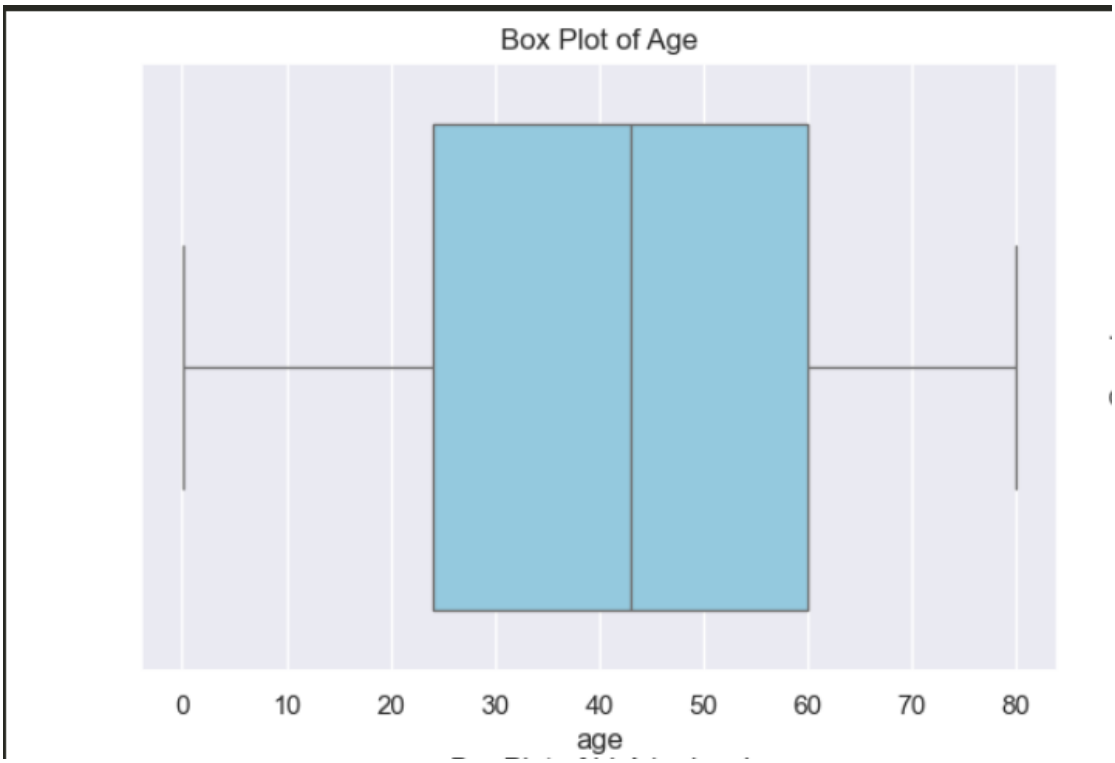
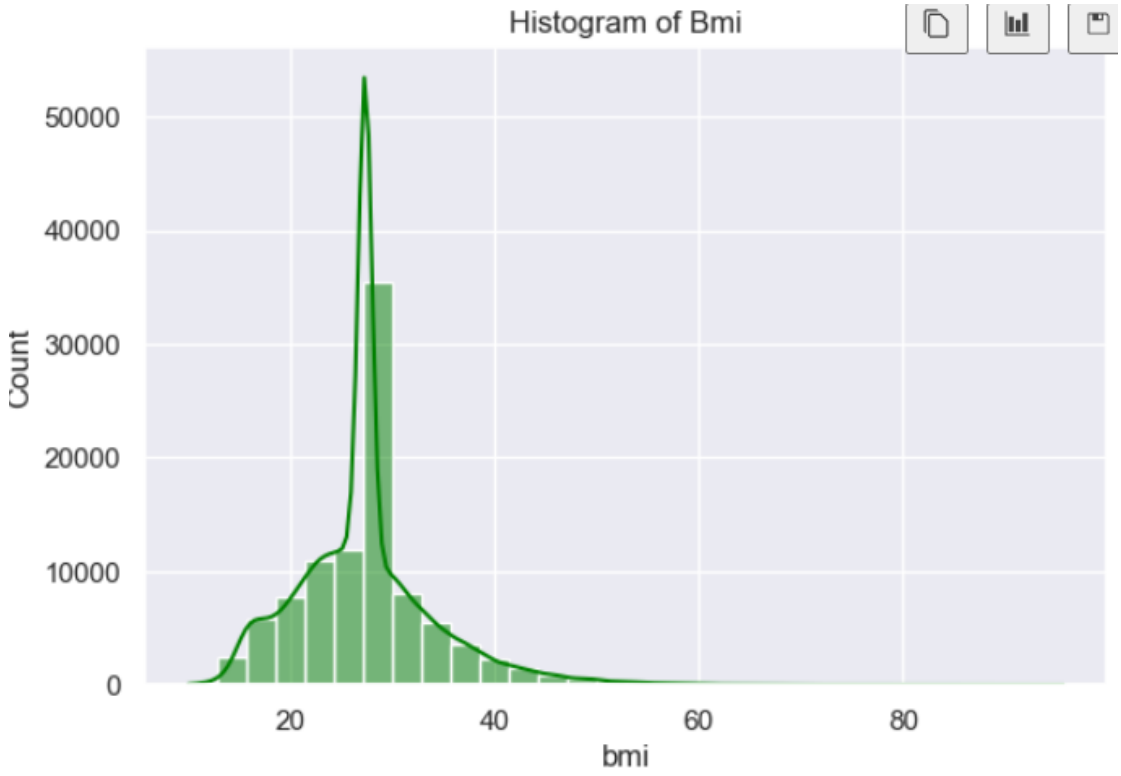
After Cleaning



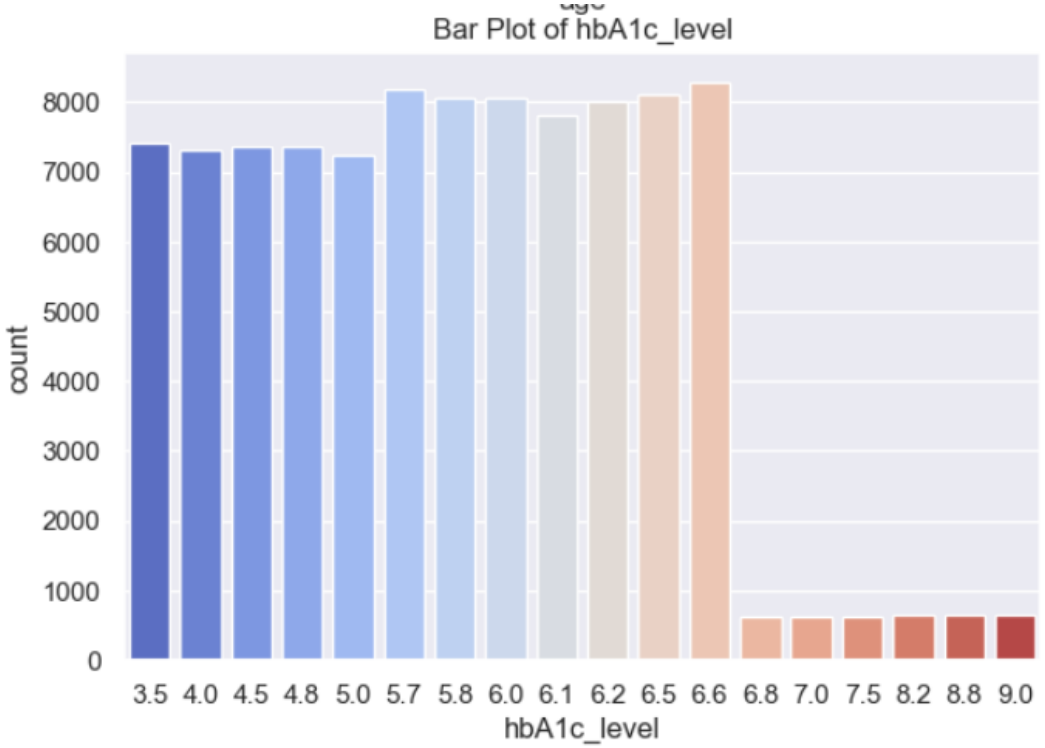
1. **Age Distribution**: The majority of individuals fall between 24 and 60 years old, indicating the dataset focuses primarily on adults.

2. **Outliers Present**: There are outlier values below 0 (e.g., 0.08 as seen in df.describe()), which are likely data errors since age cannot be negative or extremely low.

3. **Actionable Step**: Consider excluding or correcting ages below 1 year to improve data quality (e.g., df = df[df['age'] >= 1]).



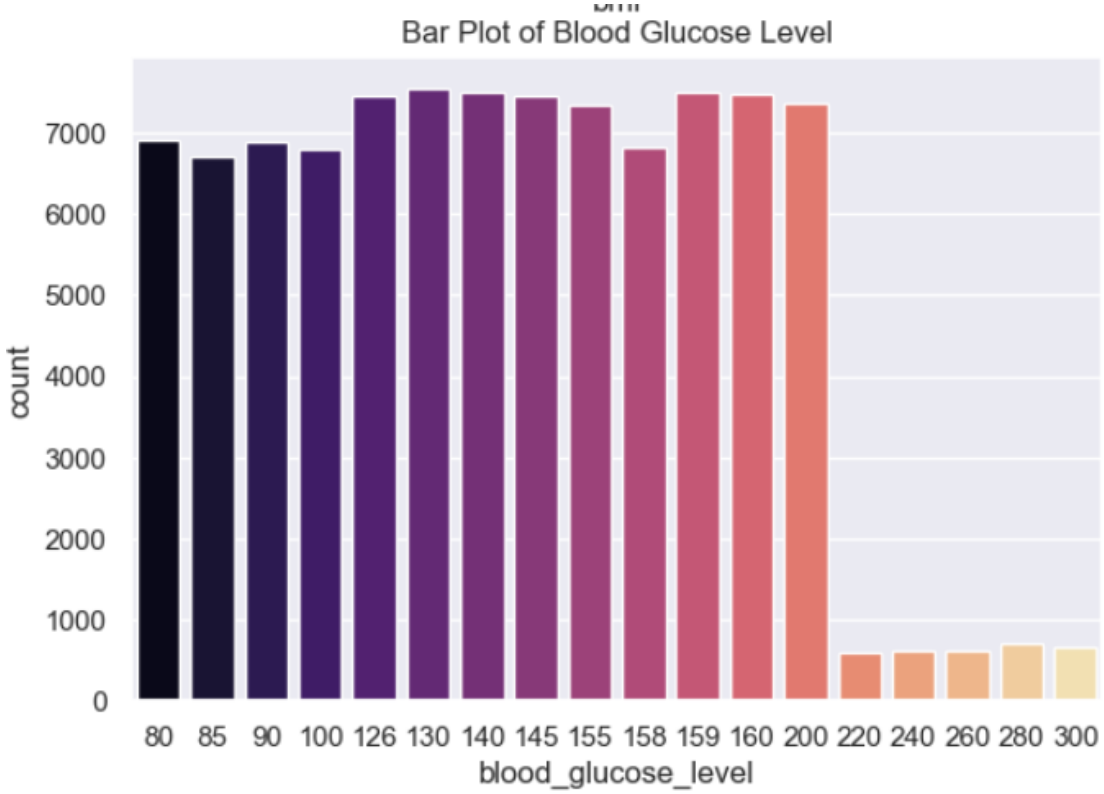
1. **BMI Distribution**: Most individuals have a BMI between 20 and 40, covering normal weight (18.5-24.9), overweight (25-29.9), and obese (30+) categories.
2. **Peak at 27-28**: The sharp peak suggests a large concentration of people in the overweight range, aligning with the mean BMI of 27.32 from df.describe().
3. **Long Tail**: Higher BMI values (60-80) are rare and may indicate outliers or data errors (e.g., maximum of 95.69).



1. **Normal Range Dominance**: The majority of values lie between 4.8 and 6.2, which is within the normal range (<5.7 normal, 5.7-6.4 prediabetes).

2. **Diabetes Indication**: Values above 6.5 (especially 6.5-9.0) decrease gradually, reflecting a lower proportion of diabetic individuals, consistent with the 8.49% diabetes rate in df.describe().

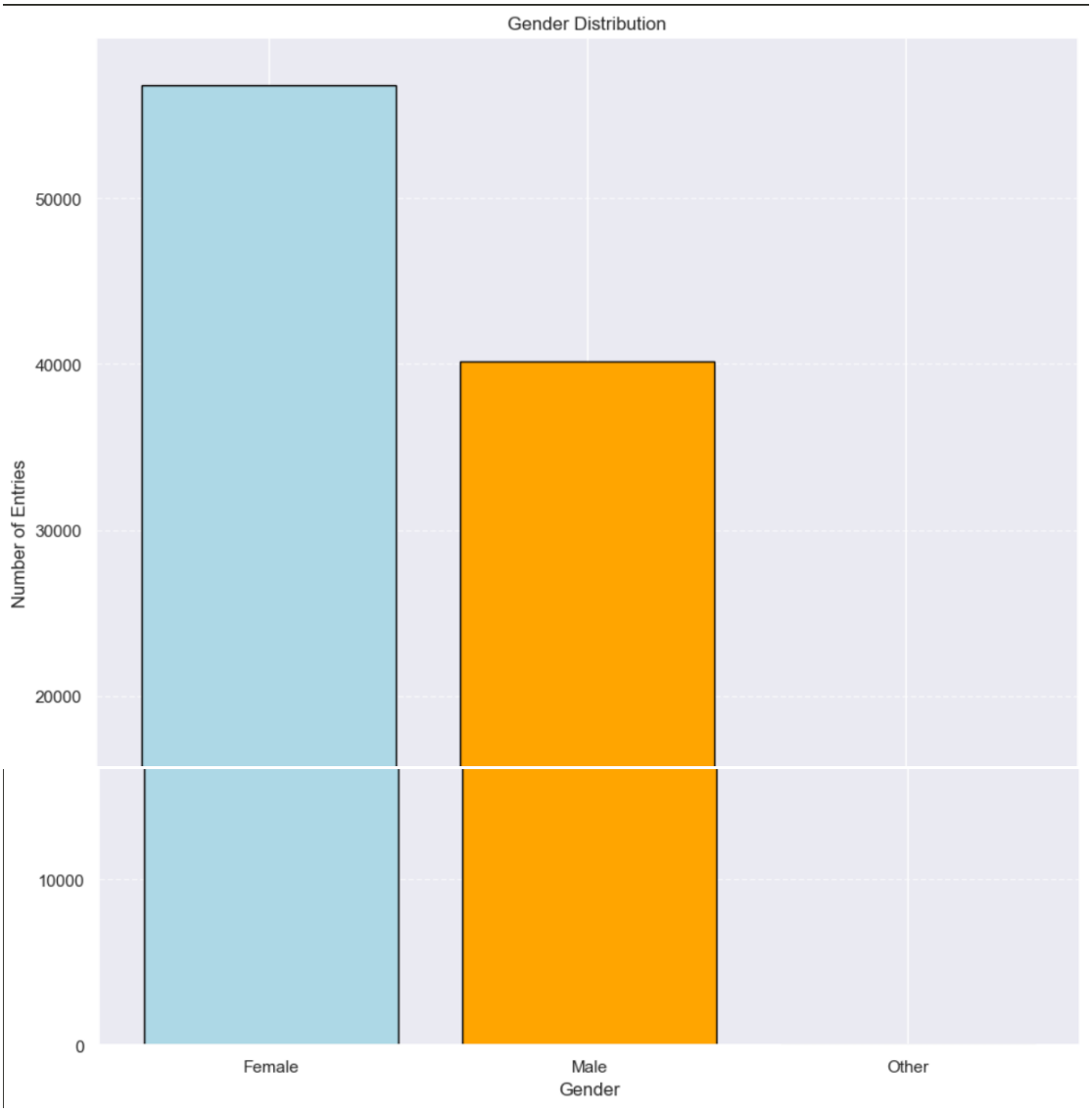
3. **Distribution Insight**: The concentration in the normal range suggests that most of the sample is not diabetic.



