**PROJECT ON STROKE**

**EXECUTIVE SUMMARY**

This project is aimed at extracting specific desired information from a dataset containing information of a sample of people both male and female, revealing their hypertension status and stroke disposition. In the course of this project, the following questions will be satisfied:

* Plot a graph to shows Age Group Distribution
* Plot a graph to show Gender Distribution
* Plot a graph to show Hypertension Distribution
* Plot a graph to show Heart Disease Distribution
* Plot a graph to show Work Type Distribution
* Plot a graph to show Smoking Status Distribution
* Build a model for this data set to see if a patient has stroke or not, the accuracy score should not be less than 96%

**INTRODUCTION**

A stroke is a medical condition in which poor blood flow to the brain causes cell death. There are two main types of stroke: ischemic, due to lack of blood flow, and hemorrhagic, due to bleeding. Both cause parts of the brain to stop functioning properly. Signs and symptoms of a stroke may include an inability to move or feel on one side of the body, problems understanding or speaking, dizziness, or loss of vision to one side. Signs and symptoms often appear soon after the stroke has occurred. If symptoms last less than one or two hours, the stroke is a transient ischemic attack (TIA), also called a mini-stroke. A hemorrhagic stroke may also be associated with a severe headache. The symptoms of a stroke can be permanent. Long-term complications may include pneumonia and loss of bladder control.

The main risk factor for stroke is high blood pressure. Other risk factors include high blood cholesterol, tobacco smoking, obesity, diabetes mellitus, a previous TIA, end-stage kidney disease, and atrial fibrillation. An ischemic stroke is typically caused by blockage of a blood vessel, though there are also less common causes. A hemorrhagic stroke is caused by either bleedingdirectly into the brain or into the space between the brain's membranes. Bleeding may occur due to a ruptured brain aneurysm. Diagnosis is typically based on a physical exam and supported by medical imaging such as a CT scan or MRI scan. A CT scan can rule out bleeding, but may not necessarily rule out ischemia, which early on typically does not show up on a CT scan. Other tests such as an electrocardiogram (ECG) and blood tests are done to determine risk factors and rule out other possible causes. Low blood sugar may cause similar symptoms

**METHODOLOGY**

Python programming language was used throughout the work which was done using Jupyter notebook. Different EDA techniques were deployed in order to achieve the desired goal of the questionnaire, but first data mining and cleaning was deployed to enable accuracy of facts pulled from the given datasets.

In the Machine learning area, ***RandomForestClassifier*** and ***MLPClassifier***were used as the first and second models respectively for the prediction.

