***PROJECT REPORT***

*HAP 618:COMPUTATIONAL TOOLS IN HEALTH INFORMATICS*

*PROBLEM DESCRIPTION OF THE PROJECT:*

*As a medical student, I have consistently observed that Cardiovascular Disease (CVD) stands out as a particularly serious health concern, affecting nearly 50% of the population. Personally, my family has experienced the consequences of this disease, amplifying my commitment to understanding its roots.*

*Two pivotal factors contributing to cardiovascular disease are high blood pressure and a sedentary lifestyle. In our project, we aim to delve into health data gathered from both patients and the general population. Our goal is not only to identify the primary reasons behind the prevalence of cardiovascular diseases but also to leverage this data to predict which age groups are most susceptible.*

*The intersection of medical science and data analytics presents a unique opportunity to unravel patterns and trends that contribute to the cardiovascular health of individuals. By scrutinizing this wealth of information, we hope to not only enhance our understanding of the disease but also pave the way for targeted interventions and preventive strategies.*

*Join with me on this journey as we navigate through the intricate web of health data, seeking insights that could potentially transform the landscape of cardiovascular health for different age demographics. Together, let's unravel the mysteries and work towards a healthier, heart-strong future for all.*

*My project revolves around performing exploratory data analysis (EDA) on cardiovascular disease data using Python. The goal is to visualize insights derived from the analysis without the use of HTML, CGI, or Unix.*

*Technologies Used:*

*Python:*

*Libraries: pandas, matplotlib.pyplot, sklearn (sci-kit),seaborn.*

*Tasks:*

*Data aggregation*

*Bar chart creation*

*Python Code Details:*

*Data Aggregation:*

*Use the pandas library to import cardiovascular disease data.*

*Data Visualization:*

*Utilize matplotlib.pyplot to create bar charts.*

*Customize charts with features like x-label, y-label, title, legend, color assignments, and x-tick marks.*

*Workflow:*

*Data Processing:*

*Python code reads cardiovascular disease data.*

*Aggregates data for analysis.*

*Data Visualization:*

*Generates bar charts displaying relevant insights.*

*Conclusion:*

*The project is focused on Python for both data analysis and visualization.*

*No HTML, CGI, or Unix components are involved, simplifying the architecture.*

*Directory Structure:*

*project\_directory/*

*data/: Store raw cardiovascular disease data (e.g., cardiovascular\_data.csv).*

*scripts/: Save Python code for data processing and bar chart generation.*

*images/: Save generated bar charts as images.*

*Execution:*

*Run Python scripts located in the scripts/ directory.*

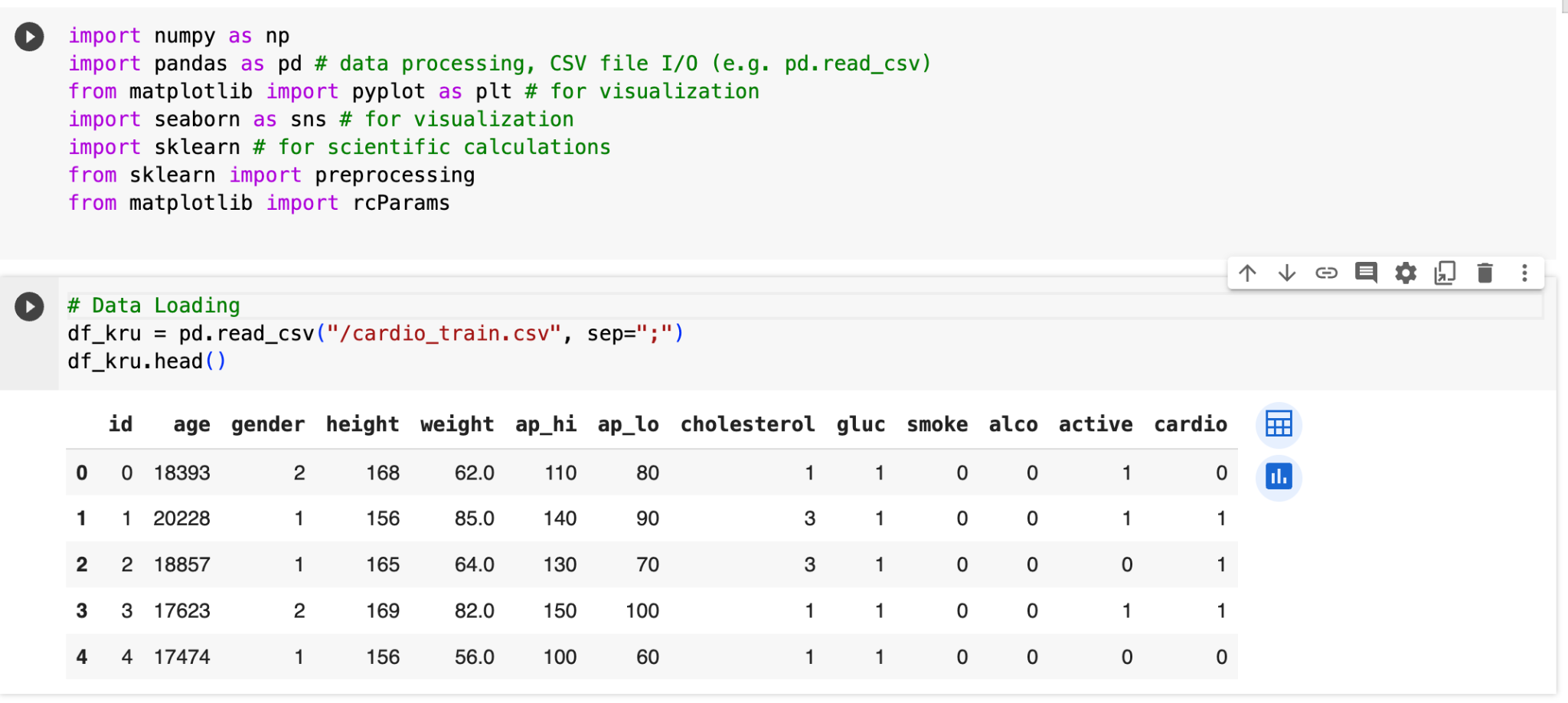
*Analyze data, create visualizations, and save generated bar charts in the images/ directory.*

*This straightforward design ensures that my project relies solely on Python for data analysis and visualization, making it more focused and easier to manage.*

