CM3010 Databases and Advanced Data Techniques: Mid-term Coursework

SOH YI JIE GABRIEL – (220516695)

UNIVERSITY OF LONDON

Introduction

In India, heart disease is becoming more prevalent among young adults and is still a major global concern [1]. Prevention and the promotion of healthy living depend on an understanding of variables such as demography, lifestyle, medical history, and clinical signs [2]. This study looks at the effects of both non-modifiable factors like family history and modifiable ones like food, exercise, smoking, and sleep. In addition to highlighting disparities by gender, socioeconomic level, and geography, it seeks to uncover high-risk profiles and investigate correlations between factors (e.g., physical activity mitigating high BMI risks). The objective is to increase understanding and create evidence-based plans to lower the risk of heart attacks and enhance cardiovascular health in this susceptible population.

Stage 1

1.1 Dataset Criteria Assessment: Heart Attack Risk Among Young Indians

(Source: https://www.kaggle.com/datasets/ankushpanday1/heart-attack-in-youth-of-india?select=heart_attack_youngsters_india.csv)

Detail

- The dataset provides a comprehensive view of heart attack risk factors, covering modifiable behaviors (e.g., diet, exercise) and non-modifiable characteristics (e.g., family history, gender).
- It makes it possible to thoroughly examine how different factors interact, such as how stress affects sleep and physical activity.
- The inclusion of clinical test results like ECG and SpO2% offers a granular perspective for advanced analysis.

Quality

- The dataset is well-organized and includes comprehensive data on young adults' risk of heart attacks, including demographics, lifestyle, medical history, and clinical test findings
- It includes 26 columns with clearly defined variables such as age, gender, socioeconomic status, smoking, and clinical indicators like blood pressure and triglyceride levels.

- Some fields, like self-reported lifestyle factors (e.g., sleep duration, physical activity), may introduce biases due to subjective responses.
- Missing or incomplete data in certain columns may require imputation or exclusion techniques during analysis.

Documentation

- Column headers are descriptive and align with the variables being measured, facilitating easy understanding and interpretation.
- Some additional metadata or documentation on data collection methods (e.g., sample selection, region-specific representation) would enhance usability.
- The dataset does not explicitly define thresholds for certain categories (e.g., “high” vs. “low” stress levels), which may need to be clarified or standardized.

Interrelation

- The dataset can be enriched by integrating with external datasets, such as:
  - Regional healthcare infrastructure data to analyze geographic disparities in outcomes.
  - Socioeconomic data for deeper insights into the role of income and education.
  - National dietary surveys to correlate with dietary preferences and triglyceride levels.
- Multi-dimensional studies of heart attack risks across populations may be made possible by such integrations.

Use

- **Machine Learning:** Train models like logistic regression or random forests to predict heart attack likelihood and identify key risk factors.
- **Public Health Policy:** Identify high-risk groups for targeted interventions based on lifestyle and medical history.
- **Behavioral Insights:** Examine role of modifiable factors, such as diet and exercise, in reducing heart attack risk.

- **Regional Analysis:** Study disparities in heart attack risk across urban and rural areas or different socioeconomic strata.

Observations and Recommendations

- **Data Cleaning:**
    - Address variables such as clinical indicators and lifestyle factors that have missing or inconsistent values.
    - Ensure consistency in categorical variables (example standardizing labels for diet types or smoking status).
- **Feature Engineering:** Create composite indicators like “overall lifestyle risk score” by combining variables such as smoking, alcohol consumption, and physical activity.
- **Questions for Future Analysis:**
    - How do physical activity levels moderate the impact of high BMI on heart attack risk [3]?
    - What is the combined effect of stress and sleep duration on clinical test results like ECG or blood pressure?
    - Are there significant differences in heart attack risk by dietary preferences (e.g., vegetarian vs. non-vegetarian)?
- **Additional Data Needs:** Additional context for evaluating results might be provided by include data on healthcare access or regional healthcare quality.

Discoverability

- Although there are other databases on heart disease, this one is especially useful because it focuses on young adults in India, a group that is frequently understudied despite being at higher risk.
- It is extremely significant for research, public health, and predictive modeling applications due to its detailed information on clinical and lifestyle aspects

1.2 Interest in the dataset

I'm interested in the dataset, especially in terms of figuring out what factors influence young individuals in India's chance of having a heart attack. This dataset offers a good chance to examine both modifiable lifestyle factors like food, physical activity, smoking, and sleep, as well as non-modifiable aspects like family history, since heart disease is growing more common in this population [4]. By looking at these variables, the study hopes to clarify high-risk profiles, gender, socioeconomic, and regional differences, as well as relationships between important elements. Evidence-based tactics for lowering the risk of heart attacks and fostering improved cardiovascular health in this susceptible group may benefit from this investigation.

As I delve into this dataset, several questions arise:

1. What are the regional disparities in heart attack likelihood among young adults in India?
Finding the areas that most require preventive actions can be facilitated by knowing the geographic distribution of heart attack risks. Analyzing differences in heart attack risk between North, South, East, and West areas is the main goal of this question.

