THE BURDEN OF CARDIOVASCULAR DISEASES ON PUBLIC HEALTH AND ECONOMY

Presented by: ***Group 8***

Dhruv Sood

Rishabh Parakh

Ravi Chadva

Ghali Tazi Hemida

Melina Azzouz

OVERVIEW

01

Business Case

02

Data Summary

03

EDA Steps

04

Payoff Matrix and Assumptions

05

Models And Metrics

06

Visulaizations and Actions

BACKGROUND

Cardiovascular diseases (CVDs) are a leading cause of mortality in the United States, claiming approximately 700,000 lives annually. This not only poses a significant health threat but also places a substantial economic burden on healthcare systems due to treatment costs and lost productivity.

BUSINESS CASE

Utilizing predictive analytics on a dataset of over 20,000 individuals, this analysis aims to **identify key risk factors for cardiovascular disease (CVD)**, such as age, gender, biometric data, blood pressure, cholesterol levels, and lifestyle habits.

By uncovering these relationships, it seeks to **inform targeted interventions and strategies** for more effective prevention and management of cardiovascular diseases, thereby reducing their impact on individuals and society.

DATA SUMMARY

Source: *NIH: Prediction of cardiovascular disease risk based on major contributing features, Mar 2023, 20,000 records*

Target Variable

'cardio' is the target variable and indicates whether a person has cardiovascular disease.

Presence of CVD?	Count	%
Yes = 1	10,027	50.135%
No = 0	9973	49.865%

Predictors

- **age:** Age of the participant in days.
- **gender:** Gender of the participant (1: female, 2: male).
- **height:** Height of the participant in centimeters.
- **weight:** Weight of the participant in kilograms.
- **ap_hi:** Systolic blood pressure (mmHg).
- **ap_lo:** Diastolic blood pressure (mmHg).
- **cholesterol:** Cholesterol level (1: normal, 2: above normal, 3: well above normal).
- **gluc:** Glucose level (1: normal, 2: above normal, 3: well above normal).
- **smoke:** Smoking status (0: non-smoker, 1: smoker).
- **alco:** Alcohol intake (0: non-drinker, 1: drinker).
- **active:** Physical activity (0: inactive, 1: active).
- **cardio:** Presence of cardiovascular disease (0: no, 1: yes).

EDA STEPS

- Removed the "**id**" column from the dataset.
- Transformed the "**age**" column into "**age_years**" by int[age]/365.
- Retained the "**gender**" column as it is.
- Transformed the "**height**" column into "**height_meters**" by dividing the values by 100.
- Retained the "**weight**" column as it is.
- Kept the "**ap_hi**" and "**ap_lo**" columns as they are.
- Transformed the "**cholesterol**" column into a categorical variable.
- Transformed the "**gluc**" column into a categorical variable.
- Kept the "**smoke**", "**alco**", **and** "**active**" columns as they are.
- Designated the "**cardio**" column as the target variable.

CONFUSION MATRIX

TP — The organization correctly identifies individuals with CVD, allowing for appropriate interventions and management.

TN — The organization correctly identifies individuals without CVD, avoiding unnecessary interventions and ensuring efficient allocation of resources.

FP — The organization incorrectly identifies individuals without CVD as having the condition, leading to unnecessary interventions or treatments, which may result in additional costs and potential harm.

FN The organization fails to identify individuals with CVD, leading to missed opportunities for early intervention and potentially worsening health outcomes.

NOTE

We've opted not to utilize a payoff matrix in this case due to the variability in treatment costs for each individual and the sensitivity of healthcare matters. Instead, our focus is on achieving optimal model performance in accurately predicting the presence of Cardiovascular disease. We prioritize the effectiveness of our model in this regard, aiming for high efficiency and reliability in identifying individuals at risk.

MODELS AND METRICS

Model	Recall	Precision	F1	ROC AUC	Accuracy
RuleFit Classifier (0.331)	0.8559	0.6555	0.7424	0.7951	0.7022
RandomForest Classifier (0.375)	0.7976	0.6924	0.7413	0.7971	0.7209
Neural Network (0.369)	0.8288	0.6653	0.7381	0.7926	0.7051
SVM (0.336)	0.8571	0.6455	0.7364	0.7905	0.6923
Logistic regression (0.389)	0.8352	0.6384	0.7237	0.7689	0.6802

RANDOMFOREST CLASSIFIER ACHIEVES BEST MODEL PERFORMANCE

Using the ROC AUC metrics to evaluate model performance, **the RandomForest Classifier is the best-performing model.**