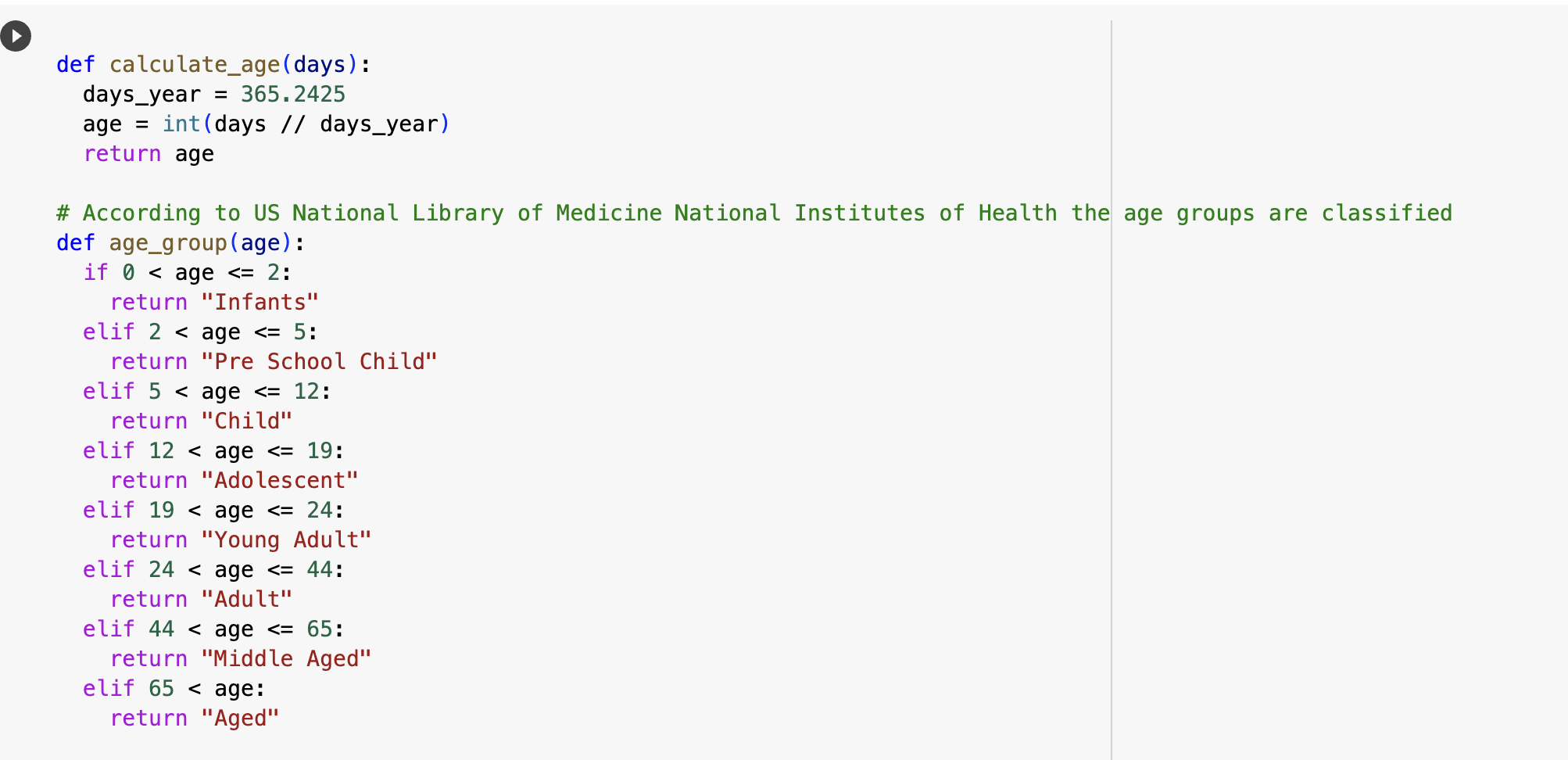
*IMPLEMENTATION:*

*Among the various components used in this cardiovascular disease Exploratory Data Analysis (EDA) project, Python code for generating bar charts plays a pivotal role. The bar charts serve as a crucial tool for data analysis, providing insights into the patterns and distributions within the dataset, which is the central focus of the project. Here is the Python code used for creating bar charts in the context of this cardiovascular disease EDA.*

*.*

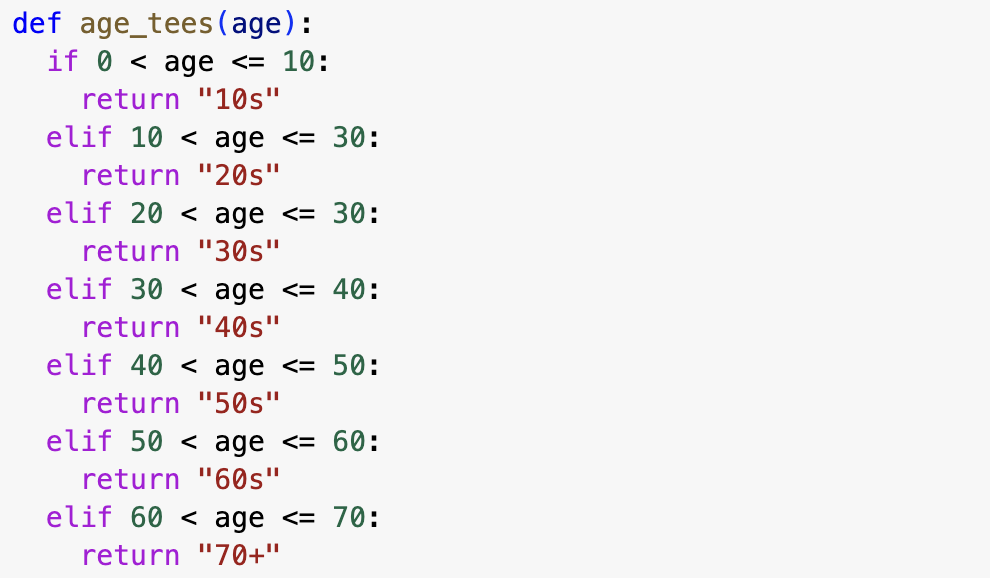
**

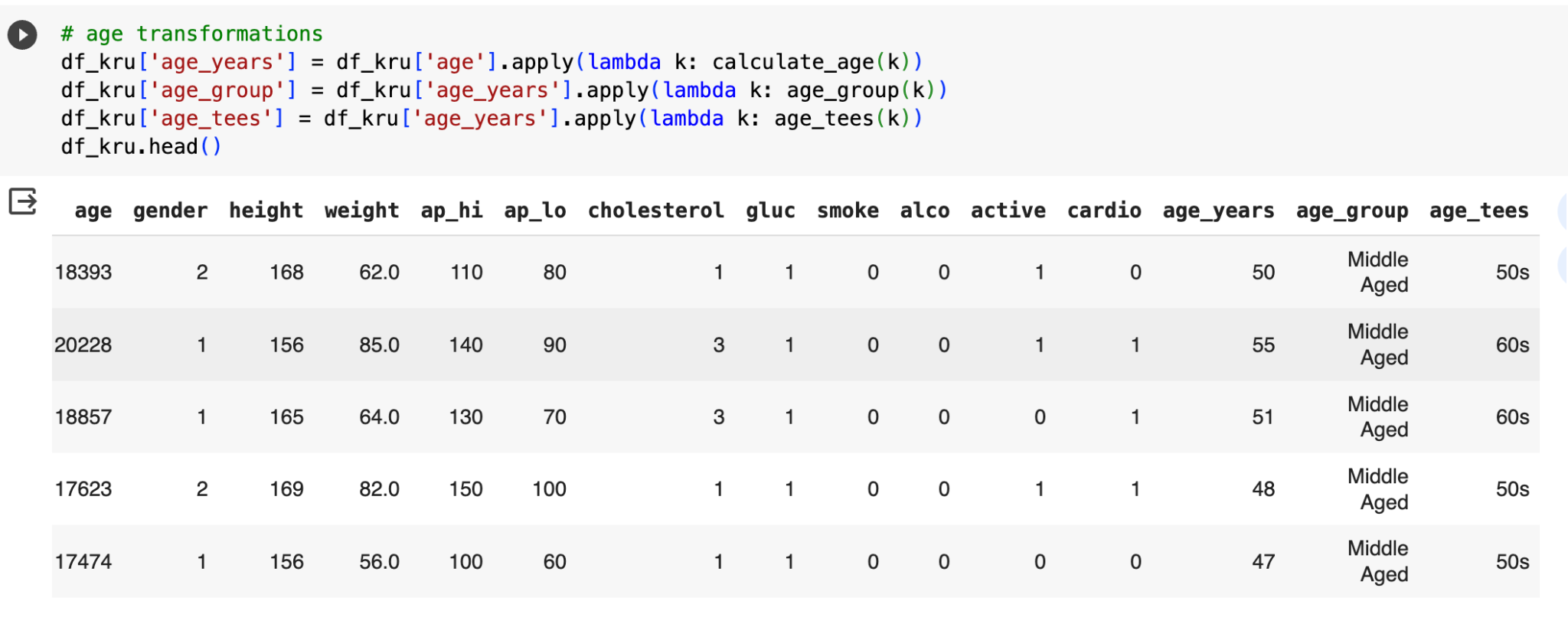
*Above is the first few rows of the data in my project I am considering only few columns like age, height, weight, ap\_hi, ap\_lo, cholestrol, active(physical activity)*

**

*In data age is calculated in days,so here in the above screenshot I am converting days into years.*