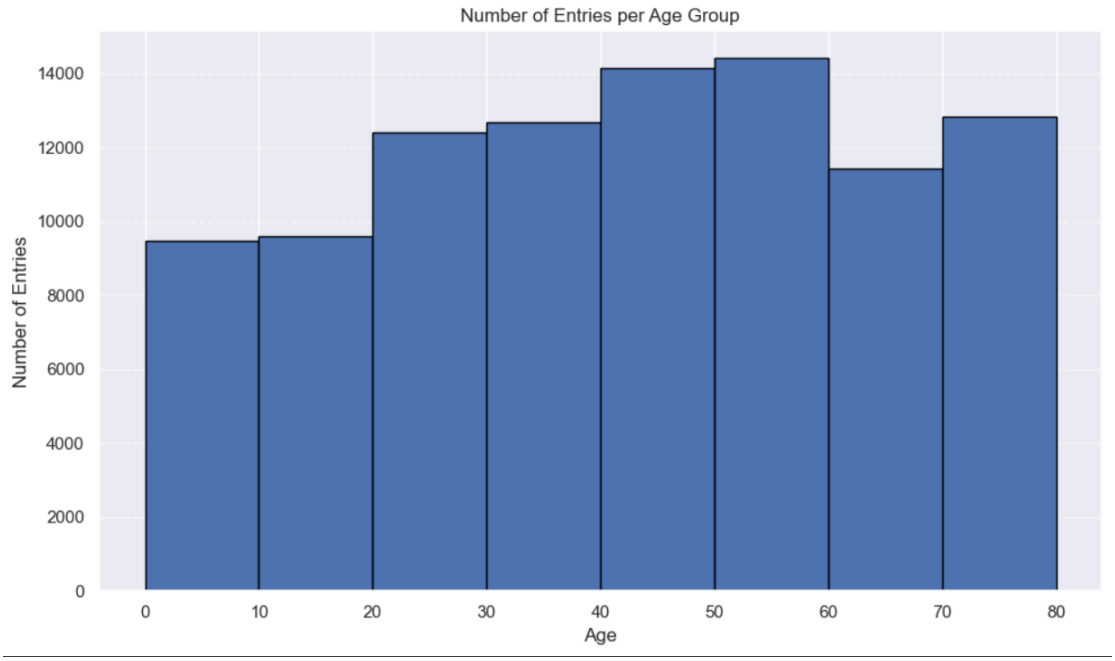
1. **Normal Range Dominance**: Most values fall between 80 and 126, which is within the normal fasting glucose range (<126 mg/dL), indicating most of the sample is not diabetic.

2. **Diabetes Indication**: Values above 200 (especially 200-300) drop significantly, suggesting a lower proportion of diabetic cases, aligning with the diabetes diagnosis threshold (>126 fasting or >200 random).

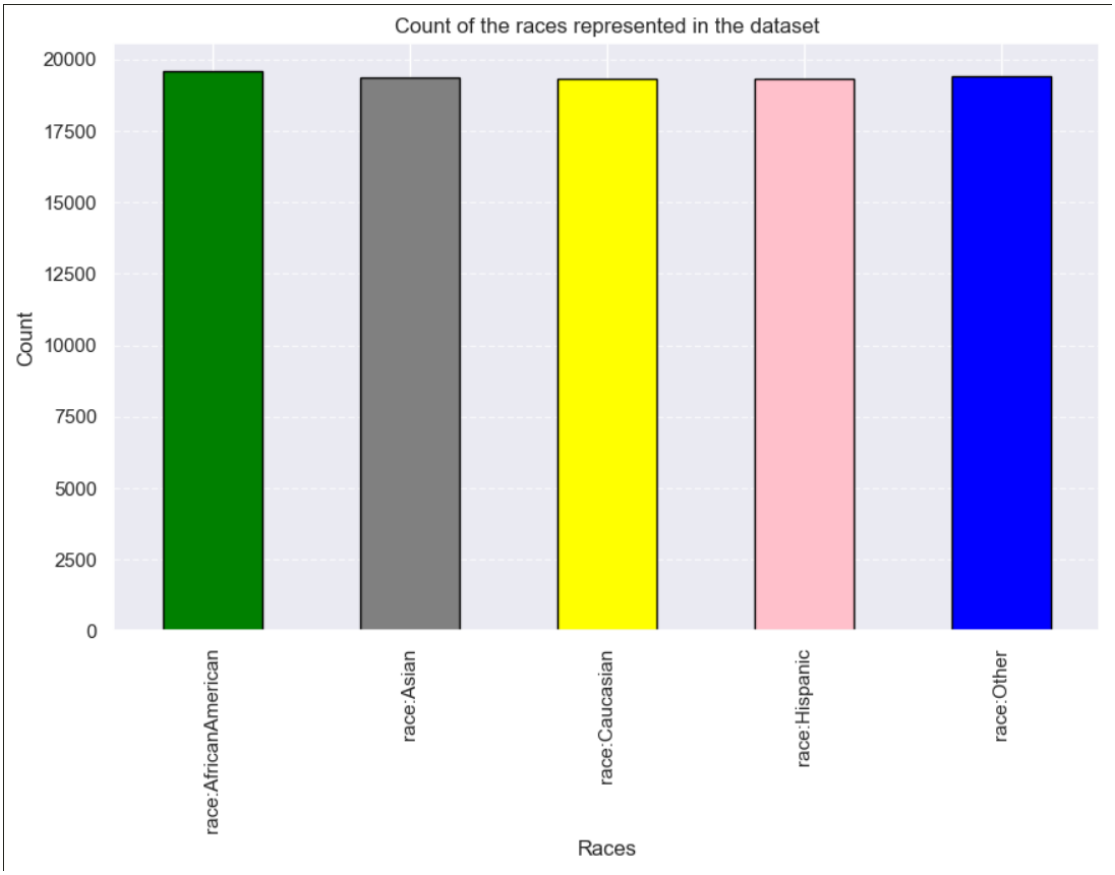
3. **Distribution Insight**: The concentration in the lower range indicates that diabetes is not highly prevalent in the sample.



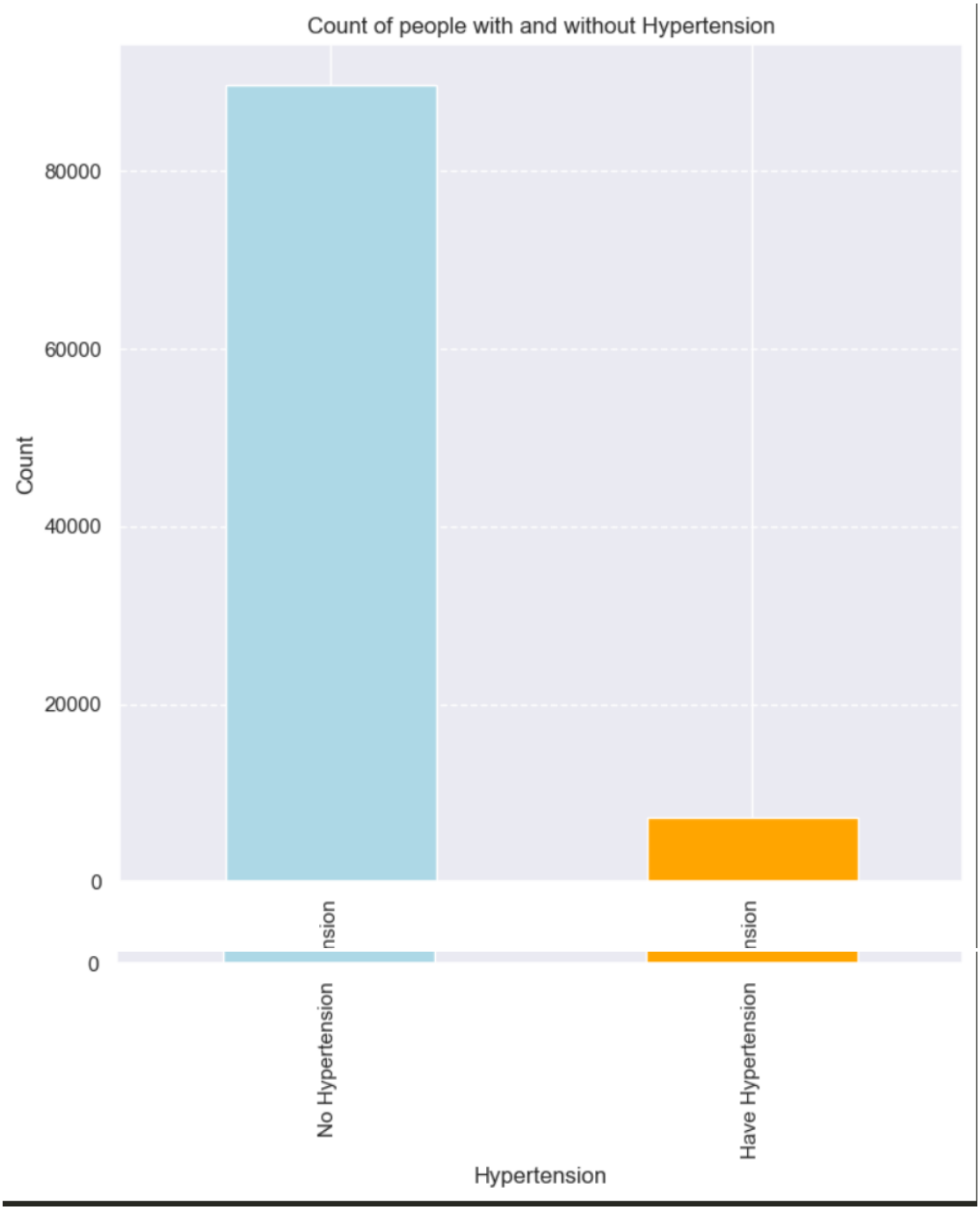
1. **Gender Proportions**:
    - The dataset contains a significant number of females (approximately 50,000 entries) and males (approximately 40,000 entries), indicating a nearly balanced gender distribution with a slight majority of females.
2. **Other Gender Category**:
    - The "Other" category has a negligible number of entries (close to 0), suggesting that this category is either not well-represented or contains missing/invalid data.
3. **Data Imbalance Consideration**:
    - The slight dominance of females (around 55-56% of the total) compared to males (around 44-45%) may influence analysis if gender has a differential impact on diabetes prevalence.



1. **Age Distribution Peak**:
    - The highest number of entries (around 14,000) is observed in the 40-50 age group, indicating this is the most represented age range in the dataset.
2. **Gradual Increase and Decrease**:
    - The number of entries increases steadily from age 0-10 (around 8,000) to a peak at 40-50, then decreases gradually through 50-80 (around 12,000 to 10,000), suggesting a bell-shaped distribution centered around middle age.
3. **Lower Representation at Extremes**:
    - Younger ages (0-20) and older ages (70-80) have fewer entries (around 8,000 to 10,000), indicating lower representation of children and the elderly.
4. **Alignment with Mean Age**:
    - The peak at 40-50 aligns with the mean age of 41.87 from df.describe(), reinforcing that the dataset is skewed toward middle-aged adults.



1. **Dominance of Non-Smokers**:
    - The "No Info" category is the largest, with around 35,000 entries, suggesting a significant portion of the dataset lacks smoking history information. The "Never" category follows with approximately 30,000 entries, indicating that non-smokers are the most represented group among those with recorded smoking history.
2. **Current and Former Smokers**:
    - "Current" smokers have around 15,000 entries, and "Former" smokers have around 10,000 entries, showing a moderate presence of smokers in the dataset.
3. **Other Categories**:
    - Categories like "Ever" and "Not Current" have much fewer entries (around 5,000 each), indicating they are less common or possibly less consistently recorded.
4. **Data Quality Concern**:
    - The high number of "No Info" entries (around 35% of the dataset) suggests missing or incomplete data for smoking history, which could impact the reliability of smoking-related analyses.



1. **Prevalence of No Hypertension**:

- Approximately 89,731 individuals do not have hypertension, making up the vast majority (around 92.5%) of the dataset, indicating that hypertension is relatively uncommon in this sample.

2. **Prevalence of Hypertension**:

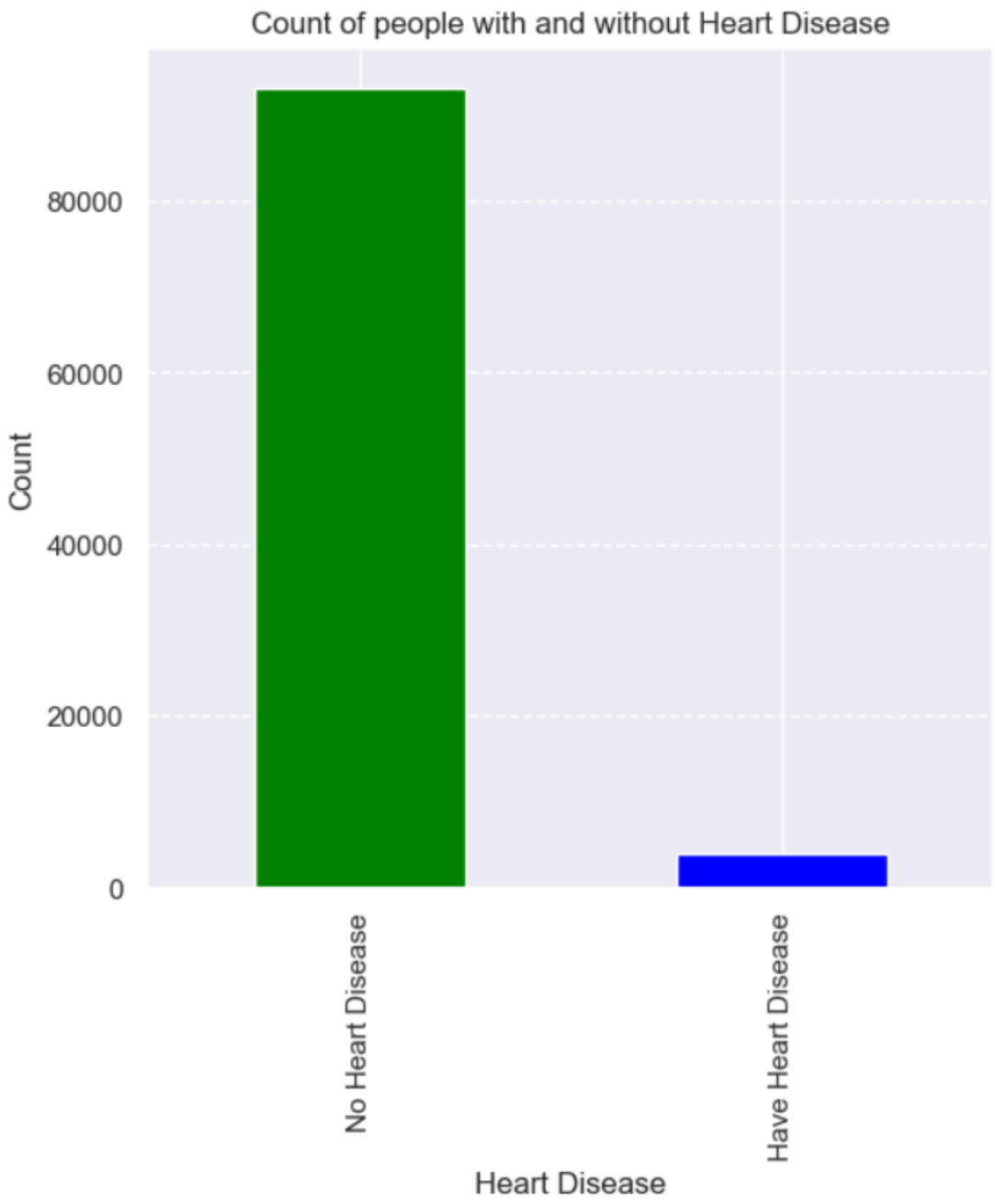
- Around 7,268 individuals have hypertension, representing about 7.5% of the dataset, which aligns with the 7.49% hypertension rate from df.describe().

