CM3010 Databases and Advanced Data Techniques

Stroke Prediction

Introduction

Strokes are a major global health concern, as they are routinely the second largest cause of mortality globally. According to estimates from the World Health Organization (WHO), strokes cause around 11% of all deaths worldwide. This highlights the urgent need for better predicting tools and prevention interventions. In a time when data-driven choices are becoming more and more important in the healthcare industry, using predictive analytics to find stroke risk factors can greatly improve early intervention efforts and lower death rates.

The dataset that is the subject of this report was carefully chosen to address this urgent health issue. The dataset functions as a fundamental instrument for forecasting an individual's risk of having a stroke, taking into account various input parameters like age, gender, medical history, and lifestyle choices. A detailed examination of the patient's information is provided by each item in the dataset, enabling a sophisticated assessment of stroke risk.

The dataset depicts human stories and the intricacies of the real world, not just a collection of numbers and categories. This dataset provides a comprehensive understanding of the variables that may influence a person's risk of stroke by including information on everything from basic demographics to specific health metrics. In order to lessen the impact of strokes on people and healthcare systems worldwide, we seek to find patterns and linkages in this data that will lead to improved prevention and treatment approaches.

Description of the Chosen Dataset:

The chosen dataset is an extensive assemblage of demographic and medical information intended to forecast the probability that a patient will suffer a stroke. A comprehensive picture of the variables that may raise a person's risk of stroke is given by the data, which encompasses a variety of personal and health-related details. These characteristics include things like age, gender, marital status, kind of occupation, type of domicile, presence of heart disease and hypertension, average blood sugar level, body mass index, and smoking status. It is a crucial tool for predictive analysis in healthcare because this data culminates into the'stroke' attribute.

URL for the Dataset Source: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

Quality Assessment: The majority of the dataset has excellent quality, with consistent measurements and formats. To verify data trustworthiness and completeness, it is noteworthy that 291 rows having 'NaN' values in the BMI attribute were found and eliminated.

Documentation Assessment: The dataset is comprehensive, providing fine-grained data on a range of characteristics related to the prognosis of stroke, such as lifestyle, health, and demographic variables.

This dataset, with its comprehensive and relevant attributes, is a valuable resource for stroke prediction. The removal of incomplete BMI data was a crucial step in maintaining its quality, though it may introduce some level of bias. The dataset's potential is somewhat hampered by the lack of detailed documentation and updated information.

Interest and Questions:

The dataset is interesting because it has the ability to discover stroke risk factors from comprehensive features, which might have a big influence on public health. Its capacity for prevention and its machine learning applicability make it very significant, providing chances for practical healthcare advancements.

List of questions you aim to answer with the database :

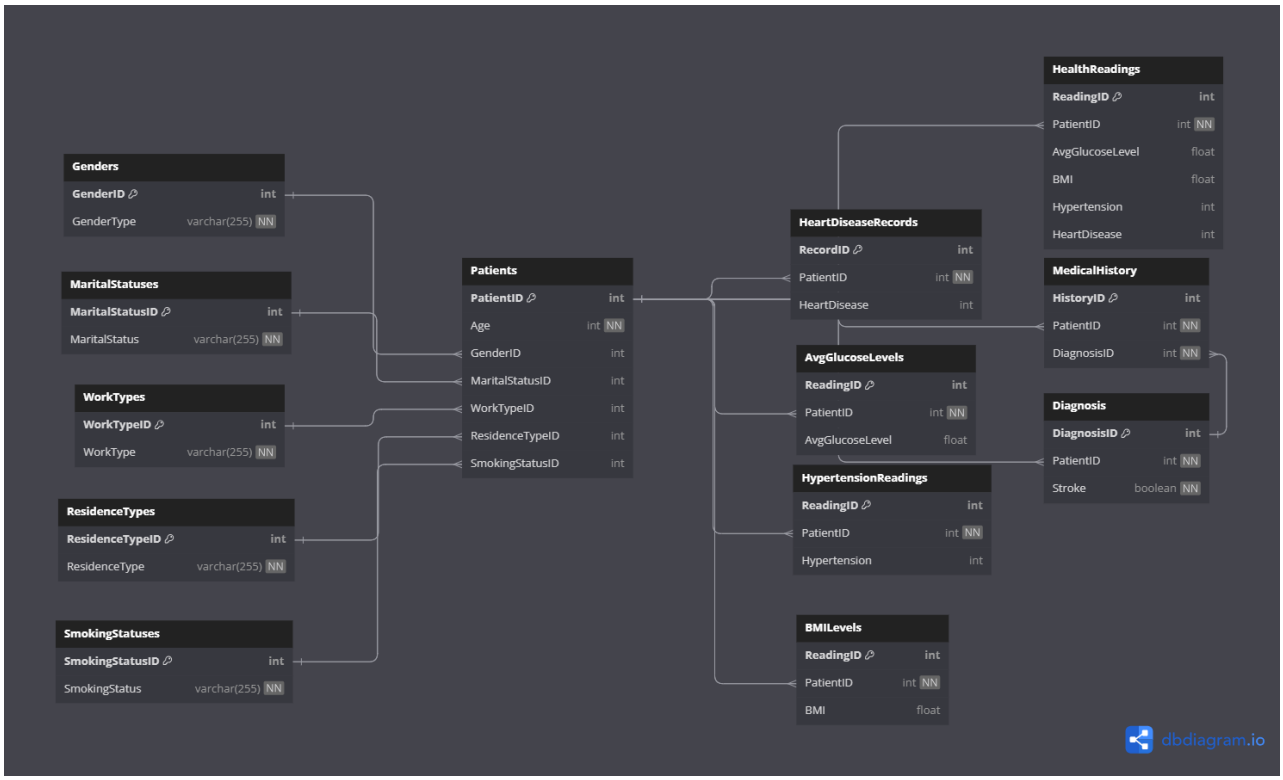
1. How does the incidence of stroke vary among different age groups, and which age cohort is most at risk?
2. What are the variations in stroke incidence across patients with different health conditions such as hypertension and heart disease?
3. What are the top 10 high-risk profiles for stroke when considering BMI, average glucose levels, and smoking status, and how do these factors interplay to escalate risk?
4. How does stroke incidence differ between genders, and are there specific risk factors that are more prevalent in one gender over the other?
5. How does socioeconomic status, inferred from factors like work type and residence, correlate with the incidence of stroke?

E/R Model :

- Lookup tables: These contain static reference data that classifies patients, eliminating duplication and guaranteeing data integrity. They are categorized by genders, marital statuses, work types, residence types, and smoking statuses.

- Table of PatientsPatients Table: It is the foundation of the database, referencing the lookup tables and compiling important demographic data.

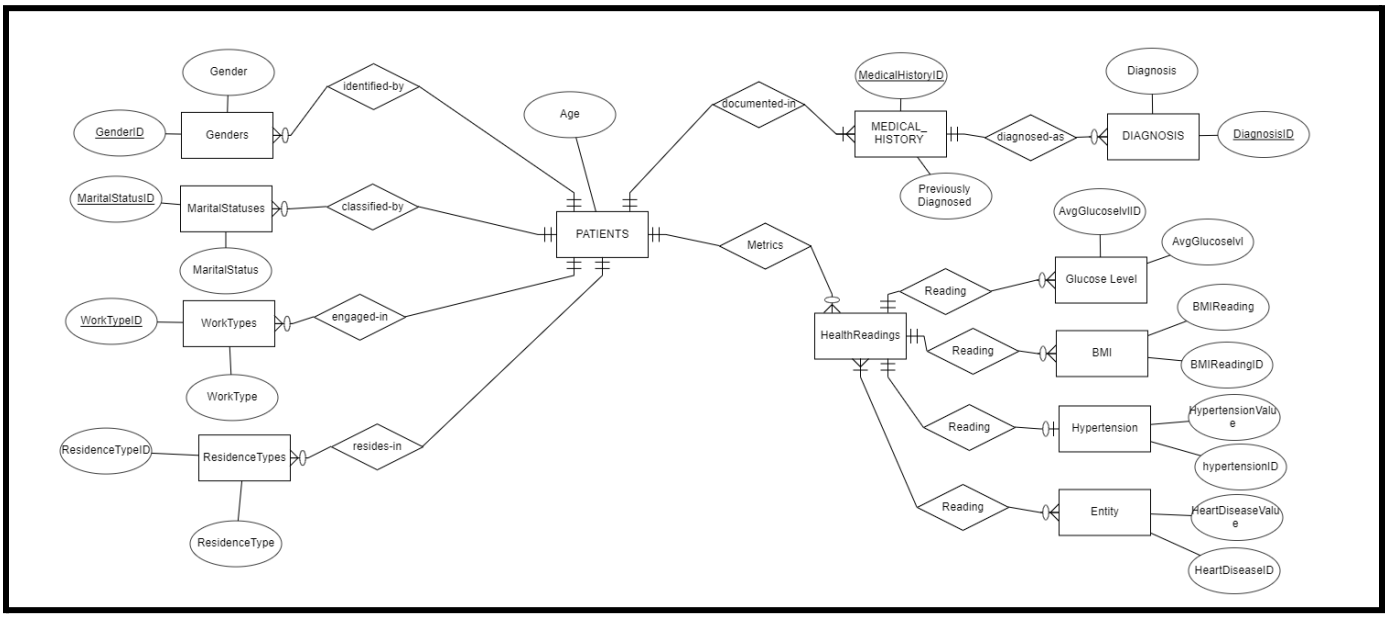
- Medical Tables (Diagnosis, HealthReadings, MedicalHistory): These offer comprehensive health-related information for medical analysis, allowing for the monitoring and association of health variables with ailments such as strokes.