2. How does physical activity mitigate the risks associated with high BMI?
This inquiry explores whether levels of physical activity mitigate the effect of elevated body mass index on the risk of heart attacks. The investigation looks at those who are inactive, moderately active, and highly active in an effort to identify the preventive benefits of exercise.

3. Is there a significant difference in heart attack likelihood based on gender and socioeconomic status?
Examining differences in the risk of heart attacks by gender and socioeconomic status can identify populations that can benefit from specialized medical care. The purpose of this inquiry is to determine whether these demographic characteristics have a substantial impact on cardiovascular outcomes.

4. How do modifiable factors like smoking and sleep duration correlate with heart attack risks?
This inquiry examines the connections among heart attack risk, sleep length, and smoking status. It looks for trends that can inform lifestyle advice aimed at lowering hazards.

5. What is the relationship between family history of heart disease and clinical signs like cholesterol levels and blood pressure?
This question delves into how a family history of heart disease influences clinical metrics like cholesterol levels and blood pressure, potentially highlighting early warning signs and areas for intervention.

1.3 Terms of Use

The MIT License governs the use of the dataset in this report. Please see the MIT License for a complete copy of the license.

1.4 Dataset preparation

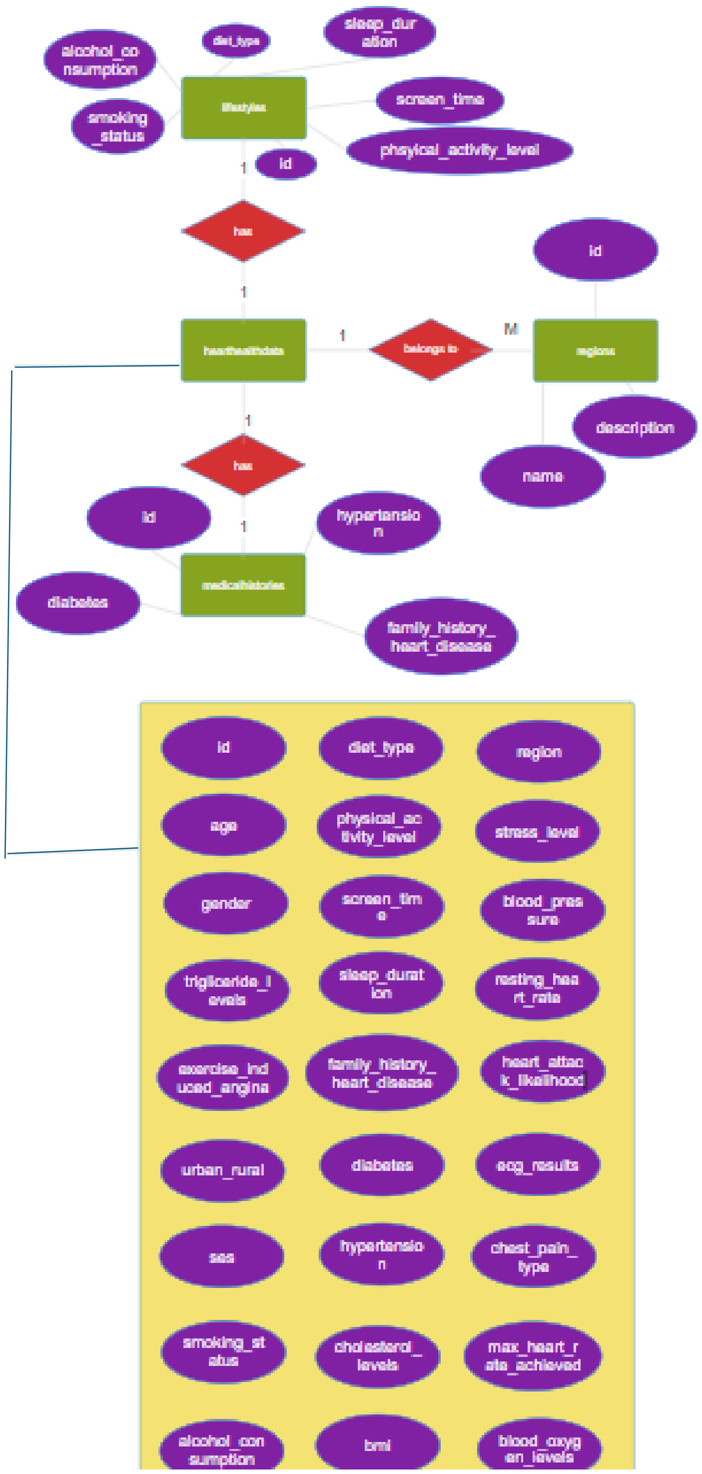
To comply with database design guidelines and make it easier to utilize them in the Entity-Relationship Diagram (ERD), the dataset's pre-existing column headers were standardized to short, meaningful names. These modifications made sure the properties were appropriate for database representation and accurately matched the underlying data model.

