

# Stroke Analysis Data Story

By

Sheikh Muhammad Saad

## Introduction:

According to the World Health Organization (WHO) 'Stroke' is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. It is a leading cause of death for Americans. The risk of having a first stroke is nearly twice as high for blacks as for whites, and blacks have the highest rate of death due to stroke.

This dataset is used to predict whether a patient is likely to get a stroke based on the input parameters like age, hypertension, BMI, various diseases, and smoking status. This report is written from an informational point of view that how our daily life work can affect us and as result of that, we get a disease like a stroke.

## Analysis Questions:

Following are the main questions for the analysis report, which can help us to understand the problems which lead to heart disease.

- Does Marriage play a part in getting heart disease?
- Which work type doesn't affect your mental peace?
- What is the average glucose level of a normal person and a heart disease patient?
- What body mass index defines a patient life or a healthy life?
- What is the relation between the age and Hypertension of a normal person and a heart disease patient?

## Libraries And Utilities:

```
In [15]: 1 import numpy as np
          2 import pandas as pd
          3 import matplotlib.pyplot as plt
          4 import seaborn as sns
          5 import matplotlib.pyplot as plt
          6 from matplotlib import pyplot
```

## Data Loading:

```
In [16]: 1 path = "/Users/saadshafiq/Desktop/IDS/ids_project/stroke_dataset.csv"
          2 df = pd.read_csv(path)
          3 df.head(15)
```

## **Exploratory Data Analysis (EDA):**

### **Overview Of Data:**

This dataset is used to predict whether a patient is likely to get a stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient.

This data is collected by “fedesoriano (Kaggle)”, who is a Data Scientist in Madrid, Spain.

### **Data Features:**

The main attribute information of this data

- 1) id: unique identifier
- 2) gender: "Male", "Female" or "Other"
- 3) age: age of the patient
- 4) hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension
- 5) heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
- 6) ever\_married: "No" or "Yes"
- 7) work\_type: "children", "Govt\_jov", "Never\_worked", "Private" or "Self-employed"
- 8) Residence\_type: "Rural" or "Urban"
- 9) avg\_glucose\_level: average glucose level in blood
- 10) BMI: body mass index
- 11) smoking\_status: "formerly smoked", "never smoked", "smokes" or "Unknown"\*
- 12) stroke: 1 if the patient had a stroke or 0 if not

## Plot a general overview of data:

### Key figures

<b>5000</b>	<b>12</b>	<b>43.6k</b>	<b>Rural/Urban</b>
Number of patients in the dataset	Number of features in the dataset	Downloads of this Dataset	Diversion of data

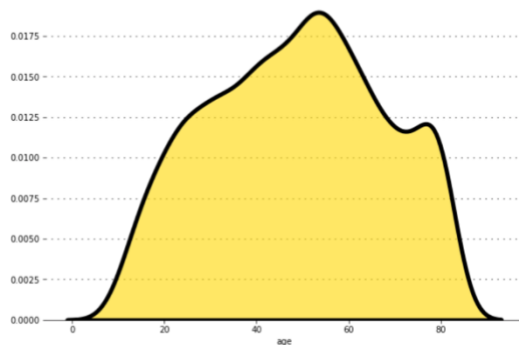
	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
5	56669	Male	81.0	0	0	Yes	Private	Urban	186.21	29.0	formerly smoked	1
6	53882	Male	74.0	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
7	10434	Female	69.0	0	0	No	Private	Urban	94.39	22.8	never smoked	1
8	27419	Female	59.0	0	0	Yes	Private	Rural	76.15	NaN	Unknown	1
9	60491	Female	78.0	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
10	12109	Female	81.0	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
11	12095	Female	61.0	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
12	12175	Female	54.0	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
13	8213	Male	78.0	0	1	Yes	Private	Urban	219.84	NaN	Unknown	1
14	5317	Female	79.0	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1

## Univariate analysis of continuous variables:

Most of the patients in the dataset are between the age of 40 – 60 years. Similarly, their Body Mass Index (BMI) are between 25 – 35.