3. **Imbalance in Distribution**:

- The significant disparity between those with and without hypertension suggests that hypertension is a minority condition in this dataset, which could affect its predictive power in models unless balanced.

4. **Health Insight**:

- The low prevalence of hypertension (7.5%) compared to the diabetes rate (8.49%) suggests that while both conditions are present, hypertension may not be as dominant a comorbidity in this population.



1. **Prevalence of No Heart Disease**:

    - Approximately 93,174 individuals do not have heart disease, making up the vast majority (around 92.7%) of the dataset, indicating that heart disease is relatively uncommon in this sample.

2. **Prevalence of Heart Disease**:

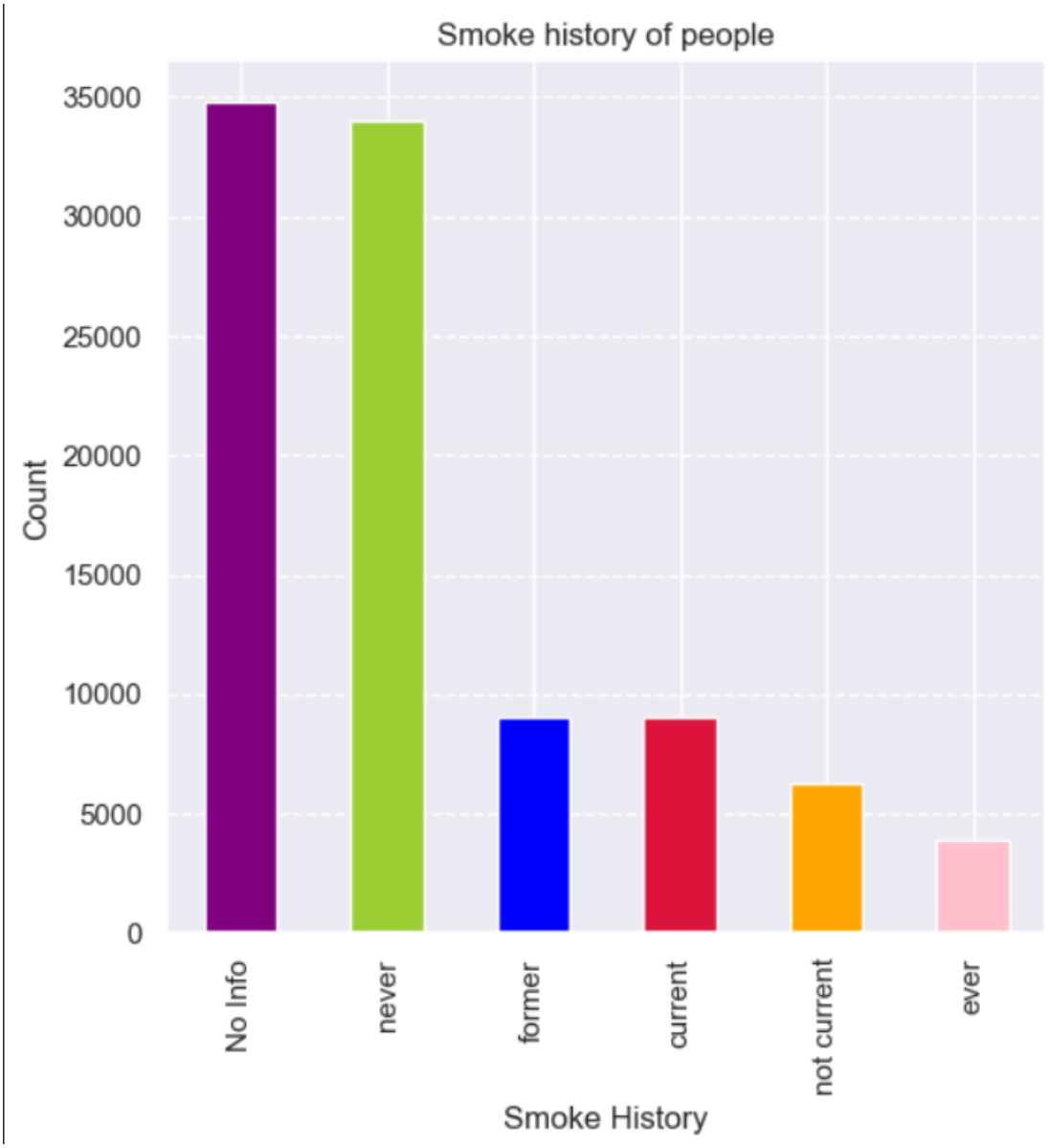
    - Around 3,825 individuals have heart disease, representing about 3.8% of the dataset, which aligns with the 3.94% heart disease rate from df.describe().

3. **Imbalance in Distribution**:

    - The significant disparity between those with and without heart disease suggests that heart disease is a minority condition in this dataset, which could affect its predictive power in models unless balanced.

4. **Health Insight**:

    - The low prevalence of heart disease (3.8%) compared to diabetes (8.49%) and hypertension (7.49%) suggests that heart disease is the least common condition among these comorbidities in this population.



1. **Dominance of Missing Data**:

    - The "No Info" category has the highest count, around 34,777 individuals (approximately 35% of the dataset), indicating a significant amount of missing smoking history data.

2. **Largest Recorded Group - Never Smokers**:

    - The "Never" category follows with approximately 30,940 individuals (around 31%), suggesting that non-smokers are the most represented group among those with recorded smoking history.

3. **Moderate Smoker Categories**:

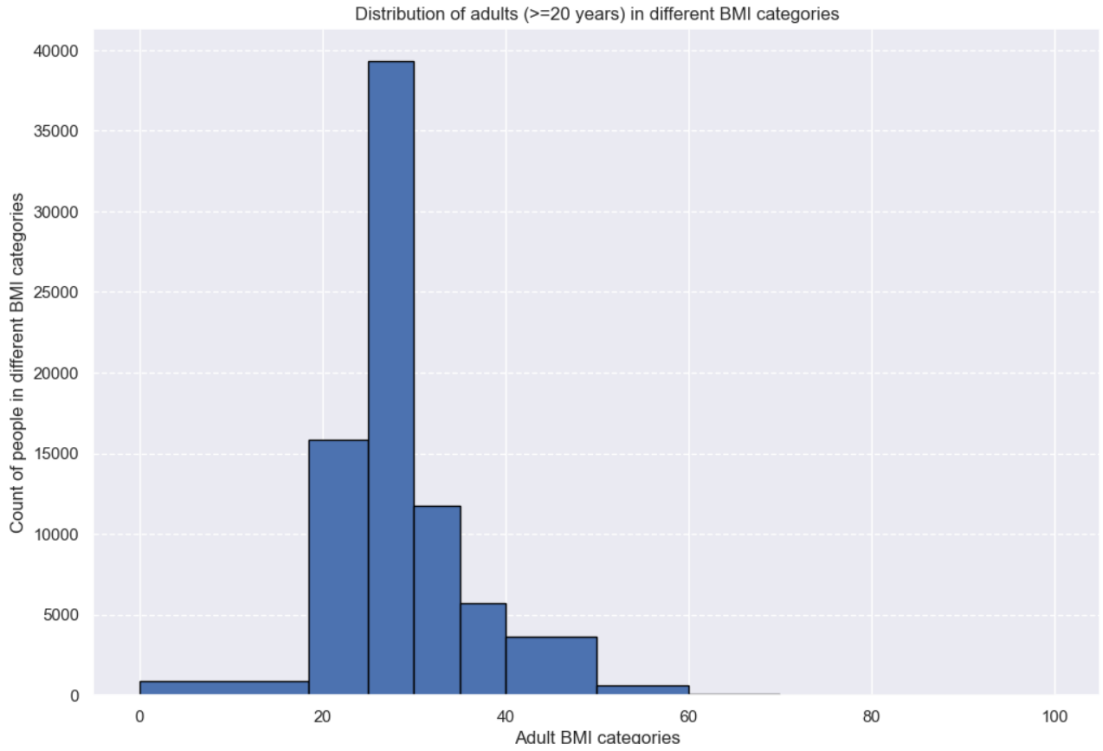
    - "Current" smokers have around 9,998 individuals (about 10%), and "Former" smokers have around 9,952 individuals (about 10%), showing a moderate presence of current and past smokers.

4. **Smaller Categories**:

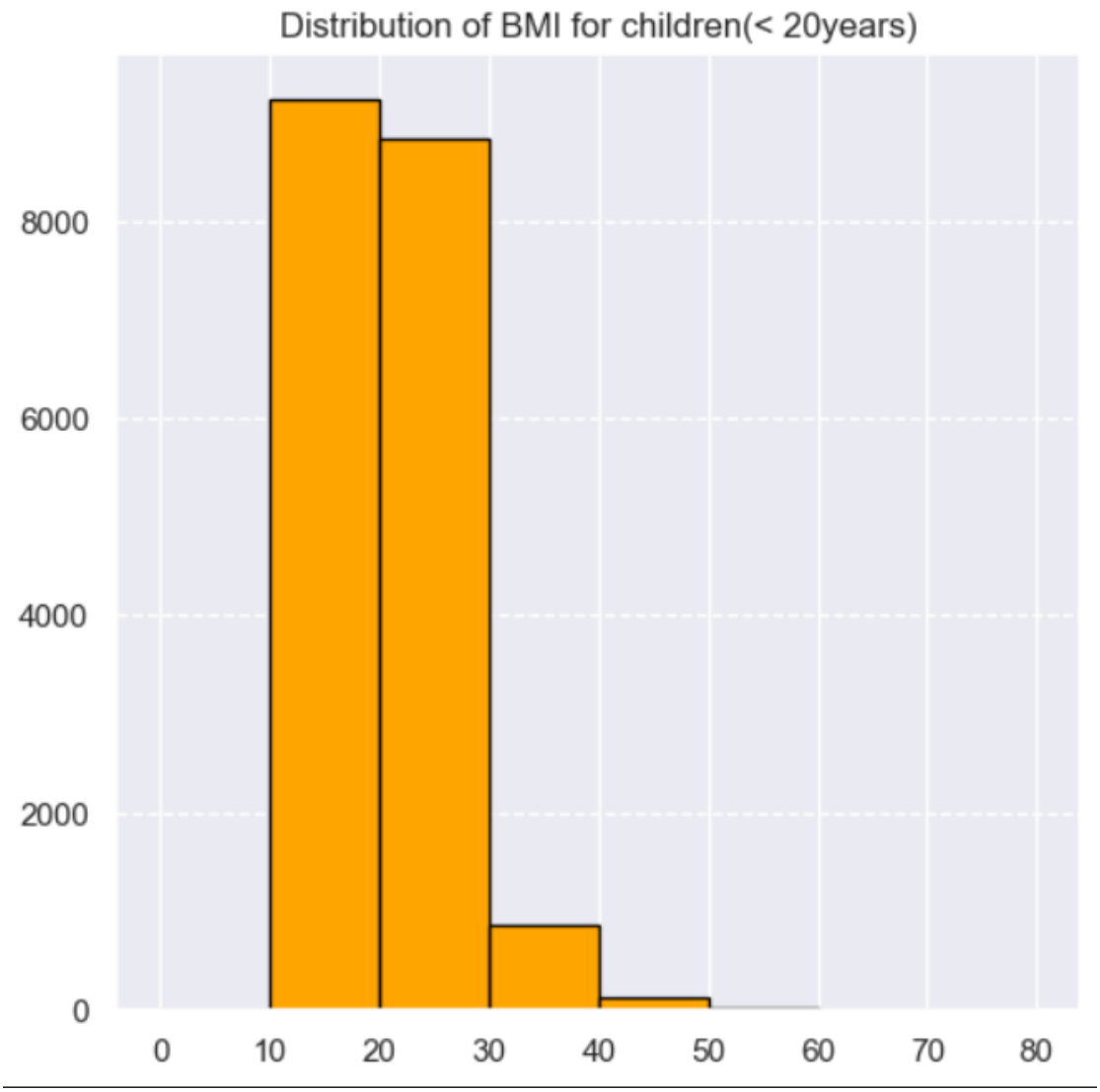
    - "Not Current" smokers have around 6,250 individuals (about 6%), and "Ever" smokers have around 3,882 individuals (about 4%), indicating these groups are less common or less consistently recorded.