The 26-column dataset was renamed to use attribute names that were both descriptive and succinct. A suitable data type was also allocated to each column according to its contents. For instance, Family History of Heart Disease became **family_history**, and Smoking and Alcohol Consumption became **smoking_alcohol**.

The renamed column headers will be shown in Section 2: Drawing a Complete E/R Model to show how they correspond with the overall data structure for clarity and to reduce clutter.

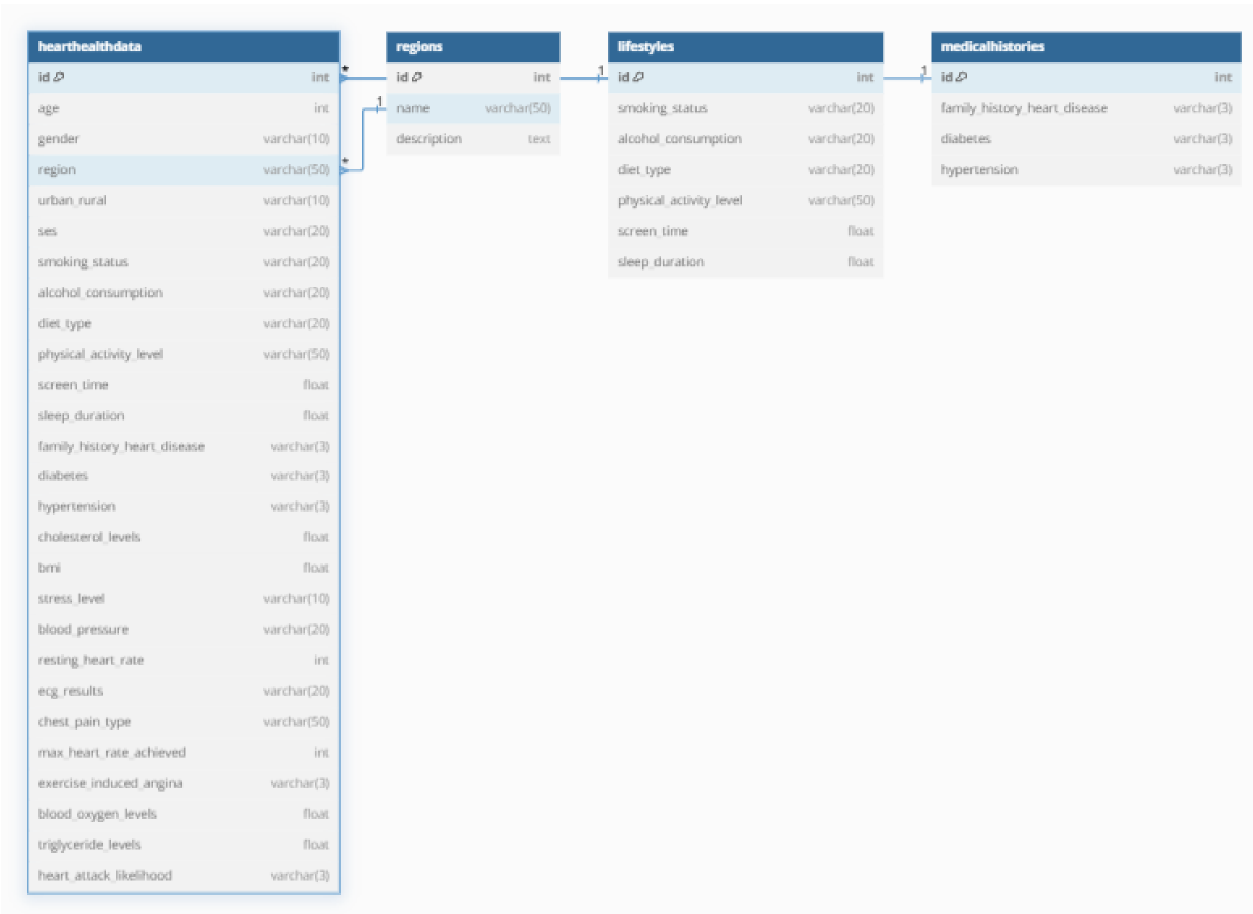
Stage 2 – Modelling the Data

E/R Model Diagram (Created with SmartDraw)



The E/R model integrates important entities including Lifestyle, MedicalStatus, HeartHealthData, and Regions to create a complete database system for heart health. Important characteristics like physical activity, medical history, smoking status, and demographic information are recorded. While accounting for a many-to-one linkage with geographic locations, the relationships show a one-to-one relationship between lifestyle, medical history, and personal health data. By linking environmental contexts, medical illnesses, and lifestyle factors, this approach facilitates a comprehensive investigation of heart health, promoting better understanding and preventative care tactics. The structure minimizes redundancy while guaranteeing scalability and data integrity.