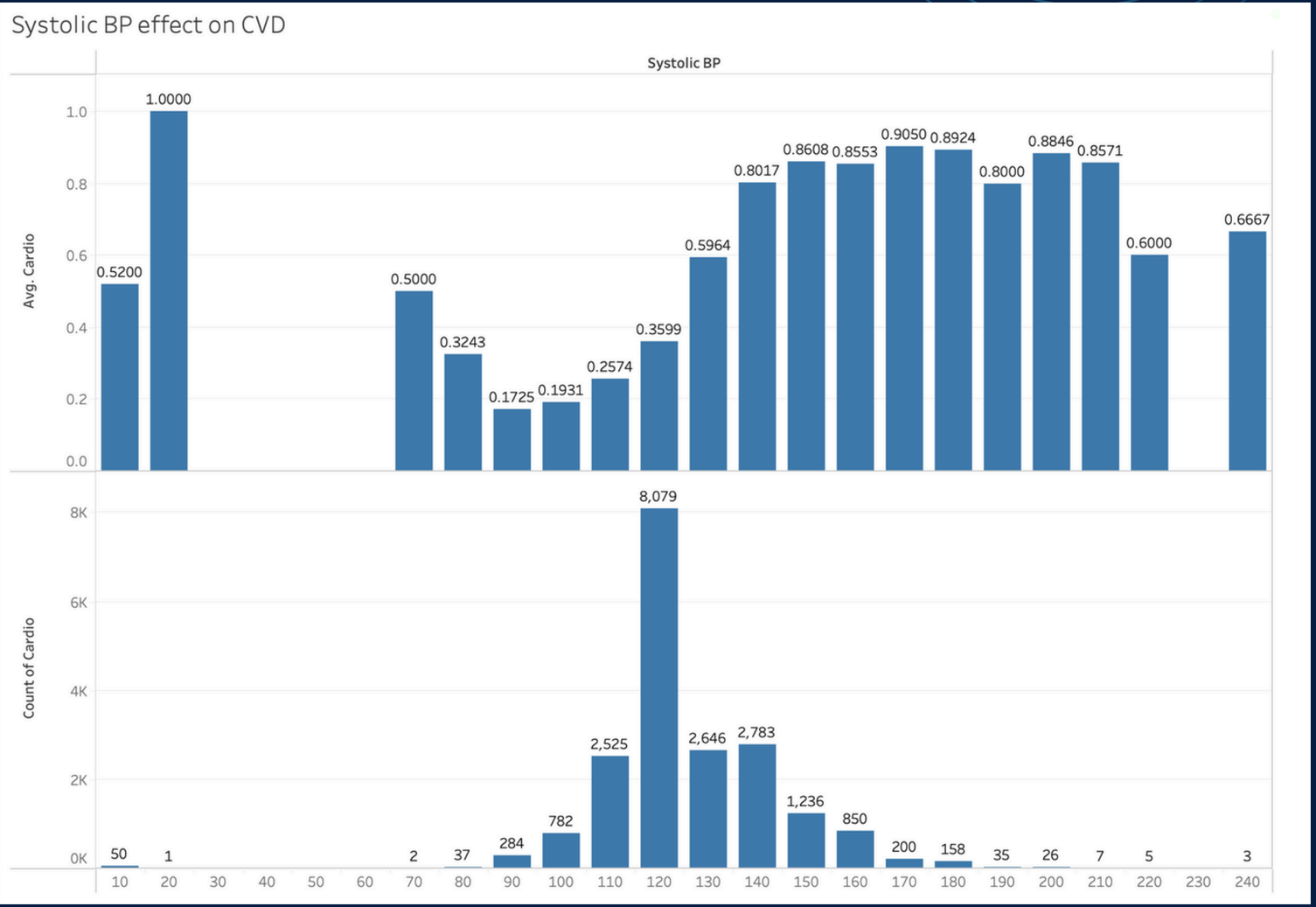
- In this analysis, our primary concern is accurately predicting the presence of disease in individuals. Therefore, we prioritize the **ROC AUC metric for model evaluation**. ROC AUC provides a comprehensive measure of the model's ability to discriminate between positive and negative cases across various thresholds, ensuring robust performance in disease detection.
- In healthcare analyses where minimizing misclassification costs is crucial, we prioritize the maximum payoff metric. However, in our focus on accurately predicting disease presence, we prioritize the ROC AUC metric.

FEATURE IMPACT ANALYSIS

Top Predictors:

- Systolic Blood Pressure
- Age
- Cholesterol Level
- Diastolic Blood Pressure
- Weight

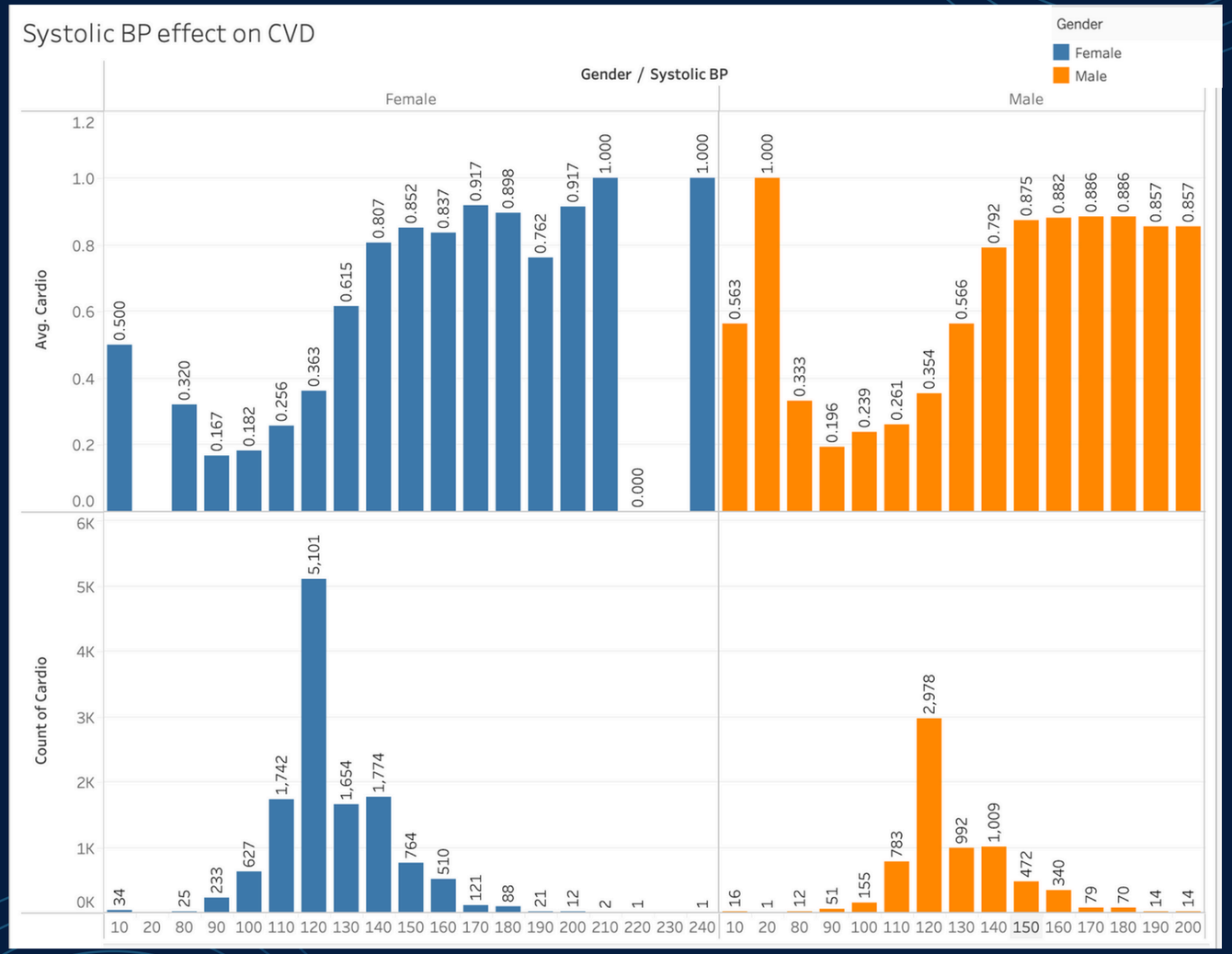
Effect of Systolic BP on CVD Prevalence



High Systolic BP and CVD Prevalence: A noticeable trend suggests higher systolic BP levels generally correlate with a higher average ratio of CVD occurrence. For instance, BP levels of 140 mm Hg and above show a consistent increase in the ratio, peaking around 160 mm Hg and slightly declining afterward but remaining relatively high compared to lower BP levels.