Cardinality: The Patients entity is at the center of most one-to-many (1:N) interactions between entities.

Every patient has a single gender, marital status, employment type, and place of residence; however, numerous patients may fall under each of these categories.
Multiple medical records, including diagnoses and health readings, may be kept on file for each patient.

Entity - Relationship Diagram :



User Tables and Fields Report

Genders Table:
GenderID: Integer (Primary Key, Auto Increment)
GenderType: Varchar (String, Not Null)

MaritalStatuses Table:
MaritalStatusID: Integer (Primary Key, Auto Increment)
MaritalStatus: Varchar (String, Not Null)

WorkTypes Table:
WorkTypeID: Integer (Primary Key, Auto Increment)
WorkType: Varchar (String, Not Null)

ResidenceTypes Table:
ResidenceTypeID: Integer (Primary Key, Auto Increment)
ResidenceType: Varchar (String, Not Null)

SmokingStatuses Table:
SmokingStatusID: Integer (Primary Key, Auto Increment)
SmokingStatus: Varchar (String, Not Null)

Patients Table:
PatientID: Integer (Primary Key)
Age: Integer (Not Null)
GenderID: Integer (Foreign Key)
MaritalStatusID: Integer (Foreign Key)
WorkTypeID: Integer (Foreign Key)
ResidenceTypeID: Integer (Foreign Key)
SmokingStatusID: Integer (Foreign Key)

Diagnosis Table:
DiagnosisID: Integer (Primary Key, Auto Increment)
PatientID: Integer (Foreign Key)
Stroke: Boolean (Not Null)

HealthReadings Table:
ReadingID: Integer (Primary Key, Auto Increment)
PatientID: Integer (Foreign Key)
AvgGlucoseLevel: Float
BMI: Float
Hypertension: Integer
HeartDisease: Integer

MedicalHistory Table:
HistoryID: Integer (Primary Key, Auto Increment)
PatientID: Integer (Foreign Key)
DiagnosisID: Integer (Foreign Key)

AvgGlucoseLevels Table:
ReadingID: Integer (Primary Key, Auto Increment)
PatientID: Integer (Foreign Key)
AvgGlucoseLevel: Float

BMILevels Table:
ReadingID: Integer (Primary Key, Auto Increment)
PatientID: Integer (Foreign Key)
BMI: Float

HypertensionReadings Table:
ReadingID: Integer (Primary Key, Auto Increment)
PatientID: Integer (Foreign Key)
Hypertension: Integer

HeartDiseaseRecords Table:
RecordID: Integer (Primary Key, Auto Increment)
PatientID: Integer (Foreign Key)
HeartDisease: Integer

The structure of each table within the database and the fields contained within them, which are designed to collect and organize comprehensive health-related data on patients.

Description of the Database Structure:

Lookup Tables: Genders, MaritalStatuses, WorkTypes, ResidenceTypes, and SmokingStatuses serve as dictionaries for other tables to reference, standardizing the data and reducing redundancy.

Central Table: Patients acts as the core table, containing demographic data and serving as the hub to which other health-related tables connect.

Health-Related Tables: Diagnosis, HealthReadings, MedicalHistory, AvgGlucoseLevels, BMILevels, HypertensionReadings, and HeartDiseaseRecords are focused on recording various health parameters and conditions, providing a comprehensive view of each patient's health status and medical history.

Foreign Key Constraints: They are used to enforce referential integrity between the tables, ensuring that relationships between the data are maintained correctly.

```
%%writefile $SCRIPT_PATH/create-tables.sql

USE stroke;
SET FOREIGN_KEY_CHECKS = 0;
-- Drop Table Statements
DROP TABLE IF EXISTS Patients;
DROP TABLE IF EXISTS Genders;
DROP TABLE IF EXISTS MaritalStatuses;
DROP TABLE IF EXISTS WorkTypes;
DROP TABLE IF EXISTS ResidenceTypes;
DROP TABLE IF EXISTS MedicalHistory;
DROP TABLE IF EXISTS SmokingStatuses;
DROP TABLE IF EXISTS Diagnosis;
DROP TABLE IF EXISTS HealthReadings;
DROP TABLE IF EXISTS AvgGlucoseLevels;
DROP TABLE IF EXISTS HeartDiseaseRecords;
DROP TABLE IF EXISTS HypertensionReadings;
DROP TABLE IF EXISTS BMILevels;
SET FOREIGN_KEY_CHECKS = 1;
```

The SQL script drops existing tables related to patients and their medical information, switches to the stroke database, and temporarily disables foreign key checks to prevent integrity constraint issues. Usually, this cleanup is carried out prior to establishing new tables in order to reset the database structure. Foreign key checks are enabled again after dropping the tables in order to guarantee the integrity of incoming data.

This SQL script defines five tables to categorize patient data:

- Genders: Stores gender types with a unique ID.
- MaritalStatuses: Holds marital statuses with a unique ID.
- WorkTypes: Records types of work with a unique ID.
- ResidenceTypes: Contains residence types with a unique ID.
- SmokingStatuses: Lists smoking statuses with a unique ID.
- Each table has an auto-incrementing primary key for unique identification and a descriptive VARCHAR field.

```sql
-- Table: Genders
CREATE TABLE Genders (
    GenderID INT AUTO_INCREMENT PRIMARY KEY,
    GenderType VARCHAR(255) NOT NULL
);

-- Table: MaritalStatuses
CREATE TABLE MaritalStatuses (
    MaritalStatusID INT AUTO_INCREMENT PRIMARY KEY,
    MaritalStatus VARCHAR(255) NOT NULL
);

-- Table: WorkTypes
CREATE TABLE WorkTypes (
    WorkTypeID INT AUTO_INCREMENT PRIMARY KEY,
    WorkType VARCHAR(255) NOT NULL
);

-- Table: ResidenceTypes
CREATE TABLE ResidenceTypes (
    ResidenceTypeID INT AUTO_INCREMENT PRIMARY KEY,
    ResidenceType VARCHAR(255) NOT NULL
);

-- Table: SmokingStatuses
CREATE TABLE SmokingStatuses (
    SmokingStatusID INT AUTO_INCREMENT PRIMARY KEY,
    SmokingStatus VARCHAR(255) NOT NULL
);
```

```sql
-- Table: Patients
CREATE TABLE Patients (
    PatientID INT PRIMARY KEY,
    Age INT NOT NULL,
    GenderID INT,
    MaritalStatusID INT,
    WorkTypeID INT,
    ResidenceTypeID INT,
    SmokingStatusID INT,
    FOREIGN KEY (GenderID) REFERENCES Genders(GenderID),
    FOREIGN KEY (MaritalStatusID) REFERENCES MaritalStatuses(MaritalStatusID),
    FOREIGN KEY (WorkTypeID) REFERENCES WorkTypes(WorkTypeID),
    FOREIGN KEY (ResidenceTypeID) REFERENCES ResidenceTypes(ResidenceTypeID),
    FOREIGN KEY (SmokingStatusID) REFERENCES SmokingStatuses(SmokingStatusID)
);
```

This SQL script creates a Patients table to store individual patient records.Each record includes a unique PatientID, the patient's Age, and foreign keys linking to various descriptive tables like Genders, MaritalStatuses, WorkTypes, ResidenceTypes, and SmokingStatuses. These foreign keys establish relationships with other tables, allowing the Patients table to reference detailed categorical data.

