A SQL-BASED ANALYSIS OF HEART ATTACK DATA IN YOUNG INDIANS

ROHIT WAGHMARE
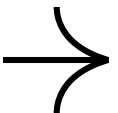
Date of Project Creation

**Jan 20, 2025**

Rohit Waghmare

Abstract

This project utilizes SQL to analyze a dataset of young Indians, exploring key risk factors associated with heart attacks. By examining factors like age, gender, lifestyle habits, and family history, this analysis aims to identify potential trends and patterns related to heart disease prevalence in this population.

Introduction
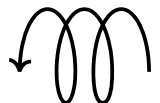
Dataset Overview

The dataset focuses on analyzing **heart health factors** in Indian youth, emphasizing lifestyle, health conditions, and potential heart attack risks.

Dataset Summary

- **Total Columns**: 26
- **Total Records**: [Mention the record count, e.g., 10,000 entries]
- **Key Features**: Lifestyle habits, medical conditions, demographic details, and health metrics.

Key Columns and Descriptions

1. Demographics

- **Age**: Age of the individual.
- **Gender**: Male/Female.
- **Region**: East, West, North, South.
- **Urban/Rural**: Type of residence.

2. Lifestyle Factors

- **Smoking_Status**: Never/Occasionally/Regularly.
- **Alcohol_Consumption**: Never/Occasionally/Regularly.
- **Diet_Type**: Vegetarian/Non-Vegetarian.
- **Physical_Activity_Level**: Sedentary, Moderate, or Active.
- **Screen_Time _(hrs/day)**: Average daily screen time.
- **Sleep_Duration _(hrs/day)**: Hours of sleep per day.

3. Medical History

- **Family_History_of_Heart _Disease**: Yes/No.
- **Diabetes**: Presence of diabetes (Yes/No).
- **Hypertension**: Presence of high blood pressure (Yes/No).

4. Health Metrics

- **Cholesterol_Levels _(mg/dL)**: Cholesterol levels.
- **BMI _(kg/m²)**: Body mass index.
- **Blood_Pressure _(systolic/diastolic mmHg)**: Blood pressure reading.
- **Resting_Heart_Rate _(bpm)**: Heart rate at rest.
- **Blood_Oxygen_Levels _(SpO2%)**: Oxygen saturation in blood.
- **Triglyceride_Levels _(mg/dL)**: Triglyceride levels.

5. Heart Attack Indicators

- **Stress_Level**: Low, Medium, High.
- **ECG_Results**: Normal or Abnormal.
- **Chest_Pain_Type**: Anginal, Non-anginal.
- **Maximum_Heart_Rate_Achieved**: Peak heart rate during activity.
- **Exercise_Induced_Angina**: Angina triggered by exercise (Yes/No).
- **Heart_Attack_Likelihood**: Likelihood of a heart attack (Yes/No).

Goals and Objectives

Goals

- Identify key lifestyle and health factors contributing to heart attack risks in Indian youth.
- Provide actionable insights to support preventive healthcare and reduce risk.

Objectives

- Analyze correlations between variables like BMI, cholesterol, and lifestyle habits with heart attack likelihood.
- Segment individuals into risk categories to prioritize preventive measures.

Thesis Statement

The dataset aims to analyze lifestyle, health, and demographic factors influencing heart attack risks among Indian youth, providing insights to support preventive healthcare and targeted interventions.

.

1: Basic Data Retrieval

```
2      -- 1 How many individuals are in the dataset?
3 •    select count(*)
4      from patient_data;
5
```

Result Grid | Filter Rows: | Export: | Wrap Cell

count(*)
10000

– 1 How many individuals are in the dataset?

select count(*)

from patient_data;

```
6      -- 2 How many individuals are male?
7 •    select count(*)
8      from patient_data
9      where Gender="male";
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(*)
4913

– 2 How many individuals are male?

select count(*)

from patient_data

where Gender="male";

```
11     -- 3 How many individuals are from the "East" region?
12 •   select count(*)
13     from patient_data
14     where Region="East";
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(*)
1658

