

The **NidaanKosha-100k-V1.0** dataset, provided by **Eka Care**, is a comprehensive collection of diagnostic test results from patients across India.

Dataset Overview

Source: [Hugging Face – EkaCare](#)

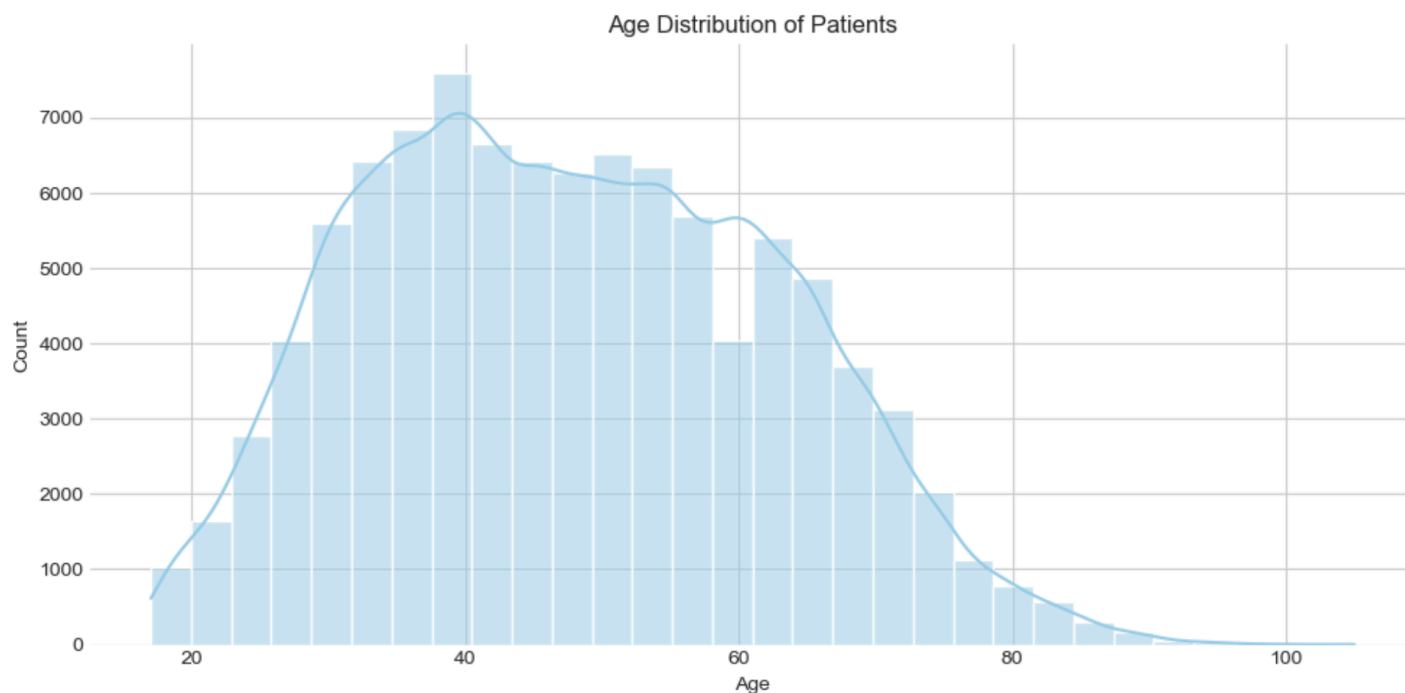
Total Records: ~6.8 million rows

Fields Included:

- **document_id:** Unique identifier for the document.
- **age:** Patient's age.
- **gender:** Patient's gender.
- **test_name:** Name of the medical test conducted.
- **display_ranges:** Reference ranges for the test results.
- **value:** Actual result of the test.
- **unit:** Measurement unit of the test result.
- **specimen:** Type of specimen analysed (e.g., blood).
- **loinc:** Logical Observation Identifiers Names and Codes (LOINC) associated with the test.

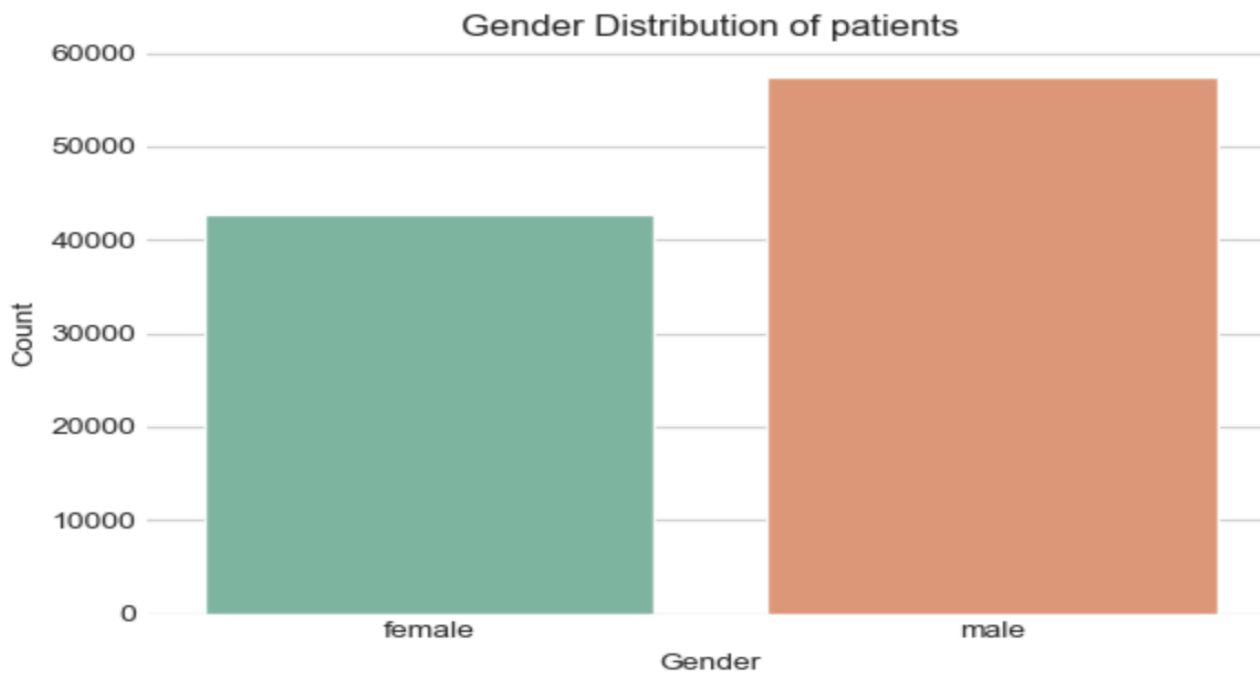
Exploratory Data Analysis

Age Distribution of Patients.



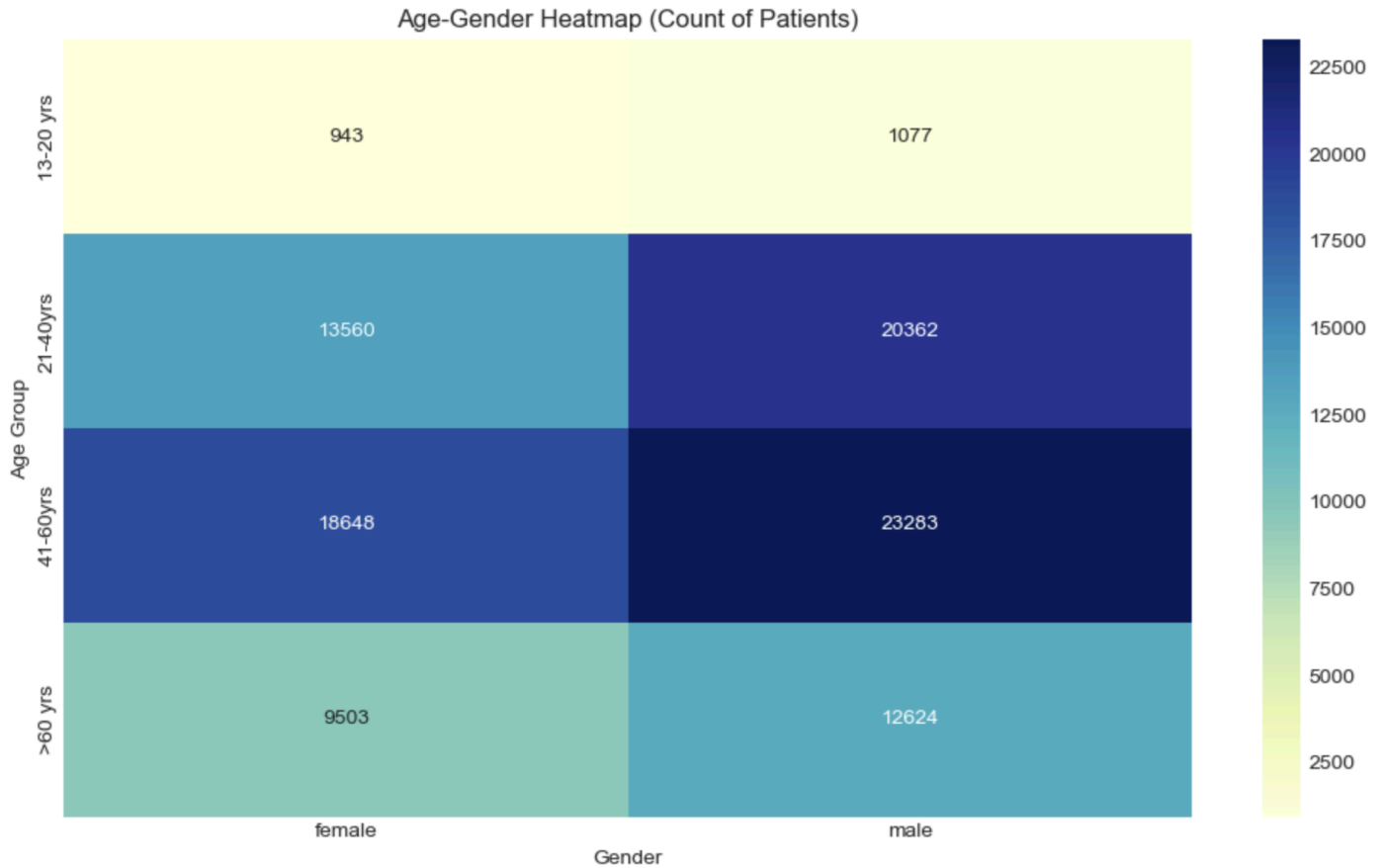
The age range of patients spans from 17 to 105 years, with the majority falling between 30 and 65 years of age.

