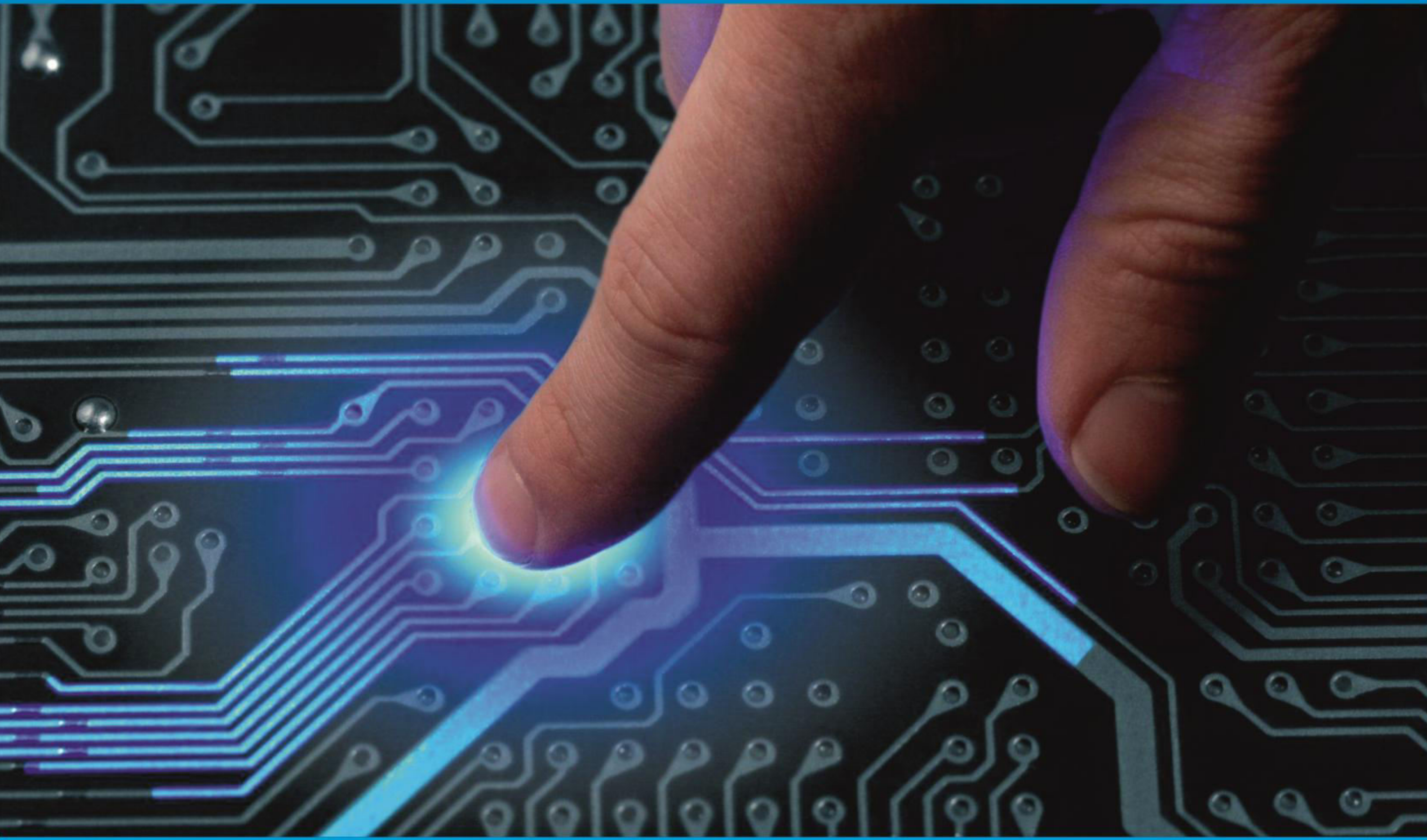
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905059 |

A Study of STROKE ANALYSIS

Soham Roy[1], Siddharth Nanda[2]

U.G Student, School of Engineering, Ajeenkya DY Patil University, Pune, Maharashtra, India [1]

Faculty, School of Engineering, Ajeenkya DY Patil University, Pune, Maharashtra, India [2]

ABSTRACT : As we all know , the human heart is responsible for keeping a person alive , it pumps blood and gives us that vital blood flow in our body that keeps us warm and alive .A stroke happens when the blood's stream to the mind is hindered, or a vein in the cerebrum bursts. Without blood, brain cells begin to kick the bucket, and the capacities constrained by that territory of the cerebrum, for example, movement or muscle development—are disabled or lost. Our body also quite similar to the data that flows and is stored and analyse for future use , The body is a place where every second Air flows in and out , the Air that goes inside does the work of Raw Data ,coming in with heavy flow but not organised .The body then organises the Air to make use of it to give us a healthy pumping heart which in turn , provides us life .So , just like our body , Raw data comes in , and is sorted first , then stored to make use of it so that we can get/predict a particular outcome via reading ,analysing and understanding the pattern of sorted incoming data .The Stroke Analysis done for this paper is through Excel

