

# **Heart Disease Analysis Project using Power BI**

---

## *Purpose and Objective*

---

This project aims to analyze medical data related to heart disease through an interactive dashboard in **Power BI**, designed to assess the likelihood of an individual developing heart disease based on variables such as **age, blood pressure, cholesterol, glucose, and body mass index (BMI)**.

The main goal is to transform medical data into clear, actionable insights that support **decision-making**, raise **awareness**, and promote **early prevention**.

It is worth noting that **heart disease remains the leading cause of death worldwide**, accounting for approximately **18 million deaths annually**.

## *Project Scope*

---

This project aims to analyze medical data related to heart disease patients through a set of interactive analytical dimensions that help in understanding the patterns of occurrence and the influencing factors.

The scope of analysis includes the following:

### **1. Demographic Differences Analysis:**

Studying the impact of gender, age, and physical factors on the likelihood of developing heart disease.

### **2. Lifestyle Analysis:**

Evaluating the effect of smoking, alcohol consumption, and physical activity levels on heart health.

### **3. Medical Indicators Analysis:**

Measuring the relationship between cholesterol, glucose, and blood pressure levels and the overall prevalence of heart disease.

### **4. Comprehensive Correlation Analysis:**

Building models that explain how different variables interact to identify the groups most at risk.

### **5. Interactive Dashboard Development:**

Designing a **Power BI Dashboard** to visually present the results, support data-driven decision-making, and enable fast and clear analysis of critical cases.

### **6. Patient Details Review:** Providing a detailed, searchable view of individual patient records, allowing for the examination of specific cases and their complete data profiles.

## *Sources and Tools*

---

An Excel file containing real medical data was obtained from an open-source dataset available online. The following tools were used for the analysis:

- **Power BI:** for data analysis and creating interactive dashboards
- **Power Query:** for data cleaning and transformation
- **Excel:** for initial review and file organization
- **DAX:** for creating measures and calculated fields

## *Project Team*

---

- Eng. Mahmoud Abdelraouf Mohamed
- Eng. Hussein Abdelhalim Badr
- Eng. Mohamed Gomaa Rabie
- Eng. Ahmed Ibrahim El-Sayed
- Eng. Asim Ahmed Khalifa
- Dr. Esraa Hamdino Abdelkader

**Supervised by:** Eng. Karim Baqli

## *Project Phases and Task Distribution*

---

Phase	Main Tasks and Activities	Expected Outcome	Responsibilities
Phase 1	<b>Data Collection</b>   Extracting the medical file from the internet and importing it into Excel	Database ready for analysis	All team members
Phase 2	<b>Data Cleaning</b>   Handling missing values, correcting data types, and transforming columns	Clean and model-ready data	Mahmoud Abd Elraof, Esraa Hamdino
Phase 3	<b>Metrics Preparation and Analysis</b>   Creating DAX measures and analyzing relationships between variables	Extraction of preliminary indicators on disease patterns	Assem Ahmed, Hussein Bedier, Ahmed Ibrahim
Phase 4	<b>Presentation File Documentation</b>   Preparing and updating presentation files (PowerPoint & PDF)	Presentation and documentation files ready for final review	Mohamed Gomaa, Mahmoud Abd Elraof
Phase 5	<b>Dashboard Development</b>   Building an interactive dashboard in Power BI	First version of the dashboard	All team members
Phase 6	<b>Review and Enhancement</b>   Reviewing design, accuracy, and validating results	Final version of the dashboard	All team members
Phase 7	<b>Documentation and Presentation</b>   Preparing the documentation report and final presentation	Project completion and results delivery	All team members

## *Dataset Overview*

---

The file contains real medical data for 68,206 patients collected from an open-source dataset available online. It includes 17 columns covering demographic, behavioral, and health aspects of the individuals. The data aims to analyze factors affecting the likelihood of developing heart diseases and to extract statistically significant indicators.

**Prepared by:** Heart Disease Analysis Project Team | 2025

## Main Categories of Variables:

- **Demographics:** Age, Gender, Height, Weight
- **Lifestyle:** Smoking, Alcohol consumption, Physical Activity
- **Health Indicators:** Cholesterol, Glucose, Blood Pressure
- **Patient Details:** Searchable patient records by ID
- **Target Variable:**

The *cardio* variable indicates the patient's condition (1 = affected, 0 = not affected).

