

Source of Data:

<https://www.kaggle.com/datasets/jillanisofttech/brain-stroke-dataset>

No Copyrights. This data set represents records of people who had or had not had a brain-stroke before. It consists of 10 columns. The columns are gender, age, hypertension, heart_disease, ever_married, work_type, Residence_type, avg_glucose_level, bmi, smoking_status and stroke.

cleaning data:

All records were filled and there are no empty cells in this data. This data set met the requirements for this assignment. However, I decided to convert 3 string columns into numerical columns because the string in these columns are binary (0 or 1), and converting them from now would enhance the efficiency of the learning model that we want to build and would reduce the number of string inputs to our model. The columns that I changed are gender, ever_married, Residence_type.

Visualizations:

- Pie chart: represent the percentage of people who had or had not had a brain-stroke before.
- Histogram: counter for age groups in the data set.
- Par Chart: Divide the people in the data into groups based on smoking_status.