# Heart Attack Risk Factors Analysis - Indonesia

**Prepared by:**

*Fatma Mahmoud*

*Data Analyst*

**Date:** July 2025

# Table Of Contents

# Introduction

- ## *Objective:*

  The primary objective of this analysis is to identify the key determinants associated with the increased risk of heart attacks in Indonesia.

  Specifically, the study investigates the impact of certain health behaviors — including smoking, hypertension, diabetes, and genetic predisposition — on heart attack prevalence.

  It also examines how demographic factors such as age and gender correlate with infection rates, with the aim of informing public health awareness campaigns and preventive interventions targeted at high-risk groups.

- ## *Significance Of the Analysis:*

  This analysis holds significant value as it provides a deeper understanding of the factors that contribute to the risk of heart attacks. By examining demographic characteristics and health-related behaviors, the study reveals patterns and relationships that are crucial for public health planning. The significance of this analysis lies in:

  Uncovering high-risk populations based on age, gender, lifestyle, and socio-economic factors.

  Highlighting the role of unhealthy behaviors such as smoking, high blood pressure, diabetes, and genetic predisposition in increasing heart attack rates.

  Providing evidence-based insights to guide awareness campaigns and prevention programs.

  Clarifying the impact of factors previously assumed to be influential, such as access to healthcare, which showed limited effect in this case.

  Enabling better targeting of healthcare resources and education based on data rather than assumptions.

  Through this analysis, stakeholders can focus on what truly matters, leading to more effective interventions and improved health outcomes.

- ## *Scope:*

  This analysis explores a wide range of factors potentially associated with heart attack risk.

  It covers geographic distribution (including states and urban vs. rural areas), behavioral factors (such as smoking, stress, high blood sugar, genetic predisposition, alcohol consumption, physical activity, sleep levels, and eating habits), and medical indicators (including BMI, cholesterol levels, and access to medical resources).
  Demographic variables were also analyzed, including age, gender, education level, marital status, income level, and employment status.
  Additionally, the analysis examines the relationship between combined factors (such as unemployment and marital status), the number of risks factors a person has, and environmental factors like the air quality index (AQI).

  It is important to note that the dataset does not include any time-related fields, and therefore this analysis does not consider trends over time. Additionally, the meaning and categorization of pollution levels are not clearly defined in the dataset, which may limit the interpretability of related insights.
  This analysis does not aim to diagnose or predict disease outcomes but rather to highlight potential patterns and areas of concern that may inform awareness and prevention efforts.

- ## *Data Sources:*

  The dataset used in this project was obtained from Kaggle, a well-known platform for open data. The data was provided in CSV format, and it did not include any temporal (time-based) components, meaning it cannot be used to assess trends or changes over time.

  Although the dataset was pre-organized, it required further preparation and cleaning. This included creating calculated columns, applying transformations such as unpivoting, and adjusting categorical variables to suit the analytical needs of the project.

- ## *Tools:*

  Microsoft Excel (Pivot Tables, Power Pivot, Power Query)

- ***Outcome:***

1- Key Findings:

The regions identified as high-risk for heart attacks include Pekanbaru, Yogyakarta, Medan, Balikpapan, Surabaya, Bandung, Palembang, and Batam.

Among health behaviors, high blood pressure and diabetes are the most strongly associated with increased heart attack risk, particularly in females.

Regarding BMI, individuals classified as overweight, obese case 1, or even normal weight have shown higher rates of infection compared to other BMI categories.

When examining cholesterol levels, extremely high cholesterol poses the greatest risk, followed by even those with normal levels, indicating other interacting factors.

The age group between 36 and 50+ shows a significant increase in risk.

Unemployment combined with being married is also associated with slightly higher heart attack risk.

When individuals possess two or three risk factors (e.g., smoking, diabetes, stress), their infection rate often exceeds the overall average.

A very unhealthy AQI level is also linked to increased heart attack occurrences.