KEYWORDS: Data , Cerebrum , Stroke

I. INTRODUCTION

Neurological issues brings harm to the focal and fringe sensory system. A portion of these illnesses can be treated while others can't be. Individuals, mostly after a certain age ,experience the ill effects of neurological issues, for example, Alzheimer's illness and Parkinson's infection, making age a critical factor for building up this illness. The causes vary from being hereditary issues, diseases, way of life to any medical conditions that may influence the cerebrum. There are more than sicknesses of the sensory system, like stroke, cerebrum tumors, epilepsy and some more. Around fifteen million individuals experience the ill effects of stroke every year. There exits very little productive methods for the patient to anticipate whether the person in question could have such sicknesses and this exploration for the most part centers around that.An simplicity of use disease forecast application that deals with appropriately examined information is a lot of required for any client to test and figure it out one's ailment. This can help in taking the essential medicines from an emergency clinic. The application moreover serves helpful in making the clients mindful of the requirement for a appropriate way of life. Around individuals were conceded because of neurological conditions in the year .The neurological conditions include Autism, Dementia, Epilepsy etc. Men are found to have a stroke at a more youthful age than ladies and stroke related passings happen more in ladies. The dataset utilized for this researchbelongs to the records of people tried for stroke with factors like sexual orientation, age, way of life credits, Body Mass Index, segment locales, etc.There are a few factors that assume a part in stroke frequency, some of which are heredity, age, sexual orientation and race, certain ailments, for example, hypertension, hypercholesterolemia, coronary illness and diabetes. Overweight, previous history of stroke can likewise build the frequency hazard of stroke. No smoking and no liquor utilization and day by day exercises can likewise be viable to decrease the danger of stroke. By utilization of the previously mentioned hazard variables, and methods of data mining, choice emotionally supportive network can be planned that other than information and experience of a doctor, can be utilized to foresee stroke. Attributable to the human need of information and expanding data volume, method improvement for computerized extraction of information from these data is unavoidable. Data mining is extraction of information and appealing examples from a huge volume of data. Data mining procedures dependent on information that can be extricated are separated into three significant gatherings: In this paper , Python is used as a tool for the above steps and for pre-processing /predicting the stroke traits .

II. LITERATURE SURVEY (10 PAPERS) :

1. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: Systematic analysis of population health data. *Lancet.* 2006;367:1747–57. [PubMed] [Google Scholar]

2. Heron M, Hoyert DL, Murphy SL, Xu J, Kochanek KD, Tejada-Vera B. Deaths: Final data for 2006. *Natl Vital Stat Rep.* 2009;57:1–134. [PubMed] [Google Scholar]

3. Lloyd-Jones D. Heart disease and stroke statistics-2010 update. A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation.* 2010;121:e1–170. [Google Scholar]

4. Brault MW, Hootman J, Helmick CG, Theis KA, Armour BS. Prevalence and most common causes of disability among adult-United States, 2005. *MMWR Morb Mortal Wkly Rep.* 2009;58:421–6. [PubMed] [Google Scholar]

5. Azarpazhooh MR, Etemadi MM, Donnan GA, Mokhber N, Majdi MR, Ghayour-Mobarhan M, et al. Excessive incidence of stroke in Iran, evidence from the mashhad stroke incidence study (MSIS), a population-based study of stroke in the middle East. *Stroke.* 2010;41:e3–10. [PubMed] [Google Scholar]

6. Chawla M, Sharma S, Sivaswamy J, Kishore L. A method for automatic detection and classification of stroke from brain CT images. *Conf Proc IEEE Eng Med Biol Soc.* 2009;2009:3581–4. [PubMed] [Google Scholar]

7. Przelaskowski A, Sklinda K, Bargieł P, Walecki J, Biesiadko-Matuszewska M, Kazubek M. Improved early stroke detection: Wavelet-based perception enhancement of computerized tomography exams. *Comput Biol Med.* 2007;37:524–33. [PubMed] [Google Scholar]

8. Tang FH, Ng DK, Chow DH. An image feature approach for computer-aided detection of ischemic stroke. *Comput Biol Med.* 2011;41:529–36. [PubMed] [Google Scholar]

9. Mroczek T, Grzymala-Busse J, Hippe ZS. A new machine learning tool for mining brain stroke data. *3rd International Conference on Human System Interactions (HSI) IEEE: Digital Object Identifier.* 2010:246–50. [Google Scholar]

10. Han J, Kamber M. *Data Mining: Concept and Techniques.* 2nd ed. California: Morgan Kaufmann Publishers; 2006. [Google Scholar]

11. Mitchell T. *Machine Learning.* New York: McGraw-Hill; 1997. [Google Scholar]

12. Witten IH, Frank E. *Data mining, practical machine learning tools and techniques.* 2nd ed. California: Morgan; 2005. [Google Scholar]

13. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: An Update. *SIGKDD Explorations.* 2009;11:10–8. [Google Scholar]