**RESULTS**

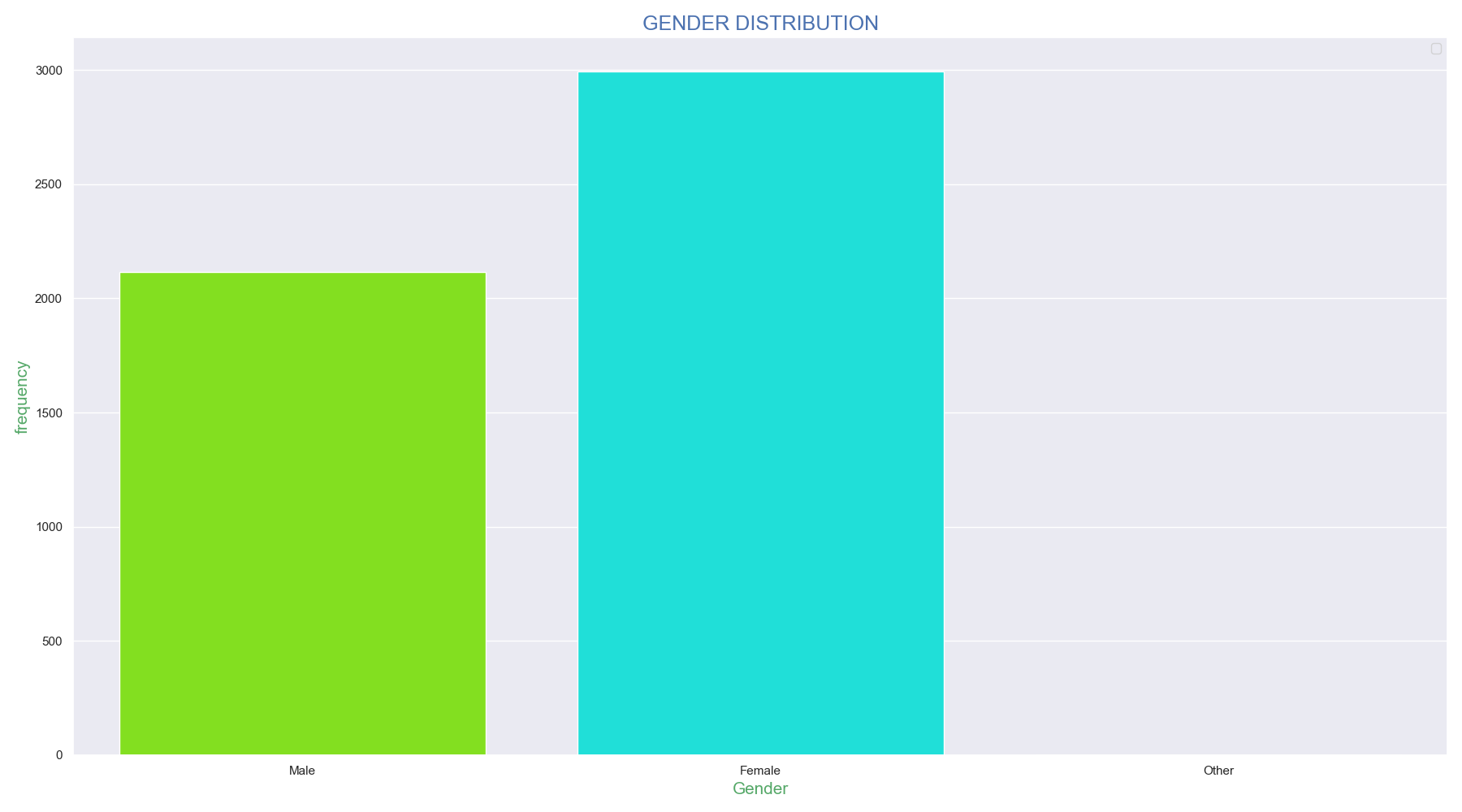
1. **Plot a graph to show Age Group Distribution:**

****

**NOTES:**

* Each histogram bar represents an age range of 0-4yrs
* The age bracket with the highest count in the dataset are 78-82 with a count of 317 persons
* The age bracket with the least count in the dataset are 10-14yrs and with a count of 317 persons

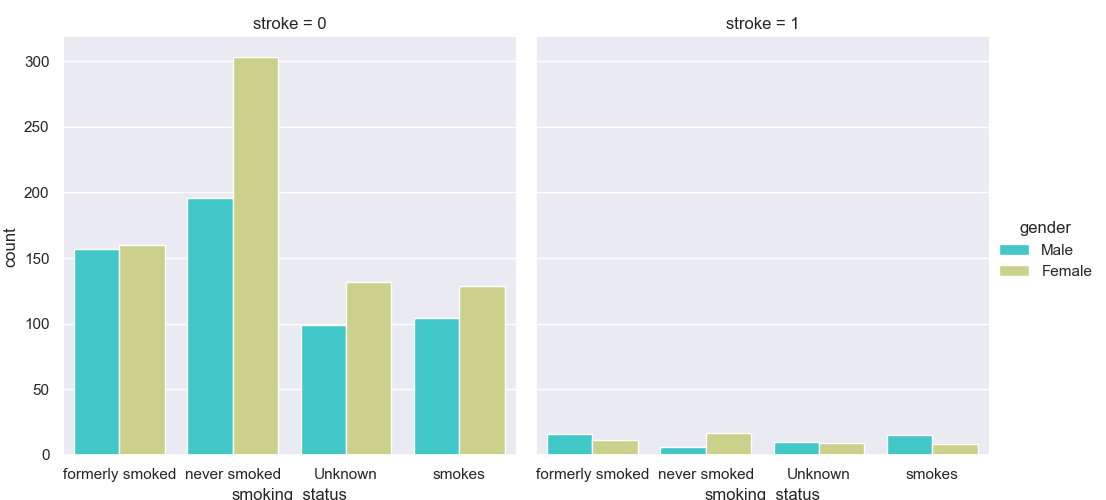
1. **Plot a graph to show Gender Distribution:**