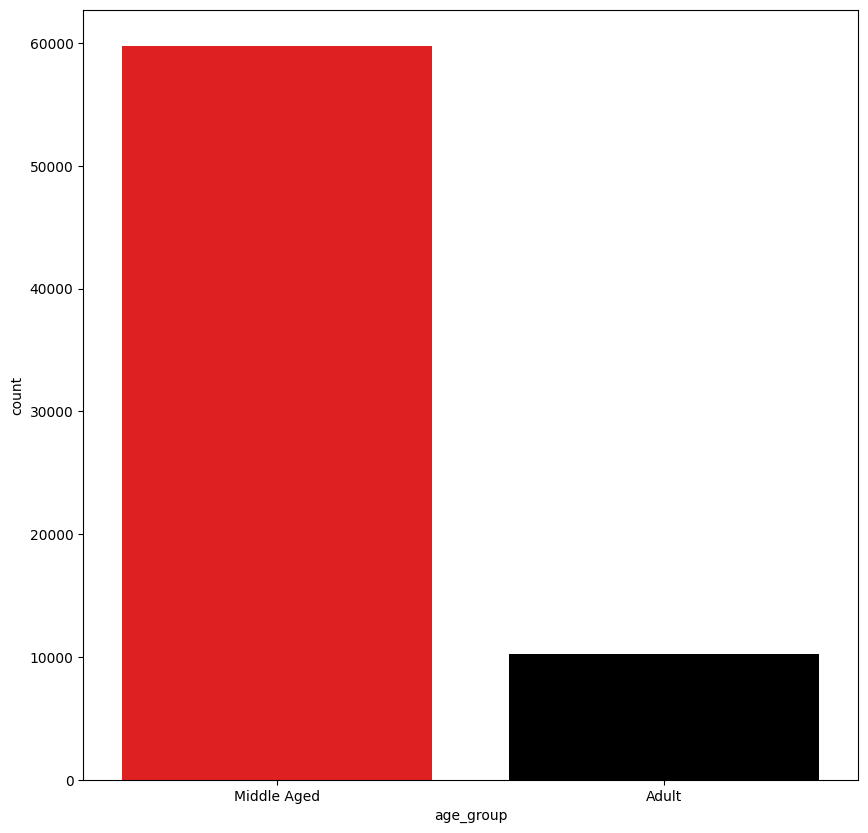
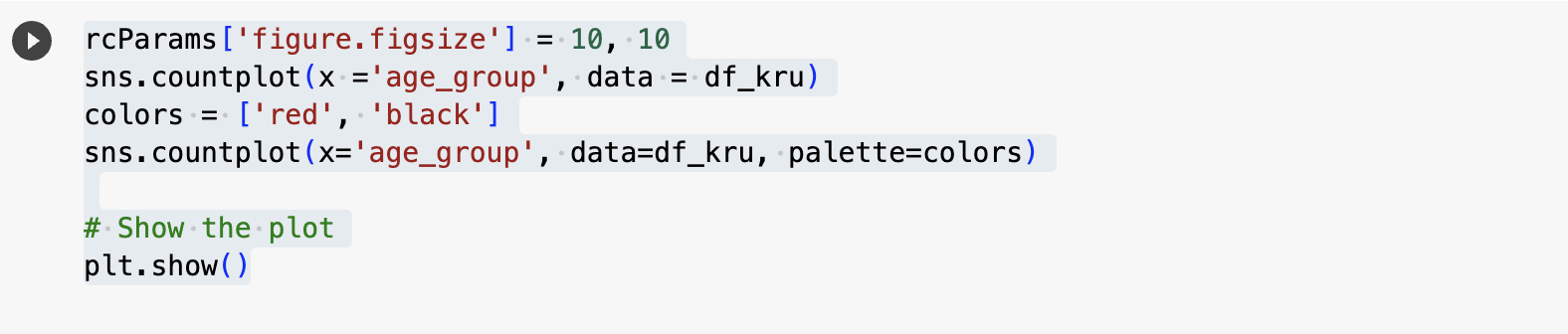
*According to the US National Library of Medicine,National Institutes of Health age groups are classified as (Infants, Preschool child, child, Adolescent, Young Adult, Adult, Middle Aged, Aged.*

**

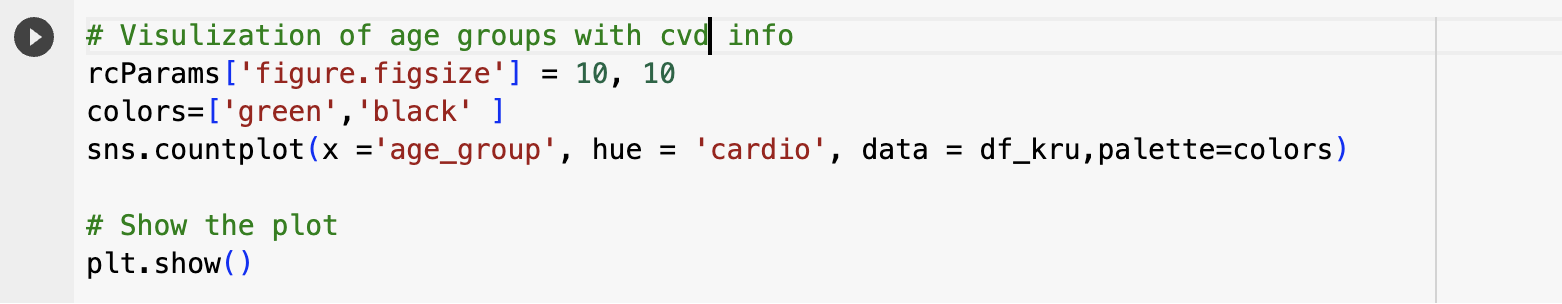
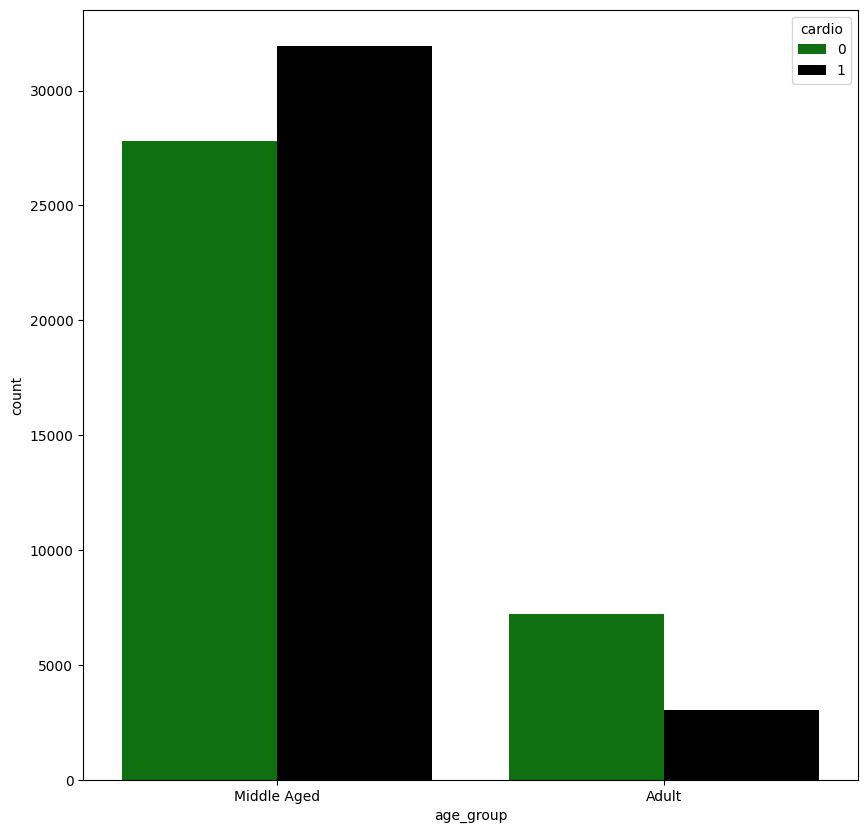
**

*In the above screenshot, a new column named "age\_tees" is added to the DataFrame, representing age groups categorized by decades.*

