Data Handling and Visualisation

NAME: Pavithira seenivasagan

STUDENT ID: 23095934

GitHub Link: https://github.com/Pavithiraseenivasagan/Final-Visualisation-Report/upload

Dataset link: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

STROKE DATA ANALYSIS REPORT

Every day, countless people around the world face the risk of stroke a silent yet devastating health condition that often strikes without warning. Behind the numbers and statistics lie human stories, each shaped by factors such as age, gender, lifestyle and environment. This report delves into a healthcare dataset of over 5,000 individuals to uncover the patterns and relationships that contribute to stroke risk. By analysing connections between age, BMI, glucose levels and residence type, we aim to understand who is most at risk and why.

The dataset provides valuable insights into key health metrics of the 5,109 individuals analysed. Below is a summary of the dataset:

Metric	Count	Mean	Std. Dev.	Min	25%	50%	75%	Max
Age	5,109	43.23	22.61	0.08	25.00	45.00	61.00	82.00
Hypertension	5,109	0.10	0.30	0.00	0.00	0.00	0.00	1.00
Heart Disease	5,109	0.05	0.23	0.00	0.00	0.00	0.00	1.00
Avg Glucose Level	5,109	106.14	45.28	55.12	77.24	91.88	114.09	271.74
BMI	5,109	28.89	7.70	10.30	23.80	28.40	32.80	97.60
Stroke	5,109	0.05	0.22	0.00	0.00	0.00	0.00	1.00

These statistics offer an overview of the population, highlighting key health indicators such as an average BMI of 28.9 and an average glucose level of 106.1 mg/dL, with notable outliers for both metrics.

Stroke Occurrence by Gender
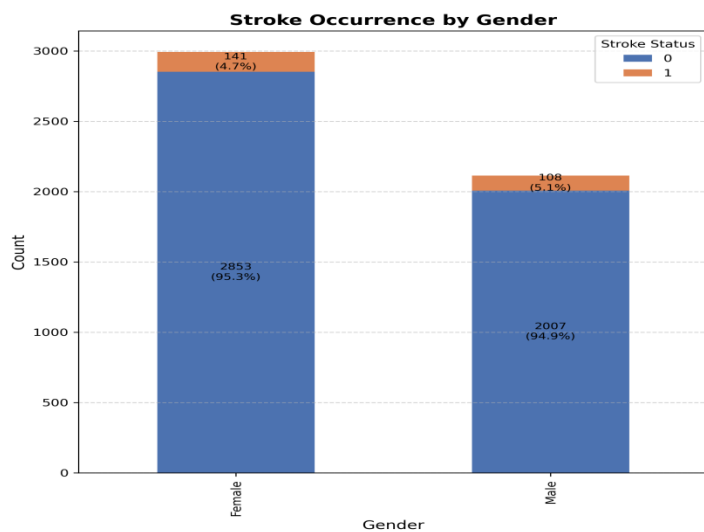
The dataset reveals several critical health patterns. Stroke occurrences accounted for 4.87% of the population, with 141 cases among women and 108 among men. While the numbers suggest a slightly higher incidence in women, this is likely due to their larger representation in the dataset. This observation, visualized in a stacked bar chart, underscores the importance of exploring potential gender-specific risk factors further.

Age tells an even more compelling story. A density plot shows that stroke risk rises sharply after the age of 60. The average age in the dataset is 43, but this masks the increased vulnerability of older individuals. The plot vividly illustrates how age is a defining factor in stroke risk, making it crucial to prioritize preventative measures for aging populations.