III. ANALYSIS APPROACH

At the point when cleaned data is present , data mining should be done to get a good understanding from it. Investigation of the factors is done in this stage to get important data about the informative factors and their connection with the reaction variable. Indeed/no visualization of stroke and depression act go about as the reaction variable in both datasets for this cycle. The point of this stage is to show and get data on how different factors are affecting and relating with our reaction variable through plots and representation. The various factors of stroke dataset such as age, BMI, glucose levels, sex, etc are checked for their connection with the stroke guess variable. On breaking down age through box plot, patients with ages 60 to 77 are found to have stroke than the patients at different ages and the mean age for stroke is discovered to be 67. Patients with a BMI of 26 upto just about 35 are found to get influenced with stroke. This shows that overweight and large individuals are well on the way to get influenced by stroke than different BMI levels. Data like this must be accomplished through the EDA stage and henceforth its significance. The normal glucose levels for a patient with stroke is seen to be between 95 to 120. The over three factors are mathematical factors and thatiswhy a box plot and histogram are utilized for investigating them. Rest of the factors thatwe are going to work upon are unmitigated factors i.e factors. Barplots are utilized for plotting these factors effectively. Male patients are found to have stroke than the female ones. 72% of the patients with coronary illness seem to have strokes more than individuals who don't have coronary illness.
In accordance with other medical services datasets, this dataset was profoundly unequal also. Just 783 patients endured a stroke while the excess 42,617 patients didn't have the experience.

Before we can continue further, we should preprocess the data, to separate significant bits of knowledge from the dataset.

1. ID attribute :
This was used to identify patients only , also did not have any meaningful inormation , hence the column needs to be dropped .

df.drop(columns=['id'], inplace=True)

2. BMI attribute :

1,458 records were recorded as NaN (not a number) in the BMI section. The previously thought was to eliminate them since they addressed a small fraction of the dataset. In any case, by examining further, it contained 140 records where patients endured a stroke. This information was valuable considering the fact that solitary 783 patients endured a stroke in this dataset. Consequently, records with void value in BMI was replaced with mean of BMI.

df[df['bmi'].isna() & df['stroke'] == 1]

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
81	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
407	Female	59.0	0	0	Yes	Private	Rural	76.15	NaN	NaN	1
747	Male	78.0	0	1	Yes	Private	Urban	219.84	NaN	NaN	1
1139	Male	57.0	0	1	No	Govt_job	Urban	217.08	NaN	NaN	1
1613	Male	58.0	0	0	Yes	Private	Rural	189.84	NaN	NaN	1
...	...	...	...	...	...	...	...	...	...	...	...
42530	Male	66.0	0	0	Yes	Self-employed	Urban	182.89	NaN	never smoked	1
42839	Female	67.0	1	0	Yes	Govt_job	Urban	234.43	NaN	never smoked	1
43007	Female	69.0	0	1	Yes	Self-employed	Rural	89.19	NaN	smokes	1
43100	Male	67.0	0	0	Yes	Self-emp	Urban	136.79	NaN	smokes	1

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
						loyed					
43339	Female	76.0	0	0	No	Private	Rural	100.55	NaN	never smoked	1

3. Smoking Status attribute :

What's more, 13,292 records or about 30.6% of the dataset had missing qualities in smoking status include . It was a tremendous extent of the dataset. Accordingly, another classification named "not known" was made to represent every one of these records, as opposed to dropping them through and through.

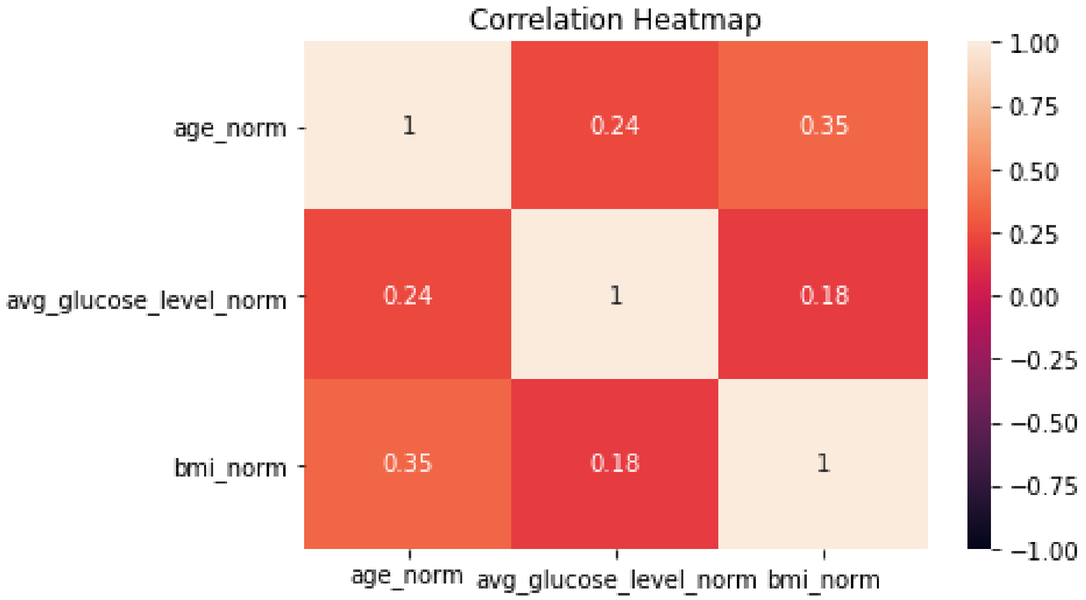
```
df['smoking_status'].fillna('not known', inplace=True)

print(df['smoking_status'].value_counts())
```

```
never smoked       16053

not known          13292

formerly smoked     7493

smokes             6562

Name: smoking_status, dtype: int64
```