Gender Distribution of Patients



Out of a total of 100,000 patients, 42% are female, and 58% are male.

Age-Gender Distribution



The age has been categorised into four groups to represent different life stages:

- **13–20 years:** Teens & Young Adults
- **21–40 years:** Adults
- **41–60 years:** Middle-aged
- **Above 60 years:** Seniors

From the heatmap above, it is evident that the majority of patients fall within the **41–60 years** age group, with **males** comprising the largest portion within this segment.

Specific Tests

We will focus on specific tests related to cardiac health and diabetes, including **Total Cholesterol**, **LDL Cholesterol**, **HDL Cholesterol**, **Triglycerides** and the **Fasting Glucose** test.

Cholesterol

Cholesterol is a waxy, fat-like substance found in all cells of the body, essential for building healthy cells and making hormones, but high levels can increase the risk of heart disease.

A cholesterol test, also known as a lipid panel, measures the following key components:

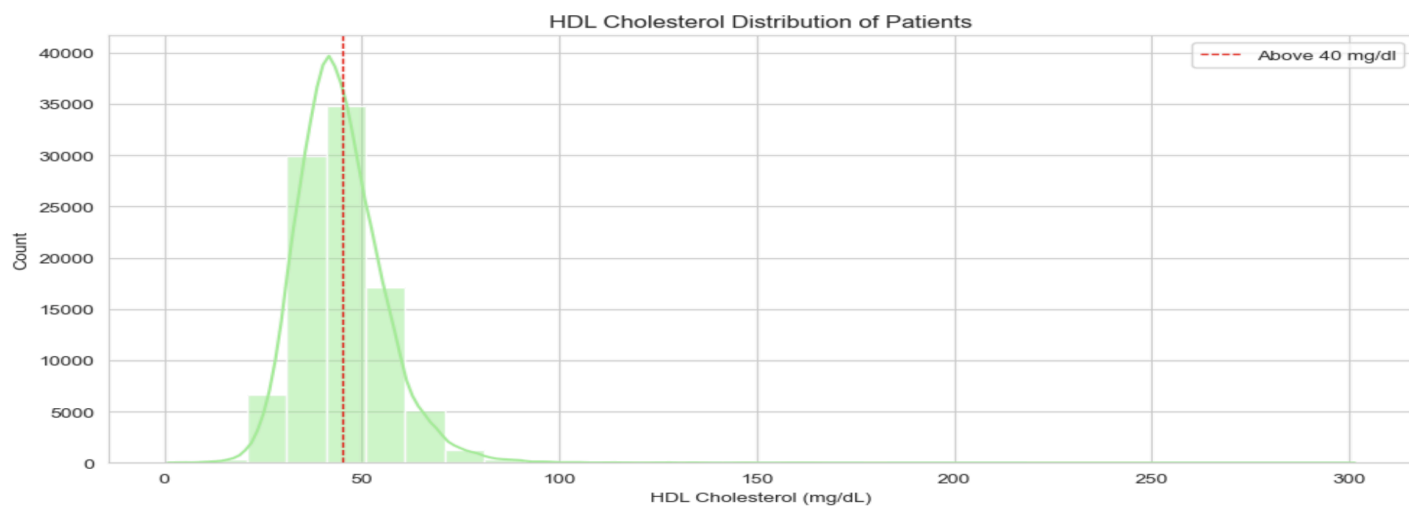
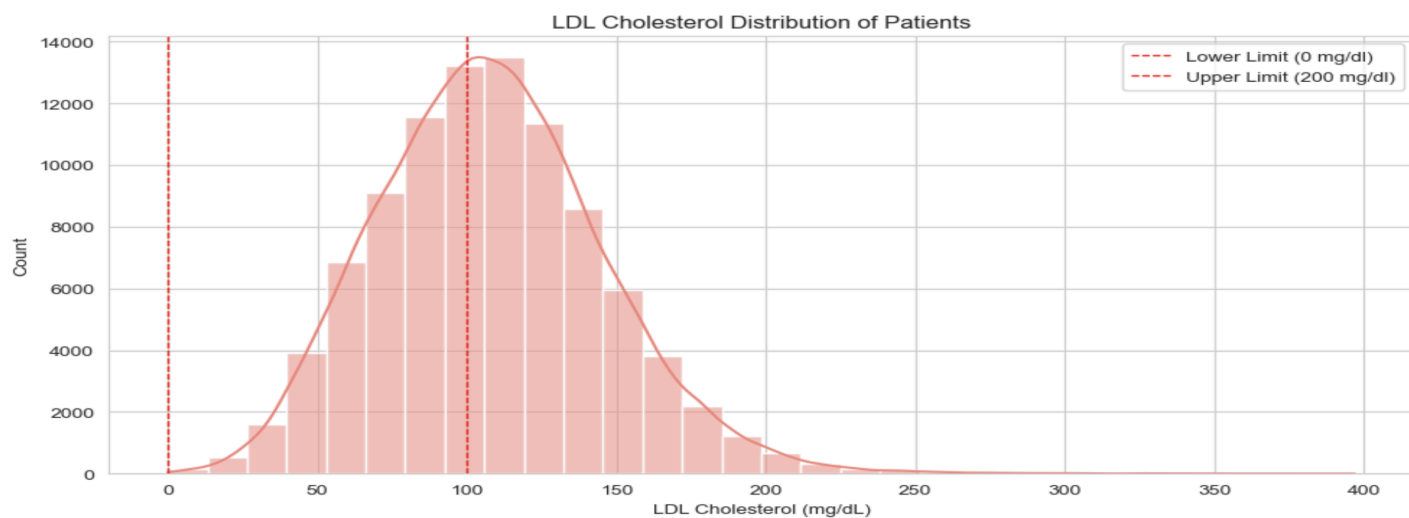
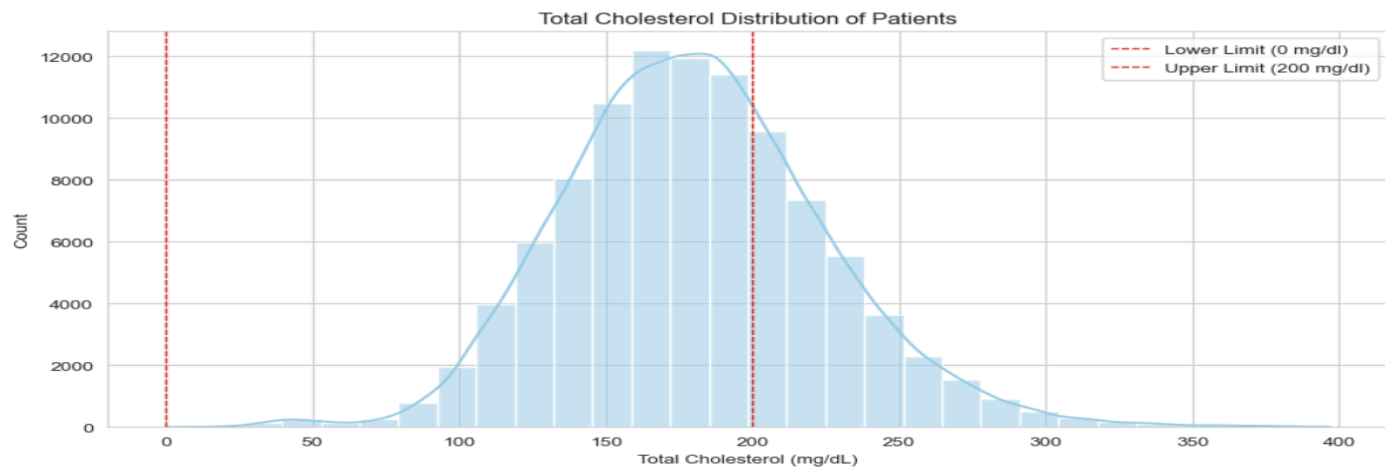
- **Low-density lipoproteins (LDL)** are known as the “bad” cholesterol. High levels of LDL can lead to plaque buildup in the arteries, increasing the risk of cardiovascular disease (CVD).
- **High-density lipoproteins (HDL)** are referred to as the “good” cholesterol. HDL helps remove excess LDL cholesterol from the bloodstream, so higher HDL levels are considered beneficial.
- **Triglycerides** are a type of fat in the blood, formed when the body stores unused calories. Elevated triglyceride levels can also raise your risk of CVD, so lower levels are preferred.
- **Total cholesterol** is calculated using the formula: $\text{HDL} + \text{LDL} + 20\% \text{ of your triglycerides}$. While you generally want this number to be lower, it should be interpreted in the context of your other cholesterol values for an accurate assessment of heart health.