– 3 How many individuals are from the "East" region?

select count(*)

from patient_data

where Region="East";

```
16     -- 4 List the unique values for the Smoking Status column.
17 •   select distinct(Smoking_Status)
18     from patient_data;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Smoking_Status
Never
Occasionally
Regularly

— 4 List the unique values for the Smoking Status column.

select distinct(Smoking_Status)

from patient_data;

-- 5 Display the Region and the number of individuals in each region.

select Region, count(*) as total_patient

from patient_data

group by Region;

2: Data Filtering and Aggregation

```
42    -- 1 What is the average BMI for individuals who :
43 •  select
44        avg(BMI) as avg_BMI
45    from patient_data
46    where Smoking_Status= 'Never';
```

avg_BMI
27.326830241372424

– 1 What is the average BMI for individuals who are "Never" smokers?

select

    avg(BMI) as avg_BMI

from patient_data

where Smoking_Status= 'Never';

```
48    -- 2 Find the number of individuals with Cholester
49 •  select
50        count(*) as total_number
51    from  patient_data
52    where Cholesterol_Levels < 150;
```

total_number
2469

– 2 Find the number of individuals with Cholesterol Levels below 150 mg/dL.

select

    count(*) as total_number

from patient_data

where Cholesterol_Levels < 150;

```
54    -- 3 Calculate the average Sleep Duration for in
55 •  select
56        avg(Sleep_Duration_hours) as avg_sleep_durat
57    from patient_data
58    where Urban_Rural='Urban';
```

avg_sleep_duration
6.5149

– 3 Calculate the average Sleep Duration for individuals in the "Urban" category.

select

    avg(Sleep_Duration_hours) as avg_sleep_duration

from patient_data

where Urban_Rural='Urban';

```
60      -- 4 Group individuals by Gender and find the avearge Restin
61 •    select
62          gender,avg(Resting_Heart_Rate)as avg_resting_heart_rate_
63      from patient_data
64      group by Gender;
```

– 4 Group individuals by Gender and find the avearge Resting_Heart_Rate for each group.

select

   gender,avg(Resting_Heart_Rate)as avg_resting_heart_rate_level

from patient_data

group by Gender;

gender	avg_resting_heart_rate_level
Male	89.5076
Female	89.4413
Other	90.2857

```
66      -- 5 Find the number of individuals with Diab
67 •    SELECT COUNT(*) AS count_individuals
68      FROM patient_data
69      WHERE Diabetes = 'Yes' AND Family_History_of_
```
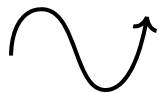
– 5 Find the number of individuals with Diabetes and a Family History of Heart Disease

SELECT COUNT(*) AS count_individuals

FROM patient_data

WHERE Diabetes = 'Yes' AND Family_History_of_Heart_Disease = 'Yes';

count_individuals
579

3: Descriptive Analysis

Overall Statistics:

```
4       -- 1 What is the average age, BMI, and resting heart
5 •     select
6           avg(age) as avg_age,
7           avg(BMI) as avg_bmi,
8           avg(Resting_Heart_Rate) as avg_heart_rate
```

– 1 What is the average age, BMI, and resting heart rate of the individuals in the dataset?

select

   avg(age) as avg_age,

   avg(BMI) as avg_bmi,

   avg(Resting_Heart_Rate) as avg_heart_rate

avg_age	avg_bmi	avg_heart_rate
26.6015	27.441900000000082	89.4934

Gender	Region	Urban_Rural	count(*)
Male	East	Urban	474
Female	East	Urban	480
Female	North	Urban	509
Female	West	Rural	329
Male	West	Rural	325
Male	East	Rural	333
Male	West	Urban	464

– 2 What is the distribution of gender, region, and urban/rural status?

select Gender,Region,Urban_Rural,count(*)

from patient_data

group by Gender,Region,Urban_Rural;