*Two other columns age\_years and age\_group are added.*

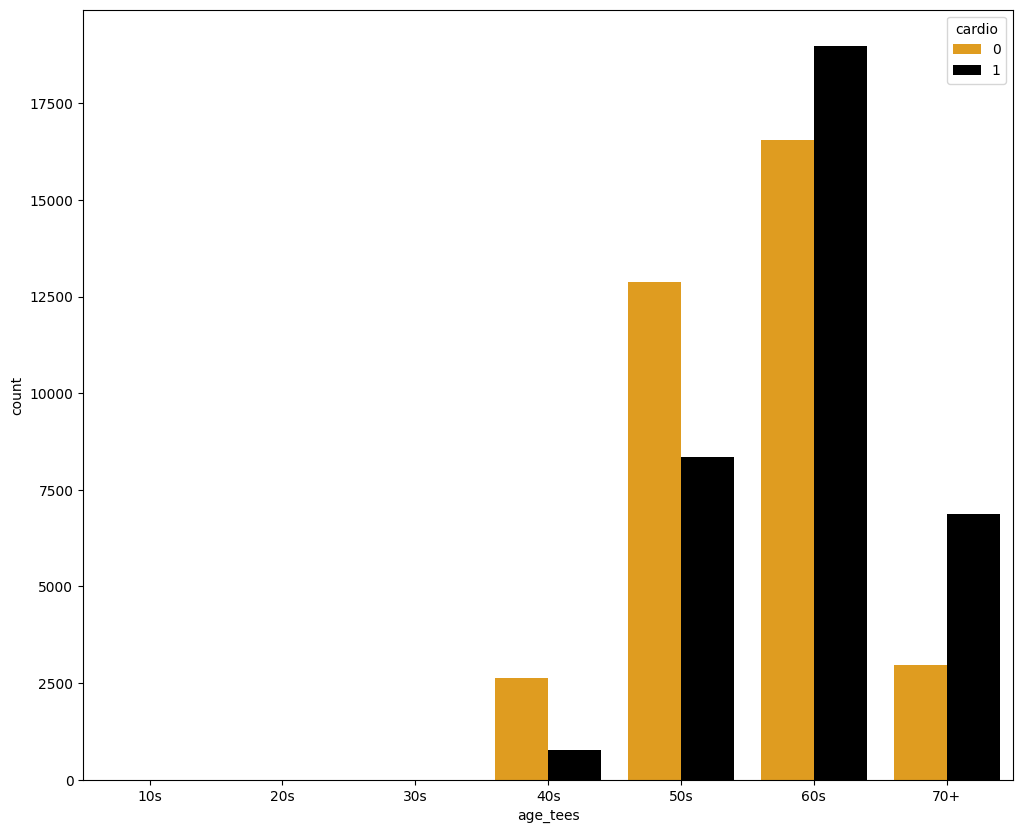
**

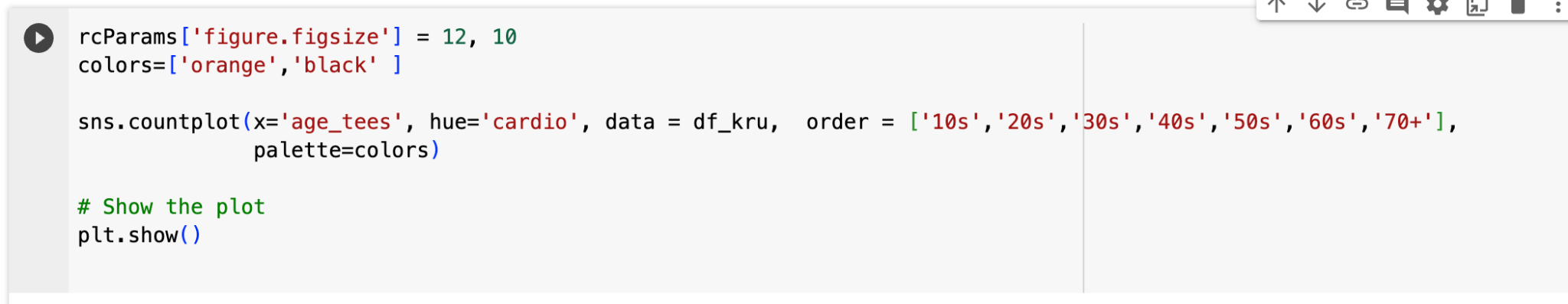
*The above code and its visual representation shows that the dataset contains majorly middle aged people and minor adult data.*

**

*1)The resulting clustered bar chart provides a quick comparison of disease presence (green) and absence (black) within each age group, aiding in the interpretation of cardiovascular disease distribution.*

*2) The graph indicates that middle aged groups are mostly affected by CardioVascular Disease when compared to adult.*



**

*The resulting visualization is a clustered bar chart where each age group is divided into two bars (orange and black) representing the counts of individuals with and without cardiovascular disease*

*The graph indicates that 60 s and 70 + are mostly affected by Cardiovascular Disease .*

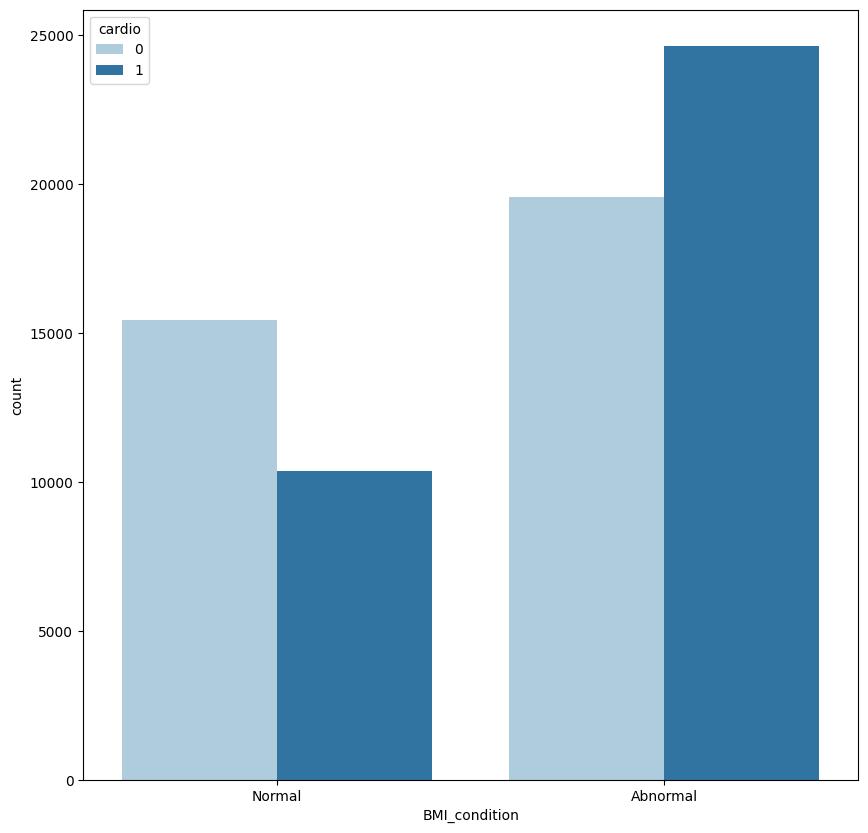
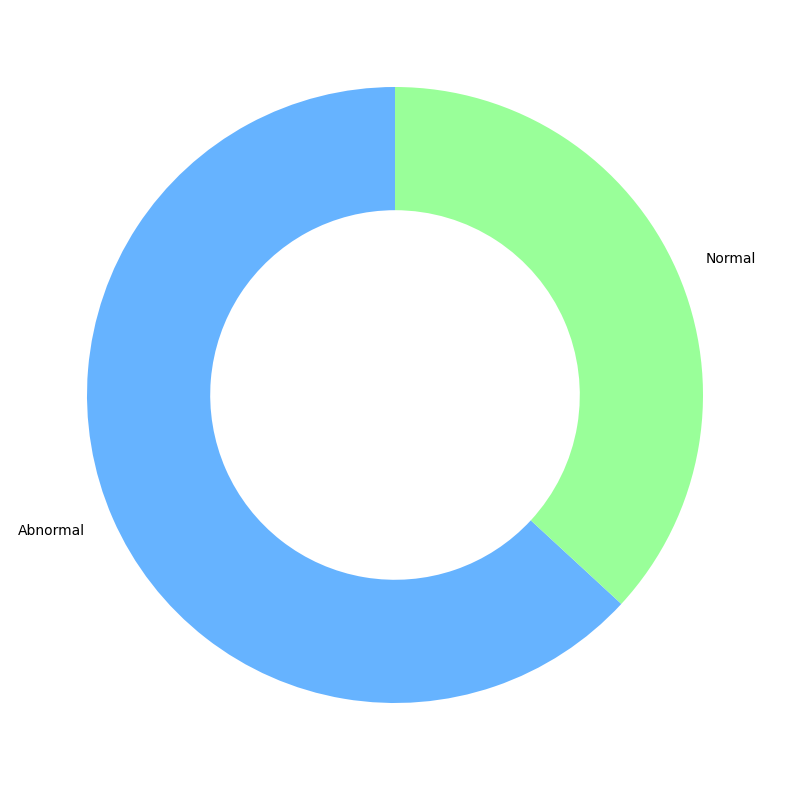
**