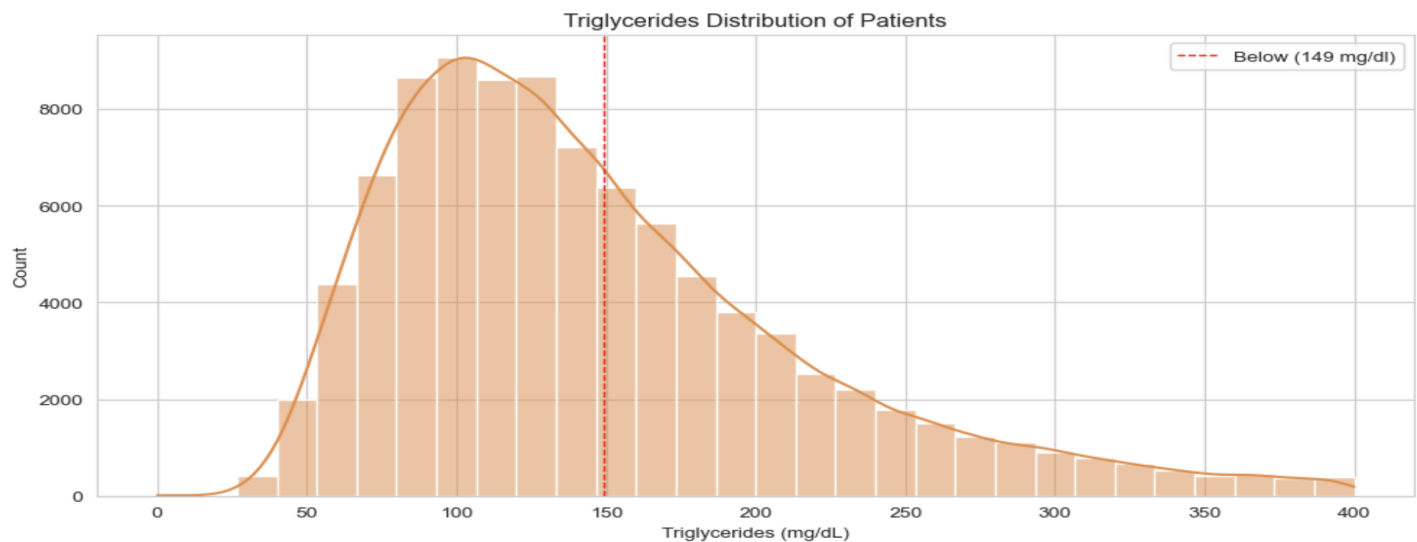
Cholesterol Distribution of Patients

The general reference ranges for key cholesterol-related tests are as follows:

- **Total Cholesterol:** Below 200 mg/dL
- **LDL (Low-Density Lipoprotein) Cholesterol:** Below 100 mg/dL
- **HDL (High-Density Lipoprotein) Cholesterol:** Above 40 mg/dL
- **Triglycerides:** Below 149 mg/dL

Maintaining these levels is essential for reducing the risk of cardiovascular diseases.

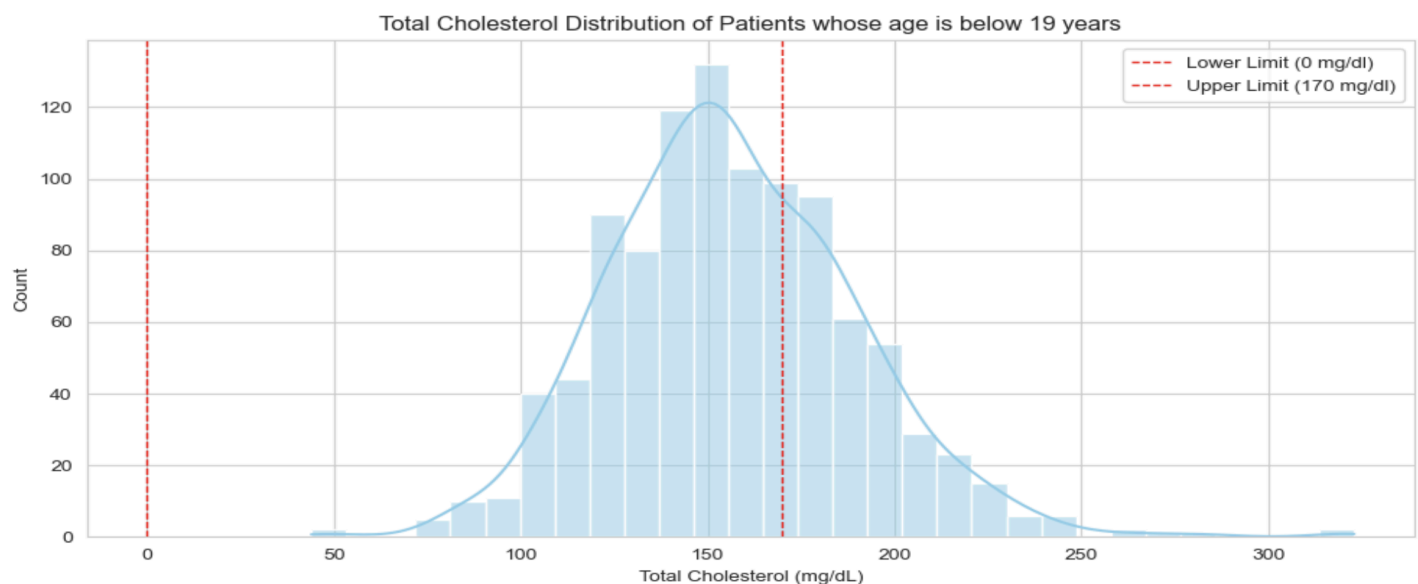




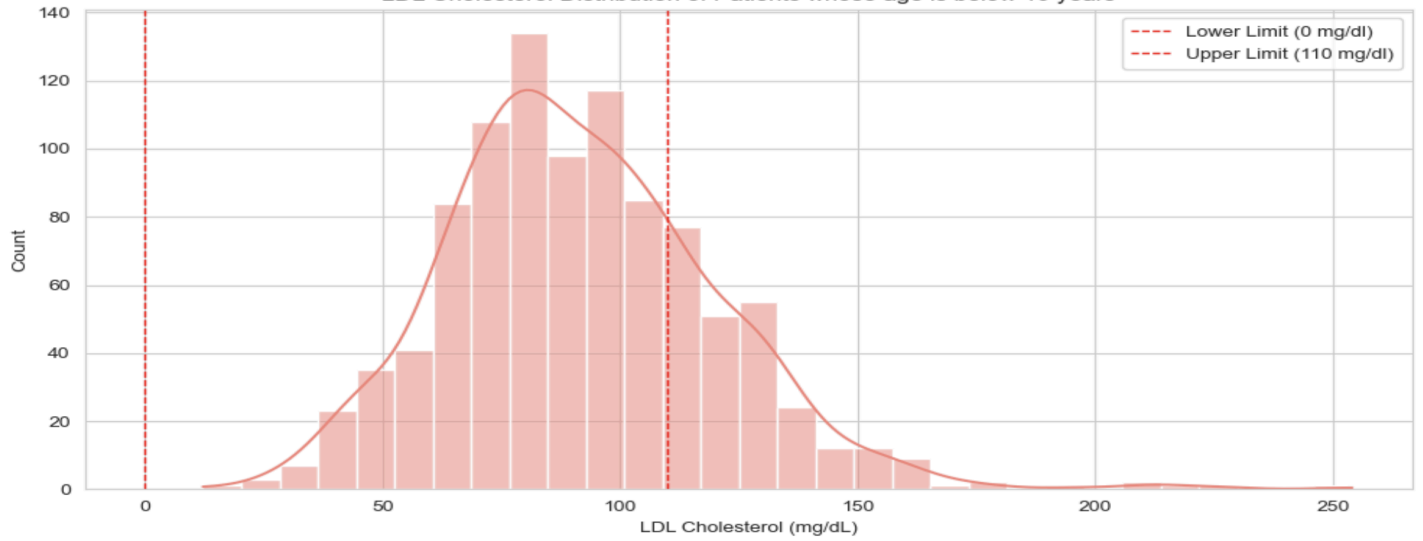
Cholesterol Distribution of Patients Below 19 Years

The general normal ranges for cholesterol levels for Patients below 19 years are as follows:

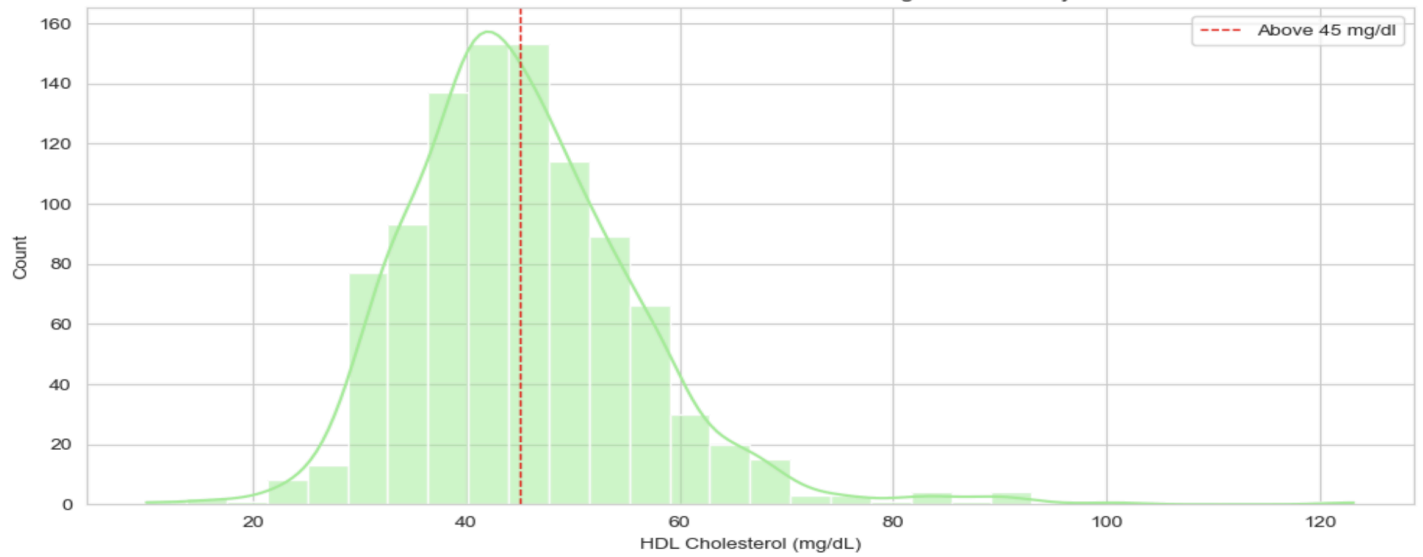
- **Total Cholesterol:** Below 170 mg/dL
- **LDL (Low-Density Lipoprotein) Cholesterol:** Below 110 mg/dL
- **HDL (High-Density Lipoprotein) Cholesterol:** Above 45 mg/dL
- **Triglycerides:** Below 90 mg/dL



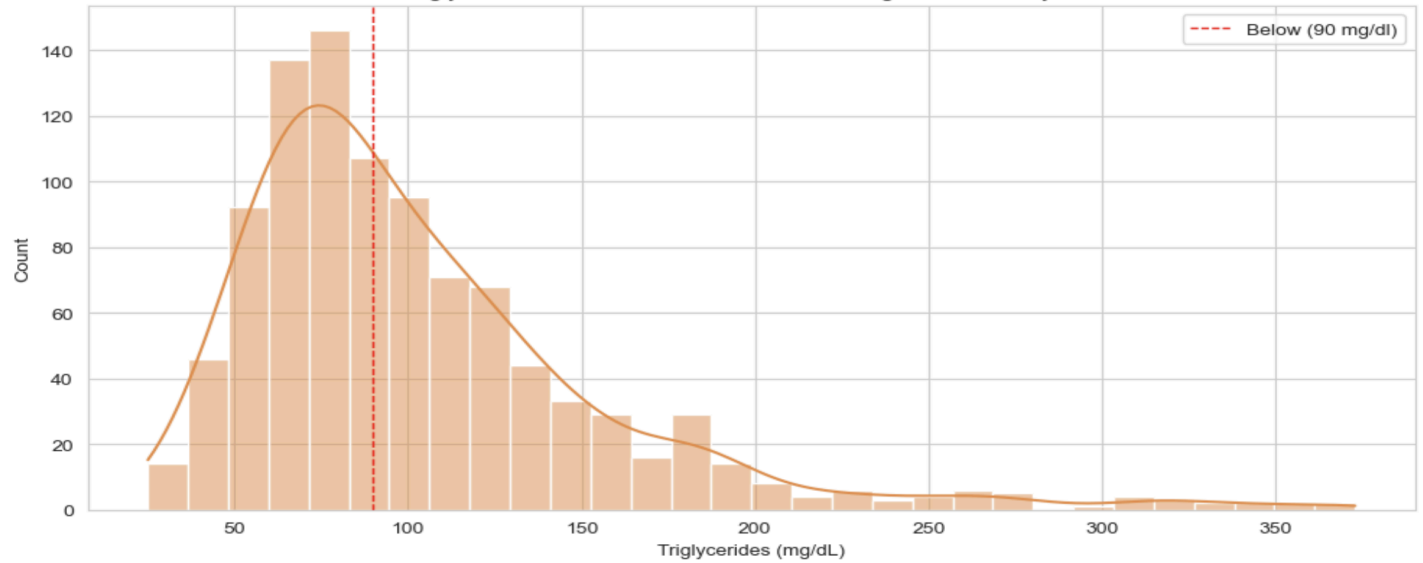
LDL Cholesterol Distribution of Patients whose age is below 19 years



HDL Cholesterol Distribution of Patients whose age is below 19 years



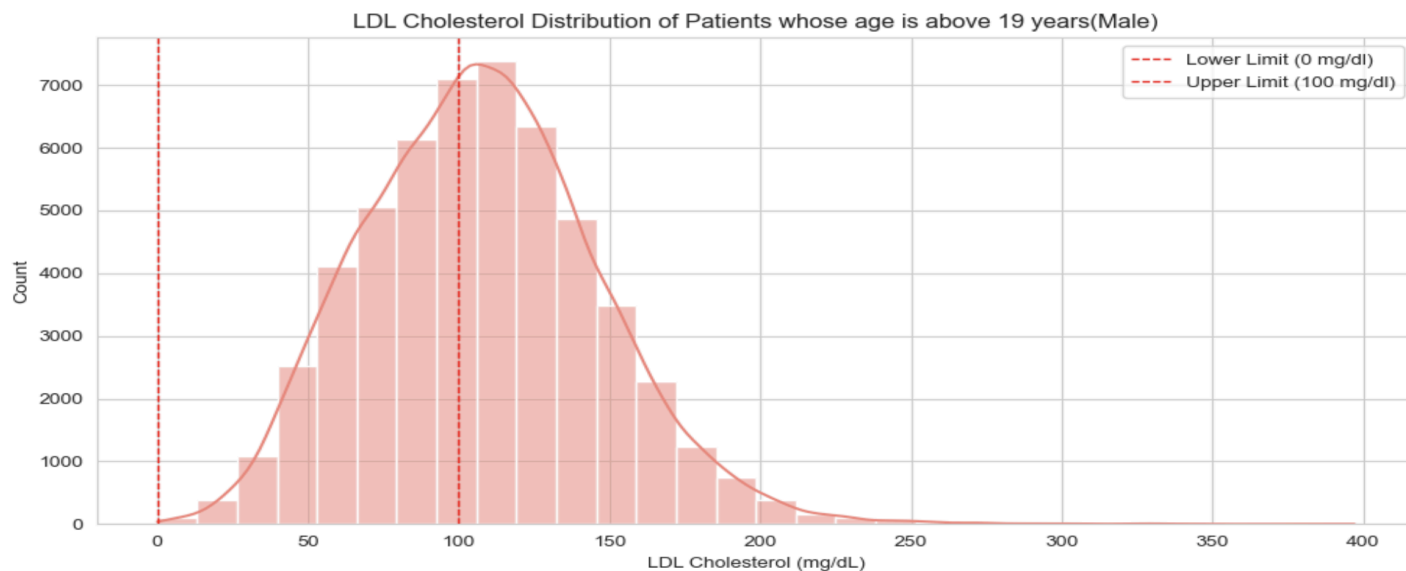
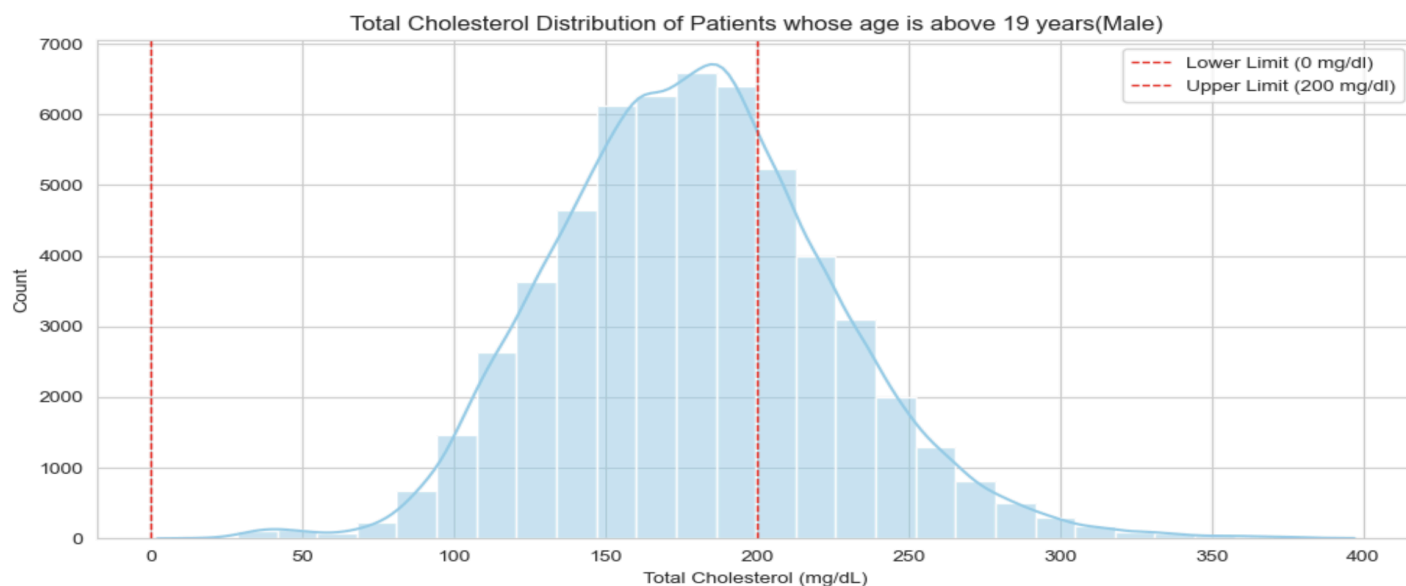
Triglycerides Distribution of Patients whose age is below 19 years