Database Fields and Tables (created with DBDiagram)



The four primary tables in the E/R diagram—hearthealthdata, regions, lifestyles, and medical histories—represent a structured relational database intended for heart health analysis. The main component is the hearthealthdata table, which collects individual-level data such as medical metrics (blood pressure, cholesterol, and heart attack likelihood), lifestyle factors (smoking status, alcohol use, and level of physical activity), and demographics (age, gender, and region). It maintains one-to-one links with lifestyles and medical histories for in-depth tracking of habits and medical history, and it connects to the regions table via a many-to-one relationship, connecting people to their geographic location. This framework facilitates thorough analysis of heart health determinants, minimizes redundancy, and encourages normalization.

Conclusion:

To sum up, the E/R model integrates important entities and their interactions to provide a well-structured framework for data organization and analysis. It retains scalability and clarity while capturing crucial information. This architecture is a useful tool for deriving significant insights and promoting well-informed decision-making since it supports thorough analysis, reduces redundancy, and guarantees data integrity.

Stage 3 – Creating the Database

Listing all the create commands used to build the database:

```sql
-- Step 1: Create and Use the Database
CREATE DATABASE IF NOT EXISTS HeartHealthDB;
USE HeartHealthDB;
```

```sql
-- Step 2: Create the Temporary Table
CREATE TEMPORARY TABLE TempHeartHealthData (
    Age INT,
    Gender VARCHAR(10),
    Region VARCHAR(50),
    Urban_Rural VARCHAR(10),
    SES VARCHAR(20),
    Smoking_Status VARCHAR(20),
    Alcohol_Consumption VARCHAR(20),
    Diet_Type VARCHAR(20),
    Physical_Activity_Level VARCHAR(50),
    Screen_Time FLOAT,
    Sleep_Duration FLOAT,
    Family_History_Heart_Disease VARCHAR(3),
    Diabetes VARCHAR(3),
    Hypertension VARCHAR(3),
    Cholesterol_Levels FLOAT,
    BMI FLOAT,
    Stress_Level VARCHAR(10),
    Blood_Pressure VARCHAR(20),
    Resting_Heart_Rate INT,
    ECG_Results VARCHAR(20),
    Chest_Pain_Type VARCHAR(50),
    Max_Heart_Rate_Achieved INT,
    Exercise_Induced_Angina VARCHAR(3),
    Blood_Oxygen_Levels FLOAT,
    Triglyceride_Levels FLOAT,
    Heart_Attack_Likelihood VARCHAR(3)
);
```

Importing Data (I used this to import the data into the database):

```sql
-- Step 3: Load Data into the Temporary Table
LOAD DATA INFILE '/home/coder/project/heart_attack_youngsters_india.csv'
INTO TABLE TempHeartHealthData
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS;
```

```sql
-- Step 4: Create the Main Table
CREATE TABLE HeartHealthData (
    ID INT AUTO_INCREMENT PRIMARY KEY,
    Age INT,
    Gender VARCHAR(10),
    Region VARCHAR(50),
    Urban_Rural VARCHAR(10),
    SES VARCHAR(20),
    Smoking_Status VARCHAR(20),
    Alcohol_Consumption VARCHAR(20),
    Diet_Type VARCHAR(20),
    Physical_Activity_Level VARCHAR(50),
    Screen_Time FLOAT,
    Sleep_Duration FLOAT,
    Family_History_Heart_Disease VARCHAR(3),
    Diabetes VARCHAR(3),
    Hypertension VARCHAR(3),
    Cholesterol_Levels FLOAT,
    BMI FLOAT,
    Stress_Level VARCHAR(10),
    Blood_Pressure VARCHAR(20),
    Resting_Heart_Rate INT,
    ECG_Results VARCHAR(20),
    Chest_Pain_Type VARCHAR(50),
    Max_Heart_Rate_Achieved INT,
    Exercise_Induced_Angina VARCHAR(3),
    Blood_Oxygen_Levels FLOAT,
    Triglyceride_Levels FLOAT,
    Heart_Attack_Likelihood VARCHAR(3)
);
```

```sql
-- Step 5: Transfer Data from Temporary Table to Main Table
INSERT INTO HeartHealthData (
    Age, Gender, Region, Urban_Rural, SES, Smoking_Status, Alcohol_Consumption,
Diet_Type,
    Physical_Activity_Level, Screen_Time, Sleep_Duration,
Family_History_Heart_Disease,
    Diabetes, Hypertension, Cholesterol_Levels, BMI, Stress_Level, Blood_Pressure,
    Resting_Heart_Rate, ECG_Results, Chest_Pain_Type, Max_Heart_Rate_Achieved,
    Exercise_Induced_Angina, Blood_Oxygen_Levels, Triglyceride_Levels,
Heart_Attack_Likelihood
)
SELECT
    Age, Gender, Region, Urban_Rural, SES, Smoking_Status, Alcohol_Consumption,
Diet_Type,
    Physical_Activity_Level, Screen_Time, Sleep_Duration,
Family_History_Heart_Disease,
    Diabetes, Hypertension, Cholesterol_Levels, BMI, Stress_Level, Blood_Pressure,
    Resting_Heart_Rate, ECG_Results, Chest_Pain_Type, Max_Heart_Rate_Achieved,
    Exercise_Induced_Angina, Blood_Oxygen_Levels, Triglyceride_Levels,
Heart_Attack_Likelihood
FROM TempHeartHealthData;
```

```sql
-- Step 6: Verify Data
SELECT * FROM HeartHealthData;
```