*1)Here after calculating BMI it is categorised into Normal and Abnormal. The new column named BMI\_condition is added to the dataframe.*

*2) The resulting donut chart displays the distribution of BMI\_conditions.*

**

*The resulting plot allows for a quick comparison of the distribution of cardiovascular disease within each BMI condition, highlighting potential associations between BMI and cardiovascular health.*

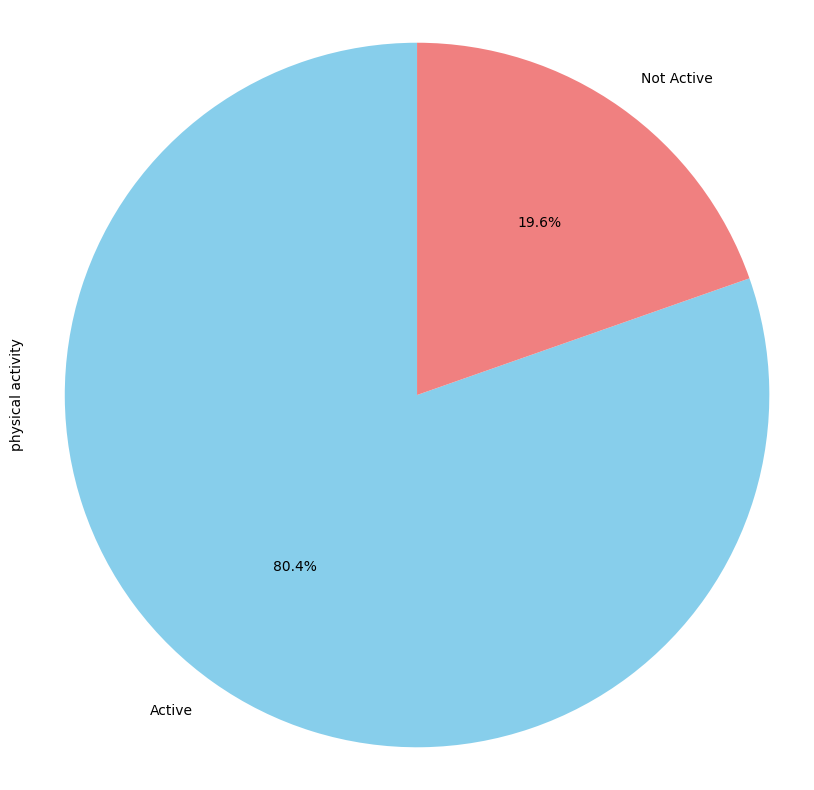
*T*

*T*

**

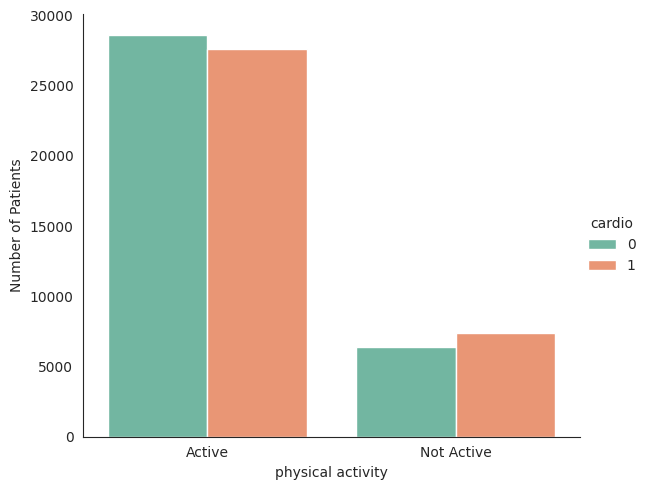
*The resulting plot creates a new column called Physical activity for the inactive person the value indicated as zero and for person with physical activity the value is indicated as one.*

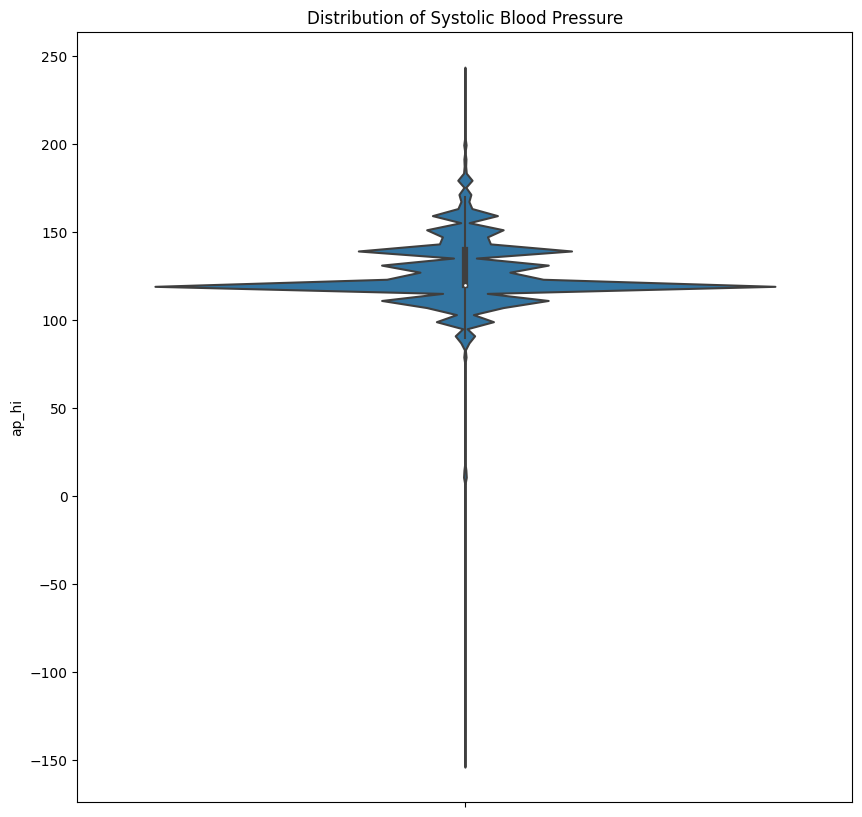
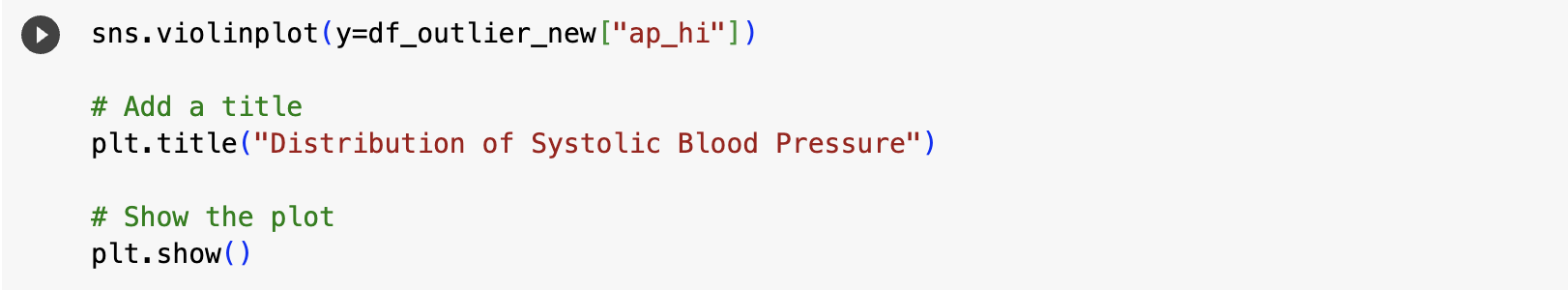
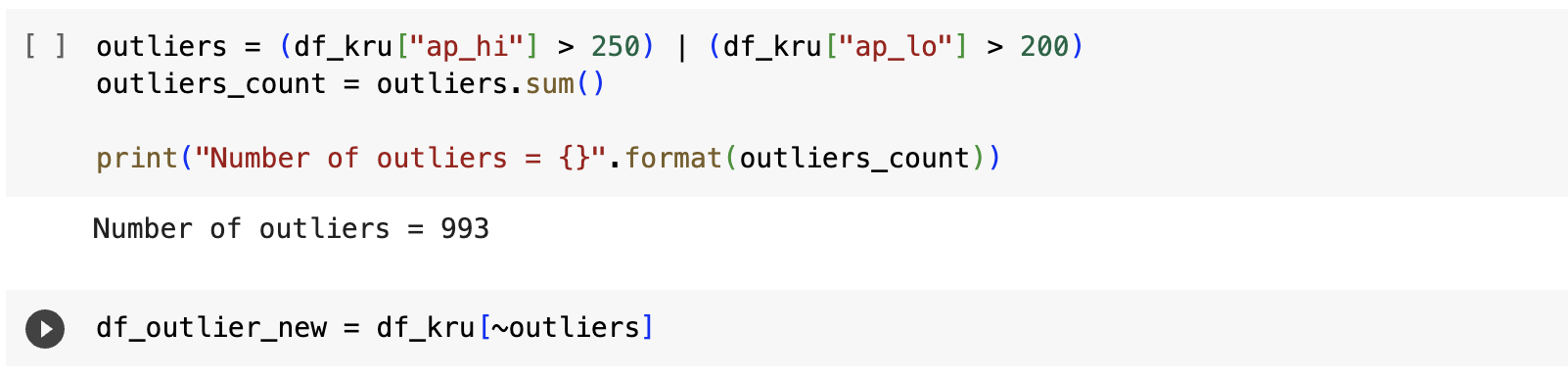
*The resulting pie chart visually represents the percentage distribution of "Not Active" and "Active" categories in the "physical activity" column*

