**What causes a Stroke?**

**A data-based approach using Kaggle’s Stroke Prediction data from 2021.**

***Introduction***

Before going into the answer to the above question, let’s first understand what is a stroke?  
A stroke is a damage to the brain, that occurs when the blood supply to part of your brain is interrupted or reduced, preventing brain tissue from getting oxygen and nutrients. Brain cells begin to die in minutes. This causes the part of the body that the injured brain controls to stop working. A stroke also is called a cerebrovascular accident, CVA, or “brain attack.”



There are many myths regarding strokes which are as follows:

Strokes only happen to elderly peopleStrokes are rare.Smoking doesn’t affect your chances of having a strokeMales are most susceptible to strokes  
Being Married will increase the risk of having a stroke  
and many more…

Therefore, I used data from Kaggle, a [Stroke Prediction Dataset](https://www.kaggle.com/fedesoriano/stroke-prediction-dataset/code) that consists of 11 clinical features for predicting stroke events and to break few myths using the features.

This Dataset contains many clinical features like Age, Heart Diseases (if any), BMI (Body Mass Index), Sugar Level, already affected cases etc. So, from this dataset, we can predict how likely a person can get a stroke based on his remaining features and also, we can break some myths which include:

1.Strokes only happen to elderly people

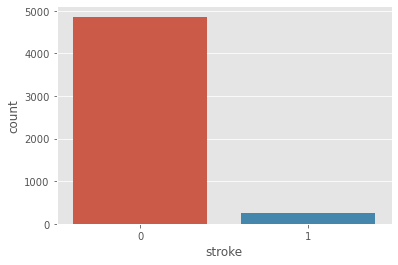
2.Smoking doesn’t affect your chances of having a stroke

3.Glucose level and BMI doesn’t contribute to stroke

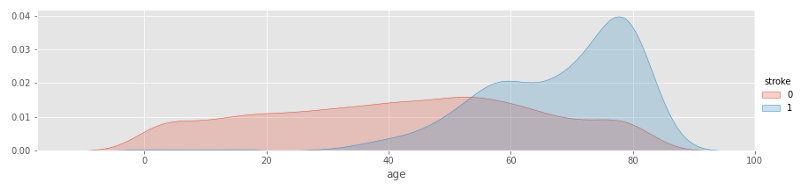
**Myth I: Strokes only happen to elderly people.**

**Fact:** It’s true that as you age your risk for stroke goes up. However, there’s also an increasing number of strokes in people between the ages of 18 and 65, so to say that strokes only occur in the elderly is false. The growing incidence of obesity and high blood pressure in ages 18 to 65 may contribute to the increased stroke risk in this population.

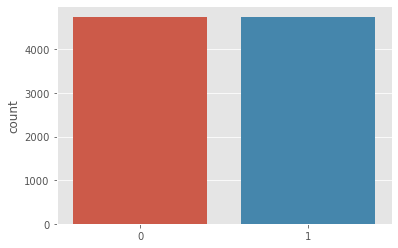
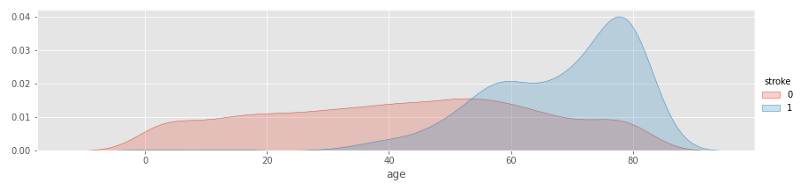
Now let us see what our data-based approach suggests:  
In our dataset, there is a total of 5110 rows and 12 columns, so there are a total of 5110 ages and Stroke/no-stroke cases recorded in our dataset. In our dataset number of people with no stroke are more compared to the number of people with a stroke which we call an imbalanced dataset.

Figure1: Count Plot on Stroke column

Now we have plotted a KDE plot to know the impact of age on the stroke column in our imbalanced dataset and we can see that stroke in people started from age 30 and continues till age 90.

Figure2: Age vs Stroke (Imbalanced)

Let’s make our dataset balanced and recheck our previous results. An Imbalanced dataset can be handled using many techniques but, in our case, we used Random Over-Sampling and after oversampling our data here is a picture of a balanced stroke column.

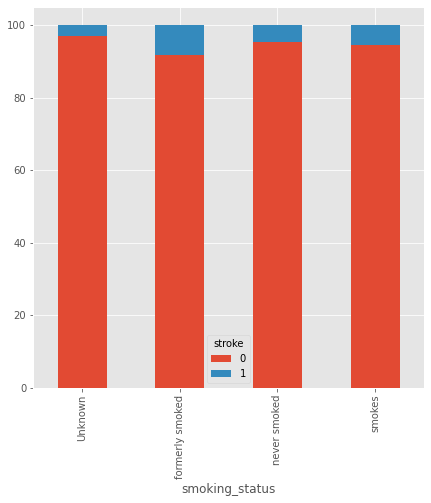
Figure 3: Balanced Count plot on Stroke Column Figure 4: Age vs Stroke (Balanced)

Now it is evident that not only elderly people but also people above age 20 are also at an equal risk of having a stroke.