Evaluation:

Building the database for heart attack risk analysis was a useful and enlightening exercise that demonstrated the value of accuracy and careful planning in database building. Starting with the CREATE DATABASE command, resilience was assured by avoiding repetition through the use of IF NOT EXISTS. It was a calculated decision to use a temporary table (TempHeartHealthData) as a staging area so that raw data could be loaded, examined, and verified before being added to the main table. Because temporary tables are not permanent, this division reduced dangers but necessitated close observation. The development of the HeartHealthData primary table brought to light the significance of creating a clear schema with proper data types and relevant column names, which improved query readability and usability.

Data integrity was strengthened by the addition of a primary key, yet query optimization and data validation could be improved because there were no restrictions like NOT NULL or indexing. Although the procedure successfully struck a balance between structure and flexibility, it highlighted the necessity of thorough planning, especially when it came to allocating data types and foreseeing future inquiries. All things considered, this experience demonstrated how important a carefully considered database structure is to guaranteeing data integrity, usability, and analysis efficiency.

Strengths/Issues of Database:

By enabling in-depth searches and upholding a clear schema design, the existing database structure has proven successful in offering insights into the dangers of heart attacks. It supports a large variety of clinical, lifestyle, and demographic data, allowing for thorough analysis. However, some design flaws, like the lack of limits, indexing, and version control systems, become more noticeable as the dataset gets bigger and more complicated. In order to ensure that the database is dependable and effective for use in the future, these gaps point to areas that should be improved to increase data integrity, query performance, and scalability.

Improving Database Design:

In retrospect, I would change the database design to guarantee data integrity by imposing constraints like NOT NULL on important fields like Age, Gender, and Heart_Attack_Likelihood. Performance would also be enhanced by adding indexing on commonly searched columns like Region and SES, particularly for bigger datasets. Version control and auditing would be improved by adding a timestamp field to track record creation and modifications. A stronger and more effective database structure would result from these modifications.

Add Indexing for Optimized Queries

```sql
CREATE INDEX idx_region ON HeartHealthData (Region);
CREATE INDEX idx_ses ON HeartHealthData (SES);
```

Update Existing Table to Add Constraints

```sql
ALTER TABLE HeartHealthData
MODIFY COLUMN Age INT NOT NULL,
MODIFY COLUMN Gender VARCHAR(10) NOT NULL,
MODIFY COLUMN Heart_Attack_Likelihood VARCHAR(3) NOT NULL;

ALTER TABLE HeartHealthData
ADD COLUMN Created_At TIMESTAMP DEFAULT CURRENT_TIMESTAMP;
```

SQL Query commands:

1. What are the regional disparities in heart attack likelihood among young adults in India?

To analyze regional disparities in heart attack likelihood among young adults in India, I used a query to calculate the total cases, high-risk cases (Heart_Attack_Likelihood = 'Yes'), and the percentage of high-risk cases for each region. The query grouped data by Region to highlight variations in risk geographically.

Key components of the query include:

- **COUNT(CASE...):** this calculates the number of high-risk individuals (RiskCount) in each region.
- **COUNT(*):** provides the total population (TotalCount) for each region.
- **COALESCE(...):** it computes the percentage of high-risk individuals (RiskPercentage) while handling null values.
- **GROUP BY:** Groups data by Region for region-wise analysis.

```
            SELECT
                Region,
                COUNT(CASE WHEN Heart_Attack_Likelihood = 'Yes' THEN 1 END) AS
RiskCount,
                COUNT(*) AS TotalCount,
                COALESCE((COUNT(CASE WHEN Heart_Attack_Likelihood = 'Yes' THEN 1
END) / COUNT(*)) * 100, 0) AS RiskPercentage
            FROM
                HeartHealthData
            GROUP BY
                Region;
```

In order to investigate contributing factors such as healthcare access, lifestyle, and socioeconomic conditions, the query identified regions with greater heart attack risks. Targeted public health initiatives can be informed by these insights.

2. How does physical activity mitigate the risks associated with high BMI?

I employed a query to study the association between physical activity levels, BMI, and the possibility of a heart attack in order to investigate how physical activity reduces the hazards associated with having a high BMI. The query determines the proportion of high-risk instances for each degree of physical activity, the total number of people with a high BMI, and the number of high-risk individuals with a BMI more than 25.