**

**

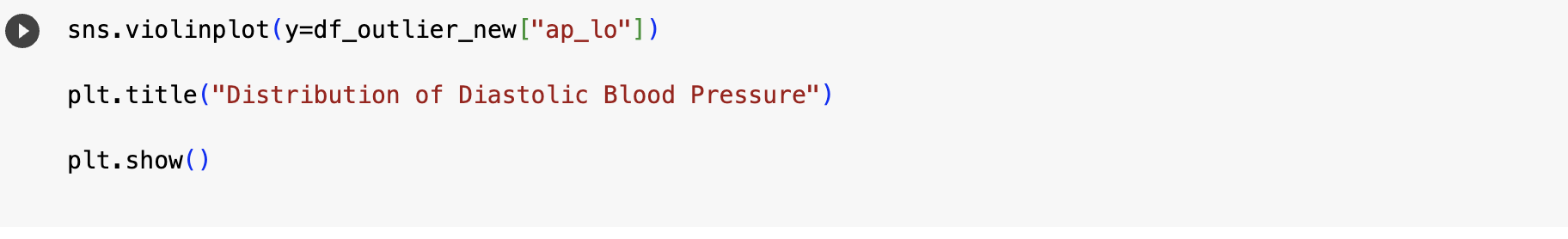
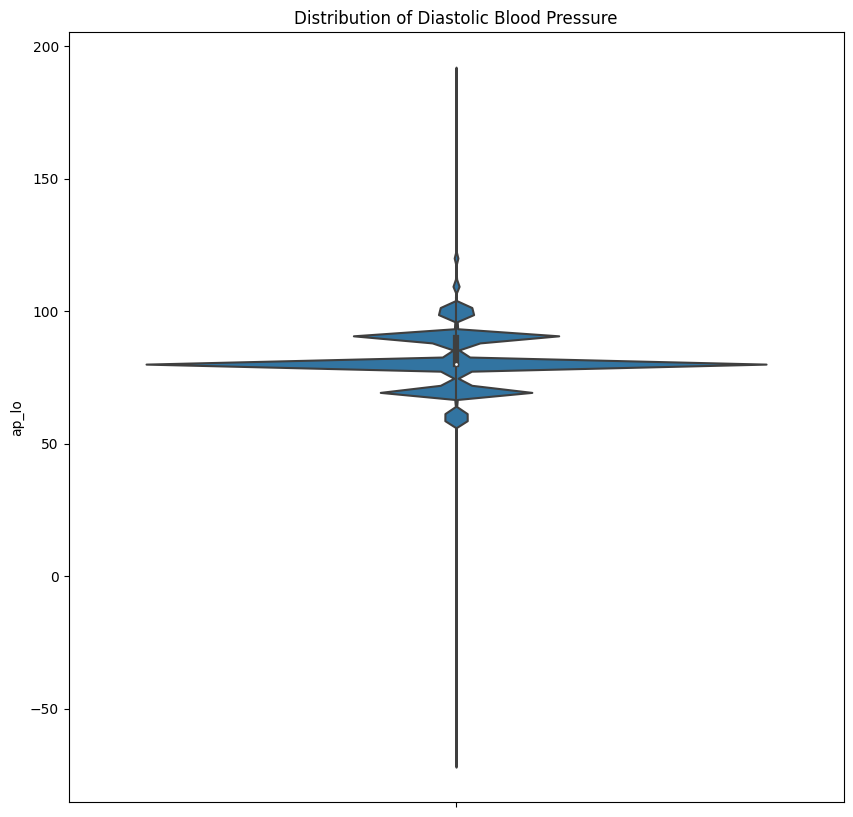
*The plot shows two bars for each physical activity category, distinguished by the "cardio" variable's hue. Different colours represent the presence or absence of cardiovascular disease*

*This graph shows that physically inactive people are affected by CVDIn the dataset there can be unusual values which are called outliers in the systolic and diastolic blood pressure. For systolic blood pressure the (ap\_hi) the values beyond 250 and for diastolic blood pressure (ap\_lo) the values beyond 200 are removed.*