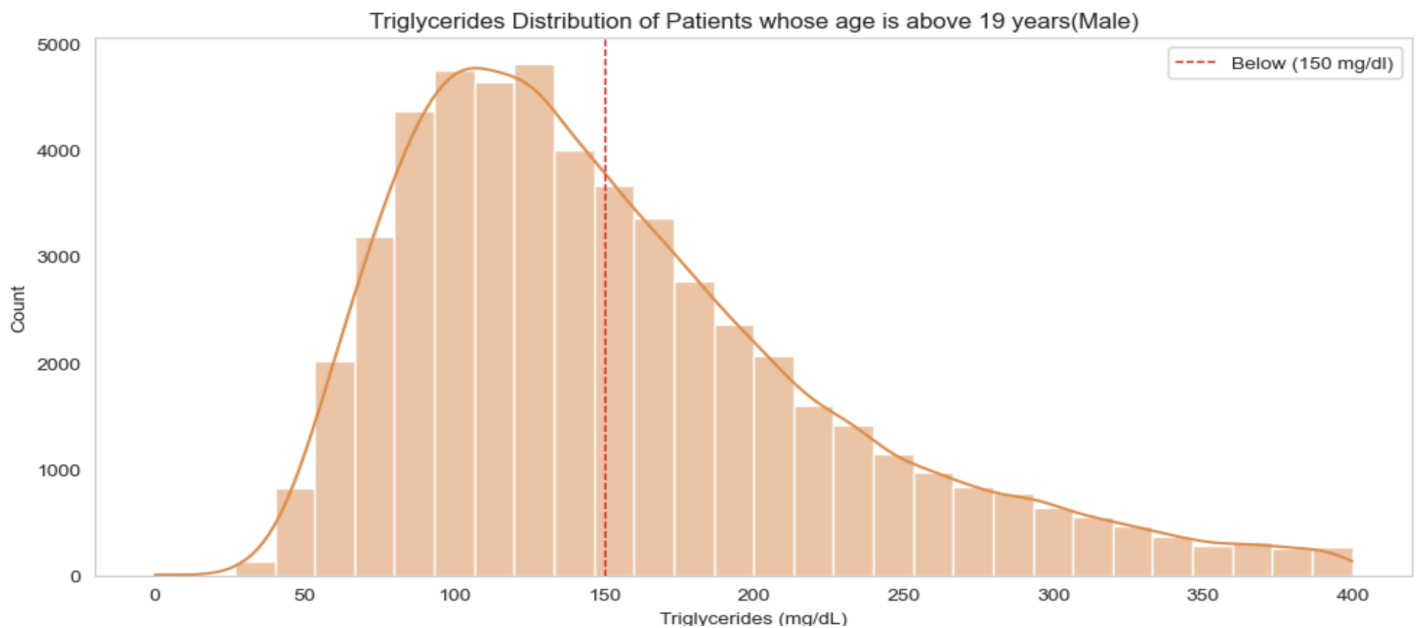
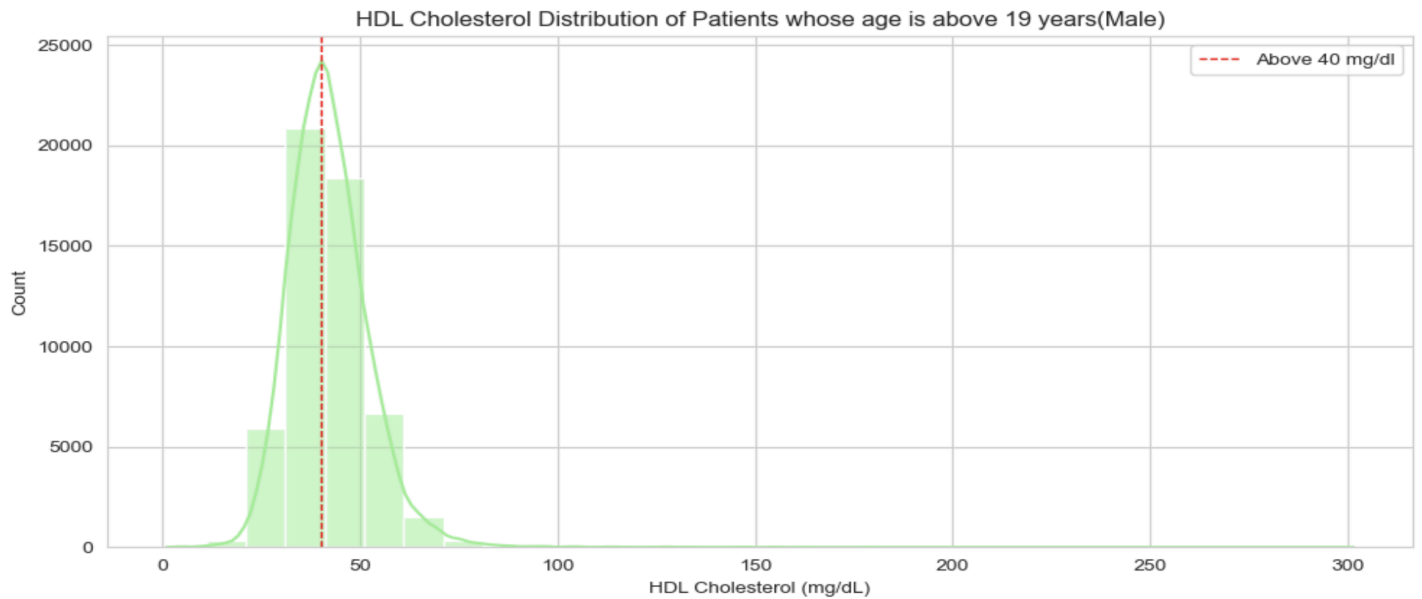


Cholesterol Distribution of Patients Above 19 Years(Male)

The general normal ranges for key cholesterol parameters for patients above 19 years(Male) are:

- **Total Cholesterol:** Below 200 mg/dL
- **LDL (Low-Density Lipoprotein) Cholesterol:** Below 100 mg/dL
- **HDL (High-Density Lipoprotein) Cholesterol:** Above 40 mg/dL
- **Triglycerides:** Below 150 mg/dL



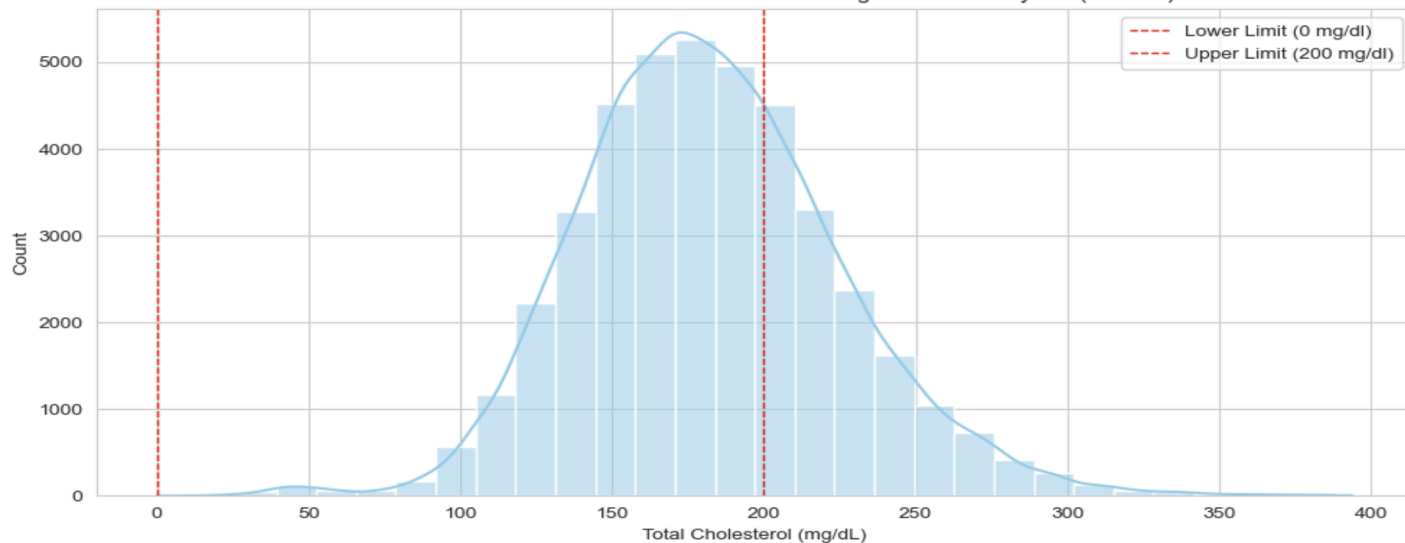


Cholesterol Distribution of Patients Above 19 Years(Female)