Key components of the query:

- **COUNT(CASE...):**
  - Calculates the number of high-risk individuals with BMI > 25 (HighRiskCount).
  - Determines the total number of individuals with BMI > 25 (TotalHighBMI).
- **COALESCE(...):** Computes the percentage of high-risk cases (RiskPercentage), ensuring null values are handled effectively.
- **GROUP BY:** Groups the results by Physical_Activity_Level to evaluate risk patterns across activity levels.

```
SELECT
                Physical_Activity_Level,
                COUNT(CASE WHEN BMI > 25 AND Heart_Attack_Likelihood = 'Yes' THEN 1
END) AS HighRiskCount,
                COUNT(CASE WHEN BMI > 25 THEN 1 END) AS TotalHighBMI,
                COALESCE((COUNT(CASE WHEN BMI > 25 AND Heart_Attack_Likelihood =
'Yes' THEN 1 END) / COUNT(CASE WHEN BMI > 25 THEN 1 END)) * 100, 0) AS
RiskPercentage
            FROM
                HeartHealthData
            GROUP BY
                Physical_Activity_Level;
```

Higher levels of physical activity are associated with lower risks of heart attacks in people with high body mass index (BMI). This finding underscores the protective function of physical activity and its significance in reducing the negative consequences of high BMI. These results can inform public health recommendations for encouraging physical activity as part of heart disease prevention strategies.

3. Is there a significant difference in heart attack likelihood based on gender and socioeconomic status?

I utilized a query that determines the number of high-risk cases (Heart_Attack_Likelihood = 'Yes'), the total number of cases, and the percentage of high-risk cases for each combination of gender and SES in order to assess the effect of these factors on the likelihood of heart attacks.

Key components of the query:

- **COUNT(CASE...):**
  - Calculates the number of high-risk cases (RiskCount).
  - Counts the total number of individuals (TotalCount) for each gender and SES group.
- **COALESCE(...):** Computes the percentage of high-risk cases (RiskPercentage), ensuring null values are handled.
- **GROUP BY:** Groups data by Gender and SES to identify patterns across these categories.

```
        SELECT
                Gender, SES,
                COUNT(CASE WHEN Heart_Attack_Likelihood = 'Yes' THEN 1 END) AS
RiskCount,
                COUNT(*) AS TotalCount,
                COALESCE((COUNT(CASE WHEN Heart_Attack_Likelihood = 'Yes' THEN 1
END) / COUNT(*)) * 100, 0) AS RiskPercentage
            FROM
                HeartHealthData
            GROUP BY
                Gender, SES;
```

Stratified by SES and gender, the query showed differences in heart attack risk, indicating which groups are more vulnerable. In order to lessen inequities in the outcomes of heart disease, these findings offer important insights for focused therapies and public health policy.

4. How Do Modifiable Factors Like Smoking and Sleep Duration Correlate with Heart Attack Risks?

To explore the relationship between modifiable factors such as smoking and sleep duration with heart attack risks, I used a query that calculates the number of high-risk cases (Heart_Attack_Likelihood = 'Yes'), the total number of cases, and the percentage of high-risk cases for each combination of smoking status and sleep duration.

Key components of the query:

- **COUNT(CASE...):**
  - Calculates the number of high-risk cases (RiskCount) for each combination of smoking status and sleep duration.
  - Counts the total number of individuals (TotalCount) in each group.
- **COALESCE(...):** Computes the percentage of high-risk cases (RiskPercentage), ensuring accurate calculations even when some groups have no data.
- **GROUP BY:** Groups the data by Smoking_Status and Sleep_Duration to assess correlations between these modifiable factors and heart attack risks.

```
        SELECT
            Smoking_Status, Sleep_Duration,
            COUNT(CASE WHEN Heart_Attack_Likelihood = 'Yes' THEN 1 END) AS
RiskCount,
            COUNT(*) AS TotalCount,
            COALESCE((COUNT(CASE WHEN Heart_Attack_Likelihood = 'Yes' THEN 1
END) / COUNT(*)) * 100, 0) AS RiskPercentage
        FROM
            HeartHealthData
        GROUP BY
            Smoking_Status, Sleep_Duration;
```

The query uncovered trends demonstrating the relationship between smoking and sleep length variations and the risk of a heart attack, underscoring the significance of these activities for cardiovascular health. These findings highlight how crucial lifestyle changes are in lowering the risk of heart attacks, including quitting smoking and getting the recommended amount of sleep.

5. What Is the Relationship Between Family History of Heart Disease and Clinical Signs Like Cholesterol Levels and Blood Pressure?

To examine the relationship between a family history of heart disease and clinical signs such as cholesterol levels and blood pressure, I used a query that calculates the average cholesterol levels and blood pressure for individuals with a family history of heart disease. The analysis focuses on cases where heart attack likelihood is high (Heart_Attack_Likelihood = 'Yes').

Key components of the query:

- **AVG(Cholesterol_Levels):** Calculates the average cholesterol levels for individuals grouped by their family history of heart disease.
- **AVG(Blood_Pressure):** Computes the average blood pressure values for the same groups.
- **WHERE Clause:** Filters the data to include only individuals with a high heart attack likelihood, ensuring the analysis is targeted.
- **GROUP BY:** Groups results by Family_History_Heart_Disease to evaluate differences between those with and without a family history.