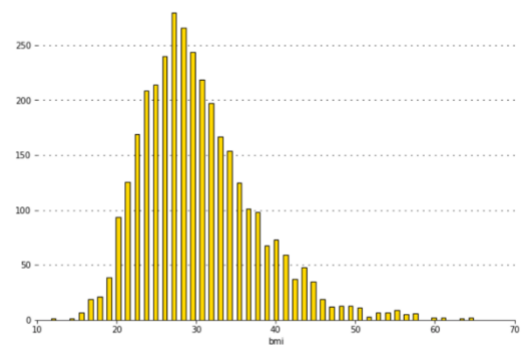
### Distribution of the age variable

Most of the people in the dataset are between 40 to 60 years old



### Distribution of the bmi variable

Most of the people in the dataset are between 25 to 35 of bmi



## Analysis:

### Feature Selection:

Principal Components Analysis (PCA) maps data of original feature dimension  $n$  to smaller dimension  $k$ . These new principal components or PCs are linear combinations of original features that carry maximal variance when data is projected onto them. The original data set is represented by only 12 features, which happen to be predictive of stroke risk according to literature. Therefore, most algorithms were done on full feature dimension. Feature selection techniques such as PCA, however, can give intuition on the most important factors in determining patient outcomes.

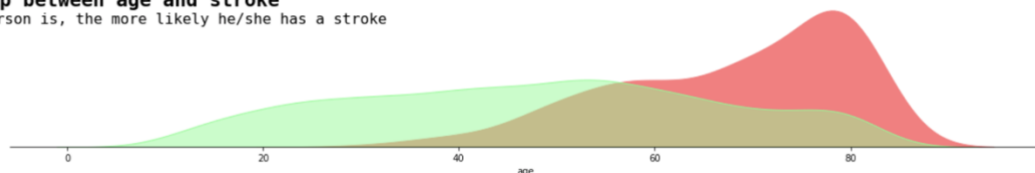
After performing PCA analysis and seeing how the original 12 features were indeed highly representative and predictive of eventual patient outcome, all supervised learning algorithms were performed on full feature dimension ( $n=12$ )

### Research Question/Analysis:

Following are the main questions that come into people minds and it also shows what is the main difference between a healthy person and a non-healthy person.

### Q: Relationship between 'Age' & 'Stroke'?

**Relationship between age and stroke**  
The older a person is, the more likely he/she has a stroke



**Analysis:** Age is just a number but as people grow older the chances of getting stroke also increase. This data tells us a story that the ratio of having a stroke in older people increases after the age of 50. According to this data, 60% of people have stroke problems after the age of 50.

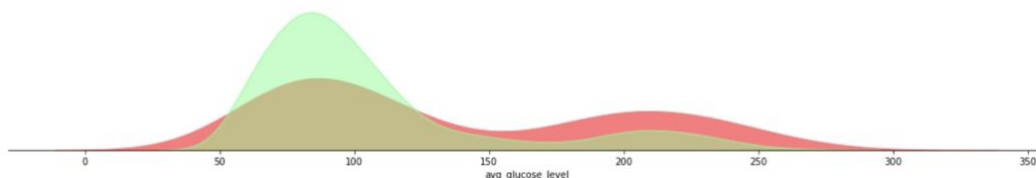
Research has indicated that blood vessels naturally 'harden' with age, losing their elasticity. This may be one explanation for why older people are more at risk of developing heart disease.

**Method:** I used the Kernel Distribution Estimation (KDE) Plot which is used for visualizing the Probability Density of a continuous variable. Its estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. A similar situation is in this case that I had the finite data samples and by this plot, we got the visualization of age and stroke.

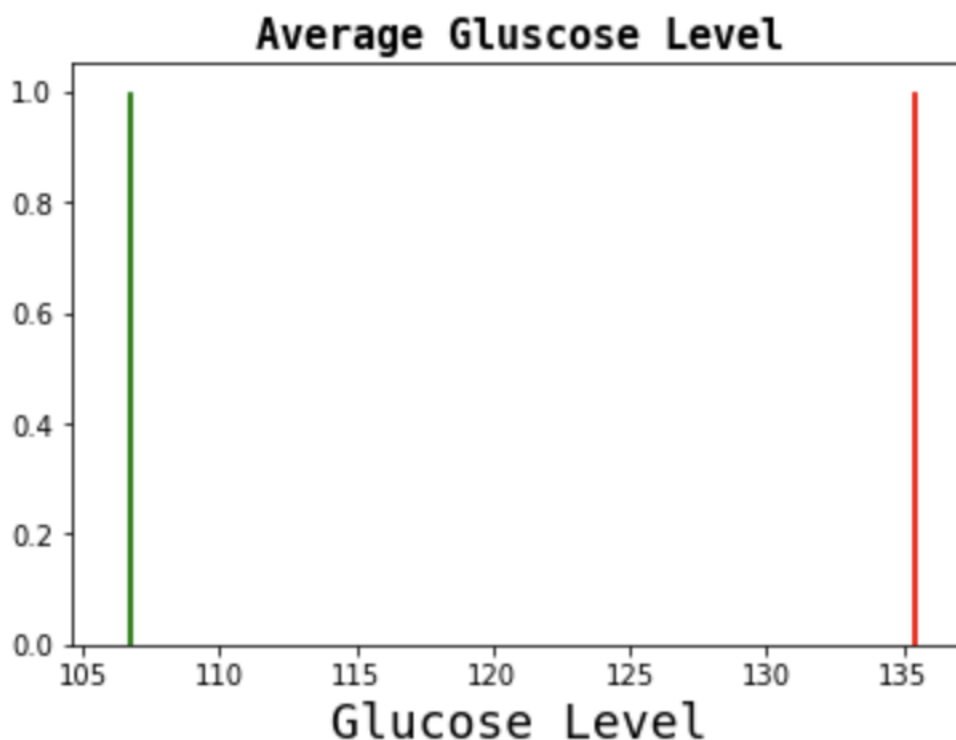
## Q: Relationship between 'Average Glucose' & 'Stroke'?