Age Distribution by Stroke Occurrence with KDE

BMI vs. Age with Regression Line
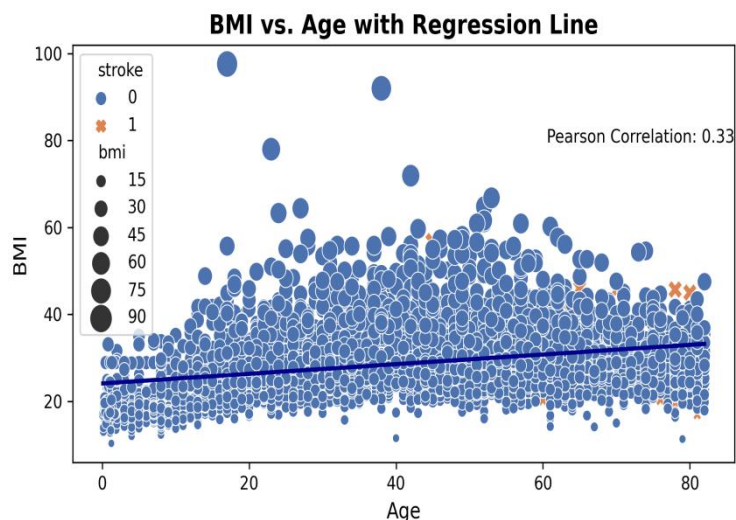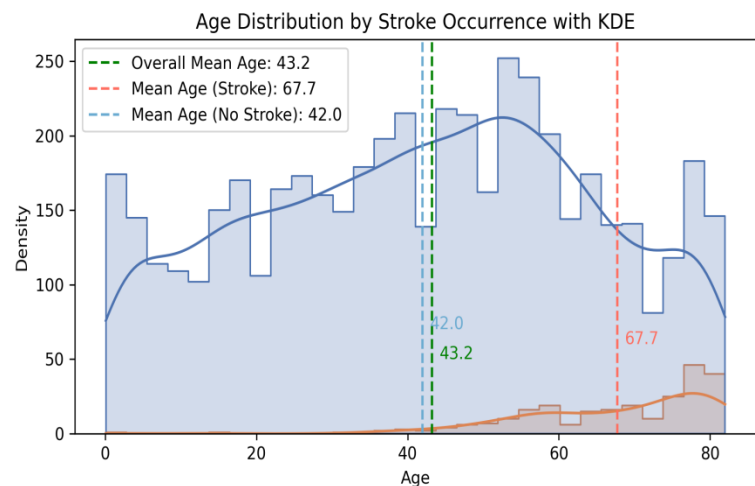
BMI adds another dimension to the narrative. The dataset's average BMI is 28.9, with stroke cases clustering among individuals with higher BMIs. A scatter plot with a regression line reveals a weak but noticeable correlation between BMI and age (Pearson: 0.33). The maximum BMI recorded is an alarming 97.6, highlighting the urgent need for weight management to reduce stroke risk.

Glucose levels emerge as a universal warning sign. A swarm plot shows that stroke patients consistently have elevated glucose levels, averaging 106 mg/dL and reaching as high as 271 mg/dL. This trend holds true regardless of whether individuals reside in urban or rural areas. This finding reinforces the critical role of glycemic control in preventing strokes.

Average Glucose Level by Residence Type and Stroke Status
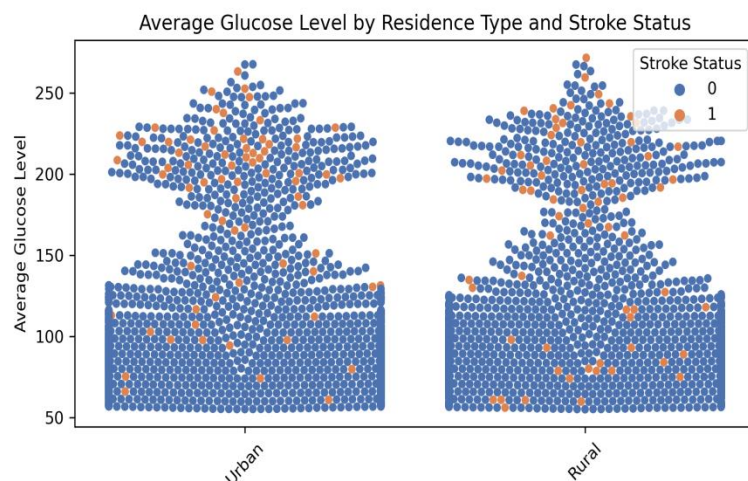
In conclusion, this analysis paints a clear picture of the silent factors driving stroke risk. Age emerges as the most critical factor, with vulnerability surging after 60, while elevated BMI and glucose levels further amplify the danger. Gender differences add nuance, suggesting subtle but meaningful dynamics. These findings stress the importance of proactive health measures like regular screenings, weight management, and glucose control. Each data point tells a story a life that could be changed or saved through timely intervention. By addressing these factors, we can turn statistics into action and create a healthier, more resilient future.