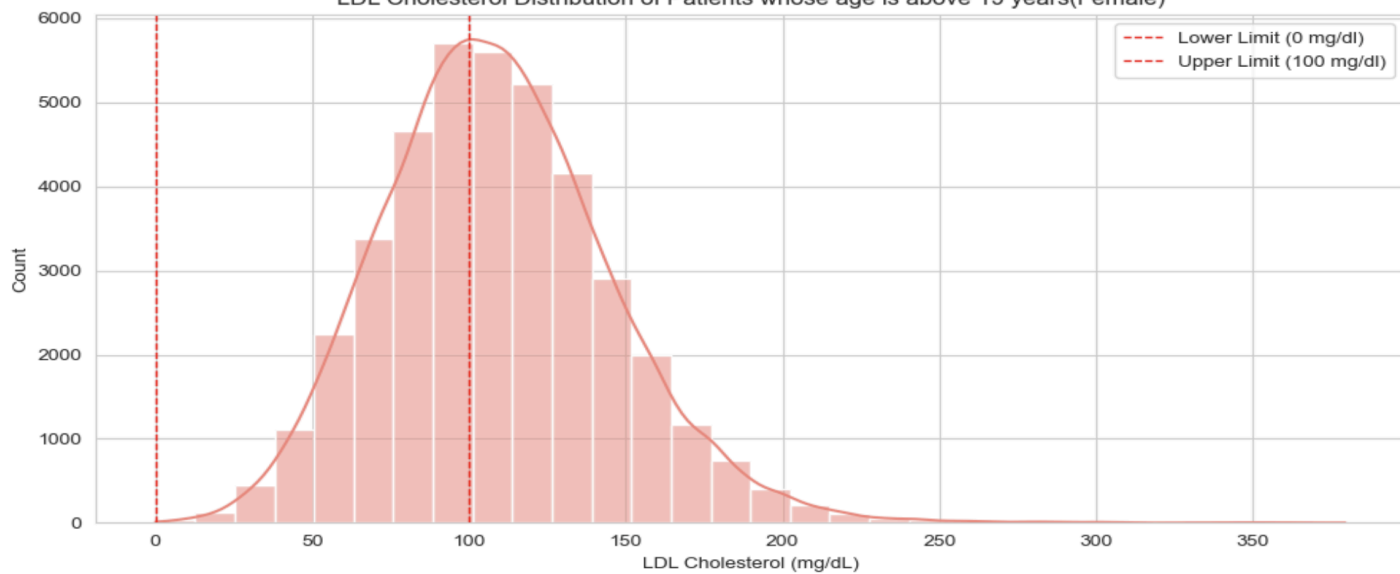
The general normal ranges of cholesterol parameters for patients above 19 years (Female) are as follows:

- **Total Cholesterol:** Below 200 mg/dL
- **LDL (Low-Density Lipoprotein) Cholesterol:** Below 100 mg/dL
- **HDL (High-Density Lipoprotein) Cholesterol:** Above 50 mg/dL
- **Triglycerides:** Below 150 mg/dL

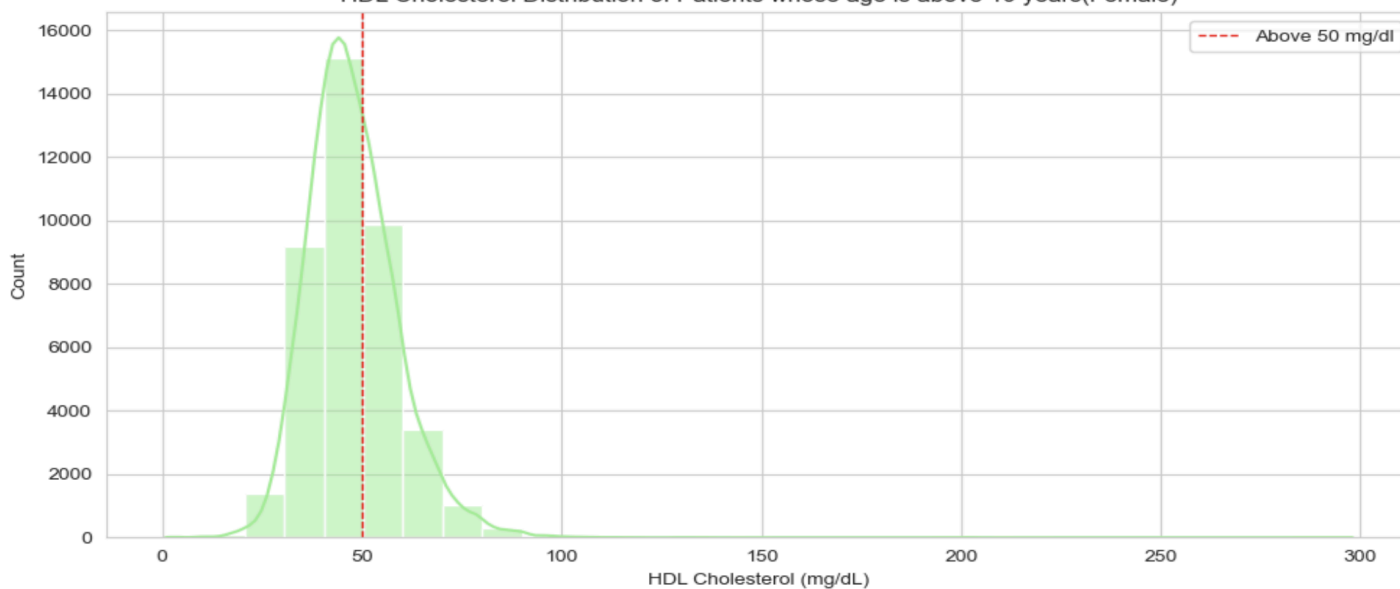
Total Cholesterol Distribution of Patients whose age is above 19 years(Female)

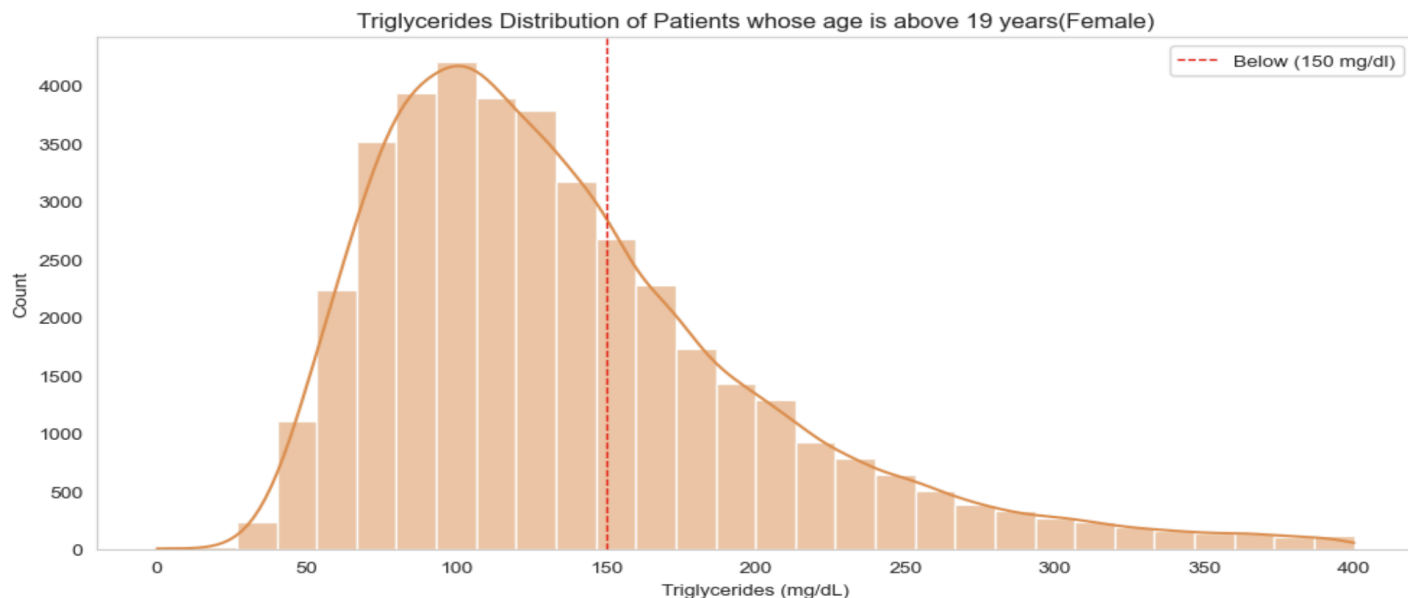


LDL Cholesterol Distribution of Patients whose age is above 19 years(Female)



HDL Cholesterol Distribution of Patients whose age is above 19 years(Female)





From the cholesterol analysis, it is evident that a significant proportion of patients have test values that lie at or near the borderline range. This indicates an elevated risk of developing cardiovascular diseases in the future if appropriate lifestyle modifications and medical interventions are not undertaken in a timely manner.

Ratio factors for cardiovascular health

Cholesterol ratio tests are valuable indicators of cardiovascular health. These ratios help assess the balance between different types of cholesterol and triglycerides in the body. Maintaining optimal levels can reduce the risk of heart disease, stroke, and other cardiovascular conditions.

