Summary - Heart Disease Risk

This project analyzes a **heart disease dataset of 920 patients** with 16 demographic, clinical, and diagnostic features. The objective was to explore disease prevalence patterns and uncover key risk indicators through descriptive statistics and visualization.

Key Findings

Demographics

- **Gender:**
  - Around **68% of patients are male** and **32% are female**.
  - Heart disease was more prevalent among men, with **~55% of male patients diagnosed** compared to **~35% of female patients**.
- **Age:**
  - Majority of cases fall between **40–65 years (≈70% of the dataset)**.
  - Patients above 60 years account for nearly **25% of all heart disease cases**, highlighting age as a strong risk factor.

Chest Pain Type (cp)

- **Asymptomatic chest pain** dominates in diseased patients (**>50%** of positive cases).
- **Typical angina & atypical angina** contribute around **35% combined**.
- **Non-anginal pain** is more common in non-diseased patients, serving as a differentiating marker.

Clinical Parameters

- **Cholesterol (chol):**
  - About **30% of patients recorded cholesterol >250 mg/dl**, with the majority of them diagnosed positive.
  - Normal-to-borderline levels (<200 mg/dl) are more frequent in disease-free patients.
- **Resting Blood Pressure (trestbps):**
  - Nearly **25% of patients have resting BP above 140 mmHg**, among which **60% are diseased**.
- **Fasting Blood Sugar (fbs):**
  - Patients with **fbs > 120 mg/dl (≈12% of dataset)** showed a notably higher rate of disease (**~65% positive**).
- **Maximum Heart Rate (thalach):**
  - Diseased patients generally achieve a lower thalach.
  - **75% of patients with thalach <140 bpm were positive for heart disease.**

Diagnostic Parameters

- **ECG (restecg):**
  - Abnormal ECG (left ventricular hypertrophy or ST-T abnormalities) appears in **>40% of diseased patients**.
- **Exercise-Induced Angina (exang):**
  - Present in **~33% of patients**, with a strong positive correlation to disease.
- **ST Depression (oldpeak):**
  - **Patients with oldpeak ≥2** form **~28% of the dataset**, of which nearly **70% are diseased**.
- **Slope of ST segment:**
  - Flat and downsloping slopes are strongly linked with disease, while upsloping is more common in healthy individuals.

Visualization Highlights

- **Bar and pie charts** reveal the **gender disparity** and chest pain distribution.
- **Histograms** clearly show differences in cholesterol and thalach between healthy vs. diseased.
- **Correlation heatmaps** highlight strong associations of age, cholesterol, blood pressure, and thalach with heart disease.

Conclusion

The study underscores that **age, male gender, high cholesterol, hypertension, abnormal ECG results, and low maximum heart rate** are critical predictors of heart disease. Asymptomatic chest pain and ST segment abnormalities further refine risk stratification.

The visualizations provide a clear narrative: **patients above 40, especially males with high cholesterol and abnormal diagnostic tests, are at significantly elevated risk.**