5. **Data Quality Concern**:

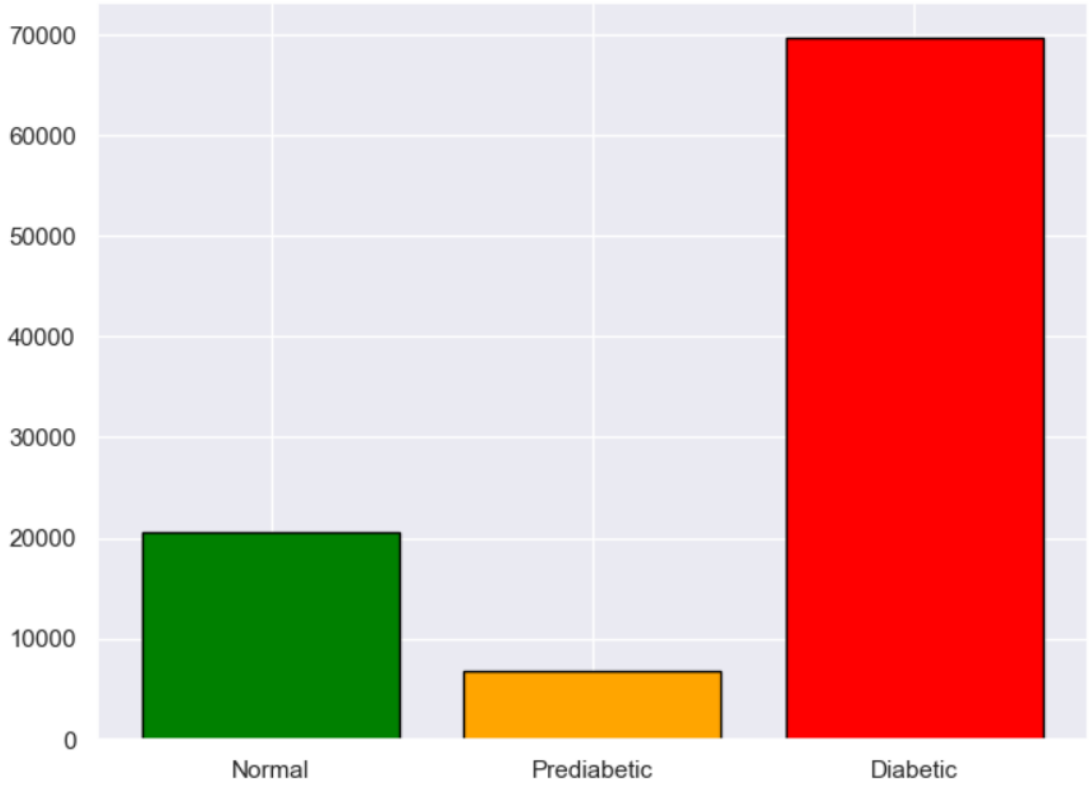
    - The high proportion of "No Info" entries (35%) suggests potential data collection issues, which could limit the reliability of smoking-related analyses.



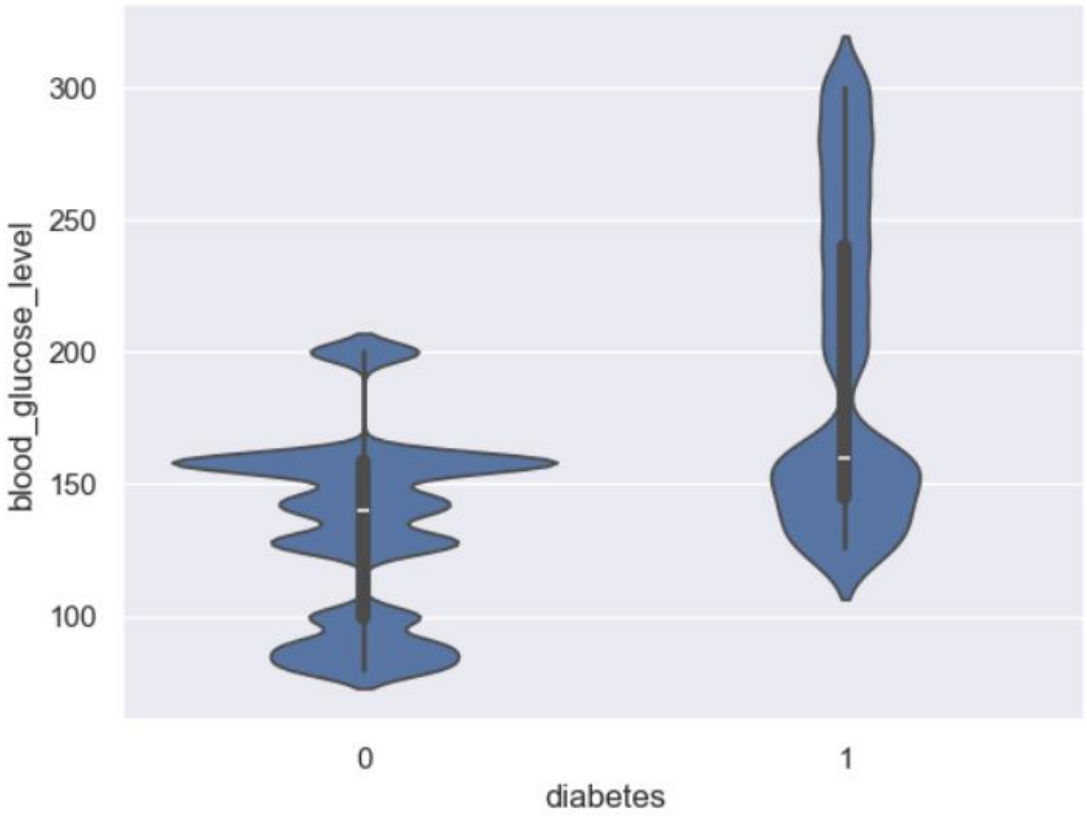
1. **Prevalence of Non-Diabetics**:
    - Approximately 91,500 individuals do not have diabetes, making up the vast majority (around 91.5%) of the dataset, indicating that diabetes is not highly prevalent in this sample.
2. **Prevalence of Diabetics**:
    - Around 8,500 individuals have diabetes, representing about 8.5% of the dataset, which aligns closely with the 8.49% diabetes rate from df.describe().
3. **Imbalance in Distribution**:
    - The significant disparity between those with and without diabetes (91.5% vs. 8.5%) suggests that diabetes is a minority condition in this dataset, which could affect predictive modeling unless balanced.
4. **Health Insight**:
    - The 8.5% diabetes prevalence is slightly higher than hypertension (7.49%) and heart disease (3.94%) in the dataset, indicating diabetes is the most common among these conditions in this population.



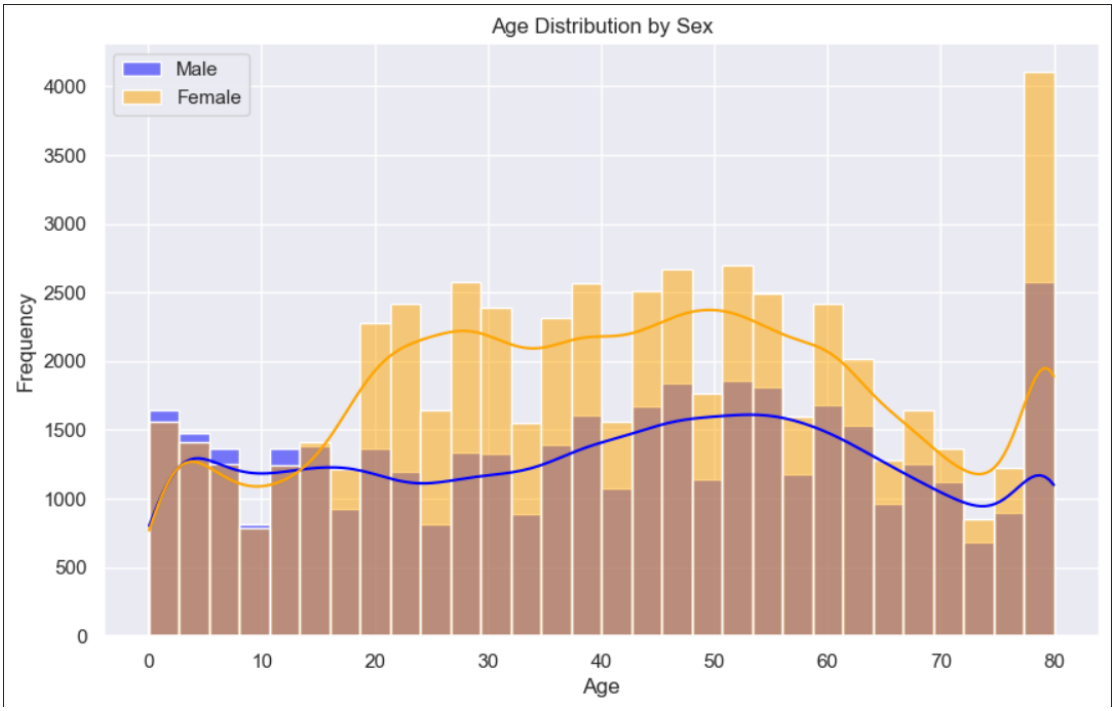
1. **Peak BMI Range**:
    - The highest number of children (around 8,000-9,000) falls within the 0-10 BMI range, indicating that the majority of children in the dataset have a low BMI, likely reflecting normal or underweight categories for this age group.
2. **Decline with Increasing BMI**:
    - The count decreases sharply beyond 10 BMI, with around 2,000-3,000 children in the 10-20 range, and further drops to near zero beyond 30, suggesting very few children have higher BMI values (e.g., overweight or obese categories).
3. **Limited High BMI Representation**:
    - BMI values above 30 are extremely rare, with counts dropping to negligible levels, indicating that obesity is uncommon among children under 20 in this dataset.
4. **Health Insight**:
    - The concentration of BMI below 20 aligns with expected norms for children, where higher BMI values are less common due to growth patterns and lower prevalence of obesity in younger populations compared to adults.



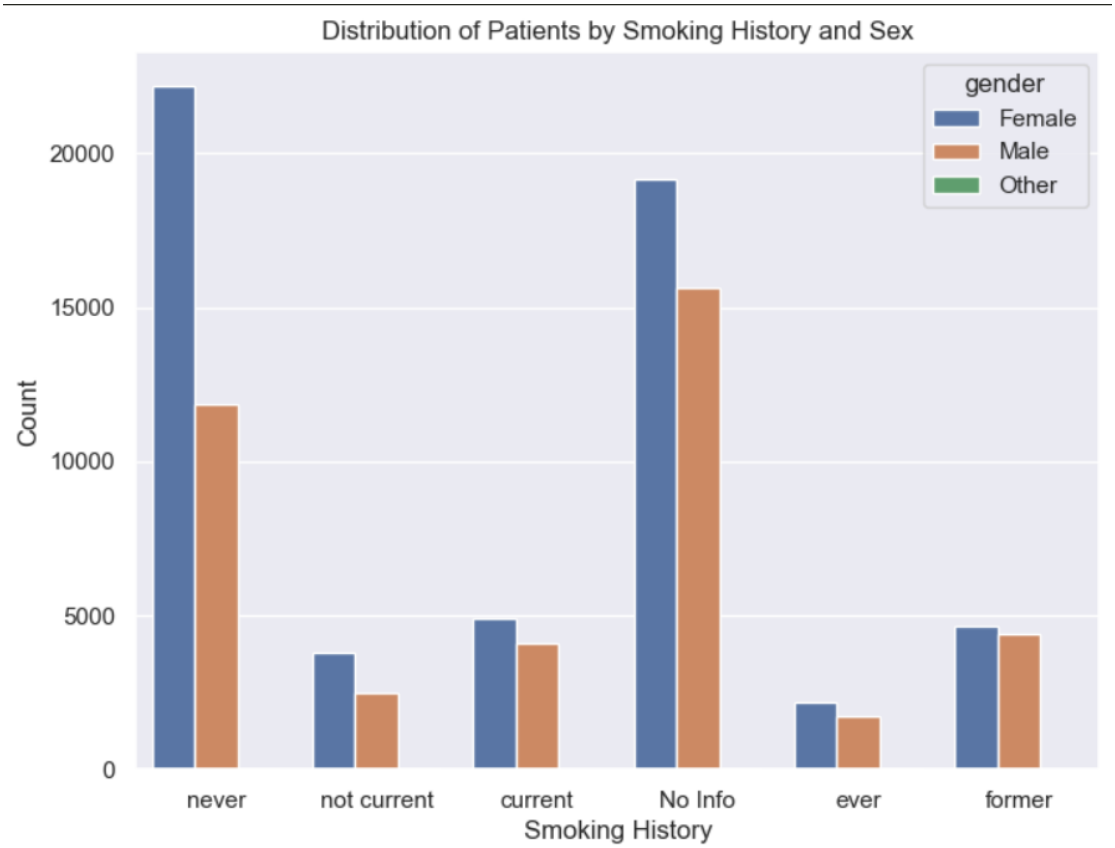
1. **Dominance of Normal Levels**:
    - Approximately 20,513 individuals have normal blood glucose levels (<100 mg/dL), making up about 20.5% of the dataset, indicating that the majority of the population falls within the normal range.
2. **Prevalence of Prediabetic Levels**:
    - Around 6,799 individuals are in the prediabetic range (100-125 mg/dL), representing about 6.8% of the dataset, suggesting a moderate presence of prediabetes.
3. **Prevalence of Diabetic Levels**:
    - Approximately 69,687 individuals have diabetic blood glucose levels (≥125 mg/dL), making up about 69.7% of the dataset, indicating a significant proportion of the population with elevated glucose levels.
4. **Imbalance in Distribution**:
    - The dataset is heavily skewed toward diabetic levels (69.7%), with a much smaller proportion in the normal (20.5%) and prediabetic (6.8%) categories, which could reflect a biased sample or a high prevalence of diabetes in the studied population.
5. **Health Insight**:
    - The high proportion of diabetic blood glucose levels (69.7%) contrasts with the earlier reported diabetes prevalence of 8.5%, suggesting that many individuals with high glucose levels may not yet be diagnosed as diabetic, highlighting a potential public health concern for undiagnosed cases.



1. **Distribution for Non-Diabetics (diabetes = 0)**:
    - The blood glucose levels for individuals without diabetes (0) are tightly clustered around 100-150 mg/dL, with a narrow and symmetrical distribution, indicating most non-diabetics have glucose levels in the normal to slightly elevated range.
2. **Distribution for Diabetics (diabetes = 1)**:
    - The blood glucose levels for individuals with diabetes (1) are much more spread out, with a peak around 200-250 mg/dL and a wider distribution extending up to 300 mg/dL, reflecting higher and more variable glucose levels among diabetics.
3. **Median Differences**:
    - The median blood glucose level for non-diabetics is lower (around 100-120 mg/dL), while for diabetics, it is significantly higher (around 200-220 mg/dL), highlighting a clear distinction between the two groups.
4. **Variability Insight**:
    - Diabetics exhibit greater variability in blood glucose levels compared to non-diabetics, suggesting less control or more severe cases among diagnosed individuals.
5. **Health Insight**:
    - The shift toward higher glucose levels in diabetics aligns with the earlier finding of 69.7% of the dataset having diabetic blood glucose levels (≥125 mg/dL), reinforcing the prevalence of elevated glucose in this population.



1. **Overall Age Distribution**:
    - The age distribution peaks between 20-50 years for both genders, with a noticeable decline after 60 years, and a sharp increase again around 80 years, suggesting a bimodal distribution with a significant elderly population.
2. **Gender Differences**:
    - Females (orange) outnumber males (blue) across most age groups, with the most pronounced difference in the 80+ age range, where the frequency exceeds 4,000 for females compared to a much lower count for males.
3. **Younger Age Groups (0-20)**:
    - The 0-20 age range has a moderate frequency (around 1,000-1,500), with males and females showing similar distributions, indicating a balanced representation of young individuals.
4. **Middle Age Groups (20-60)**:
    - The 20-60 age range shows a higher frequency for females (up to 2,500-3,000) compared to males (up to 1,500-2,000), suggesting a larger female population in the working-age and middle-age categories.
5. **Elderly Age Groups (70-80+)**:
    - There is a significant spike in the 80+ age group for females (over 4,000), while males show a much smaller increase, indicating a higher survival rate or representation of elderly females in the dataset.
6. **Health Insight**:
    - The higher proportion of females, especially in older age groups, may influence the prevalence of conditions like diabetes, hypertension, or heart disease, warranting gender-specific analysis.