****

**NOTES:**

* A total of 2,994 men were captured in the dataset
* A total of 2,115 women were captured in the dataset

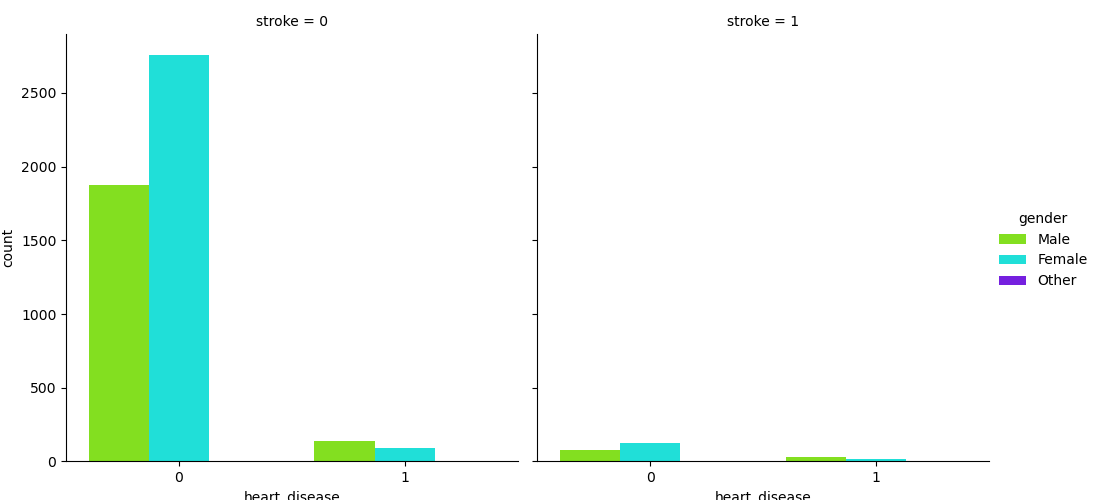
1. **Plot a graph to show Hypertension Distribution:**

****

**NOTES:**

* Over 300 women and about 200 men who have never smoked are hypertensive
* About 150 women and 150 men who have formerly smoked are hypertensive
* About 130 women and 105 men who smoke are hypertensive
* Regardless of lifestyle, more women than men in this dataset are hypertensive
* More women than men came down with stroke in the category of those who have never smoked
* In the category of those who have formerly smoked, more men than women came down with stroke
* In the category of those who smoke, there are more men than women who came down with stroke

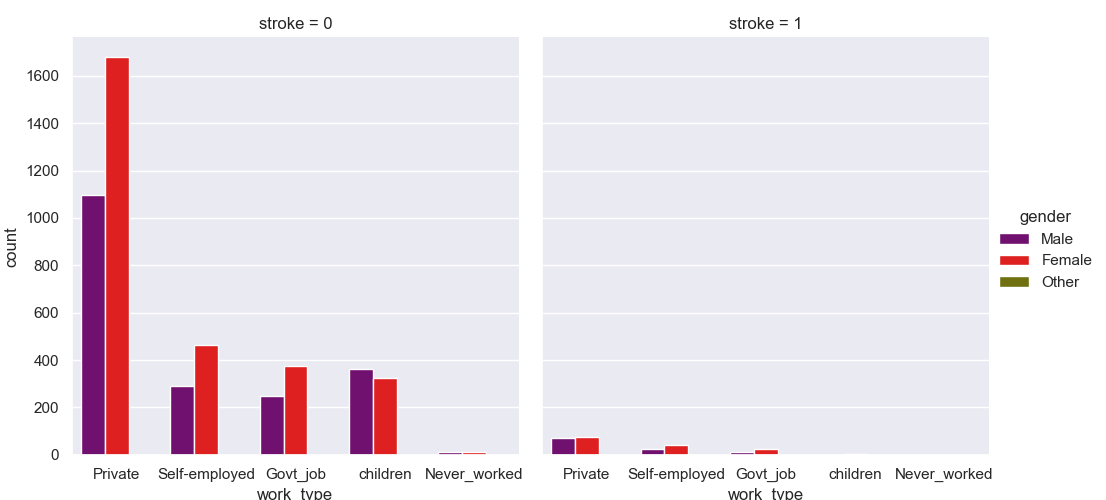
1. **Plot a graph to show Heart disease Distribution:**