The SQL script creates three tables related to patient health:

- Diagnosis: Records stroke diagnoses with a unique ID, linked to patients via PatientID.
- HealthReadings: Stores health metrics like glucose level, BMI, hypertension, and heart disease status, each reading with a unique ID and linked to patients.
- MedicalHistory: Captures the history of diagnoses for each patient, with links to the Patients and Diagnosis tables.

```sql
-- Table: Diagnosis
CREATE TABLE Diagnosis (
    DiagnosisID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    Stroke BOOLEAN NOT NULL,
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);

-- Table: HealthReadings
CREATE TABLE HealthReadings (
    ReadingID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    AvgGlucoseLevel FLOAT,
    BMI FLOAT,
    Hypertension INT,  -- Added Hypertension as an integer field
    HeartDisease INT,  -- Added HeartDisease as an integer field
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);

-- Table: MedicalHistory
CREATE TABLE MedicalHistory (
    HistoryID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    DiagnosisID INT,
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID),
    FOREIGN KEY (DiagnosisID) REFERENCES Diagnosis(DiagnosisID)
);
```

```
CREATE TABLE AvgGlucoseLevels (
    ReadingID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    AvgGlucoseLevel FLOAT,
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);

CREATE TABLE BMILevels (
    ReadingID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    BMI FLOAT,
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);

CREATE TABLE HypertensionReadings (
    ReadingID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    Hypertension INT,  -- Assuming 0 for no hypertension and 1 for hypertension
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);

CREATE TABLE HeartDiseaseRecords (
    RecordID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    HeartDisease INT,  -- Assuming 0 for no heart disease and 1 for heart disease
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);
```

The SQL script defines four tables to track specific health metrics for patients:

- AvgGlucoseLevels: Captures average glucose levels with a unique ReadingID and links to patients.
- BMILevels: Stores Body Mass Index (BMI) readings with a unique ReadingID and links to patients.
- HypertensionReadings: Records hypertension status (0 for no, 1 for yes) with a unique ReadingID and links to patients.
- HeartDiseaseRecords: Notes heart disease status (0 for no, 1 for yes) with a unique RecordID and links to patients.

```
%%writefile $SCRIPT_PATH/create-tables.sql
```

We are creating a script named create-tables.sql to execute SQL commands in Jupyter Notebook. This script will contain all the necessary SQL statements to define and create our database tables and their relationships.

Use of Denormalized Table:

```
%%writefile $SCRIPT_PATH/load-dnorm-data.sql
USE stroke;

DROP TABLE IF EXISTS denormalised;
CREATE TABLE denormalised (
    PatientID INT,
    Gender VARCHAR(255),
    Age INT,
    Hypertension INT,
    HeartDisease INT,
    EverMarried VARCHAR(255),
    WorkType VARCHAR(255),
    ResidenceType VARCHAR(255),
    AvgGlucoseLevel FLOAT,
    BMI FLOAT,
    SmokingStatus VARCHAR(255),
    Stroke INT
);

-- Load data into denormalised table
LOAD DATA INFILE "/home/coder/project/stroke/data/main_stroke_record.csv"
INTO TABLE denormalised
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS;
```

This flat table is initially populated with all of the data imported from the CSV. In addition to making data transformation and cleaning easier, it streamlines the initial data intake process before distributing the data into the normalized database structure.

The denormalized table serves as a straightforward staging area to prepare data for ingestion into a more complex, relational database design.

Data Ingestion:

```
%%writefile $SCRIPT_PATH/ingest-data.sql
USE stroke;

SET FOREIGN_KEY_CHECKS = 0;

TRUNCATE TABLE MaritalStatuses;
TRUNCATE TABLE WorkTypes;
TRUNCATE TABLE ResidenceTypes;
TRUNCATE TABLE Patients;
TRUNCATE TABLE HealthReadings;
TRUNCATE TABLE SmokingStatuses;
TRUNCATE TABLE Diagnosis;
TRUNCATE TABLE Genders;
SET FOREIGN_KEY_CHECKS = 1;
```

The script is designed to truncate (empty) certain tables containing patient and stroke data, disable foreign key checks, and then enable foreign key checks again. This procedure is used to remove old data and get the database ready to accept new entries.

```sql
-- Insert data into normalized tables from denormalized data

-- Insert data into Genders
INSERT INTO Genders (GenderType)
    SELECT DISTINCT Gender
    FROM denormalised;

-- Insert into MaritalStatuses
INSERT INTO MaritalStatuses (MaritalStatus)
    SELECT DISTINCT EverMarried
    FROM denormalised;

-- Insert into WorkTypes
INSERT INTO WorkTypes (WorkType)
    SELECT DISTINCT WorkType
    FROM denormalised;

-- Insert into ResidenceTypes
INSERT INTO ResidenceTypes (ResidenceType)
    SELECT DISTINCT ResidenceType
    FROM denormalised;

-- Insert into SmokingStatuses
INSERT INTO SmokingStatuses (SmokingStatus)
    SELECT DISTINCT SmokingStatus
    FROM denormalised;
```

```sql
-- Insert into Patients
INSERT INTO Patients (PatientID, Age, GenderID, MaritalStatusID, WorkTypeID, ResidenceTypeID, SmokingStatusID)
    SELECT DISTINCT
        D.PatientID,
        D.Age,
        G.GenderID,
        M.MaritalStatusID,
        W.WorkTypeID,
        R.ResidenceTypeID,
        SS.SmokingStatusID
    FROM
        denormalised D
    INNER JOIN Genders G ON D.Gender = G.GenderType
    INNER JOIN MaritalStatuses M ON D.EverMarried = M.MaritalStatus
    INNER JOIN WorkTypes W ON D.WorkType = W.WorkType
    INNER JOIN ResidenceTypes R ON D.ResidenceType = R.ResidenceType
    LEFT JOIN SmokingStatuses SS ON D.SmokingStatus = SS.SmokingStatus;


-- Insert into HealthReadings (Average Glucose Level, Body Mass Index, Hypertension, and Heart Disease)
INSERT INTO HealthReadings (PatientID, AvgGlucoseLevel, BMI, Hypertension, HeartDisease)
SELECT
    D.PatientID,
    D.AvgGlucoseLevel,
    D.BMI,
    D.Hypertension,
    D.HeartDisease
FROM
    denormalised D;
```

```sql
-- Insert into Diagnosis (Stroke)
INSERT INTO Diagnosis (PatientID, Stroke)
    SELECT PatientID, Stroke
    FROM denormalised;


INSERT INTO AvgGlucoseLevels (PatientID, AvgGlucoseLevel)
SELECT
    PatientID,
    AvgGlucoseLevel
FROM
    denormalised;

INSERT INTO BMILevels (PatientID, BMI)
SELECT
    PatientID,
    BMI
FROM
    denormalised;

INSERT INTO HypertensionReadings (PatientID, Hypertension)
SELECT
    PatientID,
    Hypertension
FROM
    denormalised;
```

```sql
INSERT INTO HeartDiseaseRecords (PatientID, HeartDisease)
SELECT
    PatientID,
    HeartDisease
FROM
    denormalised;
```

Data is imported by the SQL script from a "denormalized" table into a number of normalized tables. It inserts distinct values based on smoking status, work type, residence type, gender, and marital status. subsequently, it adds combined demographic data to the Patients dataset, health metrics to the HealthReadings table, and stroke data to the Diagnosis table. The presence of additional tables with records for heart disease, hypertension, BMI, and average glucose levels guarantees a standardized and organized database.

Database Reflection:

Upon reflection, the database structured from the stroke dataset generally performs well in several aspects but also reveals areas needing improvement.

Positive Features:

- **Normalization**: To prevent redundancy and preserve data integrity, distinct tables for attributes such as Genders, MaritalStatuses, and WorkTypes are created using an organized manner. This works especially well for qualities that have a small number of distinct values.
- **Referential Integrity:** Accurate and consistent relationships between various related bits of data are ensured by using foreign keys and linking tables based on these keys, which improves the database's overall reliability.
- **Referential Integrity:** By employing foreign keys and linking tables based on these keys, accurate and consistent relationships between various connected bits of data are ensured, which increases the overall reliability of the database.