### Relationship between average glucose level and stroke

From this graph, there is no clear relationship between avg\_glucose\_level and stroke



**Analysis:** In Asia, Glucose and Stroke have a very strong relation between them. But in this data, there is not any strong relation because glucose level is usually high in heart disease patients.



The green color represents Normal Patients & Red Color represents heart disease patients

According to another analysis, the normal person average glucose level is between 105 – 110. While on the other hand, the average glucose level of a heart disease patient is 136. The reason for this is that Over time, high blood sugar can damage blood vessels and the nerves that control your heart. This could increase the risk of heart disease.

**Method:** In this, I used KDE Plot and histogram. Both help us to visualize the data where the majority falls. Histogram data is generated by taking the mean of a normal person glucose level and a heart disease glucose level. The visualization code is simple as you can see that on GitHub. (Link is attached)

**Q: Does Marriage plays a part in getting heart disease?**

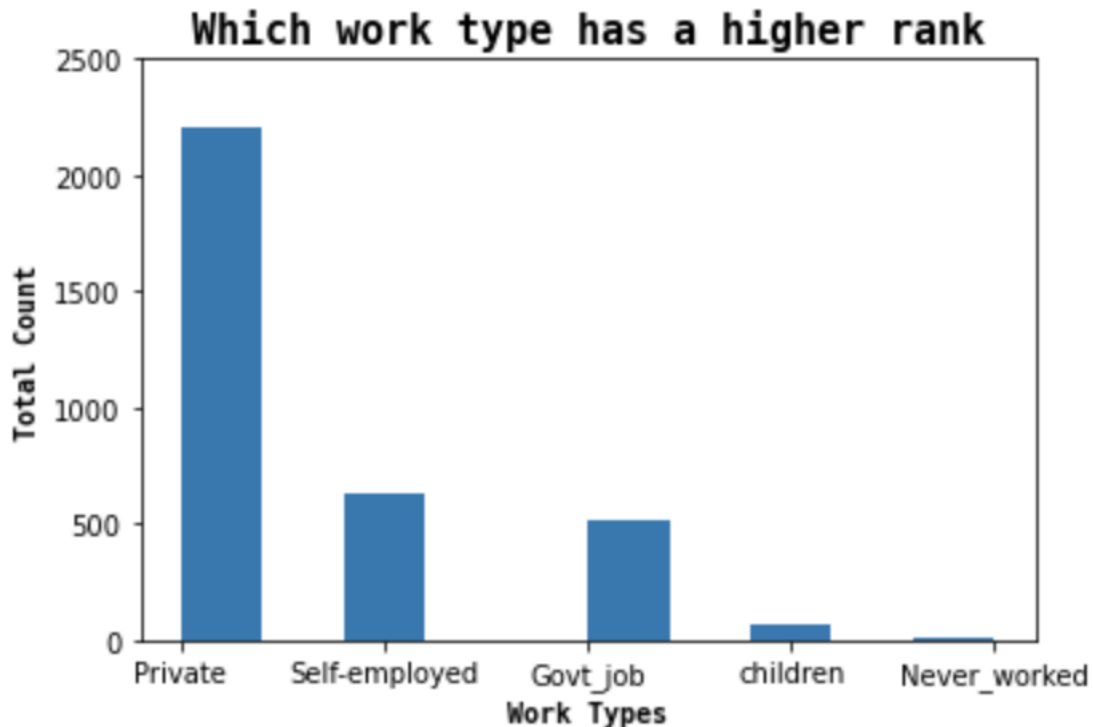


**Analysis:** Marriage is an important part of our life. Sometimes people build a healthy relationship between them or sometimes they get divorced. According to this data, 81.5% of people feel that marriage 'Doesn't play a part in getting heart disease'. Similarly, 18.5% people feel that marriage 'Play a part in getting a heart disease'. These are the 2 different types of people; all had some problems in their life due to which they feel like this.