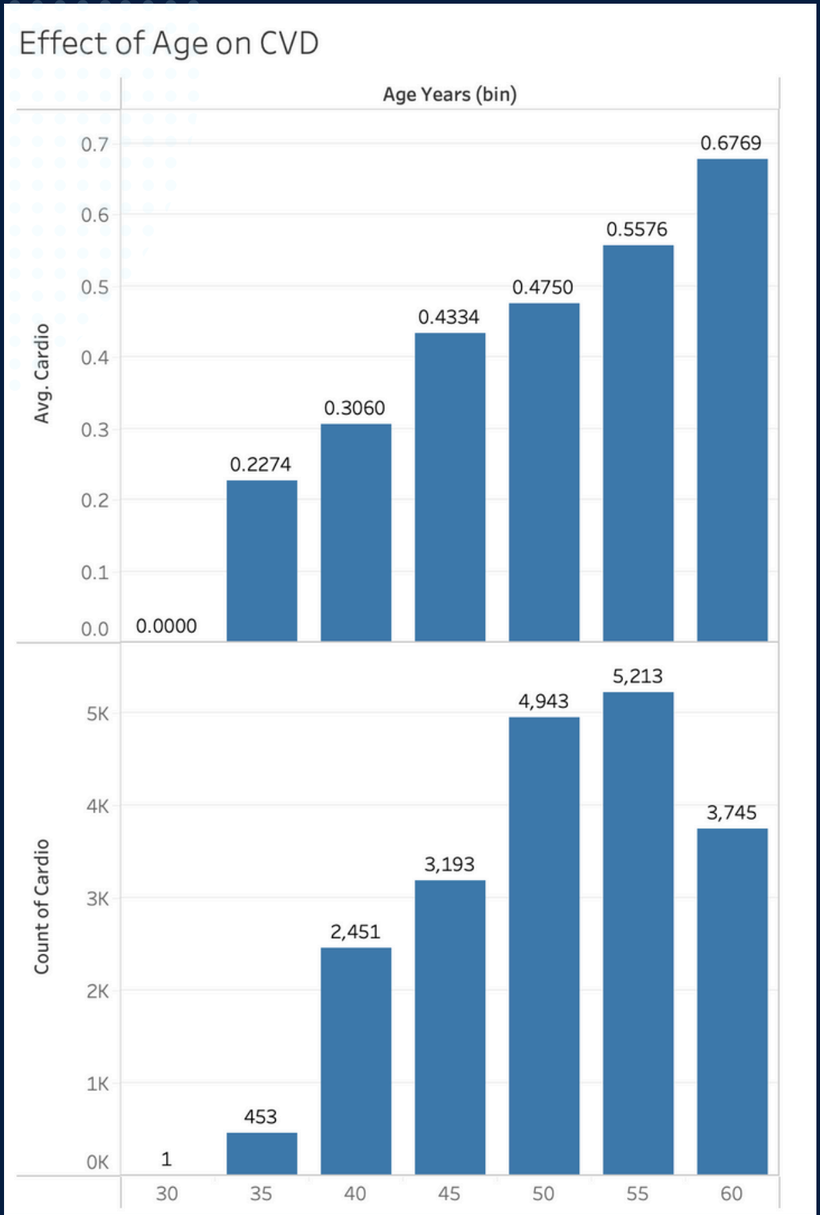
Although this visualization suggests that **systolic BP is a significant factor in predicting CVD**, the relationship is not uniform across all BP ranges, and therefore highlights the need to consider other factors that could affect CVD risk, such as weight and cholesterol level.

Effect of Systolic BP and Gender on CVD Prevalence



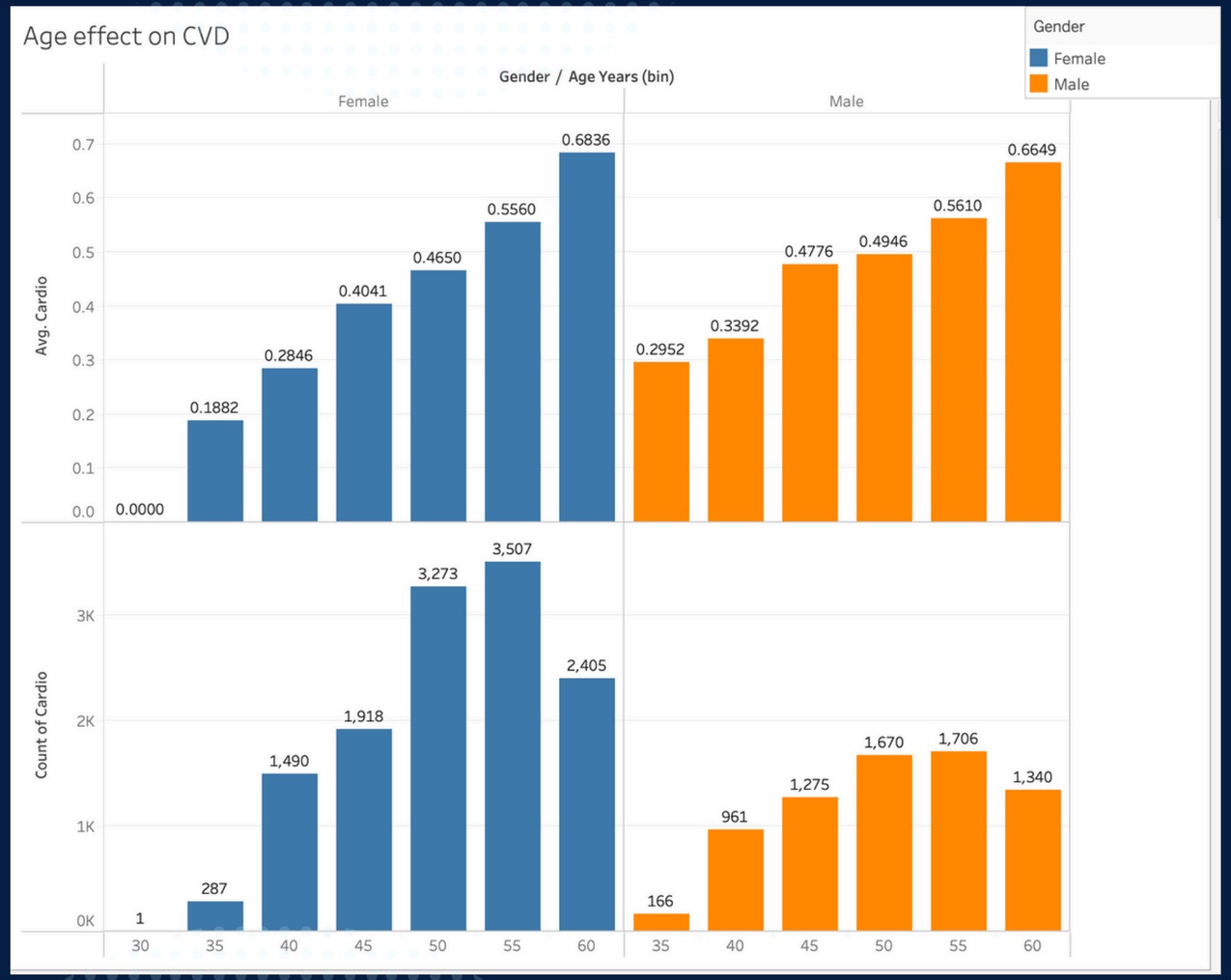
- The data reveals significant differences in cardiovascular disease (CVD) risk between genders, with **males consistently at higher risk across all blood pressure ranges**. This suggests physiological or lifestyle differences that influence heart health, underscoring the need for tailored healthcare interventions for each gender.
    - Additionally, the consistent increase in CVD risk with rising systolic blood pressure for both genders highlights the critical importance of regular blood pressure monitoring and management, particularly for those at greater risk.
- The majority of study participants have blood pressures within the 110–130 mmHg range, suggesting this is the 'normal' range for the general population. However, a notable decline in data points above 140 mmHg indicates that these higher levels are less common or potentially more hazardous. This observation is crucial for healthcare planning, particularly in enhancing screening and outreach efforts to manage high blood pressure effectively among vulnerable groups.