Areas for Improvement:

Medical History Representation: There appears to be an issue in accurately reflecting medical history, such as hypertension and heart disease status. This might be due to a design flaw in the table structure or relationships, hindering comprehensive analysis.

Web Application:
The web application, crafted with Node.js, is an interactive tool designed for detailed stroke data analysis. It's structured to provide a seamless and informative user experience.

Homepage:
The central navigation area where users are greeted with five distinct analytical options related to stroke data. This design ensures users have a clear starting point and can easily understand their options.

Analysis Selection:
Users can select from the five available analyses. Each option is tailored to offer insights into specific aspects of stroke risks, such as age, gender, or lifestyle factors. Once a choice is made, the application directs users to a dedicated page for that particular analysis.

This page explores the selected subject in detail and offers data-driven visualizations such as tables and graphs in addition to a narrative explanation. The goal is to make the analysis simply readable so that people can understand the patterns and insights found in the stroke data both visually and through reading.

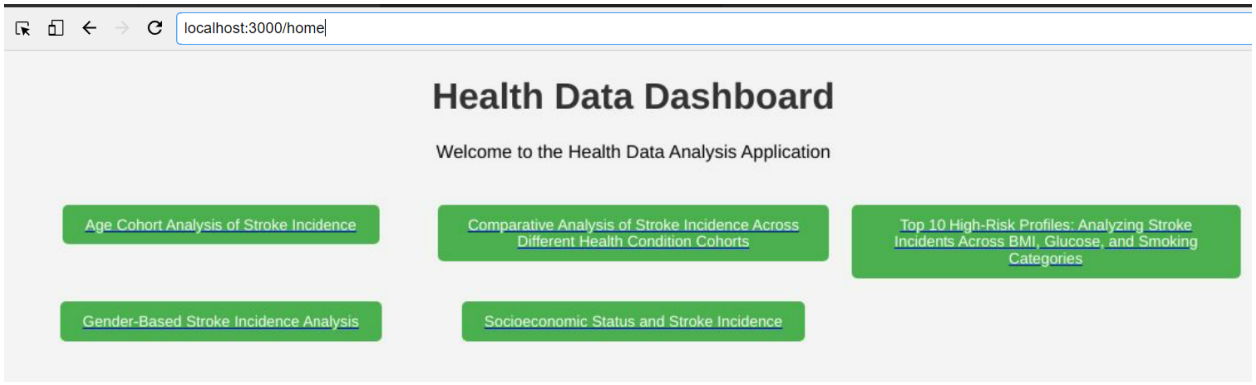
There's always a 'Home' button to make sure users can easily go back to the site and view more analytics. This feature makes navigation simple, which reduces confusion and improves the exploration experience.

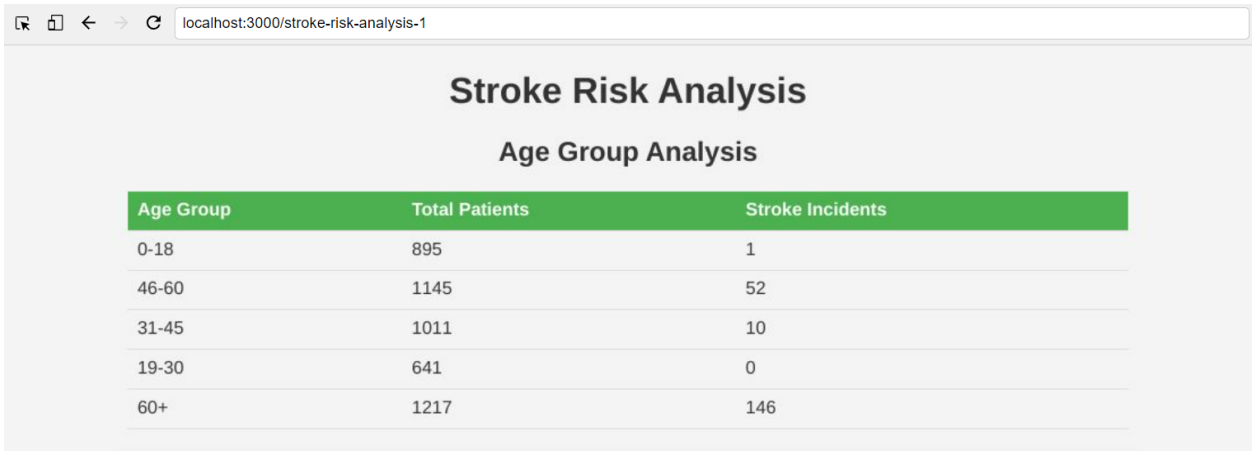
The application is set to run on port 3000.

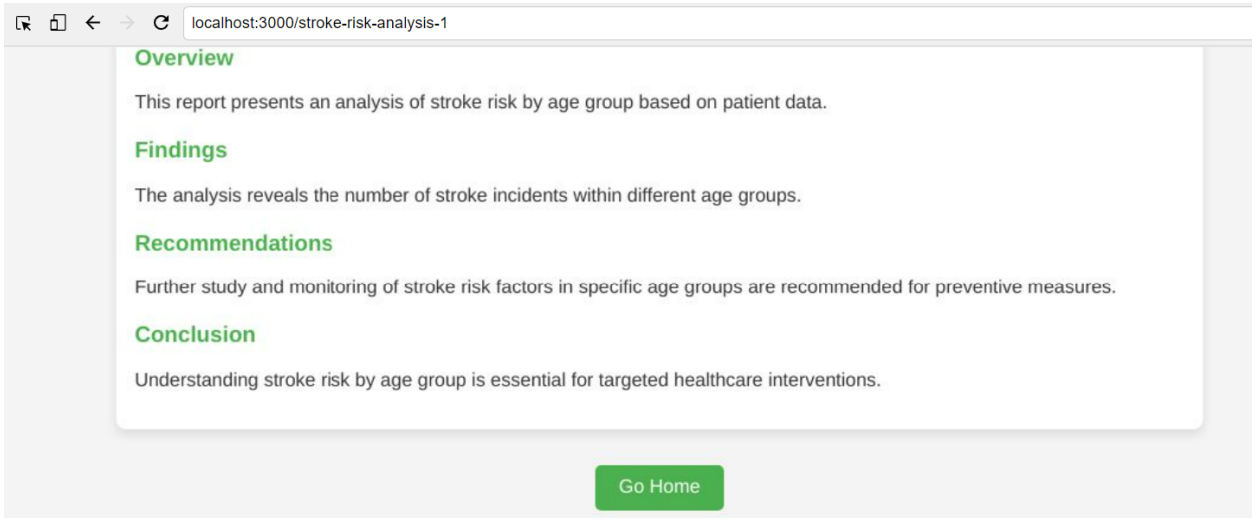
Routing and Data Handling:

The program interfaces with the MySQL database using mysql and handles HTTP requests using express. It uses mustache-express for the view engine and body-parser to parse incoming request bodies, enabling the rendering of server-side templates. By controlling environment variables, the dotenv library improves security and configurability.

Home Route: The /home route serves the homepage of the application.
Analysis Routes: There are multiple routes for different stroke risk analyses (e.g., /stroke-risk-analysis-1, /stroke-risk-analysis-2, etc.). Each route executes a specific SQL query to fetch and process data from the database. The results are then passed to the corresponding Mustache template for rendering.


localhost:3000/home
Health Data Dashboard
Welcome to the Health Data Analysis Application
Age Cohort Analysis of Stroke Incidence
Comparative Analysis of Stroke Incidence Across Different Health Condition Cohorts
Top 10 High-Risk Profiles: Analyzing Stroke Incidents Across BMI, Glucose, and Smoking Categories
Gender-Based Stroke Incidence Analysis
Socioeconomic Status and Stroke Incidence


localhost:3000/stroke-risk-analysis-1
Stroke Risk Analysis
Age Group Analysis
Age Group
Total Patients
Stroke Incidents
0-18
895
1
46-60
1145
52
31-45
1011
10
19-30
641
0
60+
1217
146


localhost:3000/stroke-risk-analysis-1
Overview
This report presents an analysis of stroke risk by age group based on patient data.
Findings
The analysis reveals the number of stroke incidents within different age groups.
Recommendations
Further study and monitoring of stroke risk factors in specific age groups are recommended for preventive measures.
Conclusion
Understanding stroke risk by age group is essential for targeted healthcare interventions.
Go Home

Evaluation and Conclusion:
The project was divided into multiple phases, the first of which was the construction of a relational database structure that was normalized using DBML and SQL. To enhance data integrity, the data was ingested from CSV files into the denormalized table and subsequently converted into normalized tables.