2- Less Significant Factors:

Access to medical resources showed no clear effect on infection rates whether access is easy or limited, the infection rates remained high.

Similarly, urban vs. rural residence showed no consistent difference in infection levels.

Some health behaviors like alcohol consumption, physical inactivity, and poor diet were generally associated with increased infection, but they affected both genders and all groups, making them broad indicators rather than targeting specific populations.

The analysis also included factors such as gender, marital status, income level, and education level, but they did not show a clear significant association with increased infection rates.

# Dashboard Overview

- This analytical dashboard is composed of five main pages, each focusing on a different category of variables that may impact the risk of heart attack:

  1. Geographic

  2. Health Behavior

  3. Health Indicators

  4. Demographic

  5. Risk Factors


- To ensure a consistent and interactive user experience, global filters have been applied across all pages. These filters allow users to explore the data based on specific groups and characteristics. The selected filters include:

  1. State
  2. Gender
  3. Age
  4. BMI Levels
  5. Cholesterol Levels
  6. Sleep Duration
  7. Number of Risk Factors

  These filters were chosen based on their relevance to the analysis objectives and their significant influence on heart attack risk.


- Each page contains a set of KPI cards at the top to provide quick insights into the dataset, including:
  1. Total Population (total number of individuals in the dataset)
  2. Total Number of Cases and the Percentage of Positive Cases
  3. Total Number of Non-Cases and the Percentage of Non-Cases
  4. Average Infection Rate (percentage of positive cases in the dataset)
  5. Top City by Infection Rate and its corresponding percentage

  These KPIs help users quickly assess the overall situation and compare subgroups across various dimensions.

- ## *Section 1: Geographic Analysis*

  The analysis of heart attack distribution across Indonesian states revealed that several regions, including Pekanbaru, Yogyakarta, Medan, Balikpapan, Surabaya, Bandung, Palembang, and Batam, have significantly higher heart attack rates compared to others. These states were classified as high-risk areas due to the noticeable percentage of affected individuals exceeding the national average. Identifying these geographic hotspots allows for more focused awareness campaigns and healthcare resource allocation.

  Additionally, when comparing urban and rural areas, the heart attack rates were found to be relatively similar. This suggests that place of residence—whether urban or rural—does not play a significant role in influencing the likelihood of heart attack occurrence. Therefore, while state-level differences are critical, urban versus rural location does not appear to be a contributing risk factor in this dataset.

- ## *Section 2: Health Behavior Analysis*

  In terms of health behaviors, diabetes and hypertension emerge as the most critical factors, as the rate of heart attack cases among individuals with these conditions exceeded the overall average incidence rate.

  On the other hand, behaviors such as Diet Type, Physical Activity, Stress Level, Sleep duration, and alcohol consumption consistently show elevated heart attack rates above the general average. This persistent trend indicates that these behaviors may be influenced by additional underlying factors that amplify their risk.

- ## *Section 3: Health Indicators Analysis*

  **Cholesterol Levels Analysis:**

  Across all cholesterol categories, the percentage of heart attack cases exceeded the overall average. However, a notable pattern was observed:

  1. Individuals with very high cholesterol had the highest incidence rate.

  2. This was closely followed by those with normal cholesterol levels.

  3. Interestingly, those with moderately high cholesterol showed only a slight increase above the average.

*Interpretation:*

This suggests that cholesterol alone may not be the sole influencing factor. The high incidence in the "normal" group may indicate the presence of other risk factors acting in combination. Therefore, individuals with normal or very high cholesterol might represent key target groups for early intervention and awareness.

## BMI Analysis:

Analysis of BMI categories revealed that the incidence of heart attacks exceeded the overall average in the following groups:

1. Normal weight individuals recorded the highest percentage of heart attack cases.
2. Followed by those in the Obese Class I and Overweight categories, with only slight differences between them.