Effect of Age on CVD Prevalence



- This visualization illustrates the percentage of individuals diagnosed with cardiovascular disease across different age groups
- It highlights a clear trend: **the prevalence of CVD increases significantly with age**.
  - Starting from a negligible presence in individuals in their early 30s, the CVD rate escalates to 67.69% by the time individuals reach their 60s. The rate increases more steeply after the age of 45, indicating this as a critical period for the onset and escalation of cardiovascular conditions.
- **Risk Factor Correlation**: **The steep increase in CVD rates with age underscores the importance of monitoring age-related factors**, possibly including blood pressure, cholesterol levels, and other metabolic changes. This data supports the need for more rigorous and frequent cardiovascular screenings as individuals age, particularly after mid-life.
- **Preventive Health Measures**: Given the marked increase in CVD prevalence from the mid-40s, implementing preventive health measures, lifestyle interventions, and possibly early medication regimes could help mitigate risk factors before they result in CVD.
- **Targeted Healthcare Interventions**: Healthcare providers and policymakers might use this data to plan for resources, including education, healthcare services, and community support targeted at older adults to manage and prevent CVD.

Effect of Age and Gender on CVD Prevalence



- The graph displays the progression of cardiovascular disease prevalence by age for males and females, represented by the average percentage of individuals with CVD within each age bin.

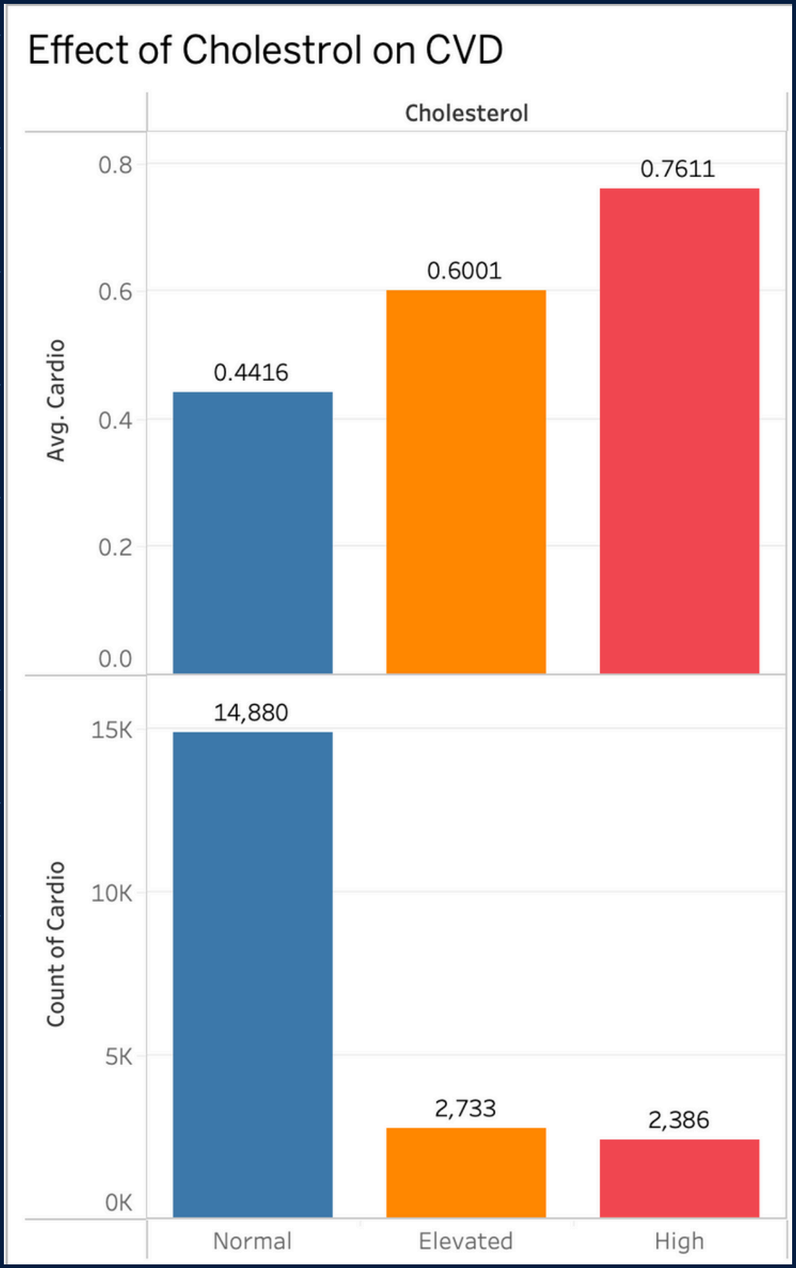
- **Female Trends:** Starting with virtually no cases in the early 30s, there is a progressive and significant increase in CVD rates among females, reaching 68.36% by their 60s. This shows a substantial escalation particularly after the age of 45.