In order to examine stroke risk factors, including age groups, BMI categories, glucose levels, hypertension, heart disease, and smoking status, SQL analysis was performed on the database. The identification of possible associations between these parameters and stroke events was made possible by these analyses.

The web application was constructed with HTML, CSS, and Express.js. It gives users access to a variety of stroke risk analyses and presents the findings in an easy-to-read manner.

Error control and data consistency were critical throughout the process. Extensive data mapping was done to guarantee data quality during the ingestion phase. Error messages were recorded in order to quickly detect and fix problems.

In summary, the database design, data ingestion, analysis, and web application development stages were all effectively completed by the project. It offers insightful information about stroke risk factors. Notwithstanding obstacles like inconsistent data, the project's product is a useful and educational instrument for comprehending and reducing the risk of stroke.

Reference:

- *Stroke Prediction Dataset*. (2021, January 26). Kaggle. https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset
- *A free database designer for developers and analysts*. (n.d.). https://dbdiagram.io/d/Final-Stroke-6556153a3be149578720d5cb
- info@erdplus.com. (n.d.). *ERDPlus*. Mosor, Inc. https://erdplus.com/
- *SQL Tutorial*. (n.d.). https://www.w3schools.com/sql/default.asp
- *Google Colaboratory*. (n.d.). https://colab.research.google.com/
- *CouRseRa | Online courses & credentials from top educators.| CourseRA*. (n.d.). https://www.coursera.org/learn/uol-cm3010-databases-and-advanced-data-techniques/ungradedLab/wNf6O/mid-term-assignment-lab

Appendices:

- Codes

```
import pandas as pd
pd.options.mode.chained_assignment = None

df_main = pd.read_csv(SRC_FILE_PATH)
df_main
```

```
df_main = df_main.dropna(subset=['bmi'])
df_main
# save the clean data to a csv file for data loading and ingestion
df_main.to_csv(f"{DATA_PATH}/main_stroke_record.csv", index=False)

%%writefile $SCRIPT_PATH/create-tables.sql

USE stroke;
SET FOREIGN_KEY_CHECKS = 0;
-- Drop Table Statements
DROP TABLE IF EXISTS Patients;
DROP TABLE IF EXISTS Genders;
DROP TABLE IF EXISTS MaritalStatuses;
DROP TABLE IF EXISTS WorkTypes;
DROP TABLE IF EXISTS ResidenceTypes;
DROP TABLE IF EXISTS MedicalHistory;
DROP TABLE IF EXISTS SmokingStatuses;
DROP TABLE IF EXISTS Diagnosis;
DROP TABLE IF EXISTS HealthReadings;
DROP TABLE IF EXISTS AvgGlucoseLevels;
DROP TABLE IF EXISTS HeartDiseaseRecords;
DROP TABLE IF EXISTS HypertensionReadings;
DROP TABLE IF EXISTS BMILevels;
SET FOREIGN_KEY_CHECKS = 1;

-- Table: Genders
CREATE TABLE Genders (
    GenderID INT AUTO_INCREMENT PRIMARY KEY,
    GenderType VARCHAR(255) NOT NULL
);
```

```
-- Table: MaritalStatuses
CREATE TABLE MaritalStatuses (
    MaritalStatusID INT AUTO_INCREMENT PRIMARY KEY,
    MaritalStatus VARCHAR(255) NOT NULL
);

-- Table: WorkTypes
CREATE TABLE WorkTypes (
    WorkTypeID INT AUTO_INCREMENT PRIMARY KEY,
    WorkType VARCHAR(255) NOT NULL
);

-- Table: ResidenceTypes
CREATE TABLE ResidenceTypes (
    ResidenceTypeID INT AUTO_INCREMENT PRIMARY KEY,
    ResidenceType VARCHAR(255) NOT NULL
);

-- Table: SmokingStatuses
CREATE TABLE SmokingStatuses (
    SmokingStatusID INT AUTO_INCREMENT PRIMARY KEY,
    SmokingStatus VARCHAR(255) NOT NULL
);
-- Table: Patients
CREATE TABLE Patients (
    PatientID INT PRIMARY KEY,
    Age INT NOT NULL,
    GenderID INT,
    MaritalStatusID INT,
    WorkTypeID INT,
    ResidenceTypeID INT,
    SmokingStatusID INT,
    FOREIGN KEY (GenderID) REFERENCES Genders(GenderID),
    FOREIGN KEY (MaritalStatusID) REFERENCES
MaritalStatuses(MaritalStatusID),
    FOREIGN KEY (WorkTypeID) REFERENCES WorkTypes(WorkTypeID),
    FOREIGN KEY (ResidenceTypeID) REFERENCES
ResidenceTypes(ResidenceTypeID),
```

```
    FOREIGN KEY (SmokingStatusID) REFERENCES
SmokingStatuses(SmokingStatusID)
);

-- Table: Diagnosis
CREATE TABLE Diagnosis (
    DiagnosisID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    Stroke BOOLEAN NOT NULL,
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);

-- Table: HealthReadings
CREATE TABLE HealthReadings (
    ReadingID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    AvgGlucoseLevel FLOAT,
    BMI FLOAT,
    Hypertension INT,  -- Added Hypertension as an integer field
    HeartDisease INT,  -- Added HeartDisease as an integer field
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);

-- Table: MedicalHistory
CREATE TABLE MedicalHistory (
    HistoryID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    DiagnosisID INT,
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID),
    FOREIGN KEY (DiagnosisID) REFERENCES Diagnosis(DiagnosisID)
);

CREATE TABLE AvgGlucoseLevels (
    ReadingID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    AvgGlucoseLevel FLOAT,
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);
```

```
CREATE TABLE BMILevels (
    ReadingID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    BMI FLOAT,
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);

CREATE TABLE HypertensionReadings (
    ReadingID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    Hypertension INT,  -- Assuming 0 for no hypertension and 1 for
hypertension
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);

CREATE TABLE HeartDiseaseRecords (
    RecordID INT AUTO_INCREMENT PRIMARY KEY,
    PatientID INT,
    HeartDisease INT,  -- Assuming 0 for no heart disease and 1 for
heart disease
    FOREIGN KEY (PatientID) REFERENCES Patients(PatientID)
);
```

```
%%writefile $SCRIPT_PATH/load-dnorm-data.sql
USE stroke;

DROP TABLE IF EXISTS denormalised;
CREATE TABLE denormalised (
    PatientID INT,
    Gender VARCHAR(255),
    Age INT,
    Hypertension INT,
    HeartDisease INT,
    EverMarried VARCHAR(255),
    WorkType VARCHAR(255),
    ResidenceType VARCHAR(255),
    AvgGlucoseLevel FLOAT,
```

```
    BMI FLOAT,
    SmokingStatus VARCHAR(255),
    Stroke INT
);

-- Load data into denormalised table
LOAD DATA INFILE
"/home/coder/project/stroke/data/main_stroke_record.csv"
INTO TABLE denormalised
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS;

%%writefile $SCRIPT_PATH/ingest-data.sql
USE stroke;

SET FOREIGN_KEY_CHECKS = 0;

TRUNCATE TABLE MaritalStatuses;
TRUNCATE TABLE WorkTypes;
TRUNCATE TABLE ResidenceTypes;
TRUNCATE TABLE Patients;
TRUNCATE TABLE HealthReadings;
TRUNCATE TABLE SmokingStatuses;
TRUNCATE TABLE Diagnosis;
TRUNCATE TABLE Genders;
SET FOREIGN_KEY_CHECKS = 1;

-- Insert data into normalized tables from denormalized data

-- Insert data into Genders
INSERT INTO Genders (GenderType)
    SELECT DISTINCT Gender
    FROM denormalised;

-- Insert into MaritalStatuses
INSERT INTO MaritalStatuses (MaritalStatus)
    SELECT DISTINCT EverMarried
```

```
    FROM denormalised;

-- Insert into WorkTypes
INSERT INTO WorkTypes (WorkType)
    SELECT DISTINCT WorkType
    FROM denormalised;

-- Insert into ResidenceTypes
INSERT INTO ResidenceTypes (ResidenceType)
    SELECT DISTINCT ResidenceType
    FROM denormalised;

-- Insert into SmokingStatuses
INSERT INTO SmokingStatuses (SmokingStatus)
    SELECT DISTINCT SmokingStatus
    FROM denormalised;

-- Insert into Patients
INSERT INTO Patients (PatientID, Age, GenderID, MaritalStatusID,
WorkTypeID, ResidenceTypeID, SmokingStatusID)
    SELECT DISTINCT
        D.PatientID,
        D.Age,
        G.GenderID,
        M.MaritalStatusID,
        W.WorkTypeID,
        R.ResidenceTypeID,
        SS.SmokingStatusID
    FROM
        denormalised D
    INNER JOIN Genders G ON D.Gender = G.GenderType
    INNER JOIN MaritalStatuses M ON D.EverMarried = M.MaritalStatus
    INNER JOIN WorkTypes W ON D.WorkType = W.WorkType
    INNER JOIN ResidenceTypes R ON D.ResidenceType = R.ResidenceType
    LEFT JOIN SmokingStatuses SS ON D.SmokingStatus = SS.SmokingStatus;


-- Insert into HealthReadings (Average Glucose Level, Body Mass Index,
Hypertension, and Heart Disease)
```

```
INSERT INTO HealthReadings (PatientID, AvgGlucoseLevel, BMI,
Hypertension, HeartDisease)
SELECT
    D.PatientID,
    D.AvgGlucoseLevel,
    D.BMI,
    D.Hypertension,
    D.HeartDisease
FROM
    denormalised D;

-- Insert into Diagnosis (Stroke)
INSERT INTO Diagnosis (PatientID, Stroke)
    SELECT PatientID, Stroke
    FROM denormalised;


INSERT INTO AvgGlucoseLevels (PatientID, AvgGlucoseLevel)
SELECT
    PatientID,
    AvgGlucoseLevel
FROM
    denormalised;

INSERT INTO BMILevels (PatientID, BMI)
SELECT
    PatientID,
    BMI
FROM
    denormalised;

INSERT INTO HypertensionReadings (PatientID, Hypertension)
SELECT
    PatientID,
    Hypertension
FROM
    denormalised;

INSERT INTO HeartDiseaseRecords (PatientID, HeartDisease)
```

```
SELECT
    PatientID,
    HeartDisease
FROM
    denormalised;



-- Insert records into MedicalHistory to link patients to their
diagnoses
-- INSERT INTO MedicalHistory (PatientID, DiagnosisID)
-- SELECT
--    P.PatientID,
--    D.DiagnosisID
--FROM
--    Patients P
--JOIN
--    Diagnosis D ON P.PatientID = D.PatientID;
```

```
%%sql

SELECT
    P.PatientID,
    P.Age,
    G.GenderType AS 'Gender',
    MS.MaritalStatus AS 'Marital Status',
    WT.WorkType AS 'Work Type',
    RT.ResidenceType AS 'Residence Type',
    AG.AvgGlucoseLevel AS 'Average Glucose Level',
    BM.BMI AS 'Body Mass Index',
    HP.Hypertension AS 'Hypertension',
    HD.HeartDisease AS 'Heart Disease',
    SS.SmokingStatus AS 'Smoking Status',
    D.Stroke AS 'Stroke'
FROM Patients P
LEFT JOIN Genders G ON P.GenderID = G.GenderID
LEFT JOIN MaritalStatuses MS ON P.MaritalStatusID = MS.MaritalStatusID
```

```
LEFT JOIN WorkTypes WT ON P.WorkTypeID = WT.WorkTypeID
LEFT JOIN ResidenceTypes RT ON P.ResidenceTypeID = RT.ResidenceTypeID
LEFT JOIN AvgGlucoseLevels AG ON P.PatientID = AG.PatientID
LEFT JOIN BMILevels BM ON P.PatientID = BM.PatientID
LEFT JOIN HypertensionReadings HP ON P.PatientID = HP.PatientID
LEFT JOIN HeartDiseaseRecords HD ON P.PatientID = HD.PatientID
LEFT JOIN SmokingStatuses SS ON P.SmokingStatusID = SS.SmokingStatusID
LEFT JOIN Diagnosis D ON P.PatientID = D.PatientID
GROUP BY P.PatientID, P.Age, G.GenderType, MS.MaritalStatus,
WT.WorkType, RT.ResidenceType, AG.AvgGlucoseLevel, BM.BMI,
HP.Hypertension, HD.HeartDisease, SS.SmokingStatus, D.Stroke;
```

