**Prompt For Data Science**

We have a Heart Disease prediction dataset with a single table which has the following attributes.

1. age - age in years

2. gender- gender (1 = male; 0 = female)

3. cp - chest pain type

-- Value 1: typical angina

-- Value 2: atypical angina

-- Value 3: non-anginal pain

-- Value 4: asymptomatic

4. trestbps - resting blood pressure (in mm Hg on admission to the hospital)

5. chol - serum cholestoral in mg/dl

6. fbs - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

7. restecg - resting electrocardiographic results

-- Value 0: normal

-- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)

-- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

8. thalach - maximum heart rate achieved

9. exang - exercise induced angina (1 = yes; 0 = no)

10. oldpeak - ST depression induced by exercise relative to rest

11. slope - the slope of the peak exercise ST segment

-- Value 1: upsloping

-- Value 2: flat

-- Value 3: downsloping

12. ca - number of major vessels (0-3) colored by flourosopy

13. thal - 3 = normal; 6 = fixed defect; 7 = reversable defect

14. num (the predicted attribute) - diagnosis of heart disease (angiographic disease status)

-- Value 0: < 50% diameter narrowing

-- Value 1: > 50% diameter narrowing  
  
**Prompts for Data Querying**

**Age Distribution**: Write an SQL query to find the minimum, maximum, and average age of patients in the dataset.

**Gender Analysis:** Write and SQL query to count the number of male and female patients in the dataset.

**Chest Pain Type Frequency:** Write an SQL query to determine the frequency of each type of chest pain (typical angina, atypical angina, non-anginal pain, asymptomatic) among patients.

**Age Group Analysis and Target Variable:** Write an SQL query to investigate the distribution of the target variable (presence or absence of heart disease) within different age groups (e.g., 20-30, 30-40, etc.).