Some similar studies also show that "The study, reported in the Archives of Internal Medicine, found the stress and anxiety of hostile, angry relationships can boost the risk of developing heart disease. Chances of a heart attack or chest pain rose by 34% compared to people on good terms with a spouse or partner."

**Method:** I used Pie Chart, which helps us divided into slices to illustrate numerical proportion. Summarized the data into parts and combined it with a pie chart for better visualization.

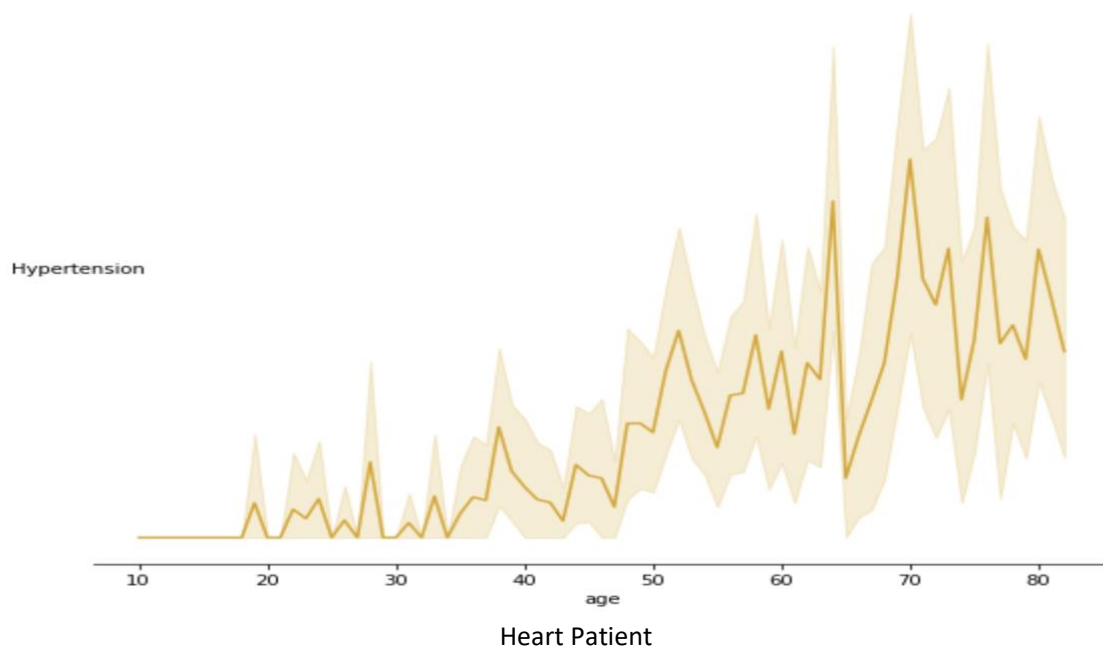
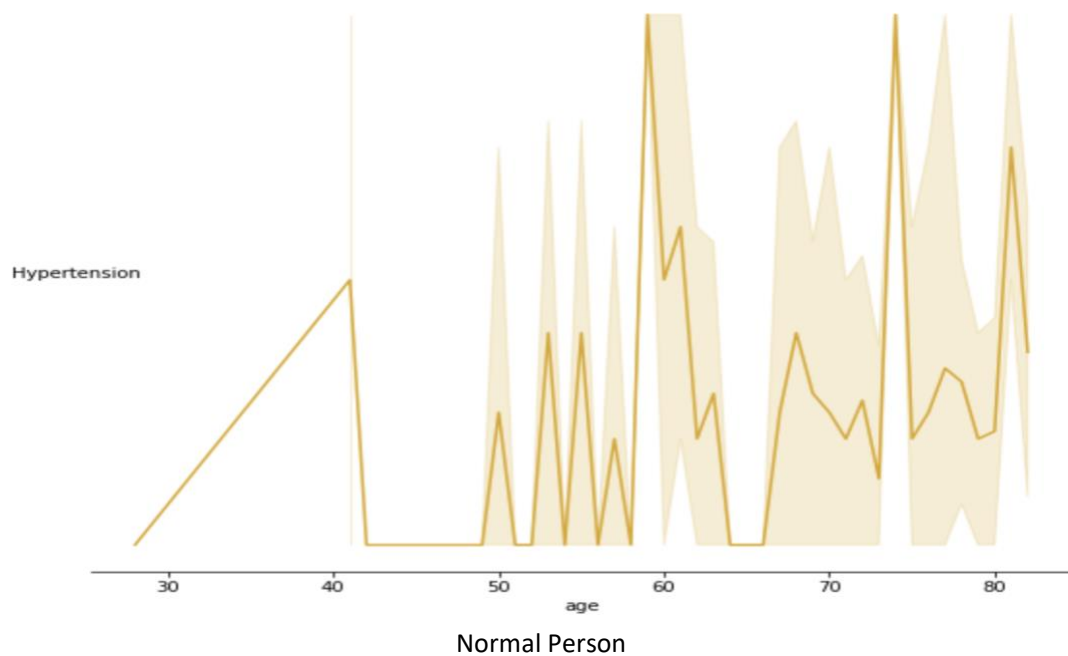
**Q: Which work type doesn't affect your mental peace?**