Code for Web App:

home.html

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8" />
    <title>Health Data Dashboard</title>
    <style>
        body {
            font-family: Arial, sans-serif;
            text-align: center;
            margin: 0;
            padding: 0;
            background-color: #f4f4f4;
        }

        h1 {
            color: #333;
        }

        .buttons-container {
            display: grid;
```

```
            grid-template-columns: repeat(3, 1fr);
            gap: 20px;
            padding: 20px;
        }
        .button {
            display: block;
            padding: 10px 20px;
            margin: auto;
            background-color: #4CAF50;
            color: white;
            text-decoration: none;
            border: none;
            border-radius: 5px;
            cursor: pointer;
        }
        .button:hover {
            background-color: #45a049;
        }
    </style>
</head>
<body>
    <h1>Health Data Dashboard</h1>
    <p>Welcome to the Health Data Analysis Application</p>
    <div class="buttons-container">
        <a href="/stroke-risk-analysis-1"><button class="button">Age
Cohort Analysis of Stroke Incidence</button></a>
        <a href="/stroke-risk-analysis-2"><button
class="button">Comparative Analysis of Stroke Incidence Across Different
Health Condition Cohorts</button></a>
        <a href="/stroke-risk-analysis-3"><button class="button">Top 10
High-Risk Profiles: Analyzing Stroke Incidents Across BMI, Glucose, and
Smoking Categories</button></a>
        <a href="/stroke-risk-analysis-4"><button
class="button">Gender-Based Stroke Incidence Analysis</button></a>
        <a href="/stroke-risk-analysis-5"><button
class="button">Socioeconomic Status and Stroke Incidence</button></a>
    </div>
</body>
</html>
```

Stroke-risk-anlaysis-1.html

```
<!DOCTYPE html>
<html>

<head>
    <title>Stroke Risk Analysis</title>
</head>

<body>
    <h1>Stroke Risk Analysis</h1>

    <!-- Age Group Analysis Table -->
    <h2>Age Group Analysis</h2>
    <table>
        <tr>
            <th>Age Group</th>
            <th>Total Patients</th>
            <th>Stroke Incidents</th>
        </tr>
        {{#ageGroups}}
        <tr>
            <td>{{AgeGroup}}</td>
            <td>{{TotalPatients}}</td>
            <td>{{StrokeIncidents}}</td>
        </tr>
        {{/ageGroups}}
    </table>

    <!-- Analysis Report Section -->
    <div class="report-section">
        <h2>Analysis Report: Stroke Risk by Age Group</h2>

        <!-- Overview -->
        <h3>Overview</h3>
        <p>This report presents an analysis of stroke risk by age group
based on patient data.</p>

        <!-- Findings -->
```

```
        <h3>Findings</h3>
        <p>The analysis reveals the number of stroke incidents within
different age groups.</p>

        <!-- Recommendations -->
        <h3>Recommendations</h3>
        <p>Further study and monitoring of stroke risk factors in
specific age groups are recommended for preventive
            measures.</p>

        <!-- Conclusion -->
        <h3>Conclusion</h3>
        <p>Understanding stroke risk by age group is essential for
targeted healthcare interventions.</p>
    </div>

    <!-- Go Home Button -->
    <div class="button-container" style="text-align: center;">
        <a href="/home" class="button">Go Home</a>
    </div>
</body>