1. **Dominance of "Never" Smokers**:

    - The "never" category has the highest count, with approximately 20,000 females and 10,000 males, indicating that the majority of both genders have never smoked, with females outnumbering males significantly.

2. **High "No Info" Prevalence**:

    - The "No Info" category shows a large number of individuals, around 20,000 females and 15,000 males, suggesting a substantial amount of missing smoking history data, with females again being more represented.

3. **Moderate "Current" Smokers**:

    - The "current" category has around 5,000 females and 5,000 males, showing a balanced but lower representation of current smokers compared to "never" and "No Info."

4. **Lower Representation in Other Categories**:

    - Categories like "not current," "ever," and "former" have much lower counts (around 2,000-5,000 for each gender), with females consistently outnumbering males in these groups as well.

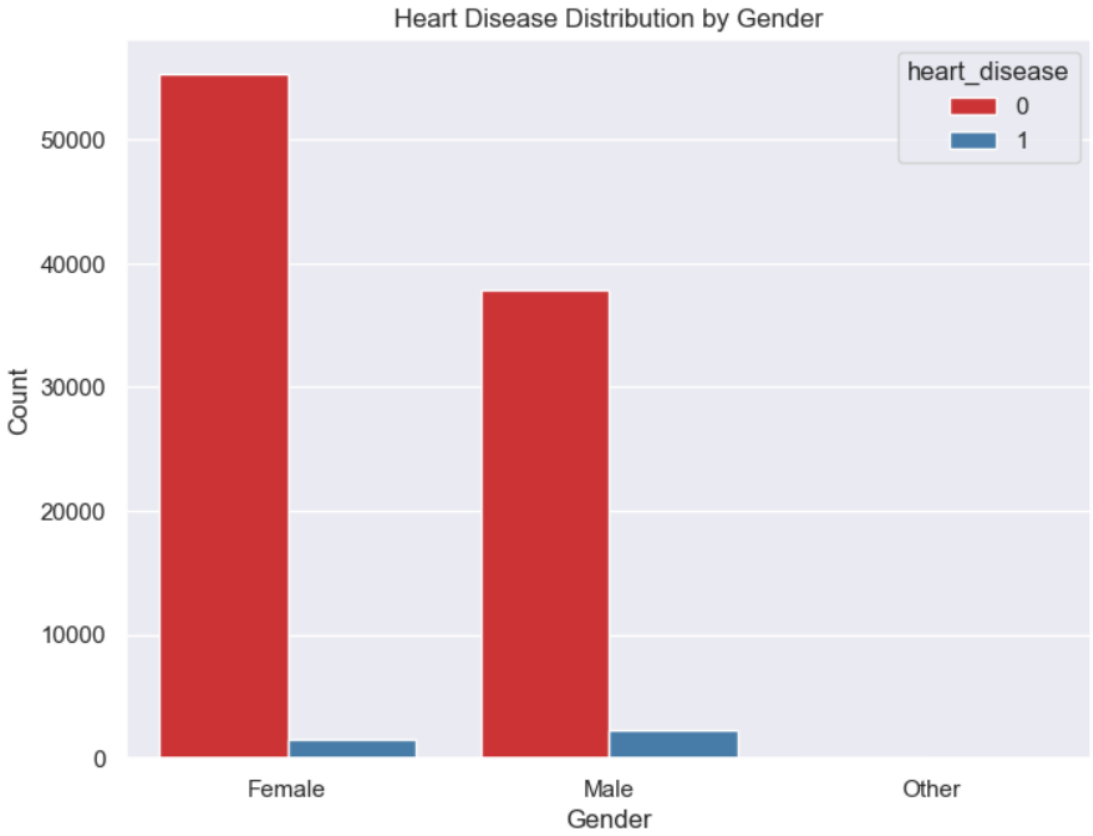
5. **Gender Distribution Trend**:

    - Across all smoking history categories, females (blue) outnumber males (orange), with the most significant gap in the "never" and "No Info" categories, suggesting a higher female representation in the dataset overall.

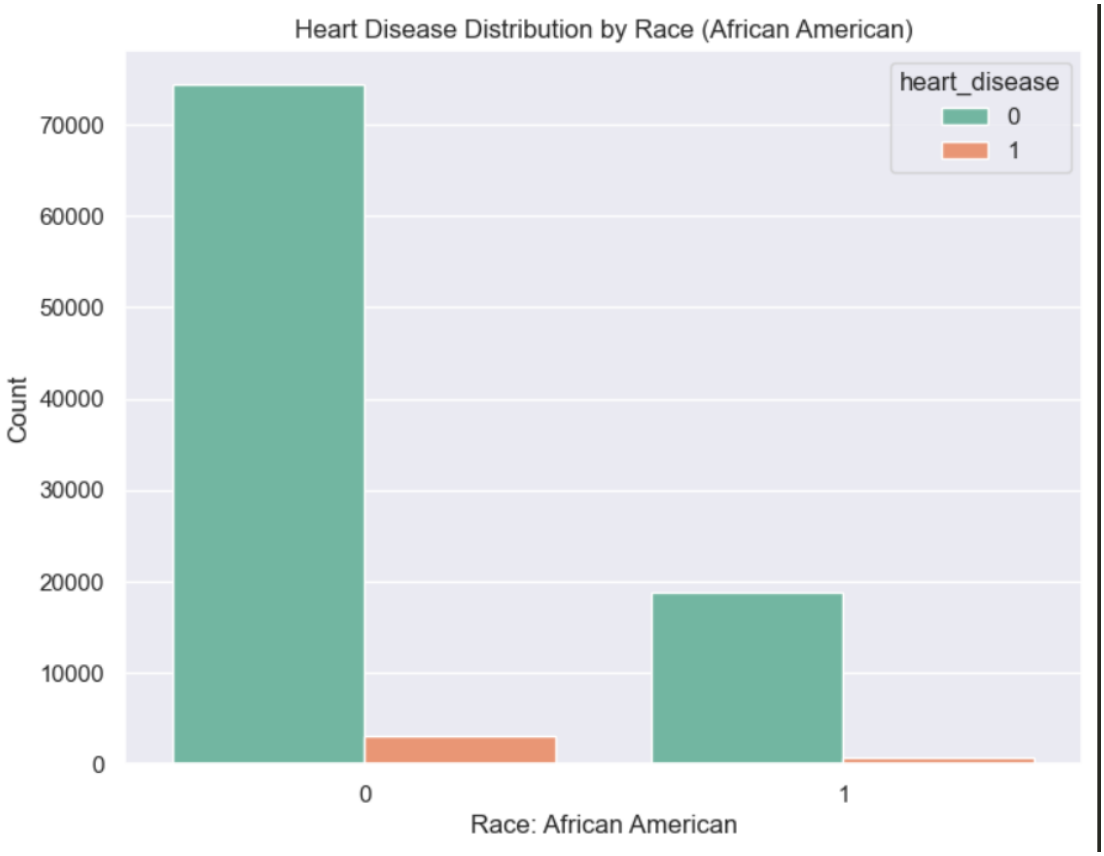
6. **Health Insight**:

    - The high proportion of "No Info" (over 35,000 individuals) indicates potential data collection issues that could affect the reliability of smoking-related health analyses.

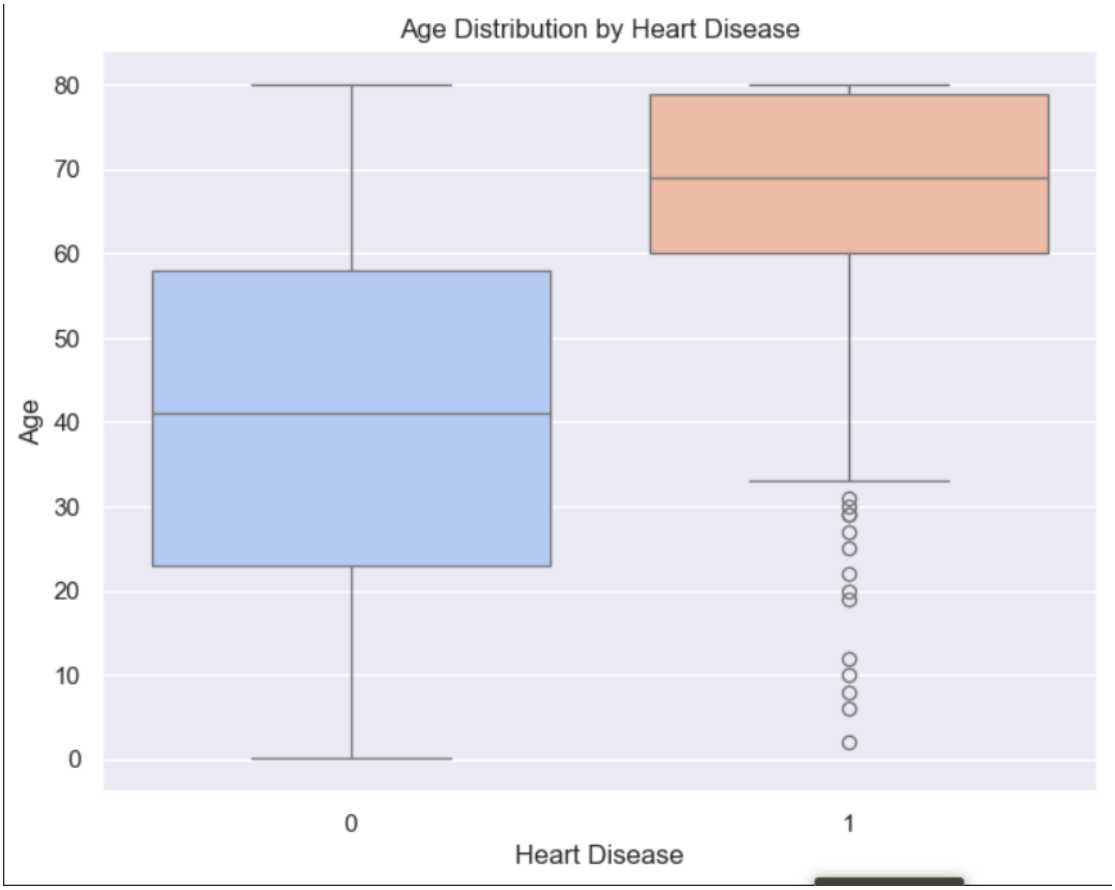
    - The dominance of "never" smokers (around 30,000 total) suggests that smoking may not be a major risk factor in this population, though this could be skewed by the missing data.



1. **Prevalence of No Heart Disease**:
    - The vast majority of individuals across all genders (around 50,000 for females, 40,000 for males, and minimal for others) do not have heart disease (hue = 0, red), indicating a low overall prevalence of heart disease in the dataset (around 92.7% as per earlier analysis).
2. **Low Prevalence of Heart Disease**:
    - Only a small number of individuals have heart disease (hue = 1, blue): approximately 2,000-3,000 for females, 2,000-3,000 for males, and negligible for the "Other" category, aligning with the earlier reported 3.8% prevalence.
3. **Gender Distribution**:
    - Females show the highest count of both no heart disease (50,000) and heart disease (2,000-3,000), reflecting their overall higher representation in the dataset (as seen in the age distribution).
    - Males have a slightly lower count of no heart disease (40,000) and a similar count of heart disease (2,000-3,000), suggesting a comparable disease rate relative to their population size.
    - The "Other" category has a minimal count for both no heart disease and heart disease, indicating very few individuals in this group.
4. **Health Insight**:
    - The low prevalence of heart disease (3.8%) across genders suggests it is not a dominant condition in this population, consistent with earlier findings.
    - The similar proportion of heart disease cases among females and males (relative to their total counts) indicates no strong gender bias in heart disease occurrence in this dataset.



1. **Prevalence of No Heart Disease**:
    - The majority of African Americans (around 70,000 for 0 and 20,000 for 1) do not have heart disease (hue = 0, teal), indicating a low overall prevalence of heart disease in this subgroup, consistent with the earlier reported 3.8% across the dataset.
2. **Low Prevalence of Heart Disease**:
    - Only a small number of African Americans have heart disease (hue = 1, orange), with approximately 2,000-3,000 individuals, aligning with the low prevalence observed previously.
3. **Binary Representation**:
    - The x-axis labels 0 and 1 likely represent the absence (0) and presence (1) of African American race, but the plot seems to show heart disease distribution within this race category. The high count for 0 (70,000) suggests a majority non-African American population, while 1 (20,000) represents African Americans, with heart disease cases being a small fraction of this group.
4. **Health Insight**:
    - The low proportion of heart disease cases among African Americans (around 10-15% of their subgroup) suggests that race alone may not be a dominant factor for heart disease in this dataset, though the sample size for African Americans (20,000) is smaller than the non-African American group (70,000).
    - The imbalance between the two groups indicates potential underrepresentation of African Americans, which could affect the reliability of race-specific conclusions.



1. **Age Range for No Heart Disease (0)**:

    - The age distribution for individuals without heart disease (blue) ranges from approximately 20 to 70 years, with a median around 45-50 years, indicating that the majority of healthy individuals are middle-aged.

2. **Age Range for Heart Disease (1)**:

    - The age distribution for individuals with heart disease (orange) ranges from approximately 30 to 80 years, with a median around 60-65 years, suggesting that heart disease is more prevalent in older age groups.

3. **Interquartile Range (IQR)**:

    - The IQR for no heart disease (approximately 35-60 years) is narrower than for heart disease (approximately 50-70 years), indicating greater age variability among those with heart disease.