- **Male Trends:** CVD prevalence among males starts higher than in females from the early 30s (29.52%) and shows a steady increase, reaching almost 49.46% by their 50s and peaking at 66.49% by their 60s. The rate of increase is more uniform compared to females.

- **Comparative Analysis**: In early age groups (30s to early 40s), males exhibit a higher prevalence of CVD compared to females. However, by the age of 60, the prevalence among females slightly exceeds that of males, indicating a potentially more aggressive progression of CVD in later years among females.

- **Gender-Specific Healthcare:** The data suggests that healthcare providers should consider gender-specific strategies for CVD prevention and treatment, especially as the progression and onset age of CVD appear to differ between genders.

Correlation Between Cholesterol Levels and CVD Prevalence



- This graph presents a clear comparison of cardiovascular disease prevalence across three cholesterol categories: normal, elevated, and high.

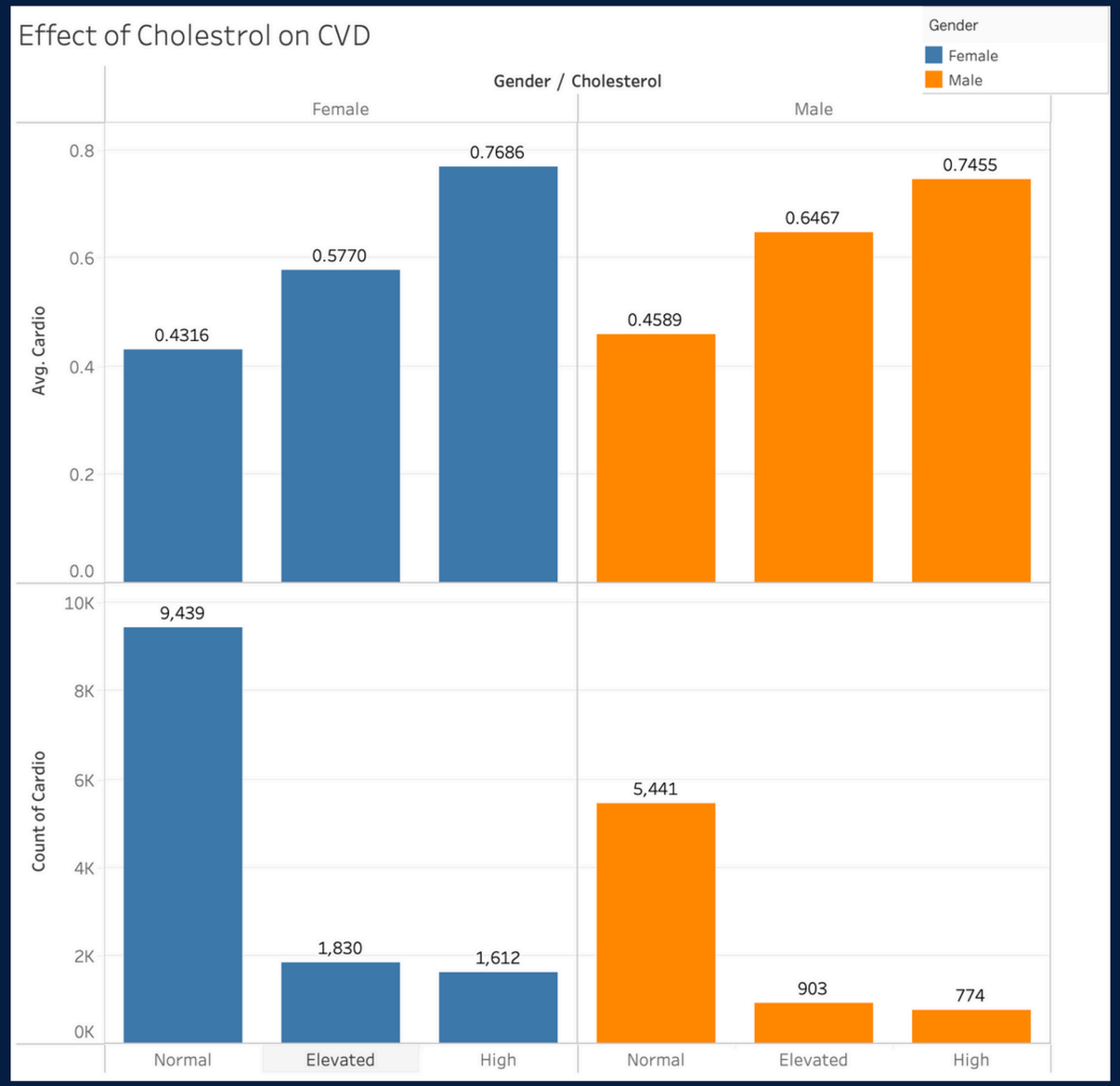
- **Increase with Cholesterol Levels**: There is a notable increase in the prevalence of cardiovascular disease as cholesterol levels rise. Approximately 44.16% of individuals with normal cholesterol levels have CVD, which escalates to 60.01% with elevated cholesterol, and further to 76.11% with high cholesterol.

- **Count Distribution:** The count of individuals varies significantly across the cholesterol categories. The majority of the surveyed population falls under the normal cholesterol category (14,880 individuals), while fewer individuals are categorized under elevated (2,733) and high (2,386) cholesterol levels.

- **Significant Risk Increase**: The data reveals a substantial risk increase in CVD with higher cholesterol levels, suggesting that cholesterol management could be crucial in preventing cardiovascular disease.

- **Targeted Interventions**: The strong correlation between elevated and high cholesterol levels with increased CVD prevalence supports the need for targeted interventions focusing on cholesterol management. Public health initiatives could include screening programs, dietary guidelines, and lifestyle interventions aimed at reducing cholesterol levels.