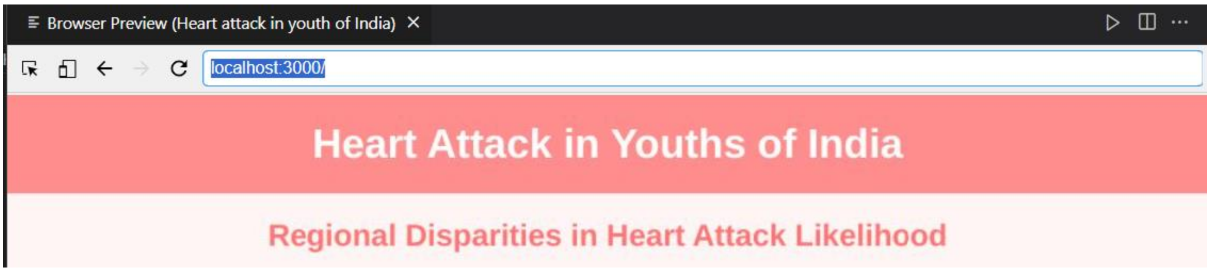
```sql
SELECT
        Family_History_Heart_Disease,
        AVG(Cholesterol_Levels) AS AvgCholesterol,
        AVG(Blood_Pressure) AS AvgBloodPressure
    FROM
        HeartHealthData
    WHERE
        Heart_Attack_Likelihood = 'Yes'
    GROUP BY
        Family_History_Heart_Disease;
```

The findings show that in people at risk of heart attacks, a family history of heart disease is associated with either greater or lower clinical markers such as blood pressure and cholesterol. These revelations highlight how crucial it is to keep an eye on clinical symptoms, especially in people who are genetically predisposed to heart disease.

Stage 4 – Creating simple web application

Here are a few screenshots of the basic web application that is operational in the Coursera lab. The technologies used to build the application include Node.js, Express, EJS, dotenv, and MySQL2, all of which are appropriate for this kind of project and with which I am familiar. A useful homepage that displays tables summarizing the outcomes and conclusions of my data analysis makes up the web application. This method makes the data easier to access and understand by enabling a clear display of the insights obtained from the database searches. The selection of these technologies made the process dependable and efficient by guaranteeing seamless integration and streamlined development.

The webpage is visually appealing due to its basic, clean design and header that focuses the title "Heart Attack in Youths of India" on a gentle pink background. Running locally on port 3000, it effectively presents the results in an understandable manner.



Regional Disparities in Heart Attack Likelihood

This table highlights the regional differences in heart attack risks among young adults. The data shows the number of individuals at isk, total population surveyed, and the risk percentage for each region. Understanding these disparities can help focus preventive neasures where they are most needed.

Region	At Risk (Count)	Total (Count)	Risk Percentage
East	349	1658	21.05%
North	353	1700	20.76%
West	349	1643	21.24%
North-East	337	1589	21.21%
Central	322	1746	18.44%
South	328	1664	19.71%

The results show significant regional differences in young adults' risk of heart attacks in India. A higher concentration of people at risk is found in the West region, which has the greatest risk percentage (21.24%), followed closely by the North-East (21.21%) and the East (21.05%). The South and North regions exhibit moderate risks with 19.71% and 20.76%, respectively, while the Central region has the lowest risk percentage at 18.44%. These differences highlight the necessity of region-specific public health initiatives, with an emphasis on awareness campaigns and preventive interventions in high-risk areas such as the North-East, East, and West. Interventions targeted at further lowering the risk of heart attacks can still be beneficial in lower-risk areas, such as the Central.

Impact of Physical Activity on High BMI Risks

This table examines how different levels of physical activity influence heart attack risks in individuals with high BMI. It provides insights into the preventive role of exercise and highlights the risk percentages for sedentary, moderately active, and highly active individuals.

Physical Activity Level	High BMI At Risk (Count)	Total High BMI (Count)	Risk Percentage
Sedentary	639	2994	21.34%
High	105	586	17.92%
Moderate	511	2362	21.63%

The results show that physical activity significantly reduces the risk of heart attacks in people with high body mass index. With a risk percentage of 21.34%, sedentary people had the highest number of high-risk cases (639), suggesting that inactivity significantly raises the risk of heart attacks. 511 high-risk instances out of 2362 people had a similar risk percentage (21.63%) to those who are moderately active. The preventive function of frequent and intensive physical activity is highlighted by the fact that highly active people have the lowest risk percentage (17.92%), with only 105 high-risk cases among 586 individuals. These findings highlight how crucial it is to encourage physical activity, especially among sedentary people, as a preventative strategy to lower the cardiovascular risks connected to high body mass index.

Heart Attack Likelihood by Gender and SES

This table analyzes the relationship between gender, socioeconomic status (SES), and heart attack risks. It reveals the at-risk counts, total population, and risk percentages for each combination of gender and SES.