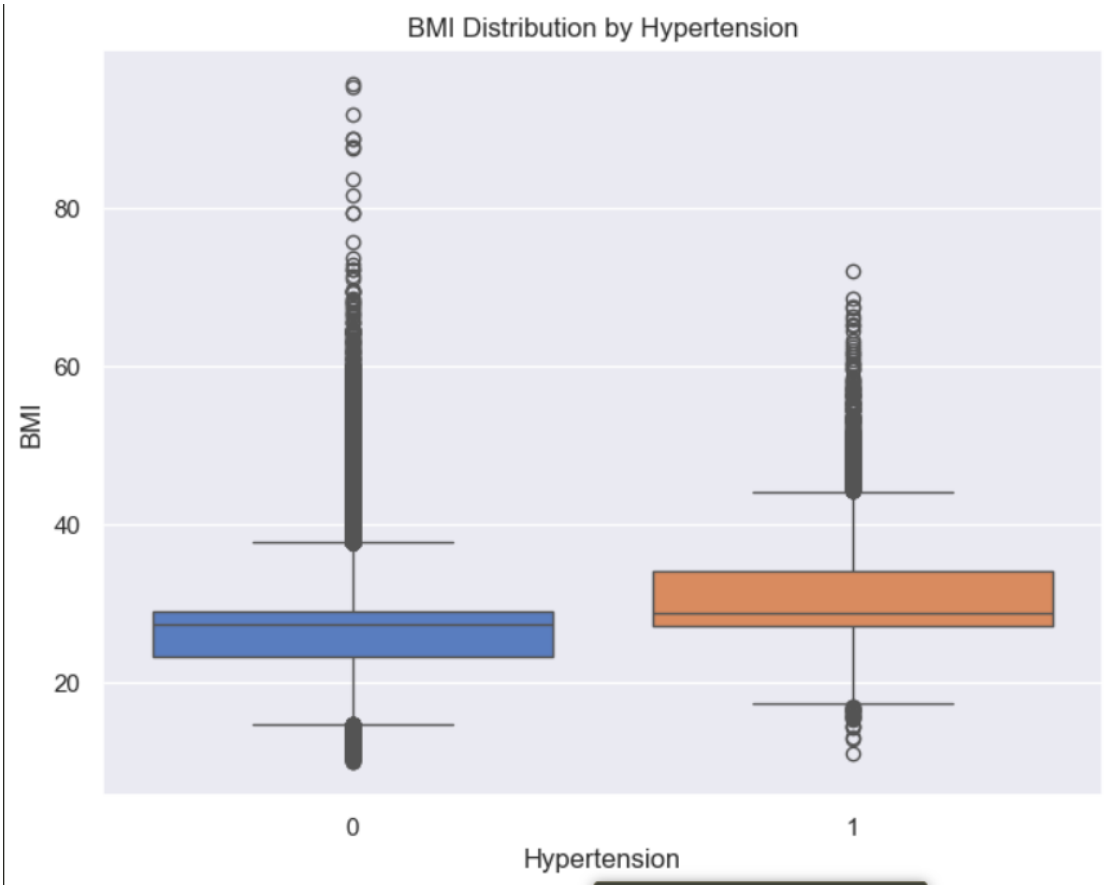
4. **Outliers**:

    - There are several outliers below 20 years and above 70 years for heart disease cases, with a few extreme values near 10 and 80 years, suggesting rare occurrences of heart disease in younger or very elderly individuals.

5. **Health Insight**:

    - The higher median age for heart disease (60-65 years) compared to no heart disease (45-50 years) confirms that age is a risk factor, with the condition becoming more common in older populations.

    - The presence of outliers in younger ages (<20 years) with heart disease may indicate congenital conditions or data anomalies worth investigating.



1. **Median BMI Difference**:

    - The median BMI for individuals without hypertension (0, blue) is around 25-30, while for those with hypertension (1, orange) it is slightly higher, around 30-35, suggesting a modest increase in BMI associated with hypertension.

2. **Interquartile Range (IQR)**:

    - The IQR for no hypertension (approximately 20-40) is similar to that for hypertension (approximately 25-45), indicating comparable variability in BMI across both groups, though the hypertension group shows a slightly wider spread.

3. **Outliers**:

    - Both groups exhibit outliers, with hypertension (1) having more extreme values above 60 and below 20, indicating a higher presence of unusual BMI values (e.g., obesity or underweight cases) among those with hypertension.

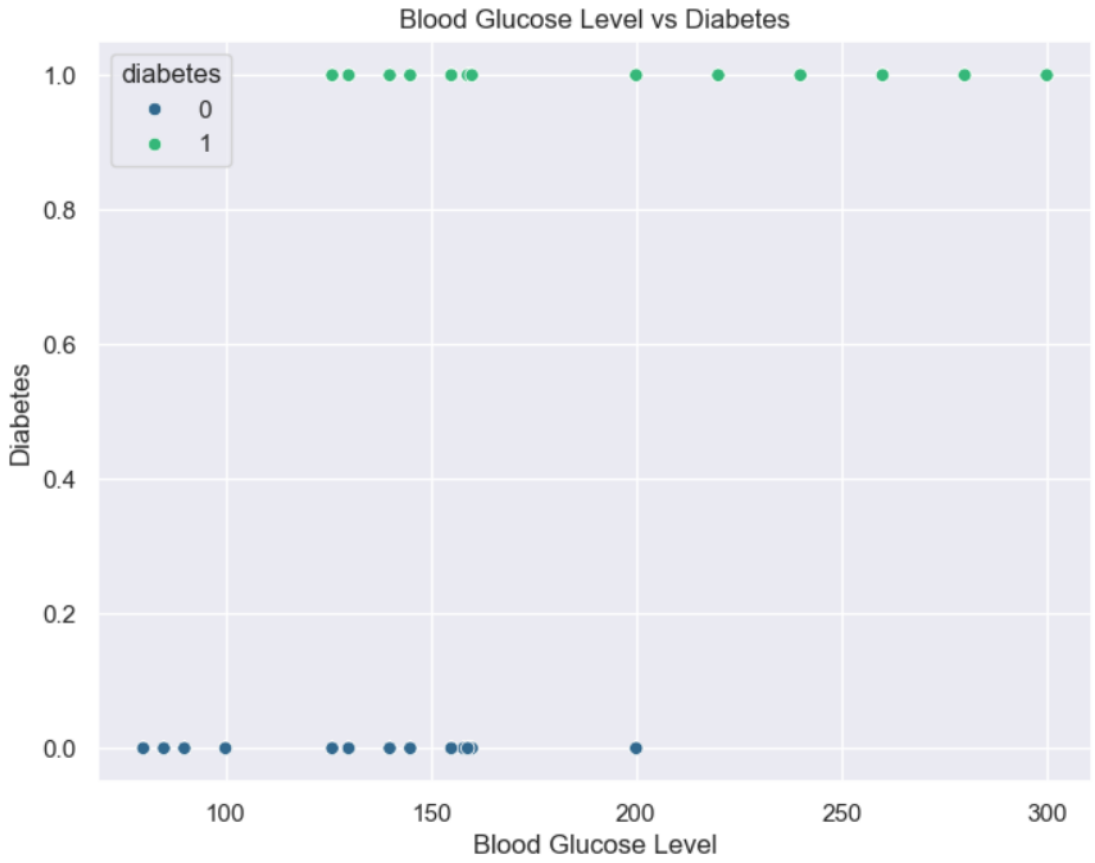
4. **Range of BMI**:

    - The BMI range for no hypertension extends from below 20 to above 60, while for hypertension, it spans a similar range but with a higher concentration of values above 40, reflecting a potential link between higher BMI and hypertension.

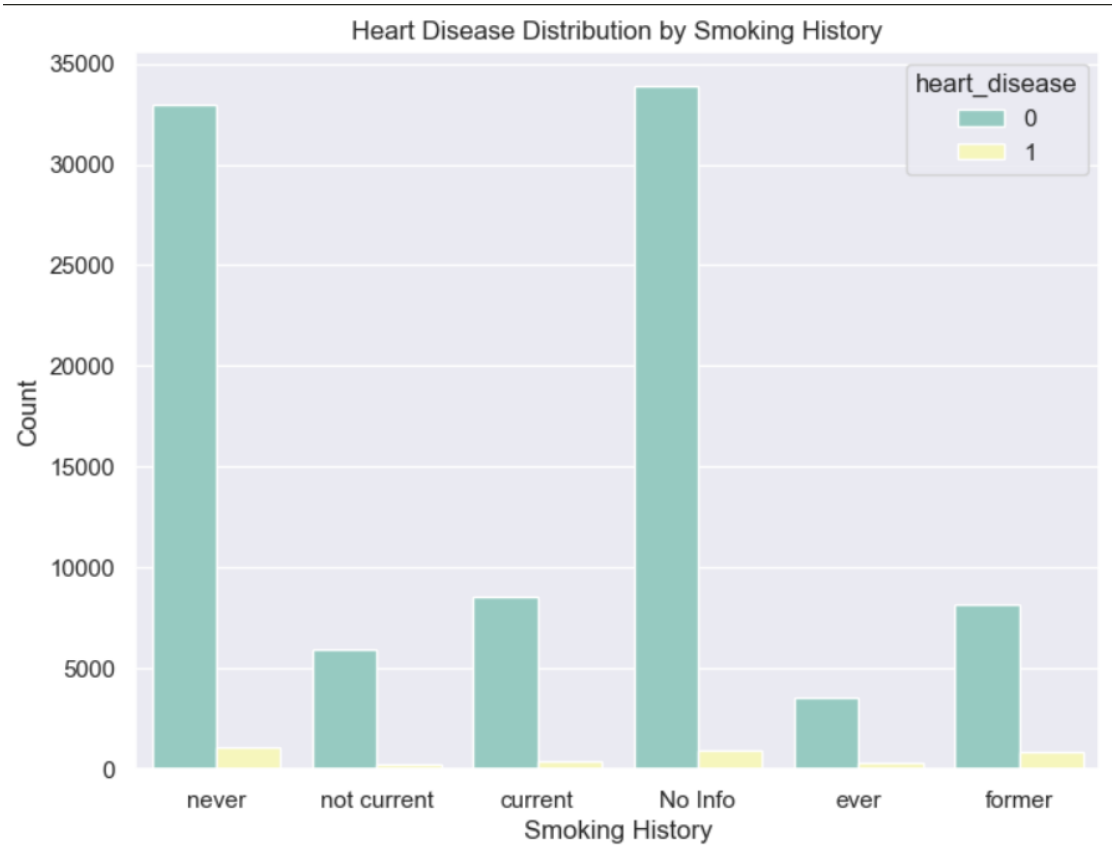
5. **Health Insight**:

    - The slightly higher median BMI in the hypertension group (30-35) aligns with the known association between obesity and hypertension, though the overlap in distributions suggests BMI alone is not a definitive predictor.

    - The presence of outliers above 60 in both groups may indicate data errors or extreme cases (e.g., severe obesity) that warrant further investigation.



1. **Glucose Levels for Non-Diabetics (0)**:
    - Individuals without diabetes (0, darker points) are mostly clustered below a blood glucose level of 150 mg/dL, with a few scattered points up to 200 mg/dL, indicating that non-diabetics typically have lower glucose levels (normal range <100 mg/dL, prediabetic 100-125 mg/dL).
2. **Glucose Levels for Diabetics (1)**:
    - Individuals with diabetes (1, lighter points) are predominantly clustered above 200 mg/dL, with some points extending to 300 mg/dL, reflecting higher glucose levels typical of diabetic individuals (≥125 mg/dL, often >200 mg/dL in uncontrolled cases).
3. **Threshold Separation**:
    - There is a clear separation around 150-200 mg/dL, where the transition from non-diabetic (0) to diabetic (1) becomes more pronounced, aligning with clinical thresholds for diabetes diagnosis (≥125 mg/dL).
4. **Overlap Region**:
    - Some overlap exists between 100-150 mg/dL, where a few non-diabetics have elevated glucose levels and some diabetics have levels closer to normal, suggesting potential misclassification or borderline cases.
5. **Health Insight**:
    - The strong clustering of diabetics above 200 mg/dL supports the earlier finding of a 0.5 correlation between blood glucose level and diabetes, indicating a reliable indicator of diabetes status.
    - The overlap in the 100-150 mg/dL range may indicate undiagnosed prediabetes or well-controlled diabetes cases, warranting further medical evaluation.

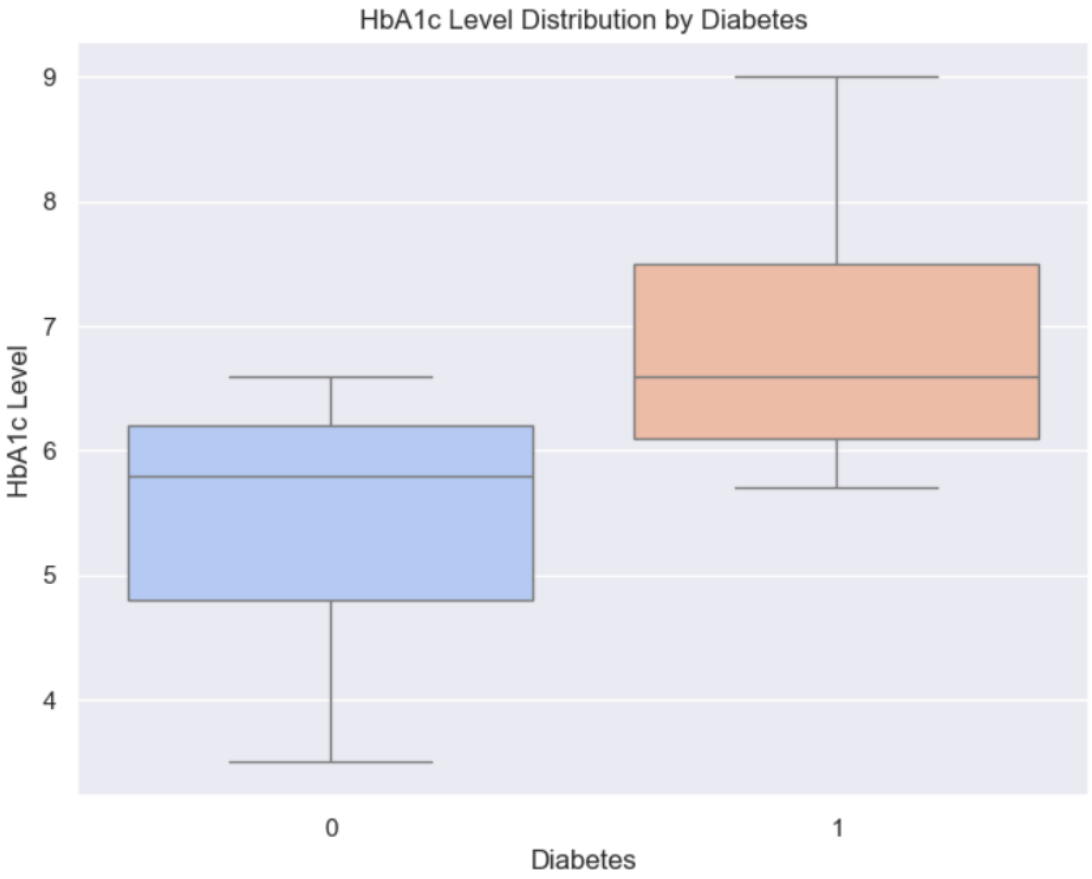


1. **Most individuals fall into the "never" or "No Info" categories:**
These groups dominate the dataset. However, the **majority within them do not have heart disease** (tall light green bars with small yellow segments).

2. **Higher heart disease rates among "current" and "former" smokers:**
While fewer people fall into these categories compared to "never" and "No Info", the **proportion of heart disease cases is visibly higher** (the yellow portion of the bar is more significant relative to total height).

3. **"Former" and "current" smokers show a noticeable risk:**
These groups demonstrate a **clear increase in heart disease prevalence** compared to "never" smokers.

4. **The "No Info" category needs careful treatment:**
Although it has a high count, its disease distribution looks similar to "never" smokers. This could suggest missing data that's **not missing at random**, or that many of these people are actually non-smokers.

