**Run the code and generate the health\_survey dataset:**

data health\_survey;

do ID = 1 to 200;

Gender = ifc(ranuni(0) < 0.5, 'Male', 'Female');

Age = int(18 + ranuni(0)\*50); /\* 18 to 68 \*/

BMI = round(18 + ranuni(0)\*17, 0.1); /\* 18 to 35 \*/

Smoking\_Status = ifc(ranuni(0) < 0.3, 'Smoker', 'Non-Smoker');

Exercise\_Freq = int(ranuni(0)\*7); /\* 0 to 6 days \*/

Cholesterol = round(150 + ranuni(0)\*100, 1); /\* 150 to 250 \*/

Diabetes = ifc(ranuni(0) < 0.2, 'Yes', 'No');

Blood\_Pressure = choose(ceil(ranuni(0)\*3), 'Low', 'Normal', 'High');

output;

end;

run;

**Answer the below questions. In case you need help from ChatGPT feel free to use it. But make sure you understand the AI generated SAS script.**

**Descriptive Statistics**

1. What is the average BMI and standard deviation for the entire dataset using PROC MEANS?
2. Find the minimum and maximum age for each gender using PROC SUMMARY.
3. Use PROC MEANS to calculate the mean cholesterol level for smokers and non-smokers.
4. How many people have diabetes in each blood pressure category using PROC FREQ?
5. Generate a frequency table for the Smoking\_Status variable.

**Contingency Table & Cross-tabulations**

1. Create a 2-way contingency table between Gender and Diabetes.
2. Use PROC FREQ to find if there's a relation between Smoking\_Status and Blood\_Pressure. Include row and column percentages.

**Statistical Tests**

1. Perform a T-TEST to compare the mean BMI between male and female respondents.
2. Use PROC TTEST to check if there is a significant difference in cholesterol between smokers and non-smokers.
3. Use PROC NPAR1WAY to compare exercise frequency between diabetes and non-diabetes groups (non-parametric test).
4. Perform a Chi-Square test to assess the association between Diabetes and Blood\_Pressure.
5. Use PROC CORR to find the correlation between Age, BMI, and Cholesterol.
6. Check if age is significantly different between people with and without diabetes using PROC TTEST.

**Data Visualization**

1. Create a bar chart showing the number of males and females.
2. Use PROC SGPLOT to create a histogram of the BMI variable.
3. Plot a scatter plot between BMI and Cholesterol with regression line.
4. Draw a box plot comparing cholesterol levels across different blood pressure categories.
5. Create a grouped bar chart of smoking status by gender.

**Advanced Summarization**

1. Using PROC SUMMARY, calculate mean, median, and standard deviation of exercise frequency for each smoking status and gender group.
2. Create a summary report using PROC REPORT to show average age, BMI, and cholesterol by diabetes status.