Stroke Risk Analysis: Impact of Smoking, Glucose Levels, and BMI

Overview

The rising incidence of stroke among individuals aged 45 and above necessitates targeted intervention. Smoking status, average glucose level, and BMI are key risk factors influencing stroke occurrence. This study aims to analyze these factors across different age groups to identify trends and correlations. The goal is to provide actionable insights for reducing stroke incidences by 12% by Q1 2025.

Success will be determined by identifying key correlations between smoking status, glucose level, BMI, and stroke incidence while developing effective recommendations to mitigate these risks. Engaging stakeholders to validate findings and influence policy is crucial, with the ultimate goal of achieving a 12% reduction in stroke incidence within the target population.

Problem Statement

By the end of Q1 2025, determine the impact of smoking status, average glucose level, and BMI on the incidence of stroke, focusing on individuals within the age brackets of 18–44, 45–64, and 65 and older, to reduce stroke incidences by 12%, particularly in individuals aged 45 and above.

Goals and Objectives

- Identify correlations between smoking status, glucose levels, BMI, and stroke incidence.

- Develop data-driven insights to inform preventive healthcare strategies.

- Provide recommendations to reduce stroke occurrences by 12%.

- Engage stakeholders to validate findings and drive policy decisions.

Tools and Technologies Used

- **Programming Languages**: Python (Pandas, NumPy, Matplotlib, Seaborn)

- **Data Visualization**: Tableau

- **Data Storage and Management**: SQL, CSV format

- **Others**: PowerPoint, Excel

Methodology/Framework: HDEIP

HDEIP stand for Hypothesis formation, Data Sourcing, Exploratory Data Analysis, Insight Generation and Presentation. To ensure a structured analytical approach, I applied the **HDEIP Framework**, which stands for:

1. **Hypothesis Formation**: Established key questions regarding the relationship between stroke and risk factors.

2. **Data Sourcing**: Acquired datasets containing relevant patient records, including stroke history, glucose levels, BMI, and smoking status.

3. **Exploratory Data Analysis (EDA)**: Conducted data cleaning, outlier detection, and preliminary visualizations.

4. **Insight Generation**: Identified correlations and key trends using statistical analysis.

5. **Presentation**: Communicated findings through executive summaries, presentations, and data visualizations.

Analytical Approach: Macro-Level and Micro-Level Analysis

Macro-Level Analysis: Identifying Broad Patterns and Trends

- **Data Aggregation**: Summarized key metrics such as averages, medians, and totals.

- **Trend Identification**: Observed overarching trends in the dataset over time.

- **Outlier Detection**: Spotted anomalies and extreme values that warranted further investigation.

- **Correlation Analysis**: Explored relationships between variables to establish broad patterns.

Application:

- Established foundational insights, such as identifying general trends in health risk factors and detecting significant variations across demographic groups.

Micro-Level Analysis: Diving into Granular Insights

- **Detailed Segmentation**: Broke down data by subgroups, such as age, gender, or health conditions.

- **Outlier Investigation**: Analyzed individual anomalies to determine their significance.

- **Hypothesis Testing**: Validated patterns through statistical tests.

- **Feature Impact Analysis**: Assessed how specific variables influenced stroke incidence.

Application:

- Identified key drivers behind macro trends, such as how specific subgroups exhibited higher risks of stroke based on lifestyle factors.

Key Insights & Strategic Implications

1. Age-Specific Stroke Risks

- Individuals aged **45–64** experience a **35% higher stroke incidence** than those aged **18–44**.

- Individuals aged **65+** see a **65% increase** compared to younger cohorts.

- Preventive strategies should focus on the 45+ age group.

2. Smoking as a Primary Risk Factor

- Individuals aged **45+** who smoke have a **30% higher stroke risk**.

- Former smokers exhibit similar risks as those who never smoked, indicating long-term cardiovascular effects.

- Targeted smoking cessation programs are recommended.

3. Glucose Levels and Stroke Risk

- Glucose levels above **200 mg/dL** are linked to a **40% higher stroke incidence** in individuals aged **45–64**.

- Diabetes management should be a critical component of intervention strategies.

4. BMI Thresholds and Stroke Risk

- A BMI over **30** correlates with a **50% higher stroke risk** in individuals aged **65+**.

- Weight management programs should be prioritized in stroke prevention efforts.

5. Combined Risk Factors

- Among individuals aged **65+**, those with high glucose, high BMI, and a smoking history have a **70% higher stroke incidence** than those with only one risk factor.
- Integrated healthcare approaches addressing multiple risk factors are needed.

6. Gender-Specific Trends

- Men aged **45–64** show a **25% higher stroke rate** than women due to smoking prevalence.
- Women aged **65+** have higher stroke rates, requiring gender-sensitive interventions.

7. Urban vs. Rural Differences

- Urban residents aged **45+** have a **20% higher stroke risk** than rural residents due to lifestyle and socioeconomic factors.
- Universal stroke prevention programs should be applicable across geographic settings.

Challenges Encountered

- **Limited dataset coverage**: Some BMI values were missing.

- **Lifestyle variability**: The dataset lacked detailed lifestyle factors beyond smoking.

- **Ethical considerations**: Ensuring data privacy and responsible reporting.

- **Time constraints**: The goal is to achieve measurable impact by Q1 2025.

Communication & Stakeholder Engagement

- **Executive Presentation**: Delivered key findings through PowerPoint presentations.

- **Tableau Dashboards**: Created interactive visualizations for easy data exploration.

Conclusion & Next Steps

- Implementation of targeted health programs for smoking cessation, weight management, and diabetes control.

- Refinement of data collection to include additional lifestyle variables.

- Ongoing monitoring to assess the impact of recommendations over time.

- Stakeholder collaboration to drive healthcare policy changes aimed at stroke prevention.