**Analysis:** Mental health includes our emotional, psychological, and social well-being. It affects how we think, feel, and act. It also helps determine how we handle stress, relate to others, and make choices. At work, Mental peace is one of the most important things. So as this data shows, 2200 people say that private work type doesn't affect your mental peace. Because in private work there is less stress as compared to Self-employed and Govt Job. In private companies, everything is organized and there are people for every task.

**Method:** A histogram is used to compare different work types. The process of visualization is very simple in this, you just need data for this. The rest of the work will be done on its own.

**Q: what is the relation between the age and Hypertension of a normal person and a heart disease patient?**



**Analysis:** The normal person graph shows that it doesn't have a strong relation between them. On the other hand, a heart patient hypertension and age are directly proportional to each other. The reason for this is "The older you are, the more likely you are to develop it."



Average blood pressure readings may vary depending on age. As people get older, they are more likely to have elevated blood pressure. This is because blood vessels become stiffer with age, which makes blood pressure rise.

According to the research:

“The American Heart Association (AHA) reports that the lifetime risk of high blood pressure from age 20 to 85 is **between 69 and 86 %**.”

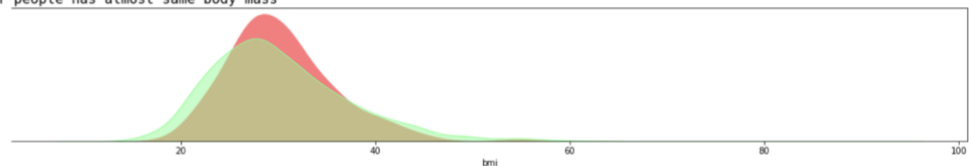
High blood pressure, also called hypertension, is blood pressure that is higher than normal. Your blood pressure changes throughout the day based on your activities. So as your age increases the chances of high blood pressure also increases.

**Method:** A line plot is used to displays data as points or checkmarks above a number line, showing the frequency of each value.

### **Q: What body mass index defines a patient life or non-patient life?**

#### **Relationship between Heart Disease and BMI**

It shows that both type of people has almost same body mass



Analysis: There's not any strong relationship between heart disease and Body Mass Index (BMI) according to this data. Both of their BMIs is between 25-35. But the majority of the people who have heart disease, their BMI falls near 30.

The research also tells us that:

According to data from the Framingham Heart Study, the rise of BMI by 1 kg/m<sup>2</sup> increases the risk of heart failure by 5% in the case of men and 7% in the case of women. Studies on heart failure show that 32%–49% of patients suffering from heart failure are obese and 31%–40% are overweight.

Method: In this, I used KDE Plot. It helped me to visualize the data where the majority falls. The visualization code is simple as you can see that on GitHub. (Link is attached)

## **Conclusion:**

In this analysis report, I demonstrated what are the main cause of Stroke and what people think about it. Stroke link with many parts of our body/life which cause heart disease problems. BMI, Hypertension, Work type, Age, Glucose Level etc how these things are interlinked with Stroke, that all are described in the analysis. All these things are risk factors for getting heart disease.