<style>
    body {
        font-family: Arial, sans-serif;
        margin: 0;
        padding: 0;
        background-color: #f4f4f4;
        color: #333;
    }

    h1,
    h2 {
        text-align: center;
        color: #333;
    }

    table {
        margin: 20px auto;
```

```
        border-collapse: collapse;
        width: 80%;
    }

    th,
    td {
        padding: 8px;
        text-align: left;
        border-bottom: 1px solid #ddd;
    }

    th {
        background-color: #4CAF50;
        color: white;
    }

    tr:hover {
        background-color: #f5f5f5;
    }

    .report-section {
        margin: 20px auto;
        width: 80%;
        padding: 15px;
        background-color: white;
        border-radius: 8px;
        box-shadow: 0 4px 8px rgba(0, 0, 0, 0.1);
    }

    .report-section h3 {
        color: #4CAF50;
    }

    .button {
        display: inline-block;
        padding: 10px 20px;
        margin: 10px;
        background-color: #4CAF50;
        color: white;
```

```
        text-decoration: none;
        border-radius: 5px;
        text-align: center;
    }

    .button:hover {
        background-color: #45a049;
    }
</style>

</html>
```

Stroke-risk-anlaysis-2.html

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8" />
    <title>Stroke Analysis</title>
</head>

<body>
    <div class="header">
        <a href="/home" class="button">Go Home</a>
        <h1>Stroke Analysis: Hypertension and Heart Disease Impact</h1>
    </div>

    <table>
        <tr>
            <th>Health Condition</th>
            <th>Patients Count</th>
            <th>Stroke Incidents</th>
        </tr>
        {{#data}}
```

```
        <tr>
            <td>{{Health Condition}}</td>
            <td>{{TotalPatients}}</td>
            <td>{{StrokeIncidents}}</td>
        </tr>
        {{/data}}
    </table>

    <div class="report-section">
        <h2>Detailed Analysis Report: Stroke Incidence and Health
Conditions</h2>

        <h3>Overview</h3>
        <p>This report focuses on the relationship between stroke
incidence and various health conditions, including
            hypertension and heart disease.</p>

        <h3>Key Findings</h3>
        <ul>
            <li><strong>Total Patient Cohort:</strong> The analysis
includes a diverse cohort of patients.</li>
            <li><strong>Health Conditions:</strong> Patients are
categorized based on different health conditions, such
                as hypertension, heart disease, or their absence.</li>
            <li><strong>Stroke Incidents:</strong> The report provides
insights into the number of stroke incidents
                within each health condition category.</li>
        </ul>

        <h3>In-depth Insights</h3>
        <p>The analysis highlights the impact of health conditions on
stroke risk and emphasizes the importance of
            managing these conditions for stroke prevention.</p>

        <h3>Recommendations</h3>
        <p>Healthcare strategies should focus on the prevention and
management of hypertension and heart disease to
            reduce stroke incidence.</p>
```

```
        <h3>Conclusion</h3>
        <p>The findings underscore the significance of addressing
underlying health conditions as part of stroke risk
            reduction efforts.</p>
    </div>
</body>

<style>
    body {
        font-family: Arial, sans-serif;
        margin: 0;
        padding: 0;
        background-color: #f4f4f4;
        color: #333;
    }

    h1,
    h2 {
        text-align: center;
        color: #333;
    }

    table {
        margin: 20px auto;
        border-collapse: collapse;
        width: 80%;
    }

    th,
    td {
        padding: 8px;
        text-align: left;
        border-bottom: 1px solid #ddd;
    }

    th {
        background-color: #4CAF50;
        color: white;
    }
```

```
    tr:hover {
        background-color: #f5f5f5;
    }

    .report-section {
        margin: 20px auto;
        width: 80%;
        padding: 15px;
        background-color: white;
        border-radius: 8px;
        box-shadow: 0 4px 8px rgba(0, 0, 0, 0.1);
    }

    .report-section h3 {
        color: #4CAF50;
    }

    .button {
        display: inline-block;
        padding: 10px 20px;
        margin: 10px;
        background-color: #4CAF50;
        color: white;
        text-decoration: none;
        border-radius: 5px;
        text-align: center;
    }

    .button:hover {
        background-color: #45a049;
    }
</style>

</html>
```

Stroke-risk-anlaysis-3.html

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8" />
    <title>Stroke Analysis</title>
</head>

<body>
    <div class="header">
        <a href="/home" class="button">Go Home</a>
        <h1>Stroke Analysis: BMI, Glucose Level, Hypertension, Heart
Disease, and Smoking Status Impact</h1>
    </div>

    <table>
        <tr>
            <th>BMI Category</th>
            <th>Glucose Category</th>
            <th>Hypertension Category</th>
            <th>Heart Disease Category</th>
            <th>Smoking Status Category</th>
            <th>Total Patients</th>
            <th>Stroke Incidents</th>
        </tr>
        {{#data}}
        <tr>
            <td>{{BMICategory}}</td>
            <td>{{GlucoseCategory}}</td>
            <td>{{HypertensionCategory}}</td>
            <td>{{HeartDiseaseCategory}}</td>
            <td>{{SmokingStatusCategory}}</td>
            <td>{{TotalPatients}}</td>
            <td>{{StrokeIncidents}}</td>
        </tr>
```

```
        {{/data}}
    </table>

    <div class="report-section">
        <h2>Detailed Analysis Report: Stroke Incidence and Risk
Factors</h2>

        <h3>Overview</h3>
        <p>This report explores the relationship between stroke
incidence and various risk factors including BMI,
            glucose
            level, hypertension, heart disease, and smoking status among
patients.</p>

        <h3>Key Findings</h3>
        <ul>
            <li><strong>Total Patient Cohort:</strong> The analysis
includes a diverse cohort of patients.</li>
            <li><strong>Categories:</strong> Patients are categorized
based on BMI, glucose level, hypertension, heart
                disease, and smoking status to assess stroke risk.</li>
            <li><strong>Stroke Incidents:</strong> The report provides
insights into the number of stroke incidents
                within each category.</li>
        </ul>

        <h3>In-depth Insights</h3>
        <p>The analysis highlights the impact of these risk factors on
stroke risk and emphasizes the importance of
            managing these factors for stroke prevention.</p>

        <h3>Recommendations</h3>
        <p>Healthcare strategies should focus on risk factor management
to reduce the incidence of stroke among
            patients.</p>

        <h3>Conclusion</h3>
        <p>The findings demonstrate the significance of BMI, glucose
level, hypertension, heart disease, and smoking
```

```
            status in assessing stroke risk and guiding preventive
measures.</p>
    </div>
</body>

<style>
    body {
        font-family: Arial, sans-serif;
        margin: 0;
        padding: 0;
        background-color: #f4f4f4;
        color: #333;
    }

    h1,
    h2 {
        text-align: center;
        color: #333;
    }

    table {
        margin: 20px auto;
        border-collapse: collapse;
        width: 80%;
    }

    th,
    td {
        padding: 8px;
        text-align: left;
        border-bottom: 1px solid #ddd;
    }

    th {
        background-color: #4CAF50;
        color: white;
    }

    tr:hover {
```

```
        background-color: #f5f5f5;
    }

    .report-section {
        margin: 20px auto;
        width: 80%;
        padding: 15px;
        background-color: white;
        border-radius: 8px;
        box-shadow: 0 4px 8px rgba(0, 0, 0, 0.1);
    }

    .report-section h3 {
        color: #4CAF50;
    }

    .button {
        display: inline-block;
        padding: 10px 20px;
        margin: 10px;
        background-color: #4CAF50;
        color: white;
        text-decoration: none;
        border-radius: 5px;
        text-align: center;
    }

