# Problem Statement

A stroke is a [medical condition](https://en.wikipedia.org/wiki/Disease) in which poor [blood flow](https://en.wikipedia.org/wiki/Cerebral_circulation) to the [brain](https://en.wikipedia.org/wiki/Brain) causes [cell death](https://en.wikipedia.org/wiki/Cell_death). It occurs when a blood vessel that carries oxygen and nutrients to the brain is either blocked by a clot or bursts (or ruptures).

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. Stroke is also the third major cause of disability and a leading cause of dementia and depression. Long term disability affects people severely, in terms of their productive life. As such, stroke poses a significant threat to global health.

Despite its enormous impact on countries’ socio-economic development, this growing crisis has received very little attention to date.

# General Objective

To analyze the factors that contribute to getting a stroke and the extent to which they increase or decrease the chances of getting a stroke

## Data Mining goals (Specific Objectives)

1. Which gender is more prone to getting a stroke?
2. Is one likely to get a stroke based on their marital status?
3. Which age group has the highest risk of getting a stroke?
4. Does the type of work affect your chances of getting a stroke?
5. Does the body mass index (BMI) level affect the chances of getting a stroke?
6. Does smoking cause a stroke?
7. Does heart disease and hypertension cause a stroke?

# Data Description

The dataset used is from <https://www.kaggle.com/fedesoriano/stroke-prediction-dataset?select=healthcare-dataset-stroke-data.csv>

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 with the following columns(attributes).

|  |  |  |
| --- | --- | --- |
| # | Column | Column Description |
| 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 |

# Data cleaning

1. Missing values:- The BMI attribute had 201 (out of 5110). The missing values were imputed by putting in the average (mean) bmi value.
2. For uniformity columns with upper case were changed to lowercase and space replaced with under\_scrore.
3. There were no data duplicates and outliers were discarded.

The cleaned data was exported in CSV format for onward analysis.

# Hypothesis Testing Procedure

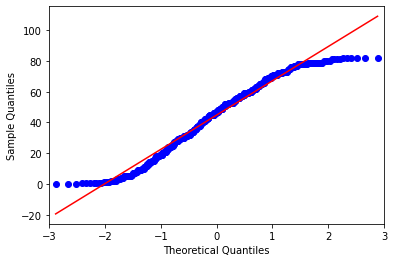
Null hypothesis (Ho): Risk of getting a stroke is evenly distributed among all age groups.

Alternative hypothesis(Ha): Individuals between the age of 75-79 have the highest risk of getting a stroke.

A level of significance of 10% with a sample size of 511 in our test statistic.

# Hypothesis Testing Results

The sample was confirmed to be a normal distribution.

****

Test Results: -

Z score:- 0.0177

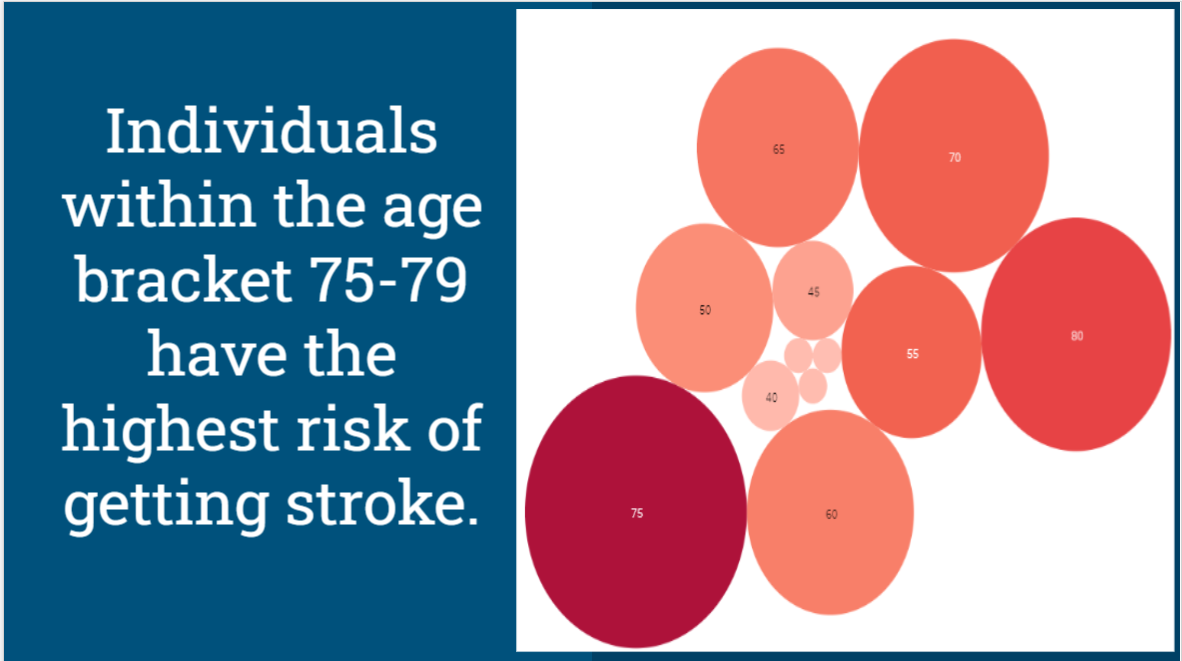
P value:- 0.5071

Since the p value is greater than 0.05, reject the null hypothesis that Risk of getting a stroke is evenly distributed among all age groups.

**Summary and Conclusions**

We concluded that the risk of getting a stroke is not the same in all age groups. It is evident that stroke mostly affects the older generation especially those between the age of 75-79 years.

There is an observation that the average level of glucose has a positive correlation with age. This is an observation for further investigation.



## Recommendations

We recommend that individuals who fall in the age bracket of 75 - 79 should:

1. Get access to quality food and practice of proper dietary habits.
2. Be given access to quality health care (Medication/physiotherapy/speech therapy/etc).
3. Have a healthy working environment.
4. Belong to social support systems to avoid being alone and depressed.