****

**NOTES:**

* In the category of patients without stroke, there are only a few with heart disease.
* Amongst those with heart disease, there are more men than women
* There are about 1,900 men and 2,800 women free of stroke and heart disease.
* In the stroke category, there are more women than men without heart disease

1. **Plot a graph to show Work type Distribution:**

****

**NOTES:**

* In the private sector, there are about 1,700 females & about 1,100 males without stroke, and about 100 males and females respectively with stroke
* In the self-employed category, there are about 460 females & about 300 males without stroke, and about 15 males and 25 females respectively with stroke
* In the Govt. Job Category, there are about 380 females & about 250 males without stroke, and about 10 males and 15 females respectively with stroke
* Amongst the children, there are about 340 girls & about 370 boys without stroke, and there are no kids with stroke
* Amongst those who have never worked, there are about 10 males and 10 females without stoke and none with stroke in same category

1. **Plot a graph to show Smoking Status Distribution**

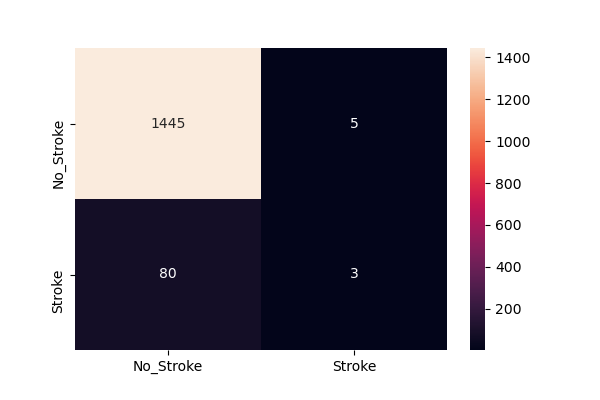
****

**NOTES:**

* In the never smoked category, there are about 1,060 females & about 680
* In the formerly smoked category, there are about 490 females & about 400
* In the category of those who smoke, there are about 480 females & about 360
* In the unknown category, there are about 840 females & about 710

1. **7.Build a model for this data set to see if a patient has stroke or not, The accuracy score should not be less than 96%**

**MODEL 1:**

****

**precision recall f1-score support**

**0 0.95 1.00 0.97 1450**

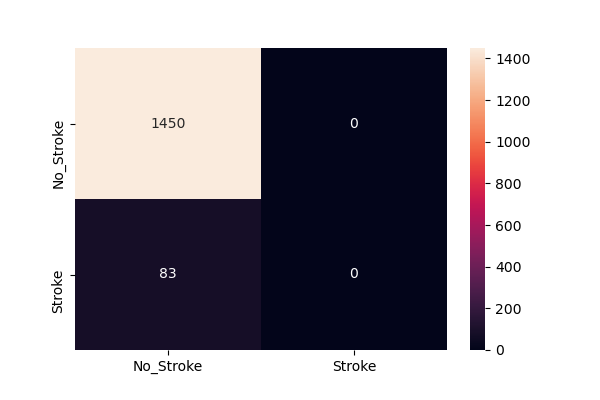
**1 0.38 0.04 0.07 83**

**accuracy 0.94 1533**

**macro avg 0.66 0.52 0.52 1533**

**weighted avg 0.92 0.94 0.92 1533**

**MODEL 2:**

****

**precision recall f1-score support**

**0 0.95 1.00 0.97 1450**

**1 0.00 0.00 0.00 83**

**accuracy 0.95 1533**

**macro avg 0.47 0.50 0.49 1533**

**weighted avg 0.89 0.95 0.92 1533**

**CONCLUSION**

From the given dataset, it is revealed that females present are prone to hypertension than males when exposed to same lifestyle and even some times regardless of life style. It also arguable that the working class (private, self-employed, and Govt. Workers) are exposed to the risk of hypertension. This could be owing to increased stress level.