*Interpretation:*

This surprising trend — especially the high incidence in the "normal weight" group — suggests that BMI alone is not a strong predictor of heart attack risk in this dataset. It indicates that other co-existing risk factors (e.g., stress, smoking, genetic predisposition) may have a greater influence in these cases. Hence, BMI should be interpreted in combination with other health indicators, rather than in isolation.

## Access to Medical Resources:

Upon analyzing the levels of access to medical resources, it was observed that heart attack incidence rates were elevated across all categories — whether access was poor, moderate, or good.

*Interpretation:*

This consistent increase across all levels suggests that access to medical resources does not appear to be a significant differentiating factor in heart attack occurrence within this dataset. In other words, the availability of healthcare services alone may not be sufficient to reduce heart attack risk without addressing underlying health behaviors and risk factors.

## • *Section 4: Demographic*

### Gender:

Analysis shows that both males and females exhibit increasing heart attack incidence rates, with no significant difference between the two.

Therefore, gender is not considered a critical influencing factor in this dataset.

### Age:

A clear trend was observed:

Heart attack incidence rates increase significantly with age, especially from 36 years old and above.

Age is a strong influencing factor and should be prioritized in awareness and prevention programs.

### Educational Level, Income Level, and Marital Status:

Although heart attack incidence rates are consistently increasing across different education levels, income groups, and marital statuses,

There is no clear distinction or significant deviation in these rates that would indicate these factors are major contributors to increased risk.

These variables may be associated with other influencing factors but do not independently indicate high risk.

### Relationship Between Marital Status and Employment Status:

Upon analyzing the intersection between marital status and employment,

The only group where the heart attack incidence exceeds the overall average is among individuals who are married and unemployed.

This suggests that the combination of marital responsibilities and lack of employment may contribute to higher stress levels or other indirect risk factors, increasing the likelihood of heart attacks within this group.

## • *Section 5: Risk Factors*

The analysis of risk factors revealed that individuals possessing two or three risk factors tend to have a higher likelihood of heart attack, with incidence rates exceeding the overall average. This suggests a cumulative effect, where the presence of multiple contributing factors significantly elevates the risk.

Moreover, regarding environmental conditions, it was observed that in areas where the Air Quality Index (AQI) was classified as "very unhealthy," the rate of heart attack cases surpassed the general average. This indicates that poor air quality may play a notable role in exacerbating cardiovascular risks.

# Conclusion

This analysis aimed to uncover the key factors contributing to the risk of heart attacks in Indonesia using survey data. The results revealed that chronic conditions like hypertension and diabetes are among the most influential risk factors. Additionally, Body Mass Index (BMI) and cholesterol levels—especially in the "very high" and "normal" ranges—showed a strong association with increased risk.

While behaviors like poor diet, physical inactivity, alcohol use, and stress showed a consistent upward trend in risk, they appear to be influenced by other contributing factors. Demographic factors such as marital status and employment revealed that married individuals who are unemployed have a slightly higher risk.

In contrast, education level, income, and gender were not found to be significant contributors.

The overall findings provide a data-driven foundation for identifying high-risk groups and prioritizing preventive health strategies.

# Recommendations

1. Launch public health campaigns focusing on managing blood pressure and diabetes.

2. Encourage healthy lifestyle changes such as regular physical activity and balanced diets.

3. Prioritize screening and monitoring in high-risk cities (e.g., Pekanbaru, Yogyakarta, Medan).

4. Develop targeted programs for unemployed and married individuals, especially above age 35.

5. Promote awareness around cholesterol and BMI management.

6. Further investigate the influence of behavioral trends on heart attack risk.

# Limitations

- ❖ The dataset used did not include a time dimension, which limits trend analysis over time.

- ❖ The air pollution Levels was ambiguous and lacked clear definitions or reference thresholds.

- ❖ Several variables, such as medical access or sleep duration, showed consistent increases in risk but require further study to confirm causality.

- ❖ The data is observational, which means relationships observed are correlational, not necessarily causal.

- ❖ The sampling distribution across states and demographic groups may not fully reflect the national population.