# Project: Patient Risk Stratification & Heart Disease Prediction

## Executive Summary

This project analysed clinical data from patient records to identify key indicators of heart disease. The objective was to move beyond standard metrics (like high cholesterol) and find specific, high-probability risk factors to improve hospital triage and resource allocation. The analysis focused on demographic profiling, symptom correlation, and the efficacy of diagnostic stress tests.

## Analysis 1: Demographic Risk Profiling (The "Volume" Danger)

**Objective:** Determine the demographic segments with the highest prevalence of heart disease to optimize preventative care targeting.

**The Interpretation:**
The analysis identified the **50–59 age bracket** as the critical "Volume Risk" group for the hospital.

- **Volume vs. Rate:** While older demographics (60+) showed a slightly higher *probability* of disease per person, the 50–59 age group accounted for the highest *total number* of positive cases.
- **Hypertension Trend:** A significant portion of this group presented with hypertension (BP > 130), distinguishing them as the primary burden on hospital resources.
- **Strategic Implication:** Preventative screenings and intervention programs (e.g., hypertension management) yield the highest ROI when targeted specifically at patients entering their 50s, rather than waiting for the 60+ threshold.

## Analysis 2: Biomarker Correlation (The Cholesterol Fallacy)

**Objective:** Evaluate if Serum Cholesterol levels are a reliable standalone indicator for distinguishing sick patients from healthy ones in the high-risk age group.

**The Interpretation:**
The data suggests that High Cholesterol is a necessary but insufficient metric for diagnosis in the 50–59 age group.

- **The "False Healthy" Baseline:** Patients *without* heart disease still exhibited an average cholesterol level of **241.8 mg/dL**, which is already above the medical "High Risk" threshold of 240 mg/dL.
- **The Disease Gap:** Patients *with* heart disease averaged **259.9 mg/dL**. While higher, the gap (~18 points) is not wide enough to be a clear binary filter.
- **Conclusion:** Relying solely on cholesterol screenings will result in excessive false positives, as the "healthy" population is statistically indistinguishable from the "at-risk" population based on this metric alone.

## Analysis 3: Symptom Analysis (The "Type 4" Red Flag)

**Objective:** Correlate reported "Chest Pain Type" (Values 1–4) with confirmed diagnoses to identify which pain presentation carries the highest risk.

**The Interpretation:**
The analysis revealed a counter-intuitive risk hierarchy regarding patient symptoms.

- **The Top Risk Factor: Chest Pain Type 4** emerged as the single most dangerous indicator. Patients presenting with this specific pain type had a confirmed disease conversion rate of over **55%** (significantly higher than Types 1, 2, or 3).
- **Triage Protocol Adjustment:** This challenges the assumption that acute/typical pain is always the primary indicator. The data proves that patients presenting with Type 4 symptoms, often associated with asymptomatic or non-typical presentation, warrant immediate admission

and testing, as they are statistically more likely to be positive for heart disease.

## Analysis 4: Diagnostic Efficacy (The Treadmill Test)

**Objective:** Quantify the predictive power of "Exercise Induced Angina" (chest pain during physical exertion) to determine if it is a viable low-cost screening tool.

**The Interpretation:**
The Exercise Stress Test was validated as a high-efficacy "First Filter" for diagnosis.

- **Quantifiable Accuracy:** The analysis yielded a **74.2% Risk Probability**. If a patient experiences angina during exercise, there is a nearly 3-in-4 chance they have heart disease.
- **Operational Value:** While not 100% conclusive, this high probability justifies the continued use of exercise stress tests as a primary screening tool before authorizing more expensive invasive procedures (like angiography).