Below is a detailed description of the main columns used in the analysis:

## Description of Key Columns:

Name Column	Description	Type Data
Patient_ID	for each patient Unique identifier	Number (P. K)
Age	birth Patient's date of	Date
Gender	(Male, 2: Female :1) Patient's gender	Number
Height	centimeters Patient's height in	Number
Weight	kilograms Patient's weight in	Number
Ap_hi	pressure Systolic blood	Number
Ap_lo	pressure Diastolic blood	Number
Cholesterol	(Normal, 2: Above Normal, 3: High :1) Cholesterol levels	Number
Gluc	(Normal, 2: Above Normal, 3: High :Glucose levels (1	Number
Smoke	(No, 1: Yes :Smoking status (0	Number
Alco	(No, 1: Yes :0) Alcohol consumption	Number
Active	(No, 1: Yes :0) Physical activity	Number
Cardio	disease (0: No, 1: Yes) Presence of heart	Number
Age_years	Age in years	Number
BMI	Weight / Height <sup>2</sup> (measures obesity) = Body Mass Index	Number
Bp_category	classification (Normal, Stage 1, Stage 2, Blood pressure Elevated)	Text
Bp_category_encoded	pressure classification (Normal, Stage 1, Stage 2, Encoded blood Elevated)	Text

Summary of Data type	Primary Key	Date	Binary Numbers	Decimal Numbers	Text
	1	1	7	6	2

## Statistical Overview

---

A preliminary statistical analysis was conducted to identify general trends and correlations among factors affecting heart diseases. The results showed the following:

**Prepared by:** Heart Disease Analysis Project Team | 2025

- The age group between 45 and 60 years is the most susceptible to heart disease.
- A clear correlation was found between high blood pressure and Body Mass Index (BMI) with an increased likelihood of disease.
- Elevated levels of cholesterol and glucose were recorded among affected patients, while regular physical activity was a contributing factor in reducing risk.
- The results also indicate that most patients have slightly increased weight and higher-than-normal blood pressure, highlighting the importance of monitoring these indicators as key factors for predicting heart diseases.

## KPIs and Analyses

---

The following key indicators (KPIs) have been identified to be included in the interactive dashboard.

KPI Title	Purpose	Recommended Visualization
Heart Disease Distribution by Gender	Compare the number of heart disease cases between males and females.	Donut Chart (Male vs. Female comparison).
Heart Disease Cases by Age Group	Identify which age groups have the highest number of heart disease cases.	Ribbon Chart (Cases by Age Group: Young, Middle, Older Adult).
Relationship Between BMI and Heart Disease	Determine if different BMI categories are linked to higher heart disease prevalence.	Clustered Bar Chart (Heart Disease presence across BMI categories).
Effect of Smoking and Alcohol Consumption on Heart Disease	Compare prevalence between smokers vs. non-smokers and those who consume alcohol vs. those who do not consume alcohol.	Ribbon Chart (Smoking and Alcohol comparison).
Impact of Physical Activity on Heart Disease Prevention	Determine if physical activity reduces heart disease rates.	Clustered Bar Chart (physically active vs. not physically active comparison).
Analysis of Combined Risk Factors	Allow interactive filtering to identify high-risk segments based on combined factors.	Slicers + Card Visual (Smoker, Cholesterol, Activity)
Impact of Cholesterol Level on Heart Disease	Examine how cholesterol levels relate to heart disease prevalence.	Stacked Bar Chart (risk proportion by cholesterol level).
Heart Disease Prevalence Rate	Present the overall percentage prevalence in the entire sample as a key performance indicator.	Card Visual (Disease Prevalence)
Trend of Heart Disease prevalence by blood pressure category	Analysis and comparison of the prevalence of heart disease across different blood pressure categories.	Slicer (BP Categories: Normal, Elevated, Hypertension Stage 1, Hypertension Stage 2).
Correlation Between Glucose Level and Heart Disease	Assess the relationship between glucose levels and heart disease risk.	Stacked Bar Chart (prevalence by glucose category).

## Conclusion

---

The phases of data collection, cleaning, and analysis have been completed, and the key indicators related to heart diseases have been extracted. The team is now preparing to move on to the development of an interactive dashboard to present the results in a more visual and analytical manner.

Sincere thanks to the Ministry of Communications and Information Technology and Supervisor Eng. Karim Baqli for their continuous support and guidance throughout the project implementation.