Now , coming to the Analysis part of this dataset , and try to get some interesting insights .

Numerical attributes :

```
 # Create the correlation heatmap

heatmap = sns.heatmap(df[['age_norm', 'avg_glucose_level_norm', 'bmi_norm']].corr(), vmin=-1, vmax=1, annot=True)

# Create the title

heatmap.set_title('Correlation Heatmap');
```



It seemed like both BMI and Age were positively correlated, though the association was not strong.

In addition, 100% stacked bar charts were plotted to discover any potential relationship between the variable and stroke. With little tweak, a new yet similar function was created to avoid duplication of codes. It takes in the name of the column and outputs the 100% stacked bar chart.def get_100_percent_stacked_bar_chart(column, width = 0.5):

```
# Get the count of records by column and stroke

df_breakdown = df.groupby([column, 'stroke'])['age'].count()

# Get the count of records by gender

df_total = df.groupby([column])['age'].count()

# Get the percentage for 100% stacked bar chart

df_pct = df_breakdown / df_total * 100

# Create proper DataFrame's format

df_pct = df_pct.unstack()

return df_pct.plot.bar(stacked=True, figsize=(6,6), width=width);
```

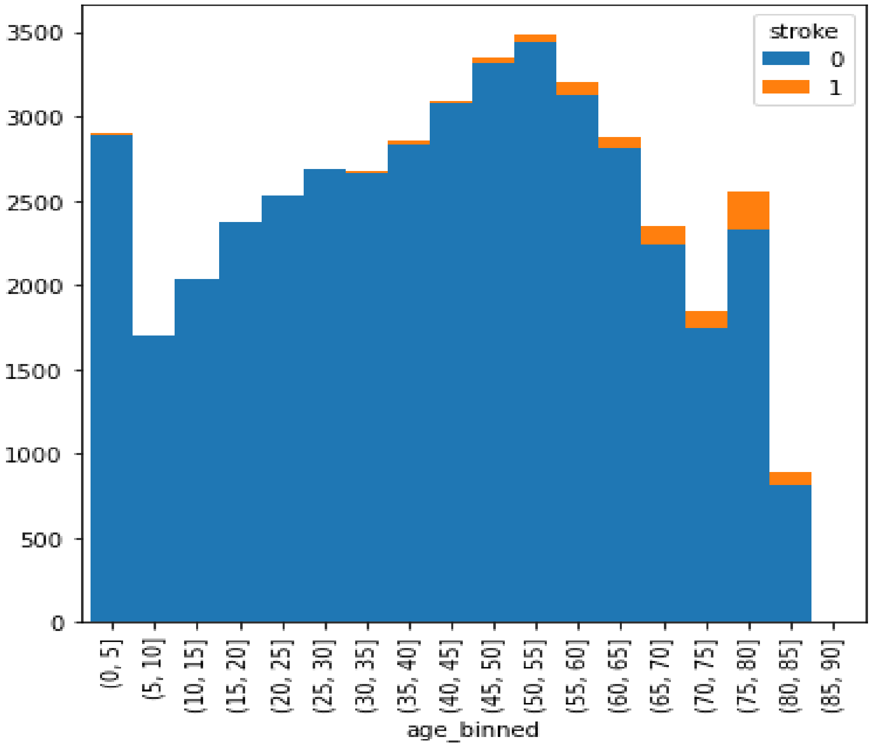
a) Age :

