003

X7 (Glazing Area): 0.27

X8 (Glazing Area Distribution): 0.09

(*For a linear regression model, these coefficients should be further explained, such as "When other independent variables fixed, every 1 unit increase in X1 will increase Y1 by 0.62 on average."*)

**Example 2**

**Background Knowledge:**

This is a dataset about how pregnancy, glucose level, blood pressure, the triceps skin fold thickness, insulin level, pedigree, BMI, and age affect the diagnosis of diabetes. Please use the pregnancy, glucose level, blood pressure, the triceps skin fold thickness, insulin level, pedigree, BMI, and age as independent variables and the diabetes as the dependent variable and use a Linear Discriminant Analysis to answer the following questions.

**Question Input:**

What is the most influential factor for diabetes?

**LLM output:**

Pregnancies: 0.213

Glucose: 1.071

BloodPressure: 0.248

SkinThickness: 0.046

Insulin: 0.201

BMI: 0.778

DiabetesPedigreeFunction: 0.231

Age: 0.421

(*The answer here should be "Glucose is the most influential independent variable" instead of simply listing the coefficients of all independent variables.*)