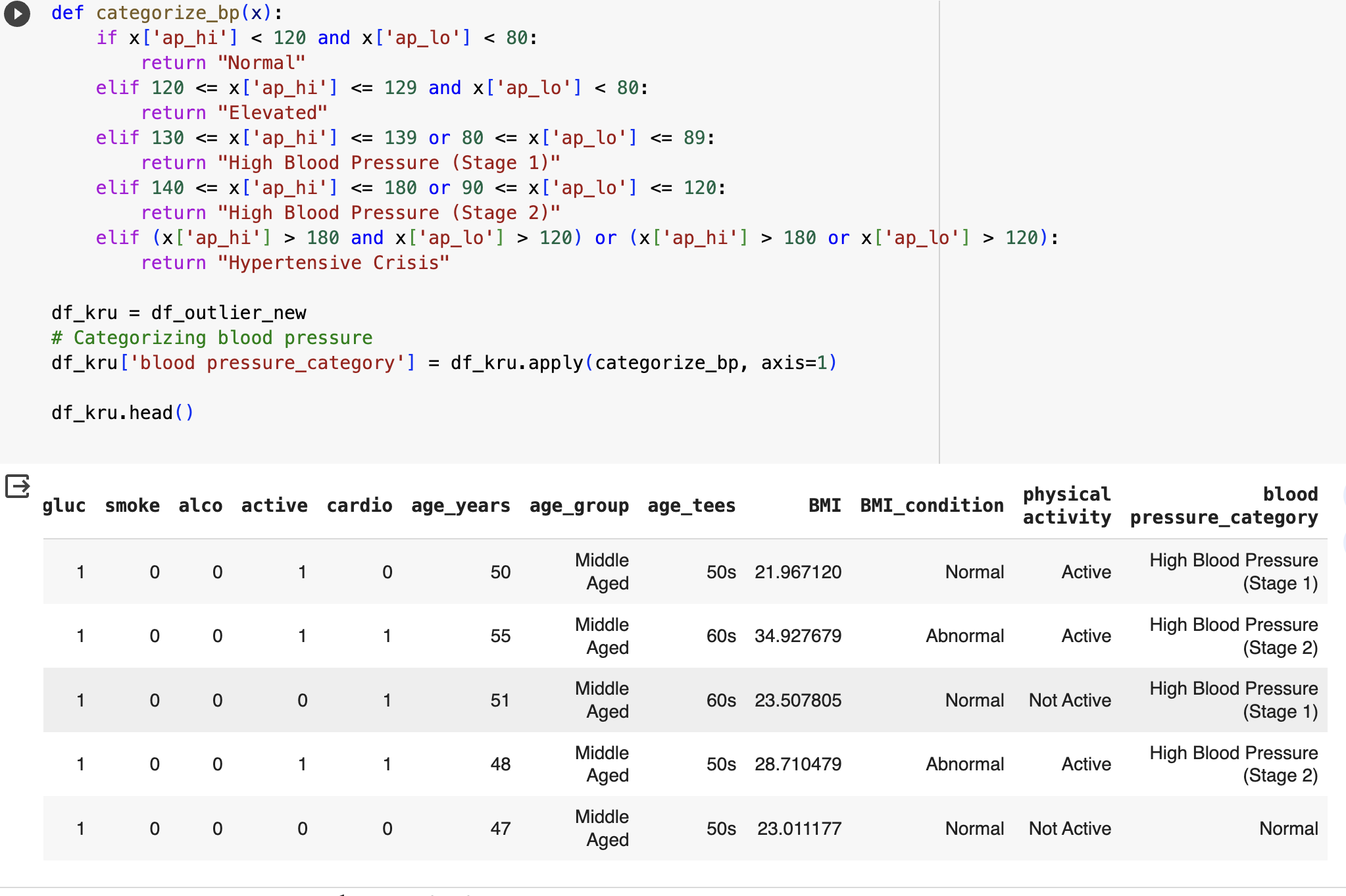
*2DLibt*

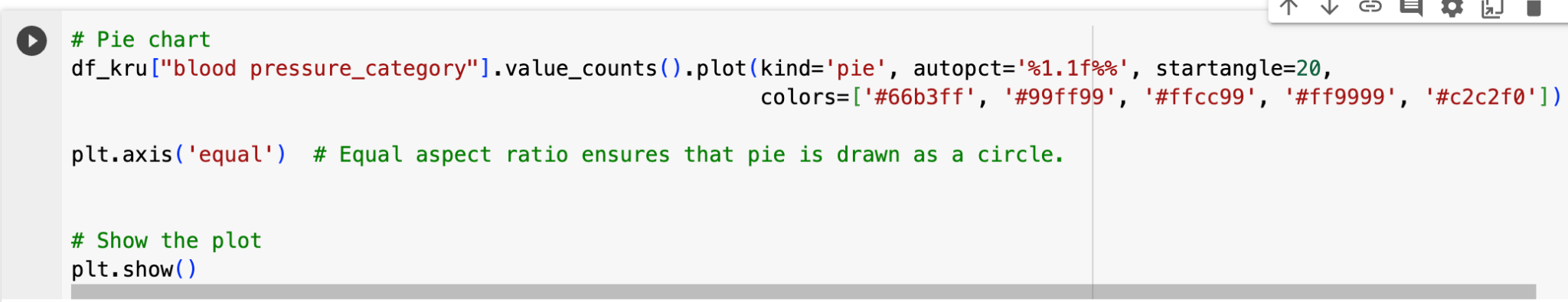
*The resulting violin plot illustrates the distribution of systolic blood pressure*

*This type of plot is particularly useful for visualizing the shape and distribution of a dataset, including the presence of multiple modes or unusual patterns*

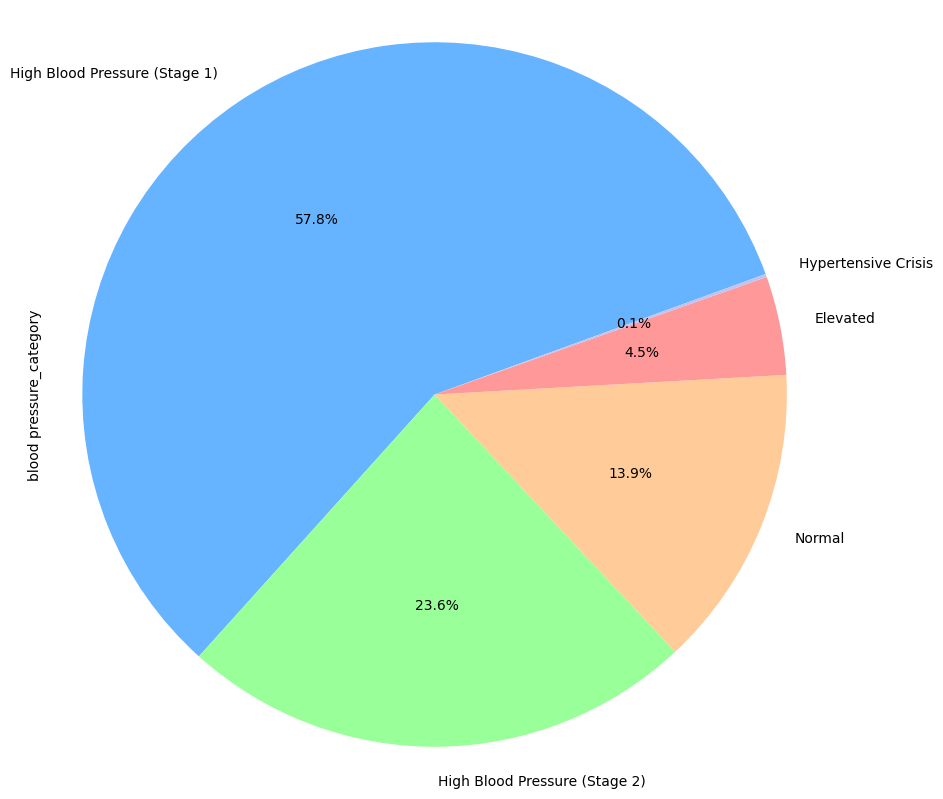
*The resulting violin plot visually represents the distribution of diastolic blood pressure*