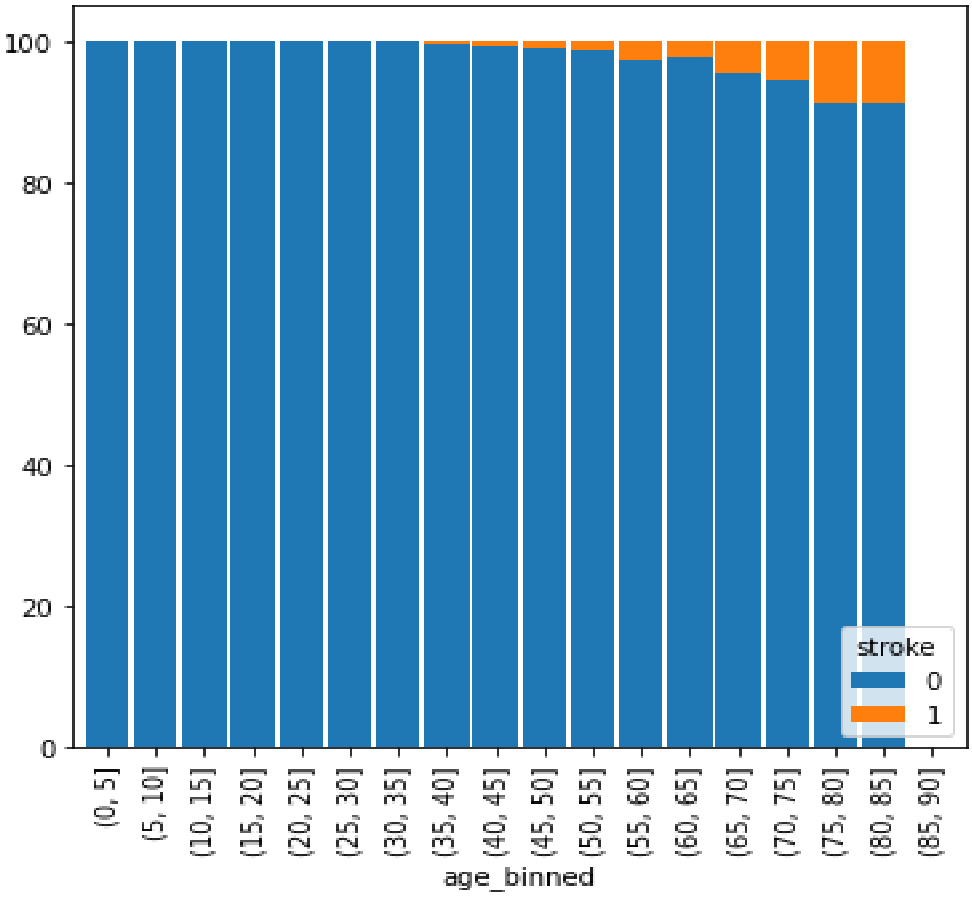
```
get_stacked_bar_chart('age_binned')

get_100_percent_stacked_bar_chart('age_binned', width = 0.9)
```

