**🫀 1. Heart Disease Dataset**

**Goal:** Predict presence of heart disease.

**📊 EDA Questions**

1. What is the distribution of age across the dataset? Does it differ between patients with and without heart disease?
2. Is there a gender imbalance in the data? How does heart disease prevalence vary across genders?
3. What is the relationship between resting blood pressure (trestbps) and cholesterol (chol)?
4. Does chest pain type (cp) show strong correlation with heart disease outcome?
5. Use a heatmap to find correlations between all numeric features. Which are most strongly correlated with target?
6. Are there patterns in maximum heart rate achieved (thalach) and exercise-induced angina (exang)?
7. Are there any obvious outliers in features like cholesterol (chol) or resting blood pressure?

**🛠️ Feature Engineering Questions**

1. Create an age group feature (young, middle-aged, senior). Does this improve model performance?
2. Encode cp, thal, and slope using one-hot encoding. How many new features are created?
3. Combine chol and trestbps to make a "cardiovascular risk index". Does it help?
4. Create an interaction feature: thalach \* (1 - exang). What insight can this give?
5. Create a binary feature high\_chol if cholesterol is above 240.
6. Normalize continuous features. Which normalization method (min-max vs z-score) works best?