Gender	SES	At Risk (Count)	Total (Count)	Risk Percentage
Male	Middle	409	1917	21.34%
Female	Low	367	1946	18.86%
Male	High	198	1011	19.58%
Female	Middle	378	1927	19.62%
Male	Low	441	1985	22.22%
Other	High	8	45	17.78%
Female	High	201	983	20.45%
Other	Middle	20	98	20.41%
Other	Low	16	88	18.18%

The results show notable differences in heart attack risks by SES and gender. Greater susceptibility is shown in these groups, as males with low SES have the highest risk percentage (22.22%), followed by males with intermediate SES (21.34%). Those who identify as "Other" have the lowest risks, with high SES at 17.78%, while women are typically at lower risk, with low SES at 18.86% and high SES at 20.45%. These findings address hazards among women in higher SES categories and emphasize the need for focused interventions, especially for males in low and middle SES groups. Reducing these gaps requires customized approaches.

Correlation Between Smoking, Sleep Duration, and Risks

This table explores how smoking habits and sleep duration affect heart attack risks. It highlights the number of at-risk individuals, tota surveyed population, and risk percentages for various smoking statuses and sleep durations.

Smoking Status	Sleep Duration	At Risk (Count)	Total (Count)	Risk Percentage
Never	8	130	651	19.97%
Occasionally	9	70	346	20.23%
Occasionally	3	72	355	20.28%
Occasionally	7	86	372	23.12%
Never	5	130	647	20.09%
Regularly	10	60	252	23.81%
Never	4	138	641	21.53%
Never	7	112	601	18.64%

The relationship between heart attack risks, sleep length, and smoking habits is highlighted in the table. The risk percentage is highest among regular smokers who get 10 hours of sleep (23.81%), and it is equally higher among occasional smokers who get 7 hours (23.12%). People who don't smoke are generally less at risk; those who sleep eight hours are 19.97% and those who sleep seven hours are 18.64%. The evidence indicates that smoking considerably raises the risk of heart attacks for smokers of all sleep durations, with longer sleep durations increasing the risk even more. These results highlight the significance of quitting smoking and the relationship between smoking and sleep duration in relation to heart health.

Family History and Clinical Metrics

This table investigates the relationship between a family history of heart disease and clinical metrics such as average cholesterol levels and average blood pressure. The data can help identify early warning signs in individuals with hereditary risk factors.

Family History	Average Cholesterol	Average Blood Pressure
Yes	200.06	140.79028006589792
No	199.36	138.67113906359188

The table illustrates the correlation between clinical measures and a family history of heart disease. It reveals that people with a family history have slightly higher average blood pressure (140.79) and cholesterol levels (200.06) than people without a family history, who have average blood pressure of 138.67 and cholesterol levels of 199.36. These results highlight the significance of routine health monitoring and early therapies for those with hereditary risk factors to reduce the risk of cardiovascular problems, as they imply that a family history of heart disease is linked to slightly higher clinical measurements.

Findings and Conclusion:

There are several factors that affect young adults' risk of heart attacks in India. According to regional differences, the Central region has the lowest dangers, while the West and Northeast are the most vulnerable. Promoting exercise is crucial because it dramatically lowers risks, especially for those with high body mass indexes. According to gender and socioeconomic differences, women in high SES categories also exhibit significant dangers, whereas men in low and intermediate SES groups are most at risk. The need for cessation programs is further supported by the substantial correlation between smoking and increased hazards across all sleep lengths. Furthermore, people with a family history of heart disease had somewhat higher blood pressure and cholesterol, which highlights the importance of routine monitoring. These findings underline the importance of targeted public health strategies addressing demographic, lifestyle, and clinical factors to reduce heart attack risks.

References

1. ET Spotlight. (2024, February 28). *Heartbeat unplugged: Addressing the silent epidemic of cardiovascular diseases through an Indian perspective* [Online]. Available at: https://economictimes.indiatimes.com/industry/healthcare/biotech/heartbeat-unplugged-addressing-the-silent-epidemic-of-cardiovascular-diseases-through-an-indian-perspective/articleshow/107849231.cms.

2. Office of Disease Prevention and Health Promotion (ODPHP). (2024, January 26). *Prevention Is Still the Best Medicine* [Online]. Available at: https://odphp.health.gov/news/202401/prevention-still-best-medicine.

3. Loukou, E. (2025, January 2). *Exactly how much exercise you need per day to lose weight AND slash risk of 19 diseases – from diabetes to cancer* [Online]. The Sun. Available at: https://www.the-sun.com/health/13202345/how-much-exercise-per-day-slash-risk-diseases/.

4. Times of India. (2023, December 16). *Why are young adults getting heart attacks these days?* [Online]. Available at: https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/why-are-young-adults-getting-heart-attacks-these-days/articleshow/106039444.cms.

Resources Used

- Panday, A. (2025). *Heart Attack in Youth of India, Version 1*. Retrieved from https://www.kaggle.com/datasets/ankushpanday1/heart-attack-in-youth-of-india?select=heart_attack_youngsters_india.csv
- Smart Draw, ER Model: https://www.smartdraw.com/entity-relationship-diagram/er-diagram-tool.htm
- DBDiagram, Entity-Relationship Diagrams: https://dbdiagram.io/