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905059 |

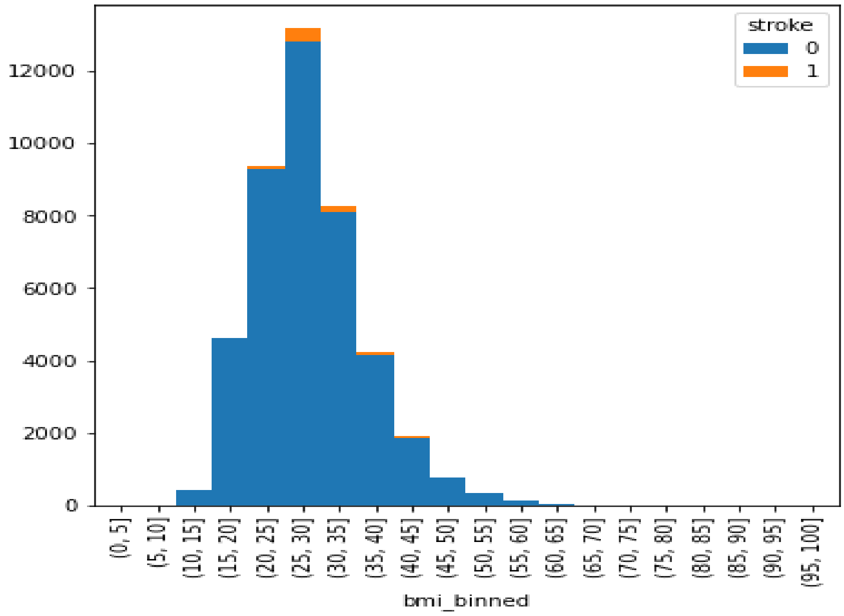


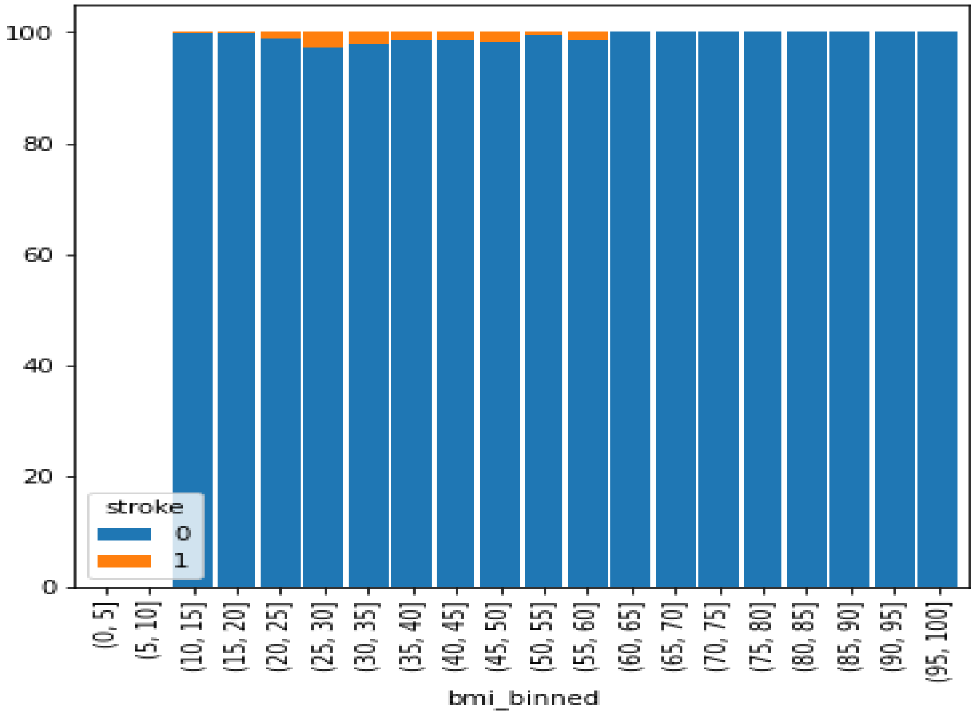


The danger of encountering a stroke increased as the patient's age progressed. More seasoned patient was bound to endure a stroke than a more youthful patient.

b) BMI

```
get_stacked_bar_chart('bmi_binned')

get_100_percent_stacked_bar_chart('bmi_binned', width = 0.9)
```





| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

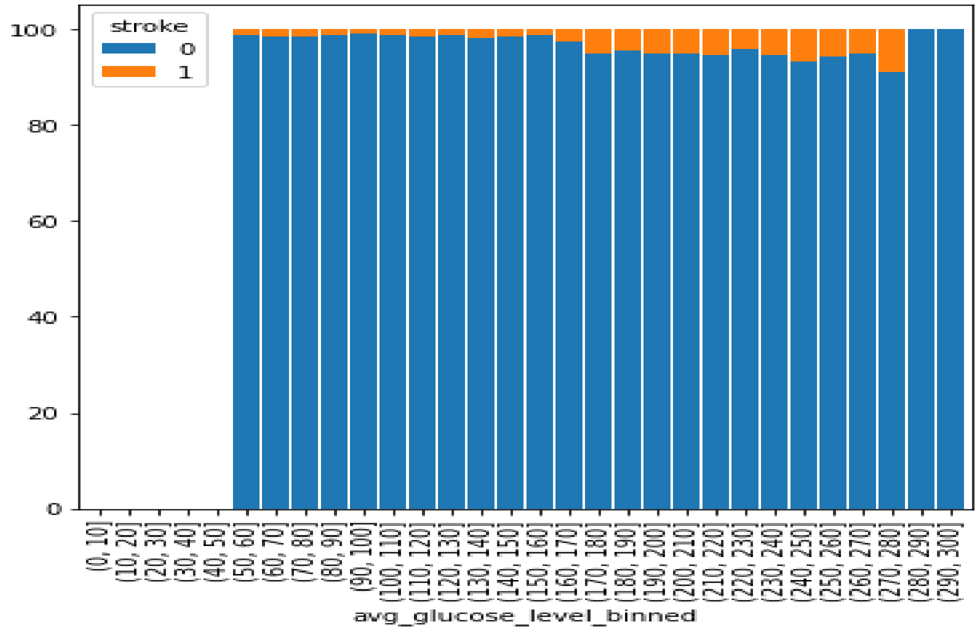
|| Volume 9, Issue 5, May 2021 ||

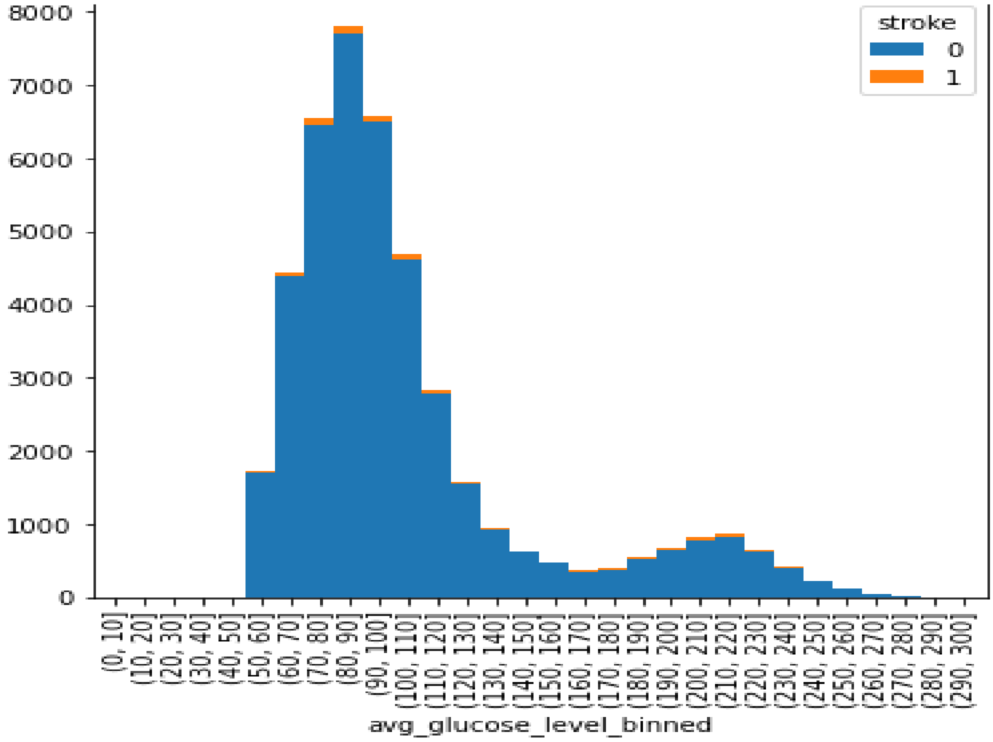
| DOI: 10.15680/IJIRCCE.2021.0905059 |

Level of patient who had BMI somewhere in the range of 25 and 35 was the most noteworthy to experience the ill effects of different gatherings.

Higher BMI doesn't expand the stroke hazard.

c) Average glucose level





The above pictures depicts that stroke frequency happened to certain patients paying little mind to the normal glucose level estimated after supper. Despite the fact that there was no stroke rate covered the keep going two segments on the right, these sections were addressed by just 3 patients, for example not huge. All things considered, higher extent of patient who had normal glucose level estimated after dinner of more than 150mg/dL (milligrams per decilitre) endured a stroke. This perception can be clarified by the presence of diabetes. Diabetes was available in persistent who had perusing of more than 200mg/dL. Pre-diabetes was likewise thought to be in persistent if the perusing was between 140–199mg/dL.

Diabetes is one of the danger factors for stroke event and prediabetes patients have an expanded danger of stroke.

d) Hypertension, Heart disease :

Higher proportion of patients who suffered from hypertension or heart disease experienced a stroke, all else being equal.

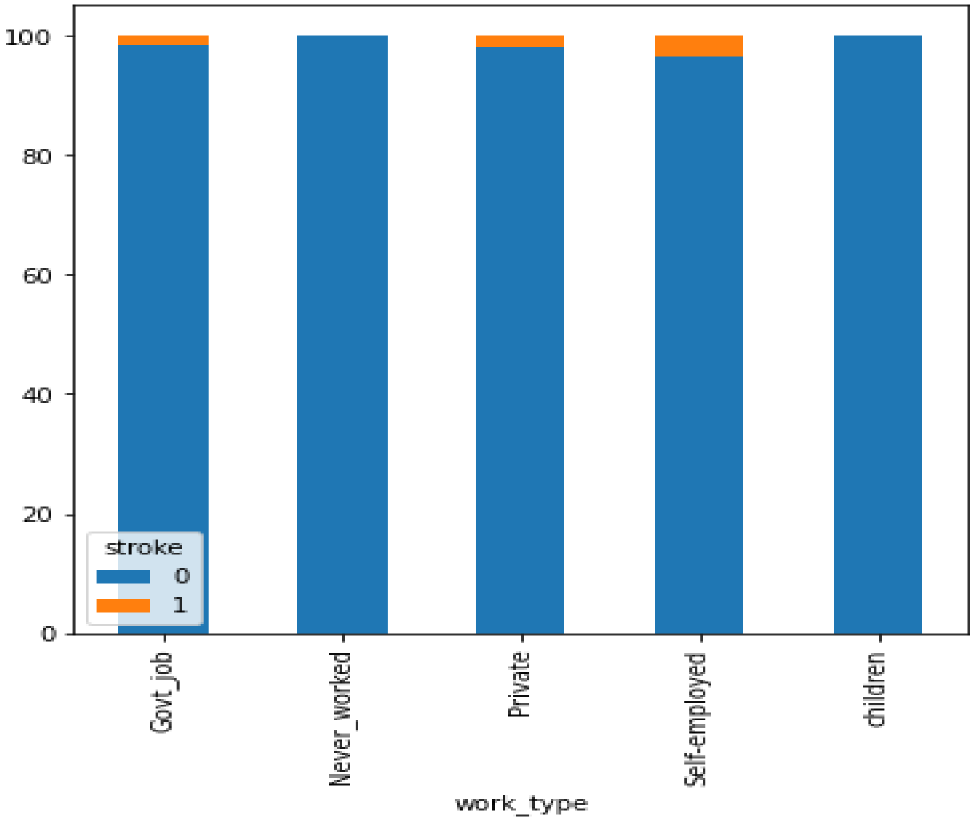
e) Gender, Residence type :

Regardless of patient’s gender, and where they stayed, they have the same likelihood to experience stroke.

f) Work type :

```
get_100_percent_stacked_bar_chart('work_type')
```

```
df.groupby(['work_type'])[['age']].agg(['count', 'mean'])
```



The above figure shows very ineteresting observation .