1. Total Cholesterol to HDL Ratio (TC/HDL)

- **Ideal Ratio:** 3.5 or lower
- **Acceptable Ratio:** 3.5 to 5
- **High Risk:** Greater than 5

A lower TC/HDL ratio indicates a healthier balance between total cholesterol and "good" HDL cholesterol, which helps remove excess cholesterol from the bloodstream.

2. LDL to HDL Ratio (LDL/HDL)

- **Ideal Ratio:** Less than 2.5
- **Acceptable Ratio:** 2.5 to 3.5
- **High Risk:** Above 3.5

This ratio compares "bad" LDL cholesterol to "good" HDL cholesterol. Lower values suggest a lower risk of plaque buildup in the arteries.

3. HDL to LDL Ratio (HDL/LDL)

- **Ideal Ratio:** Above 1.0
- **Acceptable Ratio:** 0.5 to 1.0
- **High Risk:** Below 0.5

Higher HDL relative to LDL indicates better cardiovascular protection, as HDL helps clear LDL from the blood.

4. Triglyceride to HDL Ratio (TGL/HDL)

- **Ideal:** Below 2.0 (Lower risk of cardiovascular disease)
- **Borderline High:** 2.0 to 3.0 (Moderate risk)
- **High Risk:** Above 3.0 (Increased risk of cardiovascular disease)

This ratio is an important marker of insulin resistance and heart disease risk. Lower ratios are associated with better metabolic and heart health.

Classification of Lipid Ratios

To evaluate cardiovascular health, key cholesterol (Total cholesterol, LDL and HDL) and triglyceride ratios have been categorised into three risk levels: **Ideal**, **Acceptable**, and **High Risk**. These classifications help identify individuals who may be at risk for heart disease and related conditions.

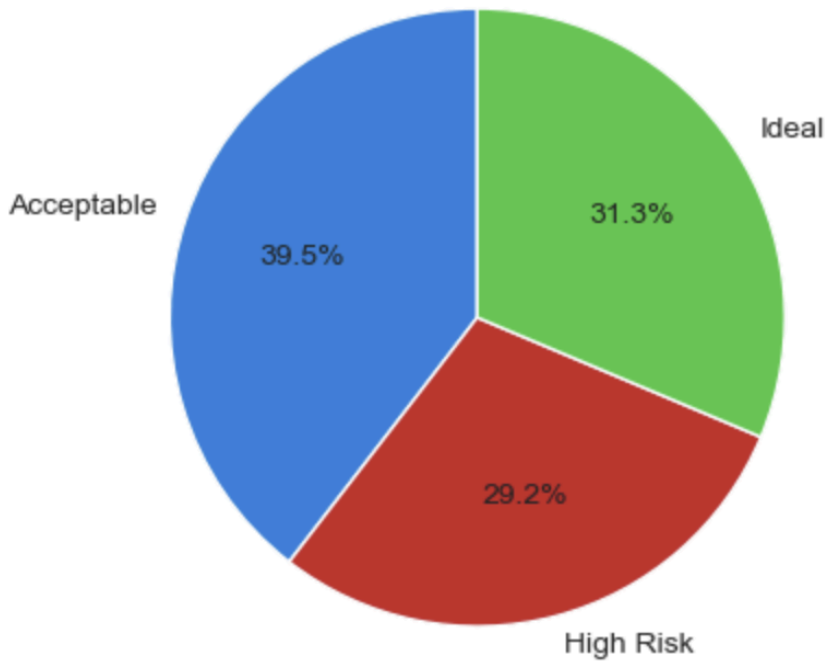
Each ratio is classified using specific threshold values:

- **Ideal:** Indicates a healthy balance and low risk of cardiovascular disease.
- **Acceptable:** Indicates moderate risk, often manageable with lifestyle changes.
- **High Risk:** Suggests an increased likelihood of cardiovascular issues.

These classifications are implemented in functions that evaluate a given ratio and return the corresponding risk category.

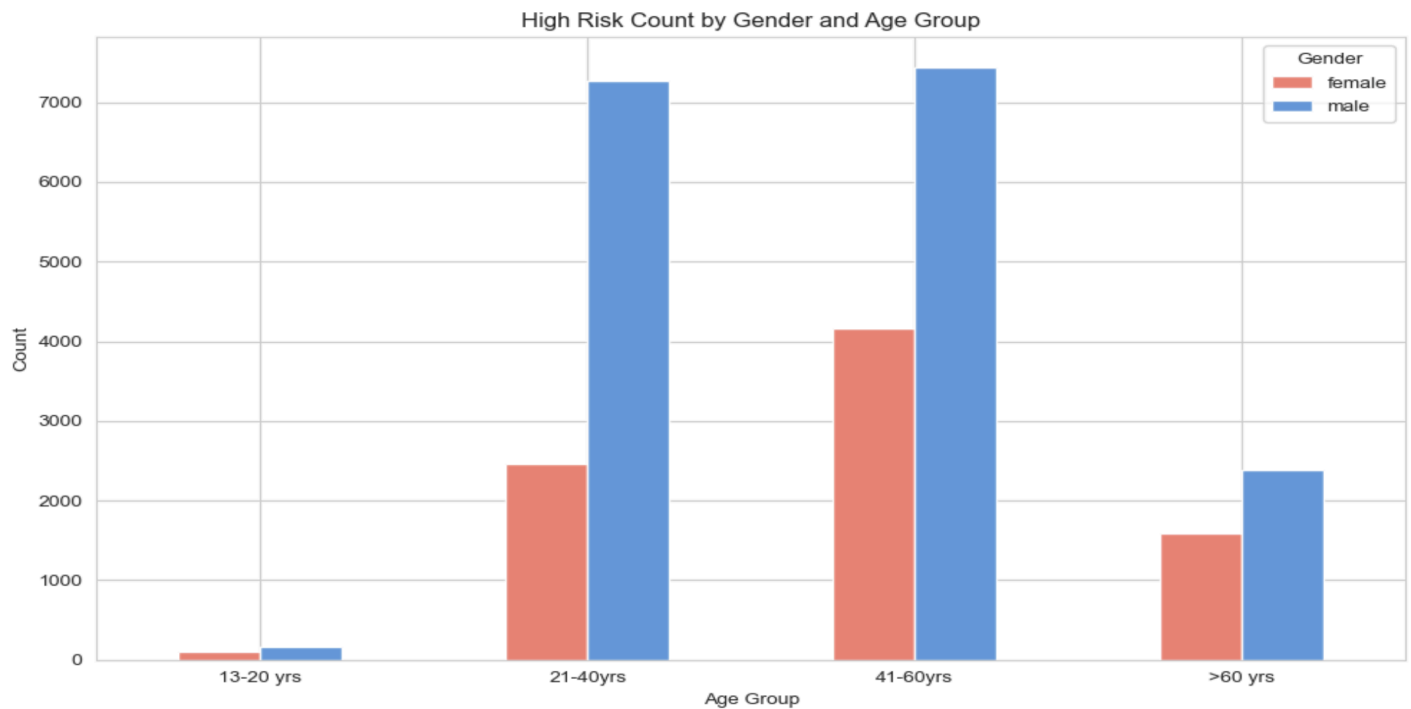
Below is the graph indicating the distribution of Cardiovascular Risk Factors in Patients.

Distribution of Cardiovascular Risk Factor



The figure shows that **29.2%** of patients are in the high-risk category, **39.2%** fall within the acceptable range, and **31.3%** are at the ideal level.

High-risk count of patients by age and gender



Fasting Glucose

Fasting Glucose refers to the level of glucose (sugar) in the blood after at least 8 hours of fasting. It's a simple and commonly used test to assess how well the body regulates blood sugar.

Why the Test is Performed:

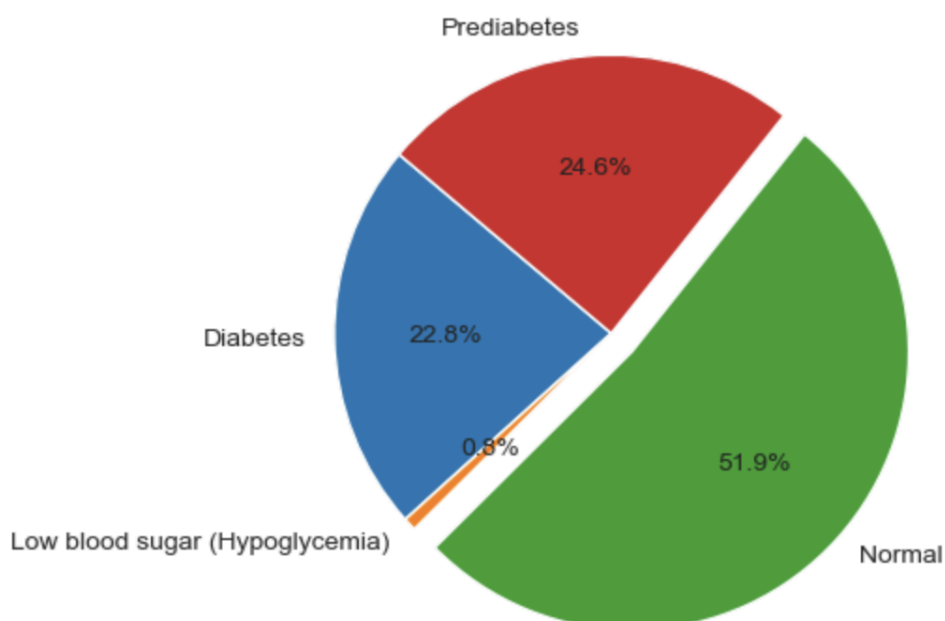
- To **screen** for diabetes or prediabetes
- To **monitor** glucose control in individuals with diabetes
- To **detect** early signs of insulin resistance or metabolic issues

Keeping fasting glucose within a healthy range is essential to prevent long-term complications such as heart disease, kidney problems, and nerve damage.