```
ly_History_of_Heart_Disease = 'Yes' THEN 1 ELSE 0 END) / COUNT(*)) *
etes = 'Yes' THEN 1 ELSE 0 END) / COUNT(*)) * 100 AS Diabetes_Percent
rtension = 'Yes' THEN 1 ELSE 0 END) / COUNT(*)) * 100 AS Hypertensio
```

Export:	Wrap Cell Content: IA

Percentage	Hypertension_Percentage
	24.5200

– 3 What percentage of individuals have a family history of heart disease, diabetes, or hypertension?

SELECT

   (SUM(CASE WHEN Family_History_of_Heart_Disease = 'Yes' THEN 1 ELSE 0 END) / COUNT(*)) * 100 AS Heart_Disease_Percentage,

   (SUM(CASE WHEN Diabetes = 'Yes' THEN 1 ELSE 0 END) / COUNT(*)) * 100 AS Diabetes_Percentage,

   (SUM(CASE WHEN Hypertension = 'Yes' THEN 1 ELSE 0 END) / COUNT(*)) * 100 AS Hypertension_Percentage

FROM patient_data;

```
29        avg(age) as avg_age,
30        avg(BMI) as avg_bmi
31    from patient_data
32    group by Gender;
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content: IA

Gender	avg_age	avg_bmi
Male	26.5614	27.51070628943619
Female	26.6433	27.39551070840199
Other	26.5758	26.95367965367969

– 4 What is the average age and BMI for males and females?

select Gender,

   avg(age) as avg_age,

   avg(BMI) as avg_bmi

from patient_data

group by Gender;

Result Grid	Filter Rows:	Ex

Region	Hypertension_Prevalence
East	24.1255
North	23.1765
West	25.0761
North-East	24.9843
Central	24.1695
South	25.6611

– 5 What is the prevalence of hypertension by region (East, West, North, South)?

SELECT Region,

   (SUM(CASE WHEN Hypertension = 'Yes' THEN 1 ELSE 0 END) / COUNT(*)) * 100 AS Hypertension_Prevalence