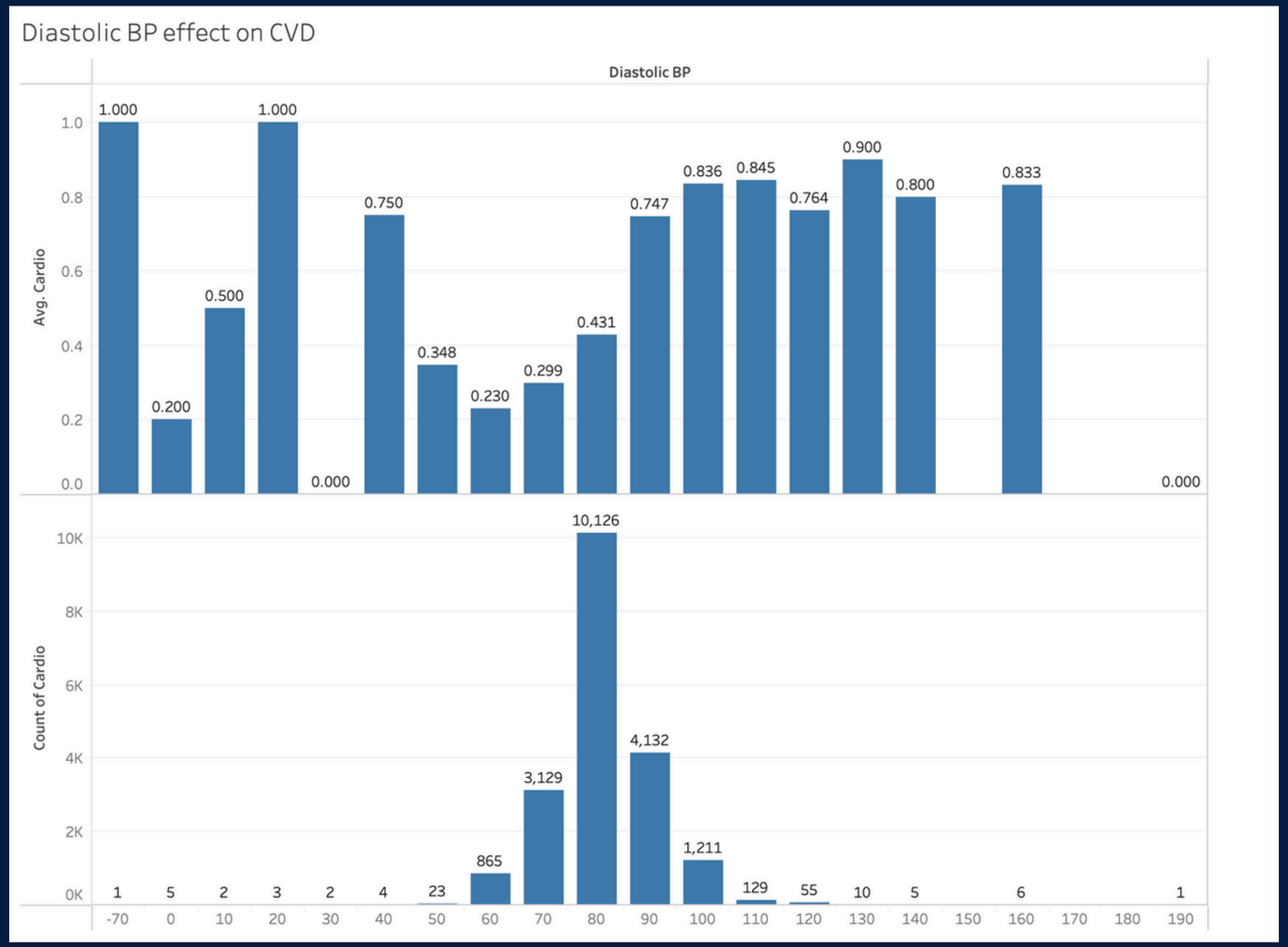
Gender-Specific Impact of Cholesterol Levels on CVD Prevalence



This chart provides a clear visualization of how cholesterol levels impact the prevalence of cardiovascular disease, broken down by gender.

- Cholesterol Impact on CVD:
  - Females: There is a progressive increase in CVD prevalence with rising cholesterol levels. The average rate of CVD among females is 43.16% for those with normal cholesterol, which jumps to 57.70% for those with elevated levels and peaks at 76.86% for those with high cholesterol.
  - Males: Similarly, males show an increase in CVD rates from 45.89% at normal cholesterol levels to 64.67% at elevated levels and 74.55% at high cholesterol levels.
- Comparison by Gender:
  - Females generally show a higher rate of CVD than males at high cholesterol levels. However, for normal and elevated cholesterol levels, males have slightly lower rates of CVD than females.
- Population Distribution:
  - There are more females with normal and high cholesterol compared to males, while more males have elevated cholesterol levels.

Effect of Diastolic BP on CVD Prevalence



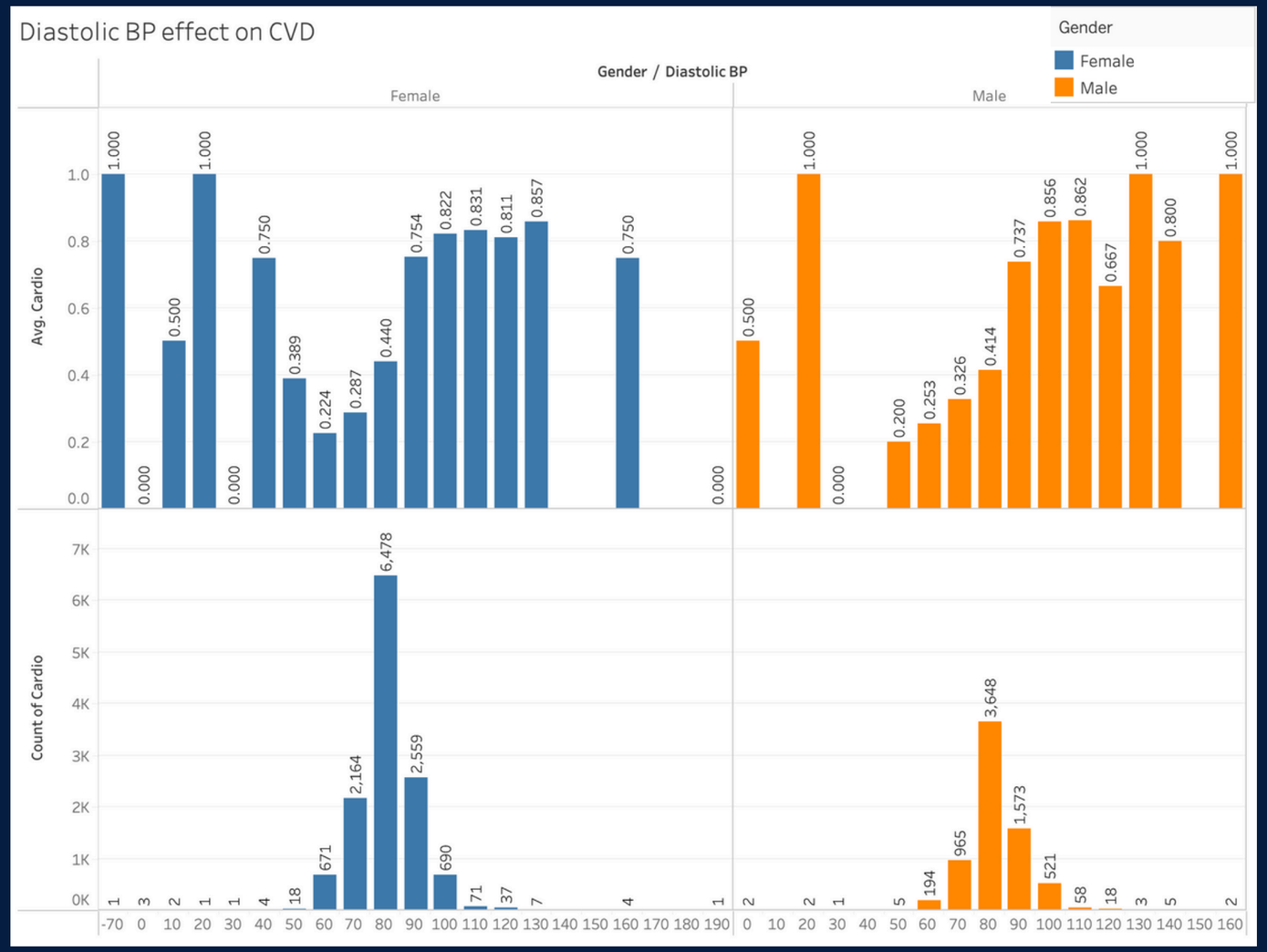
The chart indicates that:

- Extreme diastolic blood pressures, both very low (below 20 mmHg) and very high (above 180 mmHg), are linked to the highest risks of cardiovascular disease (CVD), indicating serious health risks at these levels.

- Moderate diastolic BP levels, particularly from 60 to 90 mmHg, present variable but generally safer CVD risk, suggesting this range as optimal for maintaining cardiovascular health.

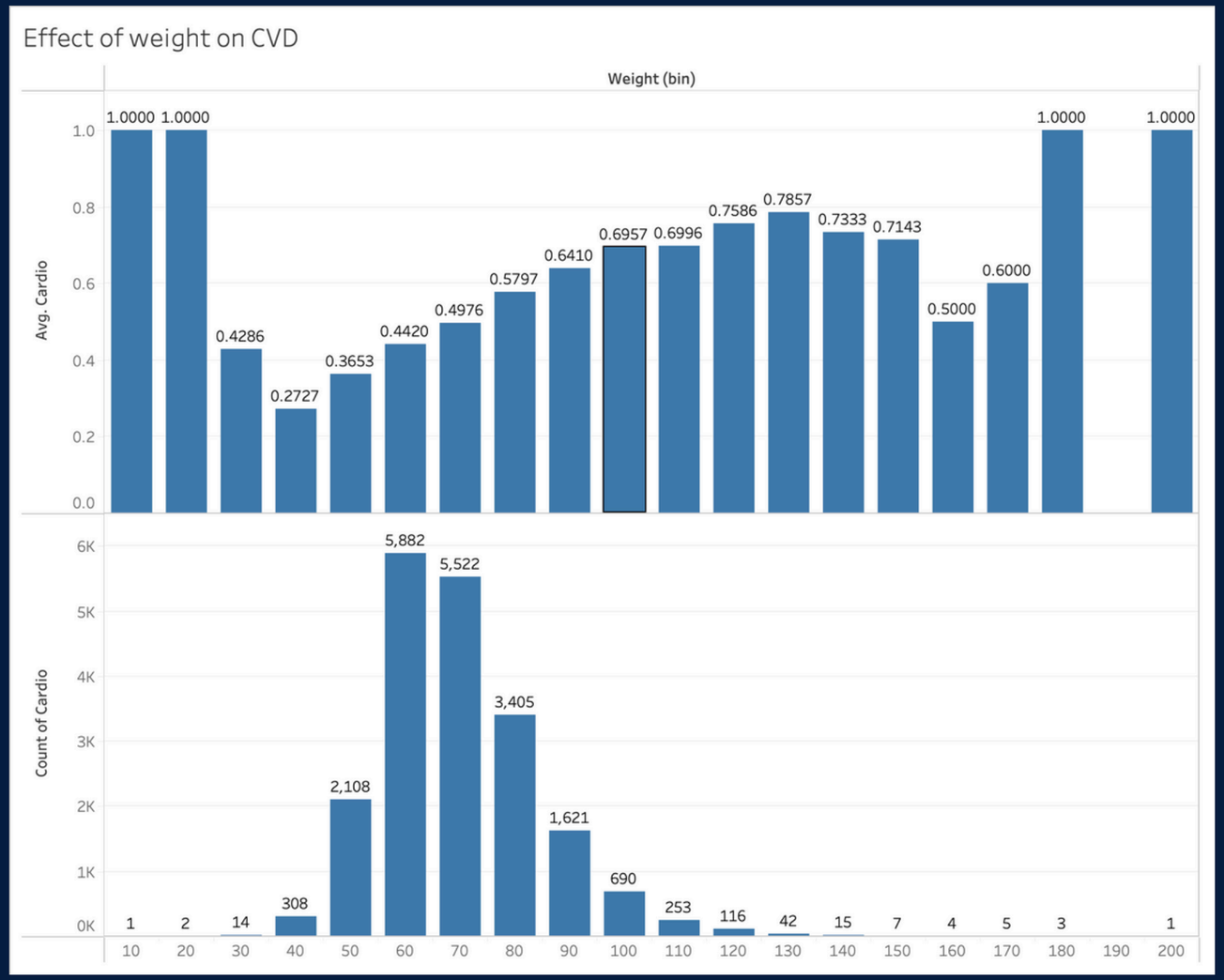
- As diastolic BP increases beyond this moderate range, the risk of CVD begins to escalate, highlighting the importance of managing BP to prevent progression into higher risk categories.

Effect of Diastolic BP on CVD Prevalence by Gender



- Men show a consistently higher average risk of cardiovascular disease at all diastolic blood pressure levels, suggesting underlying biological or lifestyle factors that impact gender differently.
- Normative diastolic blood pressure ranges are around 70 mmHg for women and 80 mmHg for men, providing a baseline for identifying and addressing elevated cardiovascular risk.
- There is a strong correlation between rising diastolic blood pressure and increasing cardiovascular disease risk, especially significant beyond 80 mmHg for women.
- Maintaining diastolic blood pressure within optimal ranges is crucial for effectively mitigating cardiovascular risks, emphasizing the need for careful management and monitoring.