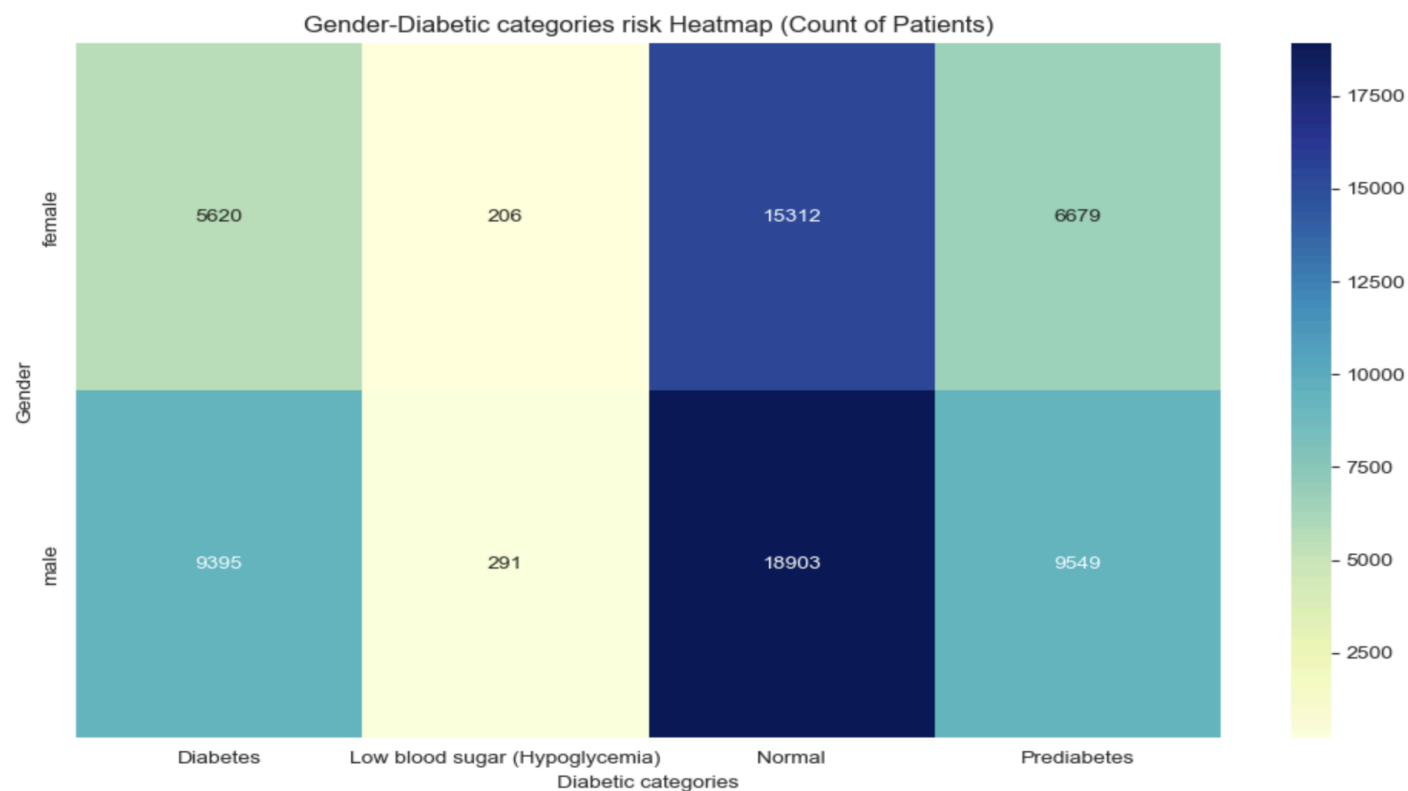
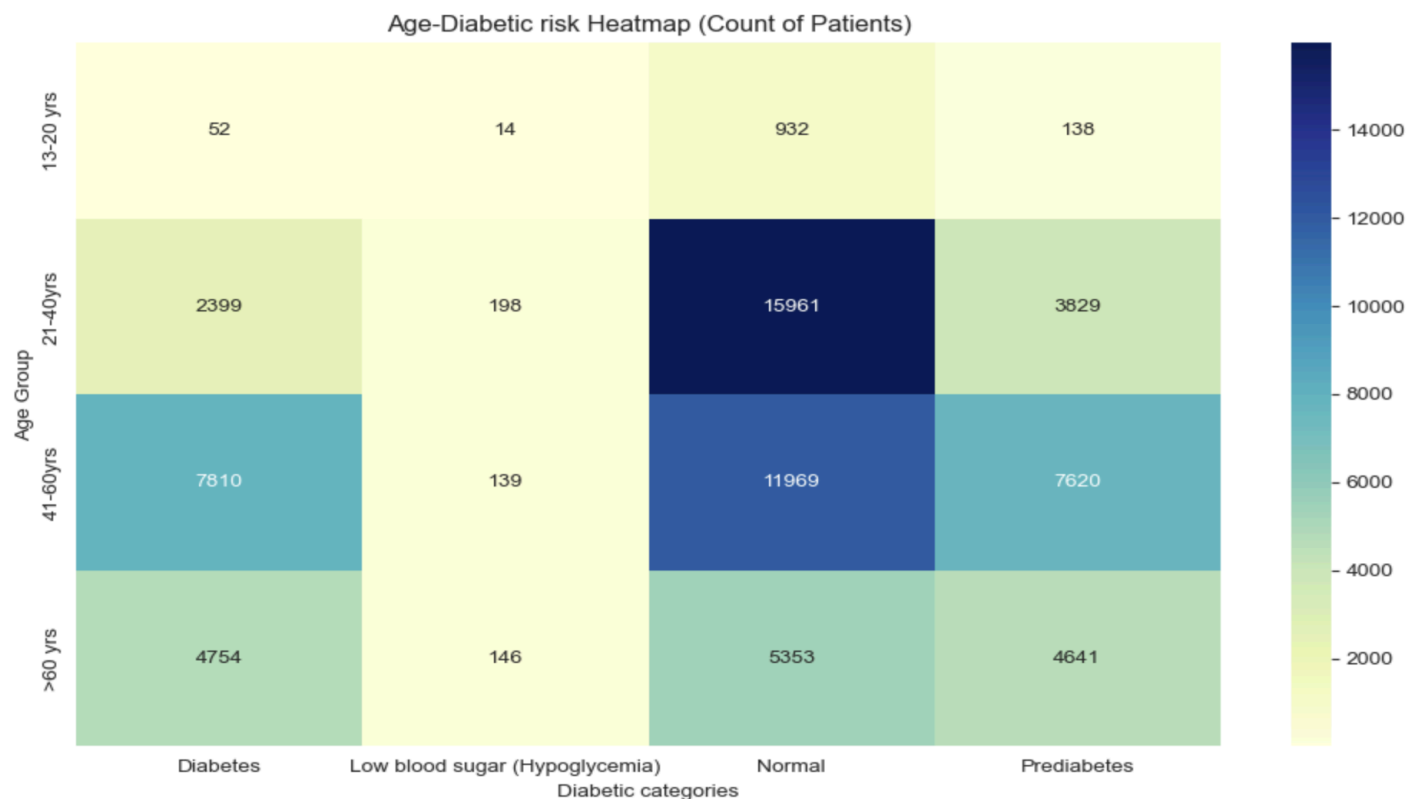
The usual range of values are

- **Low Blood Sugar (Hypoglycemia):** Less than 70 mg/dL
- **Normal:** 70–99 mg/dL
- **Prediabetes:** 100–125 mg/dL
- **Diabetes:** 126 mg/dL and above

The following figure presents the percentage of patients in each category:



Distribution of Diabetic categories by age and gender



The two heatmaps indicate that the highest number of patients with diabetes fall within the age group of 41–60 years, and that, overall, a majority of diabetic patients are male.

Conclusion

The analysis of the **NidaanKosha-100k-V1.0** dataset provides critical insights into the diagnostic profiles of patients across India, with a particular focus on cardiac and diabetic health indicators. A significant proportion of patients were found to have cholesterol and glucose values that lie at or near borderline thresholds. These patterns highlight a considerable risk of cardiovascular disease and diabetes within the population, especially among individuals aged 41–60 years, with a higher prevalence observed in males.

The distribution of lipid ratios further emphasises that approximately 29.2% of patients fall in the high-risk category for cardiovascular conditions, while 39.2% lie within an acceptable range. Similarly, the fasting glucose analysis indicates that a notable portion of the population is either prediabetic or diabetic, underscoring the importance of early intervention.

These findings underscore the urgent need for widespread awareness, timely medical screening, and preventive healthcare measures. If leveraged appropriately, this dataset can play a pivotal role in driving targeted health interventions, improving clinical decision support systems, and guiding policy frameworks aimed at mitigating chronic disease risks in India.