    .button:hover {
        background-color: #45a049;
    }
</style>

</html>
```

Stroke-risk-anlaysis-4.html

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8" />
    <title>Stroke Risk Analysis</title>
</head>

<body>
    <div class="header">
        <a href="/home" class="button">Go Home</a>
        <h1>Stroke Risk Analysis by Gender</h1>
    </div>

    <table>
        <tr>
            <th>Gender</th>
            <th>Total Patients</th>
            <th>Stroke Incidents</th>
        </tr>
        {{#data}}
        <tr>
            <td>{{Gender}}</td>
            <td>{{TotalPatients}}</td>
            <td>{{StrokeIncidents}}</td>
        </tr>
        {{/data}}
    </table>

    <div class="report-section">
        <h2>Gender-Based Stroke Risk Analysis</h2>

        <p>This analysis provides insights into the stroke risk based on
gender:</p>
```

```
        <ul>
            <li><strong>Male:</strong> Total Patients - {{MaleTotal}},
Stroke Incidents - {{MaleStroke}}</li>
            <li><strong>Female:</strong> Total Patients -
{{FemaleTotal}}, Stroke Incidents - {{FemaleStroke}}</li>
            <li><strong>Other:</strong> Total Patients - {{OtherTotal}},
Stroke Incidents - {{OtherStroke}}</li>
        </ul>

        <p>It's important to note that gender-based stroke risk analysis
is part of a broader assessment of stroke
            risk factors. The data presented here does not imply
causation, and further research is needed to explore
            the relationship between gender and stroke risk.</p>
    </div>
</body>

<style>
    body {
        font-family: Arial, sans-serif;
        margin: 0;
        padding: 0;
        background-color: #f4f4f4;
        color: #333;
    }

    h1,
    h2 {
        text-align: center;
        color: #333;
    }

    table {
        margin: 20px auto;
        border-collapse: collapse;
        width: 80%;
    }

    th,
```

```
td {
    padding: 8px;
    text-align: left;
    border-bottom: 1px solid #ddd;
}

th {
    background-color: #4CAF50;
    color: white;
}

tr:hover {
    background-color: #f5f5f5;
}

.report-section {
    margin: 20px auto;
    width: 80%;
    padding: 15px;
    background-color: white;
    border-radius: 8px;
    box-shadow: 0 4px 8px rgba(0, 0, 0, 0.1);
}

.report-section h3 {
    color: #4CAF50;
}

.button {
    display: inline-block;
    padding: 10px 20px;
    margin: 10px;
    background-color: #4CAF50;
    color: white;
    text-decoration: none;
    border-radius: 5px;
    text-align: center;
}
```

```
    .button:hover {
        background-color: #45a049;
    }
</style>

</html>
```

Stroke-risk-anlaysis-5.html

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8" />
    <title>Work and Residence Analysis</title>
</head>

<body>
    <div class="header">
        <a href="/home" class="button">Go Home</a>
        <h1>Work and Residence Analysis for Stroke Patients</h1>
    </div>

    <table>
        <tr>
            <th>Work Type</th>
            <th>Residence Type</th>
            <th>Total Patients</th>
            <th>Stroke Incidents</th>
        </tr>
        {{#data}}
        <tr>
            <td>{{WorkType}}</td>
            <td>{{ResidenceType}}</td>
            <td>{{TotalPatients}}</td>
```

```
            <td>{{StrokeIncidents}}</td>
        </tr>
        {{/data}}
    </table>

    <div class="report-section">
        <h2>Work and Residence Analysis for Stroke Patients</h2>

        <p>This analysis provides insights into the relationship between
work type, residence type, and the incidence of
            stroke:</p>

        <ul>
            <li>
                <strong>Work Type:</strong>
                <ul>
                    <li><strong>Private Job:</strong> Total Patients -
{{PrivateJobTotal}}, Stroke Incidents -
                        {{PrivateJobStroke}}</li>
                    <li><strong>Self-Employed:</strong> Total Patients -
{{SelfEmployedTotal}}, Stroke Incidents -
                        {{SelfEmployedStroke}}</li>
                    <li><strong>Govt Job:</strong> Total Patients -
{{GovtJobTotal}}, Stroke Incidents -
                        {{GovtJobStroke}}</li>
                </ul>
            </li>
            <li>
                <strong>Residence Type:</strong>
                <ul>
                    <li><strong>Urban:</strong> Total Patients -
{{UrbanTotal}}, Stroke Incidents - {{UrbanStroke}}</li>
                    <li><strong>Rural:</strong> Total Patients -
{{RuralTotal}}, Stroke Incidents - {{RuralStroke}}</li>
                </ul>
            </li>
        </ul>

        <p>This analysis aims to understand if there is any correlation
```

```
between work type, residence type, and the risk
            of stroke among patients. Further research and statistical
analysis may be required to establish any
            significant relationships.</p>
    </div>
</body>

<style>
    body {
        font-family: Arial, sans-serif;
        margin: 0;
        padding: 0;
        background-color: #f4f4f4;
        color: #333;
    }

    h1,
    h2 {
        text-align: center;
        color: #333;
    }

    table {
        margin: 20px auto;
        border-collapse: collapse;
        width: 80%;
    }

    th,
    td {
        padding: 8px;
        text-align: left;
        border-bottom: 1px solid #ddd;
    }

    th {
        background-color: #4CAF50;
        color: white;
    }
```

```
    tr:hover {
        background-color: #f5f5f5;
    }

    .report-section {
        margin: 20px auto;
        width: 80%;
        padding: 15px;
        background-color: white;
        border-radius: 8px;
        box-shadow: 0 4px 8px rgba(0, 0, 0, 0.1);
    }

    .report-section h3 {
        color: #4CAF50;
    }

    .button {
        display: inline-block;
        padding: 10px 20px;
        margin: 10px;
        background-color: #4CAF50;
        color: white;
        text-decoration: none;
        border-radius: 5px;
        text-align: center;
    }

    .button:hover {
        background-color: #45a049;
    }
</style>

</html>
```

Screen Shot of Anlaysis:

Age Group Analysis

Age Group	Total Patients	Stroke Incidents
0-18	895	1
46-60	1145	52
31-45	1011	10
19-30	641	0
60+	1217	146

Analysis Report: Stroke Risk by Age Group

Overview

This report presents an analysis of stroke risk by age group based on patient data.

Findings

The analysis reveals the number of stroke incidents within different age groups.

Recommendations

Further study and monitoring of stroke risk factors in specific age groups are recommended for preventive measures.

Conclusion

Understanding stroke risk by age group is essential for targeted healthcare interventions.

Stroke Analysis: Hypertension and Heart Disease Impact

Health Condition	Patients Count	Stroke Incidents
Hypertension, No Heart Disease	393	49
Hypertension and Heart Disease	58	11
No Hypertension/Heart Disease	4273	120
Heart Disease, No Hypertension	185	29

Detailed Analysis Report: Stroke Incidence and Health Conditions

Overview

This report focuses on the relationship between stroke incidence and various health conditions, including hypertension and heart disease.

Key Findings

- **Total Patient Cohort:** The analysis includes a diverse cohort of patients.
- **Health Conditions:** Patients are categorized based on different health conditions, such as hypertension, heart disease, or their absence.
- **Stroke Incidents:** The report provides insights into the number of stroke incidents within each health condition category.

In-depth Insights

The analysis highlights the impact of health conditions on stroke risk and emphasizes the importance of managing these conditions for stroke prevention.

Recommendations

Healthcare strategies should focus on the prevention and management of hypertension and heart disease to reduce stroke incidence.

Conclusion

The findings underscore the significance of addressing underlying health conditions as part of stroke risk reduction efforts.

Stroke Risk Analysis by Gender

Gender	Total Patients	Stroke Incidents
Female	2897	120
Male	2011	89
Other	1	0

Gender-Based Stroke Risk Analysis

This analysis provides insights into the stroke risk based on gender:

- **Male:** Total Patients - , Stroke Incidents -
- **Female:** Total Patients - , Stroke Incidents -
- **Other:** Total Patients - , Stroke Incidents -

It's important to note that gender-based stroke risk analysis is part of a broader assessment of stroke risk factors. The data presented here does not imply causation, and further research is needed to explore the relationship between gender and stroke risk.

Work and Residence Analysis for Stroke Patients

Work Type	Residence Type	Total Patients	Stroke Incidents
Private	Urban	1405	66
Private	Rural	1406	61
Self-employed	Urban	404	28
Self-employed	Rural	371	25
Govt_job	Urban	328	15
Govt_job	Rural	302	13
children	Rural	333	1

Work and Residence Analysis for Stroke Patients

This analysis provides insights into the relationship between work type, residence type, and the incidence of stroke:

- **Work Type:**
  - **Private Job:** Total Patients - , Stroke Incidents -
  - **Self-Employed:** Total Patients - , Stroke Incidents -
  - **Govt Job:** Total Patients - , Stroke Incidents -
- **Residence Type:**
  - **Urban:** Total Patients - , Stroke Incidents -
  - **Rural:** Total Patients - , Stroke Incidents -

This analysis aims to understand if there is any correlation between work type, residence type, and the risk of stroke among patients. Further research and statistical analysis may be required to establish any significant relationships.