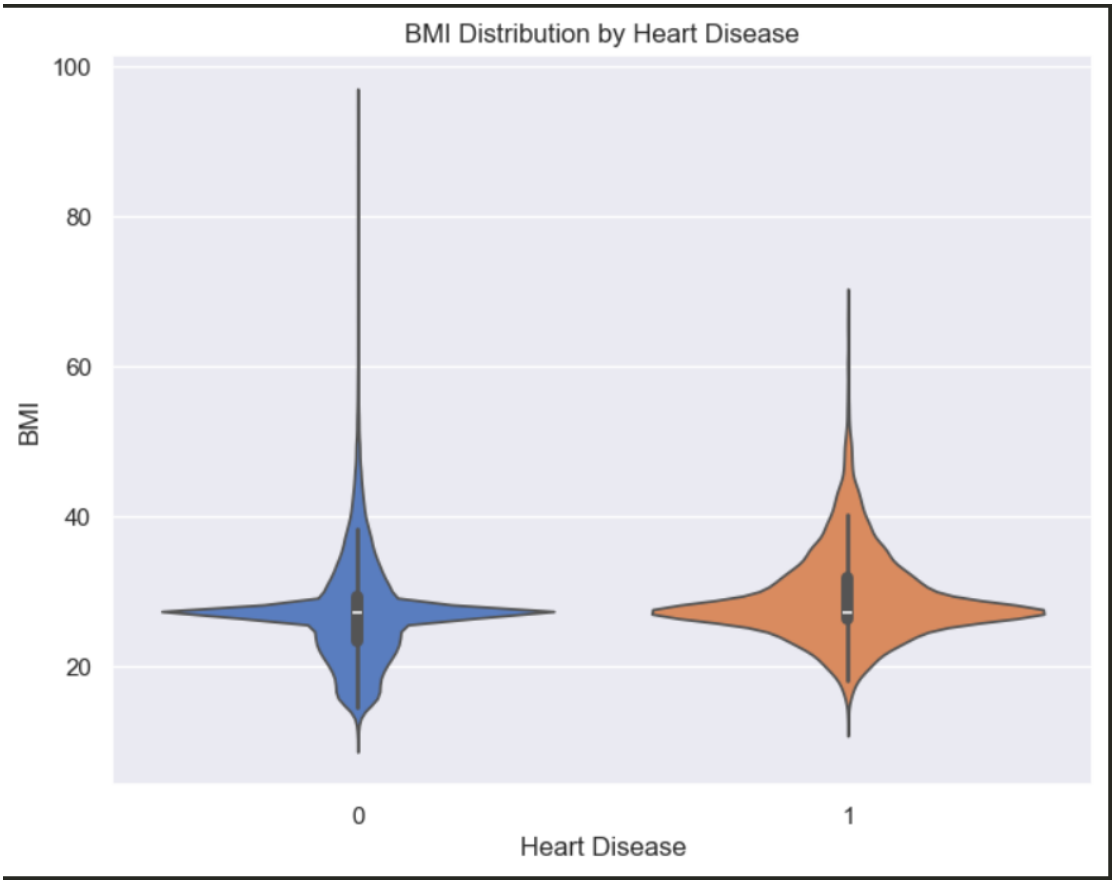


1. **Higher HbA1c levels among diabetics (diabetes = 1):**
The median (central line) for diabetics is noticeably higher than for non-diabetics. This is expected since HbA1c is a key indicator of blood sugar control, and elevated levels are diagnostic of diabetes.

2. **Wider spread in diabetic group:**
The interquartile range (IQR) for diabetics is broader, indicating more variability in HbA1c levels among diabetics. Some may have tight control, while others are poorly controlled.

3. **Non-diabetics have lower and tighter HbA1c range:**
Their values mostly fall between ~4.8 to 6.4, which aligns with normal or prediabetic ranges.

4. **Potential overlap exists:**
There is a small overlap in HbA1c levels between diabetics and non-diabetics, especially around the 6.0–6.5 range — this may include newly diagnosed or borderline individuals.



1. Different BMI Distributions for the Two Groups:

- For **group 0** (no heart disease), BMI values are more tightly clustered around the median with some high outliers (some values even exceed 90).
- For **group 1** (with heart disease), the BMI distribution appears more spread out and skewed, showing greater variability.

2. People with Heart Disease Tend to Have Higher BMI:

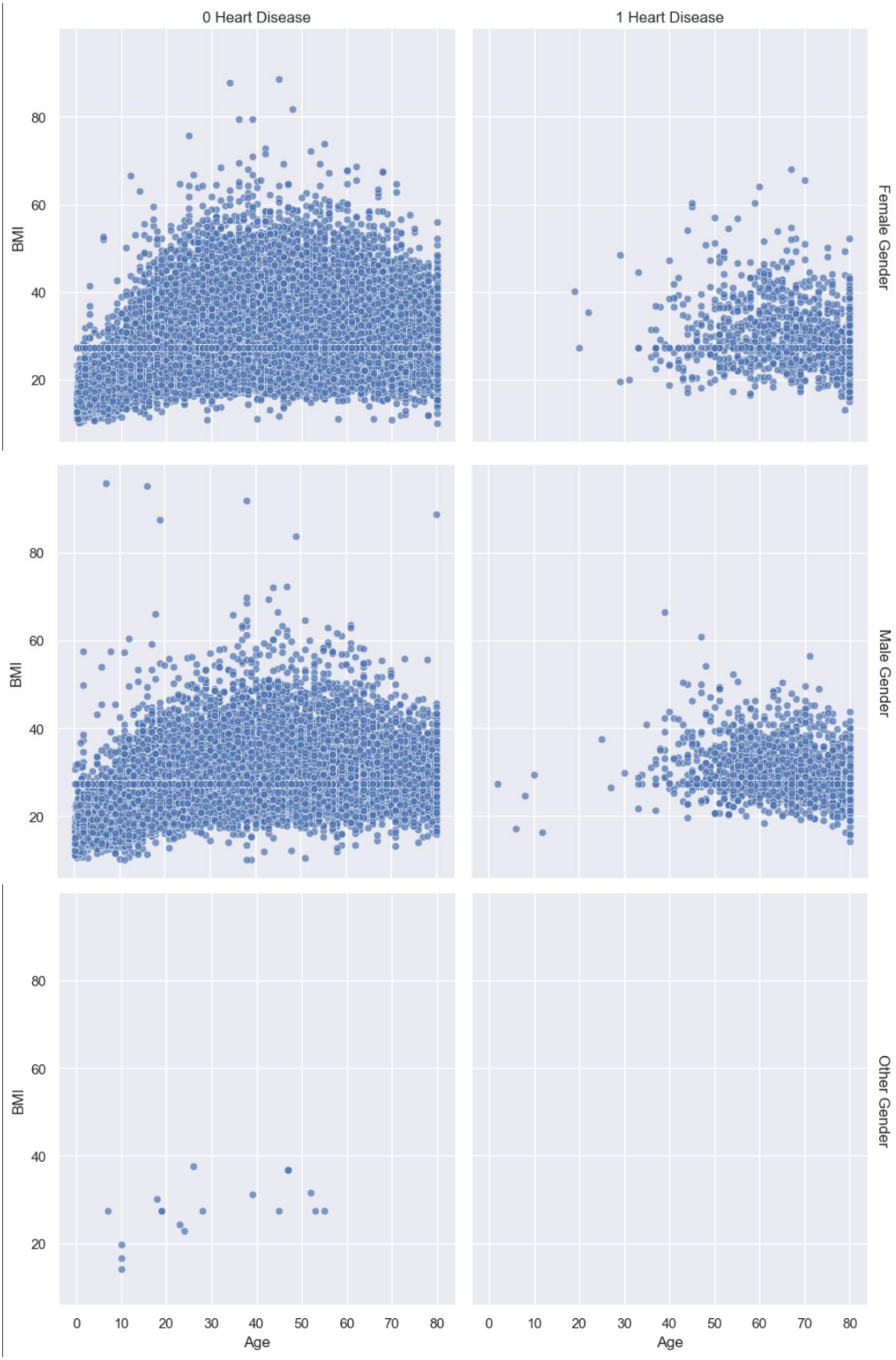
- The median BMI (white line inside the box) for heart disease patients (1) is slightly higher than that of those without the condition (0).
- This suggests a potential link between **higher BMI (overweight/obesity)** and heart disease.

3. Presence of Outliers:

- Both groups contain outliers, especially in group 0 (no heart disease), where some BMI values are extremely high. These might need to be reviewed or cleaned.

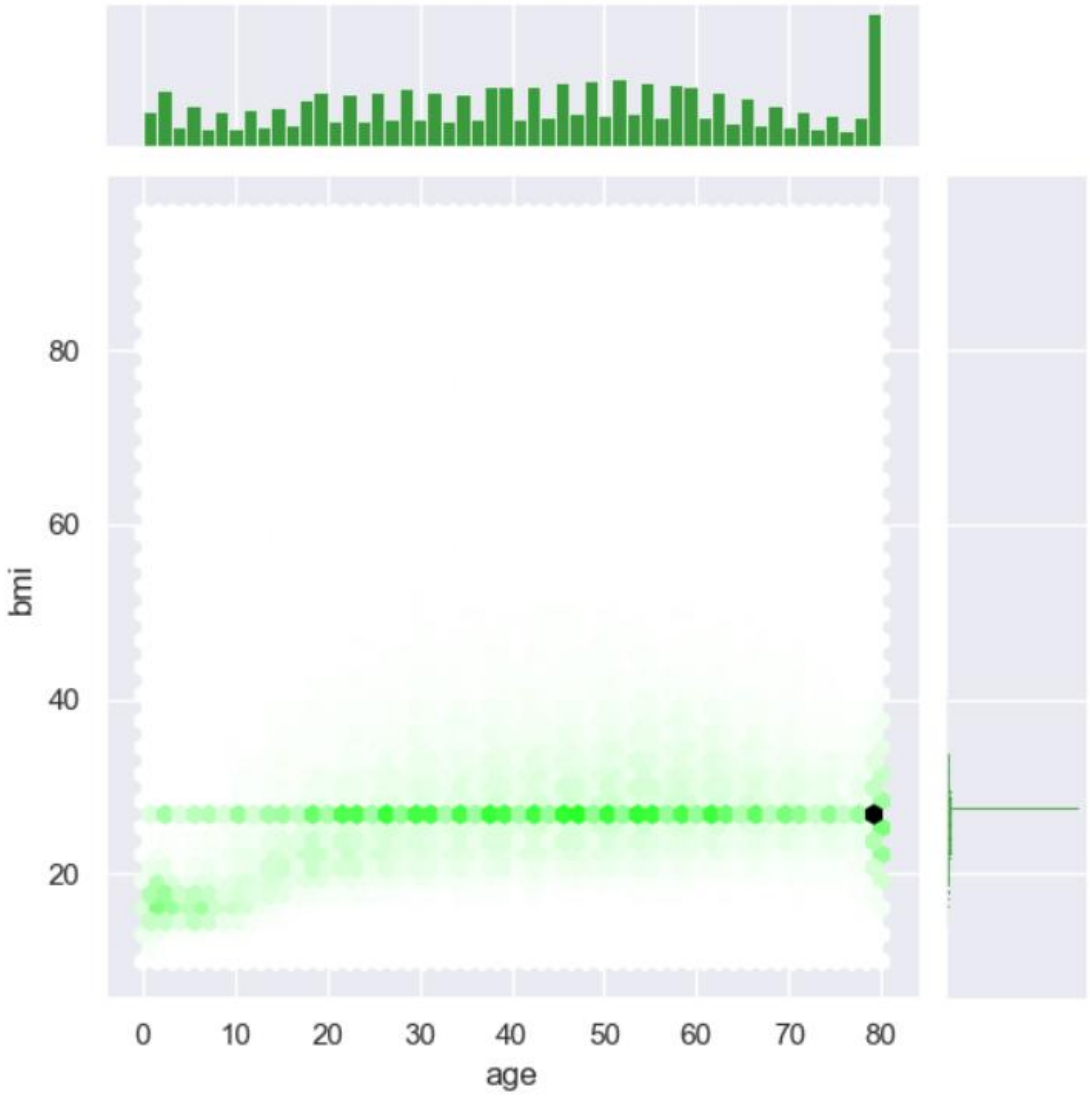
4. Low Density at Extremes:

- The thin tails of the violin indicate that very few individuals have extremely low or extremely high BMI values. Most values are concentrated between 20 and 40.


0 Heart Disease
1 Heart Disease
Female Gender
Male Gender
Other Gender
BMI
Age
0
10
20
30
40
50
60
70
80
20
40
60
80

1. **BMI and Heart Disease:**
    - For individuals without heart disease (left side of the plots), the BMI distribution is more centered around lower to moderate BMI values, with some cases reaching higher values.
    - For individuals with heart disease (right side of the plots), there is a higher concentration of BMI in the higher ranges compared to those without the disease, suggesting that individuals with heart disease tend to have a higher BMI.
2. **Gender Impact:**
    - For females (first set of plots), it can be observed that the average BMI is higher among individuals with heart disease compared to those without.
    - For males (second set of plots), the same pattern appears, but the rate is higher in the heart disease group.
    - In the "Other" gender category (third set of plots), the data is sparse, making it difficult to extract clear insights compared to the other genders.
3. **Age and BMI Relationship:**
    - There are a few younger individuals with heart disease who show a higher BMI, but the total number of these cases is quite low.
    - Older individuals appear more prominently in the heart disease category, suggesting that heart disease is more prevalent among older age groups.

- Higher BMI is also more common in the heart disease category among older individual



1. **Low BMI across ages:**

- The plot shows a strong concentration of data points along the lower end of the BMI scale (around 20). This suggests that the majority of the data points (for both younger and older individuals) have a BMI within a narrow, lower range, which could indicate a prevalence of healthy BMI levels.

2. **Age distribution:**

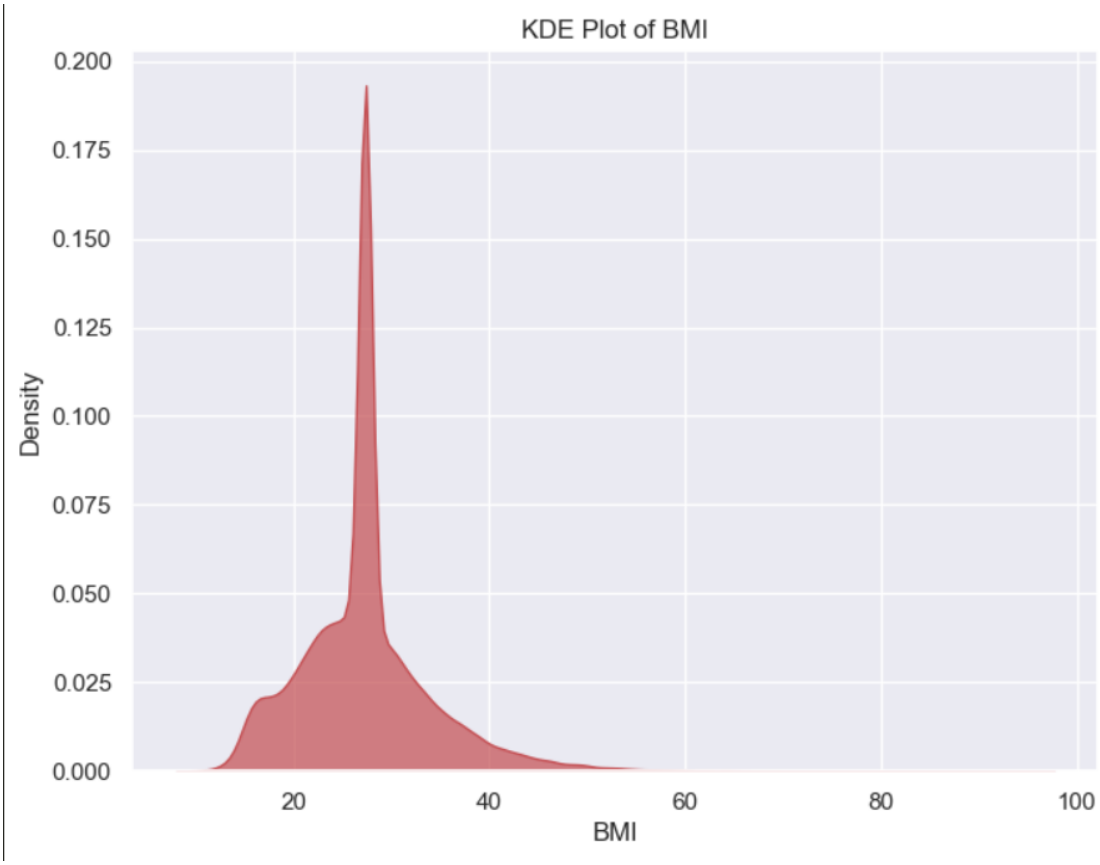
- The age distribution appears more spread out, with a wider range of ages represented, but there is little to no significant increase in BMI as age increases. This could suggest that BMI does not significantly increase with age in this dataset, or that most age groups have relatively low BMI.