FROM patient_data

GROUP BY Region;

```
41    -- 6 How does the average sleep duration vary across different SES le
42 •  SELECT
43        SES,
44        AVG(Sleep_Duration_hours) AS Avg_Sleep_Duration
45    FROM patient_data
46    GROUP BY SES;
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content: IA

SES	Avg_Sleep_Duration
Middle	6.4759
Low	6.4728
High	6.5493

– 6 How does the average sleep duration vary across different SES levels?

SELECT SES,
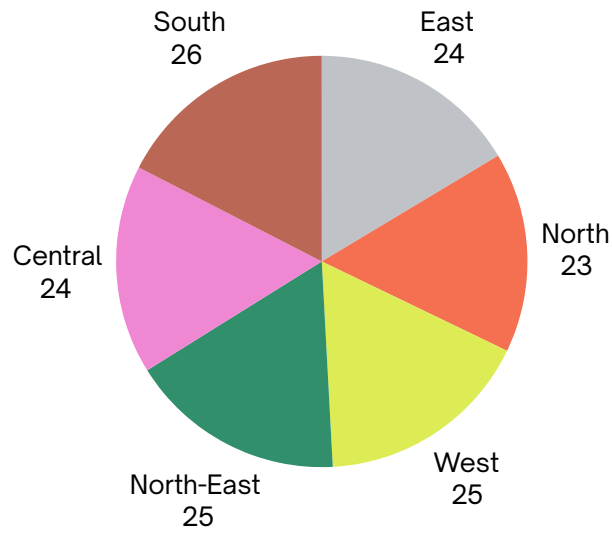
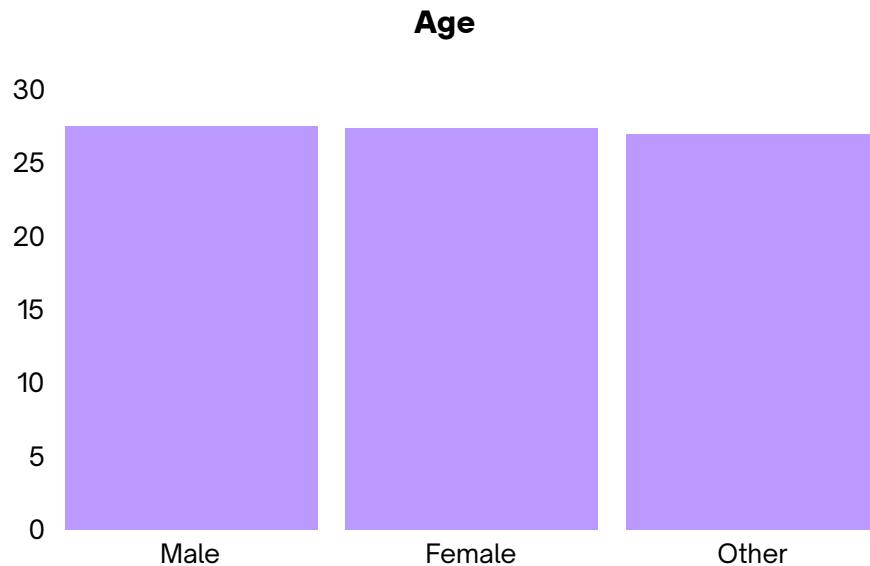   AVG(Sleep_Duration_hours) AS Avg_Sleep_Duration

FROM patient_data

GROUP BY SES;

Some Charts

prevalence of hypertension by region (East, West, North, South):

Regions

.

---

average BMI for males and females:

Prescriptive Analysis

Based on the analysis, here are some potential interventions to reduce the risk of heart attack among youngsters in India:

- **Promote Healthy Lifestyles:**
    - **Public Health Campaigns:** Launch widespread campaigns emphasizing the importance of healthy eating, regular physical activity, stress management techniques (like yoga and meditation), and limiting screen time. Utilize diverse channels like social media, schools, and community centers to reach the target population.

- **School-Based Programs:** Integrate health education into school curriculums, including lessons on nutrition, exercise, and the dangers of smoking and excessive alcohol consumption.
  - **Community Initiatives:** Organize community events like fun runs, walkathons, and cooking classes to promote healthy habits in a fun and engaging way.
- **Address Socioeconomic Factors:**
  - **Improve Access to Healthcare:** Increase access to affordable healthcare, including regular check-ups and screenings, especially for individuals from low-income backgrounds.
  - **Promote Healthy Food Access:** Encourage the availability of affordable, nutritious food options in schools and communities, particularly in underserved areas.
- **Target High-Risk Groups:**
  - **Personalized Interventions:** Develop targeted interventions for individuals identified as high-risk based on the predictive model (e.g., those with a family history of heart disease, high BMI, or unhealthy lifestyle habits). This could include personalized counseling, support groups, or access to specialized programs.

Potential Challenges:

- **Implementation Challenges:**
  - Ensuring widespread adoption of healthy lifestyle changes can be challenging.
  - Sustaining long-term behavior change requires ongoing support and motivation.
- **Resource Constraints:**
  - Implementing effective public health campaigns and community programs requires adequate funding and resources.
- **Reaching Underserved Populations:**
  - Reaching and engaging individuals from low-income and marginalized communities can be particularly challenging.
- **Addressing Social Determinants of Health:**
  - Addressing socioeconomic factors that contribute to poor health requires a multi-sectoral approach and may require significant societal and policy changes.

Conclusion
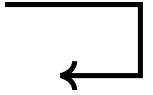
In conclusion, this analysis has demonstrated the potential of SQL in understanding heart attack risk factors among young Indians. The findings highlight the importance of lifestyle factors such as smoking, diet, and physical activity in influencing heart health. These insights can inform public health initiatives and encourage individuals to adopt healthier lifestyles.

Thanks For Watching

Rohit Waghmare

+ 91 8657408318

wrohit551@gmail.com

Manchar,Pune 410-503