**Myth II: Smoking doesn’t affect your chances of having a stroke.**

**Fact:** Smoking is one of the biggest risk factors for stroke, especially in younger people. This is true for both ischemic and haemorrhagic strokes, as well as first-time and recurrent strokes  
**Ischemic Stroke**: This type of stroke is caused by a blockage in an artery that supplies blood to the brain. The blockage reduces the blood flow and oxygen to the brain, leading to damage or death of brain cells.  
**Haemorrhagic stroke:** A **haemorrhagic stroke** is also called an intracerebral haemorrhage, or an ICH. An ICH occurs when a blood vessel ruptures and blood accumulates in the tissue around the rupture. This puts pressure on the brain and causes a loss of blood to the surrounding areas.

Now let us see what our data-based approach suggests:  
As per our analysis, it is proven that no matter what his current smoking status is, if the person is a smoker, then the chance of getting a stroke increase.

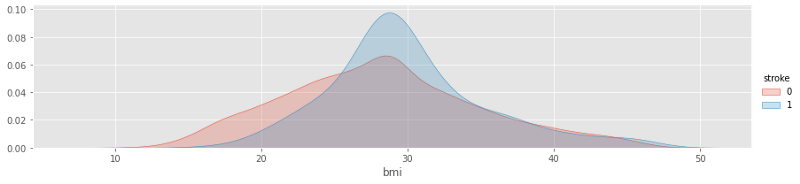
Figure 5: Smoking Status vs Stroke

As we can see here the picture suggests that people who once smoked and people who are currently smoking are having more percent of stroke cases compared to people who never smoked and people who don't know their current status.

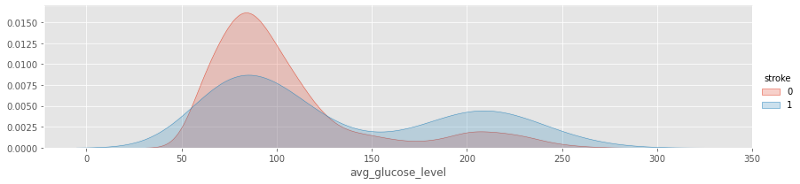
This busts the myth and proves that Smoking is one of the biggest risk factors for stroke.

**Myth III: Glucose level and BMI doesn’t contribute to stroke.  
Fact:** Diabetes is a well-established risk factor for **stroke**. It **can cause** pathologic changes in blood vessels at various locations and **can lead to stroke** if cerebral vessels are directly affected. Additionally, mortality is higher and poststroke outcomes are poorer in patients with **stroke** with uncontrolled **glucose levels.** Another **effect** of being **overweight** is that the body’s metabolism changes in ways that lead to an excess of circulating lipids, high cholesterol, and elevated blood glucose, all of which, over time, harm the blood vessels of the brain and the heart and lead to the formation of **stroke**-causing blood clots in the heart and brain.

Now let us see what our data-based approach suggests:  
So first we analysed the BMI column in our dataset and found out that people with overweight are more susceptible to stroke.

Figure 6: BMI vs stroke

As we can see people with a BMI of more than 28 are the ones with more risk of having a stroke. So the average glucose levels are as follows

Figure 7: Average Glucose level vs stroke

The picture depicts that people with low glucose levels and also people with higher glucose levels are at equal risk of having a stroke. **Hypoglycaemia (**Lower glucose level) is defined as a blood **glucose** level less than 70 mg/dl. It's a common occurrence in diabetic patients receiving pharmacologic treatment. Transient **hypoglycaemia** is well known to **produce a stroke**-like picture with hemiplegia and aphasia. The **effect** of being **overweight** is that the body’s metabolism changes in ways that lead to an excess of circulating lipids, high cholesterol, and elevated blood glucose. Due to this elevated blood glucose risk of having a stroke raises exponentially.

So, this busts the myth and proves that glucose levels and BMI play an important role in stroke occurrence.

**Conclusion**

In this article, we took a look at What causes a stroke according to Kaggle’s 2021 healthcare data.

* First, we have busted a myth that says “Strokes only happen to elderly people”. It’s true that as you age your risk for stroke goes up. However, there’s also an increasing number of strokes in people between the ages of 20 and 90.
* Second, we have busted another myth that says “Smoking doesn’t affect your chances of having a stroke”. we proved that Smoking is one of the biggest risk factors for stroke, especially in younger people.
* Finally, we have proved that Glucose level and BMI both together play an important role in raising the risk of having a stroke.

A stroke is a medical emergency, and prompt treatment is crucial. Early action can reduce brain damage and other complications.

So, this is my analysis of how a brain stroke can occur. but the real question is:

Are YOU healthy enough for not having a stroke?

To see more about this analysis, see the link to my GitHub available [here.](https://github.com/lakshman533/Brain-Stroke)