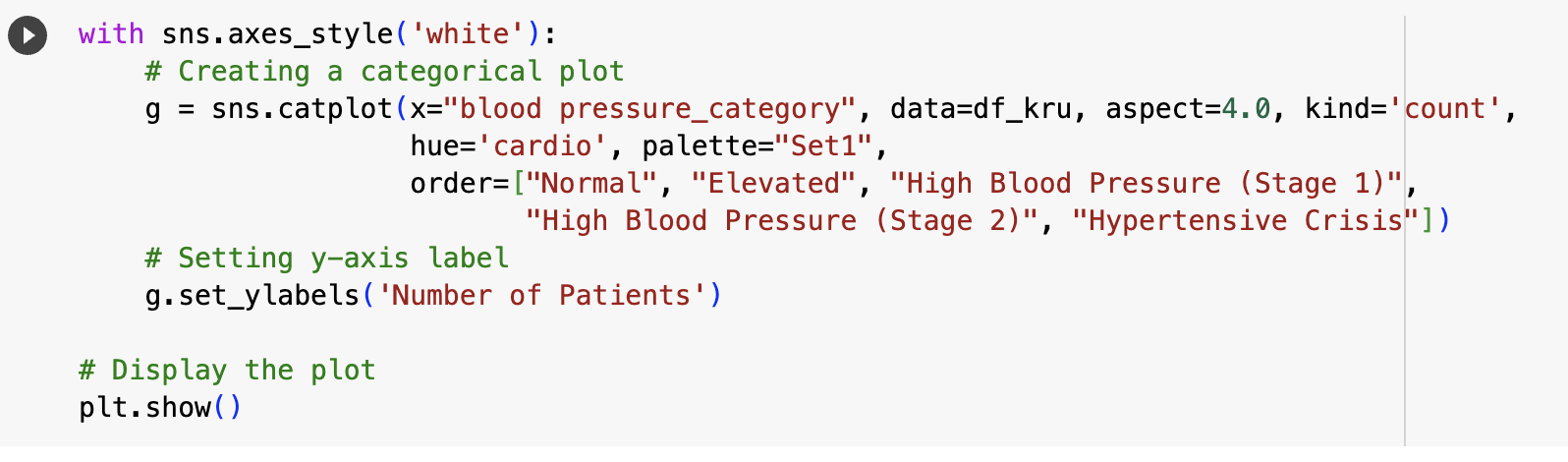
*This type of plot helps in understanding the spread and shape of the diastolic blood pressure data*

*categorizing the blood pressure and adding a new column which is blood\_pressure category.*

**

*The provided code generates a pie chart to visually represent the distribution of blood pressure categories in the DataFrame df\_kru.*

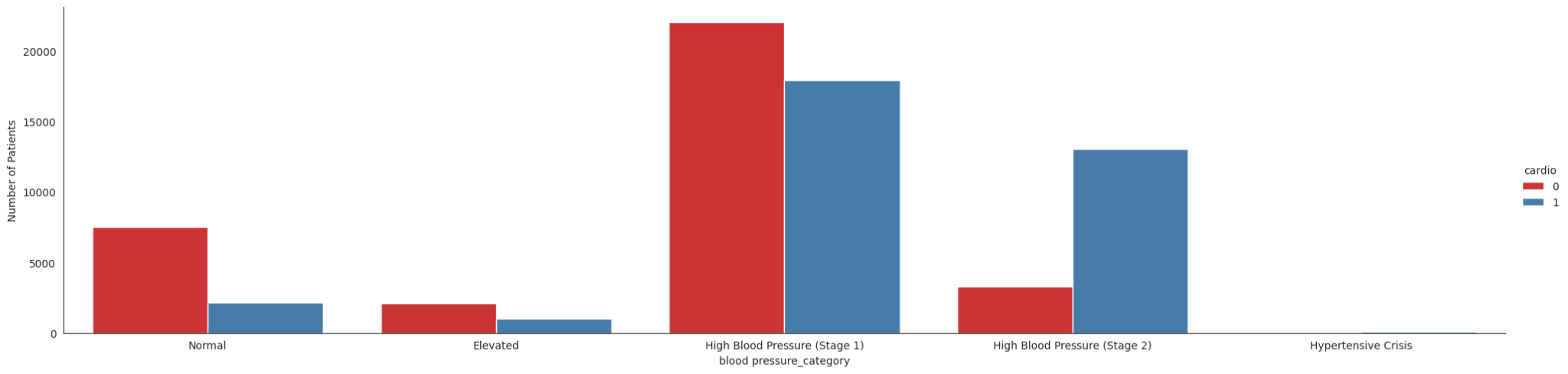
**

**

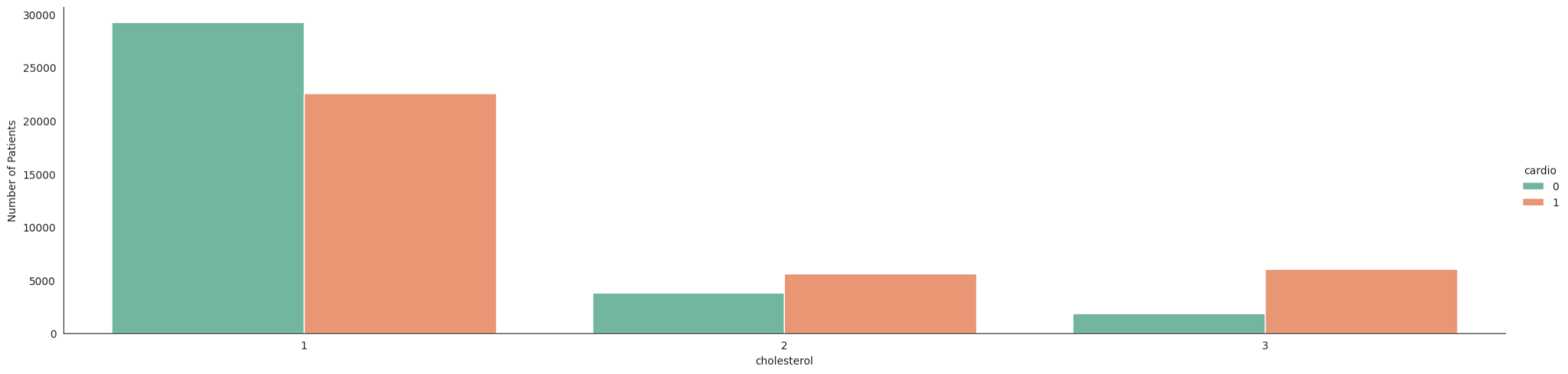
*The provided code generates a pie chart to visually represent the distribution of blood pressure categories in the DataFrame df\_kru*

*The resulting plot is a grouped bar chart where each bar represents the count of patients in different blood pressure categories, segmented by the presence or absence of cardiovascular disease. This visualization helps understand the distribution of cardiovascular disease across distinct blood pressure levels…*

*The graph indicates that people with High BloodPressure (stage-2) are more affected with CardioVascular disease.*

**

**

**

*The resulting plot is a grouped bar chart where each bar represents the count of patients in different cholesterol levels, segmented by the presence or absence of cardiovascular disease*

*The graph shows that people with very high cholesterol levels are mostly affected with cardiovascular disease when compared to Normal and High cholesterol levels*

***CONCLUSION:***

*The exploratory data analysis (EDA) conducted on the cardiovascular disease dataset revealed insightful patterns and correlations, particularly through visualizations such as bar charts, pie charts, donut charts. The identified relationships between age groups and cardiovascular disease presence showcased the potential of age as a significant factor in disease prevalence. The distinct patterns observed, where certain age groups exhibited higher rates of cardiovascular disease, provide valuable insights for further investigation and intervention strategies.*

*The EDA on BMI condition, cholesterol, and blood pressure underscores their significant associations with cardiovascular disease. Abnormal BMI, elevated cholesterol levels, and higher blood pressure contribute to increased cardiovascular risk. Monitoring these factors is crucial for preventive healthcare. Lifestyle interventions, regular screenings, and effective management are key in mitigating cardiovascular risks. Further research in these areas can enhance preventive strategies and improve overall cardiovascular health outcomes.*

***REFERENCES:***

DATASET SOURCE:INTERNET

* [*https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)*](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
* [*https://www.heart.org/en/health-topics/high-blood-pressure*](https://www.heart.org/en/health-topics/high-blood-pressure)
* [*https://www.uptodate.com/contents/high-cholesterol-and-lipids-beyond-the-basics/print*](https://www.uptodate.com/contents/high-cholesterol-and-lipids-beyond-the-basics/print)
* [*https://www.geeksforgeeks.org/data-visualisation-in-python-using-matplotlib-and-seaborn/*](https://www.geeksforgeeks.org/data-visualisation-in-python-using-matplotlib-and-seaborn/)
* [*https://seaborn.pydata.org/generated/seaborn.catplot.html*](https://seaborn.pydata.org/generated/seaborn.catplot.html)
* [*https://seaborn.pydata.org/generated/seaborn.violinplot.html*](https://seaborn.pydata.org/generated/seaborn.violinplot.html)