3. **A slight rise in BMI at higher ages:**

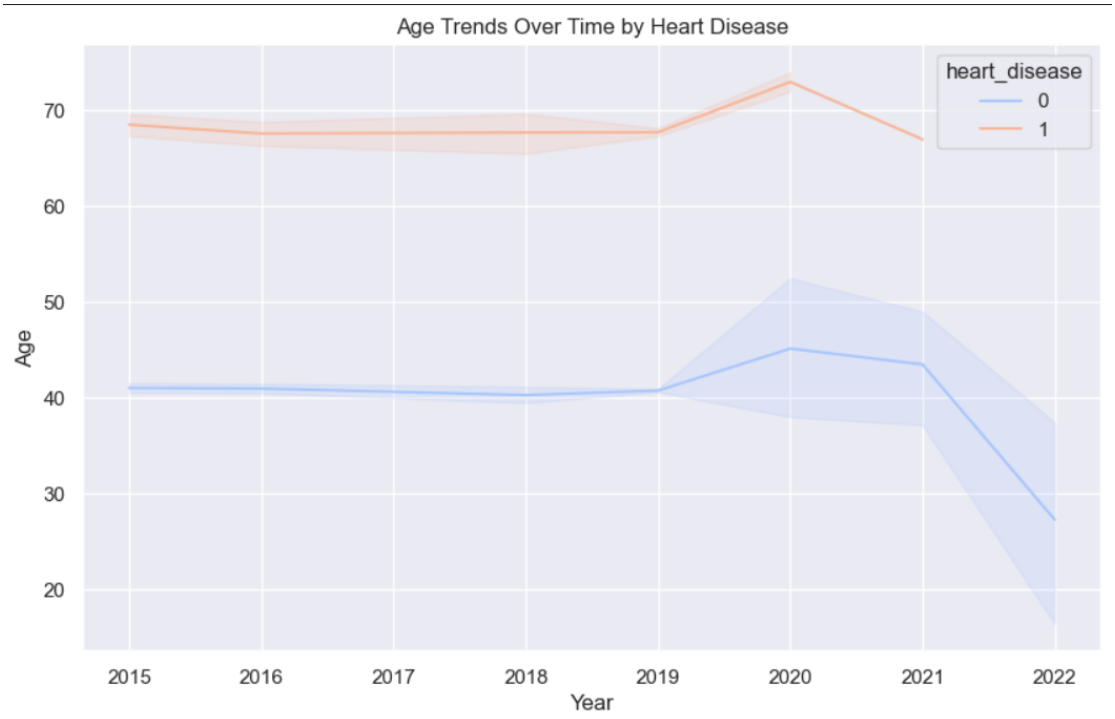
- There is a small concentration of data points at the higher age range (around 50-80 years) that shows an increase in BMI, but it is still relatively low. This could imply that, while there may be a slight trend for higher BMI with age, the dataset does not show a strong pattern of BMI increasing dramatically as age increases.

4. **Outliers:**

- A few extreme outliers are visible at the higher end of the BMI scale (above 60), especially at the lower age range. These could represent individuals with unusually high BMI values at younger ages.



1. **Peak at Low BMI:**
    - The plot shows a very sharp peak around a BMI of 20, indicating that the majority of the data points are concentrated in the lower BMI range. This suggests that a large portion of the population in the dataset has a relatively low BMI, likely within a healthy or normal range.
2. **Long Tail towards Higher BMI:**
    - The plot shows a long tail extending towards higher BMI values (above 40), but the density decreases as BMI increases. This suggests that although there are some individuals with high BMI, they are much less frequent than those with a lower BMI.
3. **Skewed Distribution:**
    - The distribution appears to be highly skewed to the right, which means there are a few individuals with much higher BMI, but they are not common. This is common in many real-world datasets where most individuals have average values and a smaller portion of the population has extreme values.

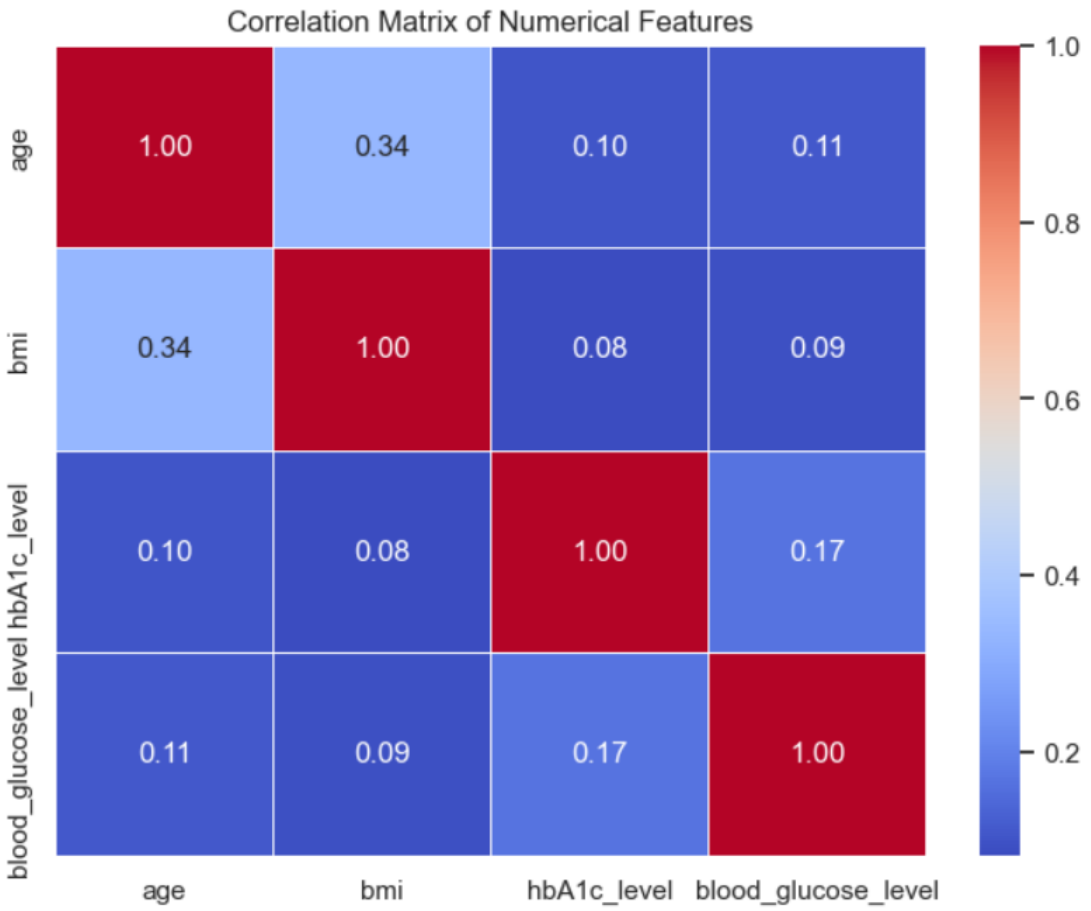


1. **Age Distribution Over Time:**

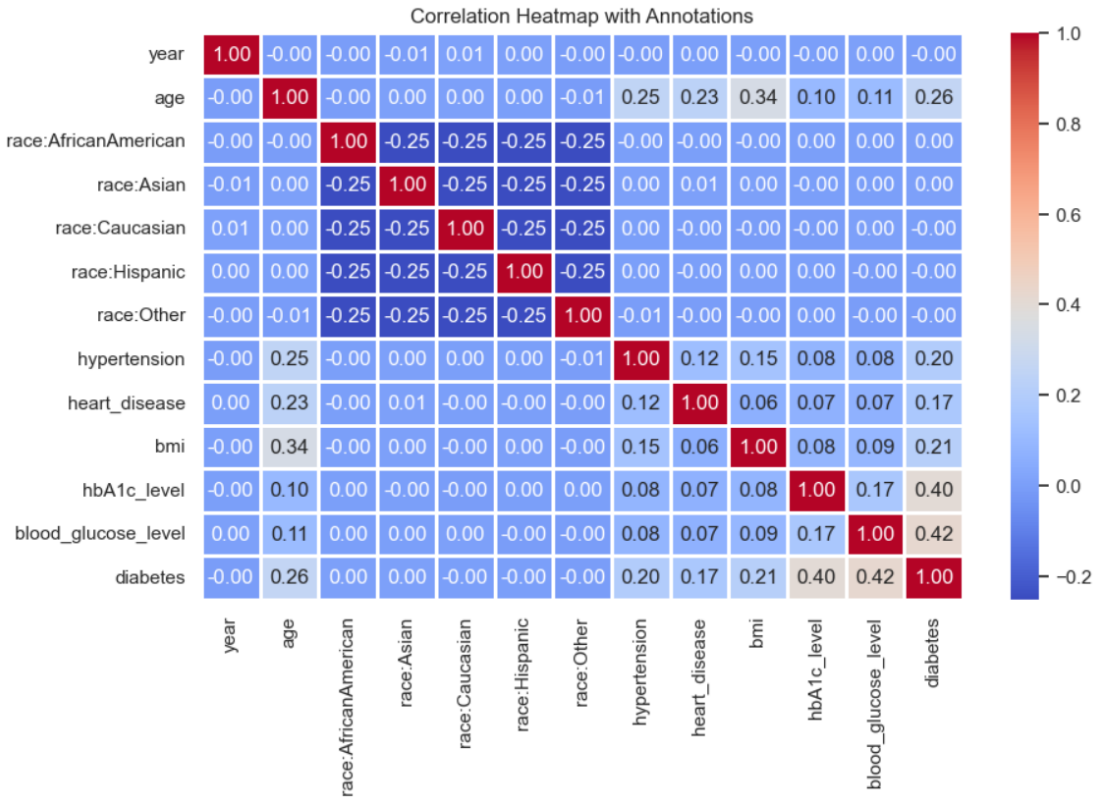
   - **For individuals without heart disease (blue line):**
     - The age appears to be relatively stable over time, with minor fluctuations. There is no significant upward or downward trend in the average age of individuals without heart disease.
   - **For individuals with heart disease (orange line):**
     - The age of individuals with heart disease appears to increase over time. This suggests that, on average, individuals diagnosed with heart disease are getting older as the years progress.
     - The increase in age becomes more pronounced towards the later years (2021 and 2022), indicating that heart disease is becoming more common in older populations over time.

2. **Trend Comparison:**

   - The gap between the two lines is growing, indicating that heart disease is being diagnosed more frequently among older individuals as time passes.
   - The average age for those without heart disease stays relatively constant, while the average age for those with heart disease tends to rise, which could reflect the aging population's greater susceptibility to heart disease.



1. **Age and BMI (Correlation: 0.34):**
    - There is a moderate positive correlation between age and BMI. This suggests that as individuals get older, their BMI tends to increase slightly, but the relationship is not very strong.
2. **Age and HbA1c Level (Correlation: 0.10):**
    - The correlation between age and HbA1c level is very weak. This implies that age does not have a significant relationship with HbA1c levels in this dataset.
3. **Age and Blood Glucose Level (Correlation: 0.11):**
    - Similar to HbA1c, the correlation between age and blood glucose level is also weak. This suggests that age does not significantly influence blood glucose levels in this dataset.
4. **BMI and HbA1c Level (Correlation: 0.08):**
    - There is a very weak positive correlation between BMI and HbA1c level. This indicates that BMI is not strongly related to HbA1c levels in this dataset.
5. **BMI and Blood Glucose Level (Correlation: 0.09):**
    - The correlation between BMI and blood glucose level is also weak. It suggests that BMI has a minimal effect on blood glucose levels.
6. **HbA1c Level and Blood Glucose Level (Correlation: 0.17):**
    - There is a slightly stronger but still weak positive correlation between HbA1c levels and blood glucose levels. This indicates that as HbA1c levels increase, blood glucose levels also tend to increase, but the relationship is still not very strong.



1. **Correlation Between Demographics and Health Indicators:**

   - **Age and hypertension (0.25)**: There is a moderate positive correlation between age and hypertension, suggesting that as individuals get older, they may be more likely to experience hypertension.
   - **Age and heart disease (0.23)**: Age also shows a moderate positive correlation with heart disease, indicating that older individuals are more likely to be diagnosed with heart disease.
   - **Hypertension and heart disease (0.12)**: There is a weak positive correlation between hypertension and heart disease, meaning that individuals with hypertension may have a slightly higher likelihood of heart disease, but the relationship is not strong.

2. **BMI and Health Indicators:**

   - **BMI and heart disease (0.15)**: A weak positive correlation between BMI and heart disease, suggesting that individuals with higher BMI might have a slightly higher risk of heart disease, but this relationship is weak.
   - **BMI and hypertension (0.15)**: A weak positive correlation with hypertension, suggesting that higher BMI may be slightly linked to higher rates of hypertension.

3. **Blood Glucose and Other Health Indicators:**

   - **Blood glucose and diabetes (0.42)**: A moderate positive correlation, indicating that as blood glucose levels increase, the likelihood of having diabetes also increases.
   - **Blood glucose and heart disease (0.17)**: A weak positive correlation, suggesting a slight relationship between elevated blood glucose levels and heart disease.

4. **HbA1c and Blood Glucose:**

   - **HbA1c level and blood glucose (0.40)**: A moderate positive correlation, indicating that as HbA1c levels increase, blood glucose levels tend to increase as well.

5. **Correlations with Race:**

   - The correlation between **race categories (African American, Asian, Caucasian, Hispanic, Other)** and other health indicators is quite low, suggesting that race is not strongly correlated with most health measures like age, BMI, or blood glucose levels in this dataset.