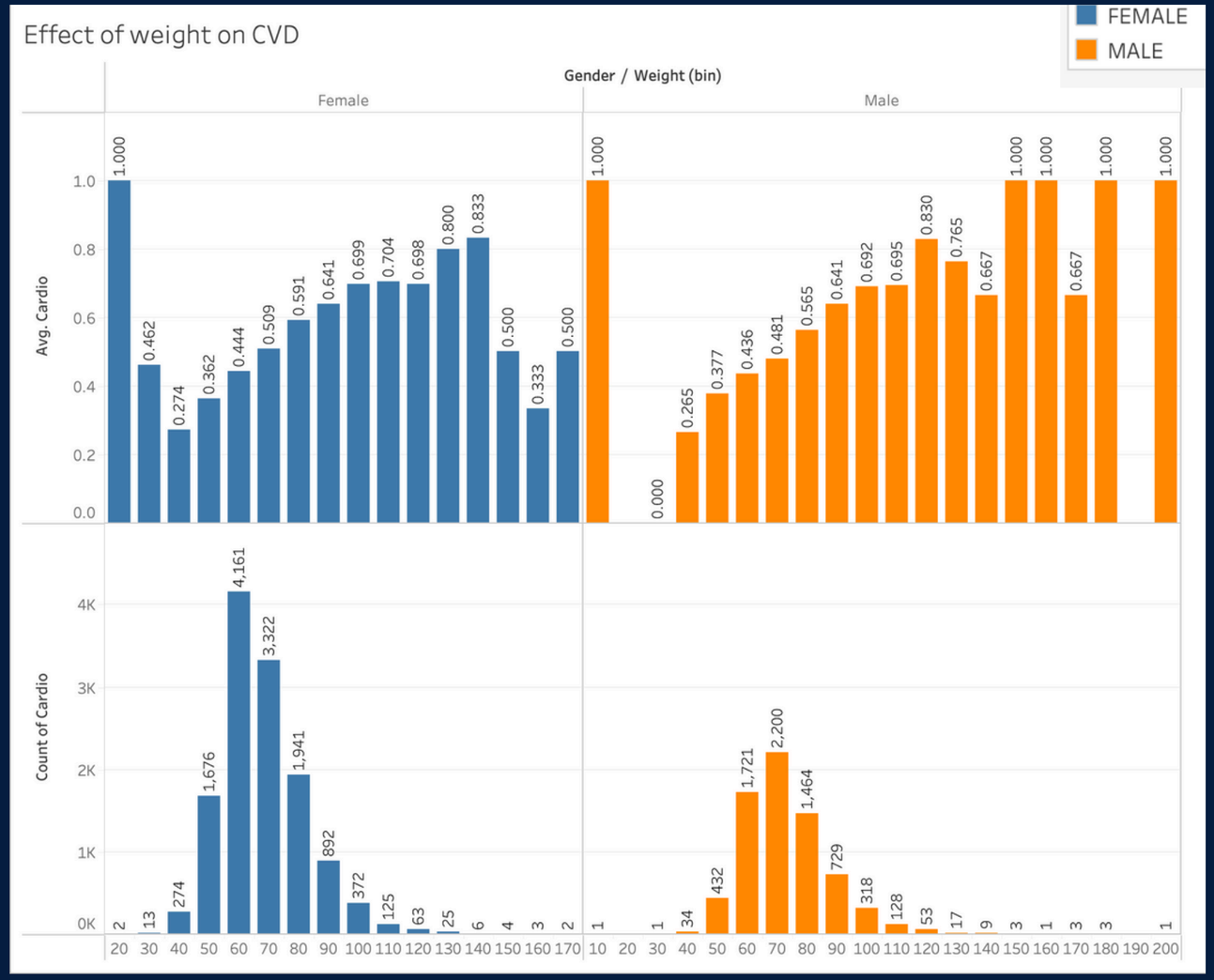
Effect of Weight on CVD Prevalence



The graph depicts that :

- Extreme weight categories—both significantly underweight and overweight—are associated with the highest cardiovascular disease (CVD) risk, emphasizing the dangers of weight extremes.
- The risk of CVD decreases towards more typical weight ranges, suggesting that maintaining a moderate weight could enhance cardiovascular health, although this relationship is complex.
- The majority of the population falls within the 50-70 kg weight range, highlighting it as the most common weight category in the study.
- Observations decline sharply for weights above 130 kg, suggesting research and healthcare coverage gaps for individuals at higher weights and emphasizing the need for inclusive health studies

Effect of Weight on CVD Prevalence by Gender



- Females exhibit significant variability in cardiovascular disease (CVD) risk with weight, peaking at 90 kg, which suggests a strong correlation between weight fluctuations and increased CVD risk for women.
- Males show a steadier CVD risk profile with less variability, maintaining a generally higher risk as weight increases, indicating the need for continuous weight management.
- The data reveals that females tend to cluster in the lower weight ranges, particularly around 50 kg, while males have a broader distribution with a notable concentration at 70 kg.
- These observations emphasize the necessity for gender-specific health interventions, particularly focusing on the most common weight ranges to effectively manage and reduce CVD risks.

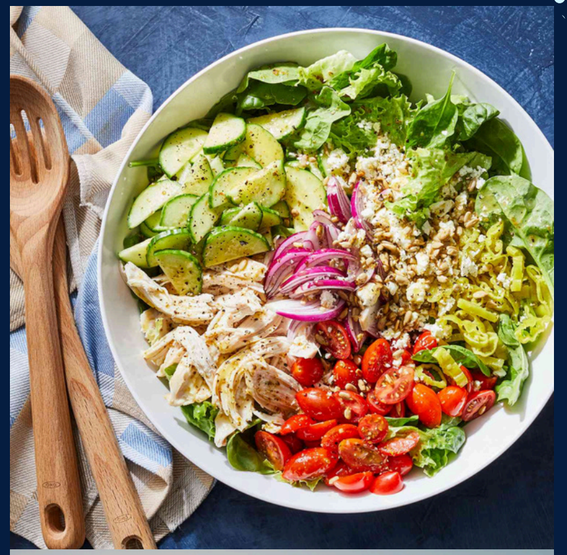
RECOMMENDED ACTIONS

It is hard to recommend any actions to be taken for each specific individuals as we are not medical experts.

While looking at the top predictors for our model, it is safe to say that promoting healthy habits and lifestyle could lead to some potential improvements concerning patients overall health, but will not lead to a significant decrease for Cardio Vascular Diseases.

While looking at certain predictors such as the systolic and diastolic blood pressures, some medications and best practices exist to reduce them. Sometimes this high blood pressure can exist because of another disease therefore it needs to be treated first.

Adapted medications also exist for treating high cholesterol levels, as well as radical surgical interventions such as a LDL apheresis (procedure that removes the artery-clogging LDL cholesterol from the blood).



Each case is individual, therefore we need medical consultancy

Thank you!

Any quesitons?