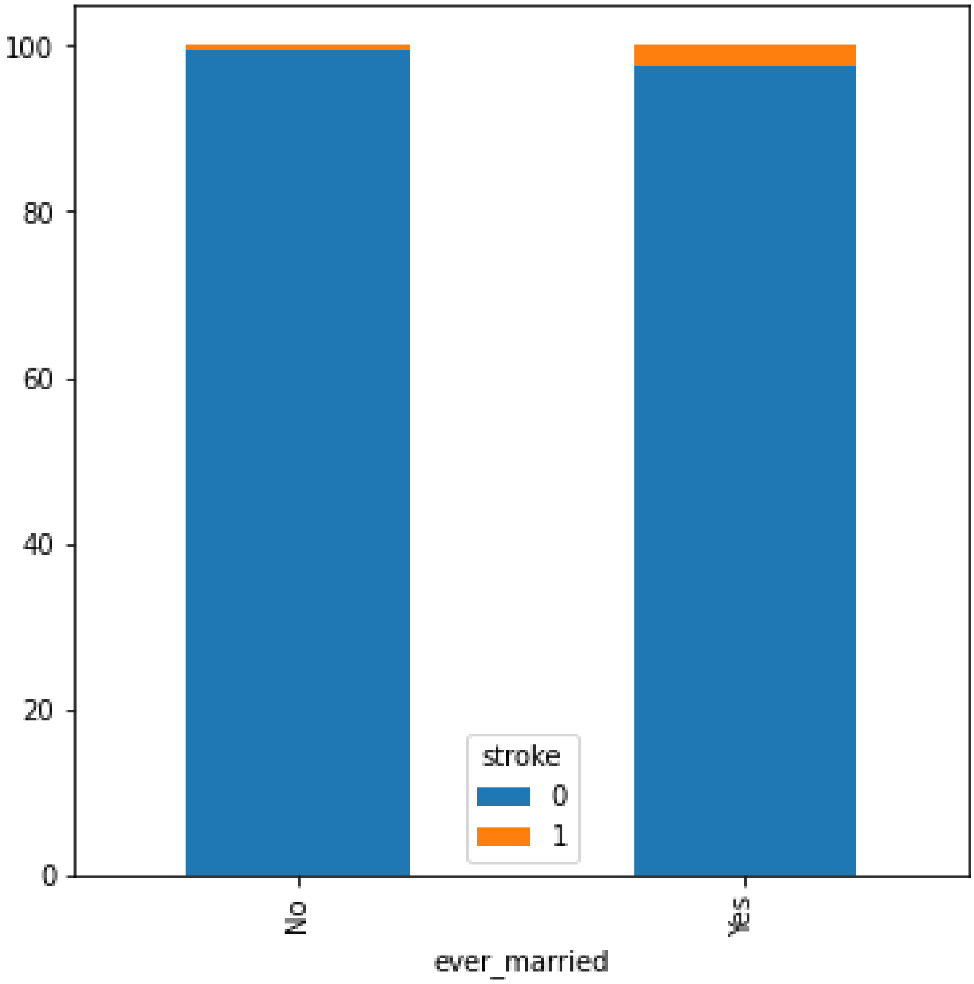
From the outset, extent of patient who was independently employed and endured a stroke was moderately higher than different classes. Be that as it may, this variable was profoundly connected with age. Both never worked and youngsters classifications were really clear as crystal. Practically non-existent stroke was recorded because of below age. Then again, the mean period of patients who were independently employed was 59.3 years old. It was the most noteworthy among all classes.

Work type variable was profoundly connected with age.

g) Ever Married :

get_100_percent_stacked_bar_chart('ever_married')

df.groupby(['ever_married'])[['age']].agg(['count', 'mean'])



The above graph shows a comparative observation as the work type variable. In any case, the top diagram shows the distinct contrast in mean of age of the two classifications. Hence , Marital status variable was exceptionally connected with age.

III. CONCLUSION

There a sum of 8 insights found in the stroke dataset :

- It does looked like the two BMI and Age were positively corresponded, however the association was not strong.
- More seasoned patient was bound to suffer a stroke than a more youthful patient.
- Higher BMI does not increase the stroke risk.
- Diabetes is one of the risk factors for a stroke occurence event and pre-diabetes patients have an increased risk of stroke.
- Higher extent of patients who suffered from hypertension or coronary heart illness encountered a stroke, all else being equivalent.
- Regardless of patient's sex, and where they stayed, they have the same probability to encounter stroke .
- Work type variable was highly associated with age.
- Marital status variable was highly associated with age.

REFERENCES

1. https://www.ahajournals.org/doi/full/10.1161/01.STR.28.6.1138
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3104713/
3. https://www.hindawi.com/journals/rerp/2019/1726964/
4. https://bmcneurol.biomedcentral.com/articles/10.1186/s12883-019-1409-0
5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271469/
6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5325924/
7. https://www.ahajournals.org/doi/10.1161/strokeaha.106.474080
8. https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/methods-of-sampling-population
9. https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-019-3222-x

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

☎ 9940 572 462 ✆ 6381 907 438 ✉ ijircce@gmail.com

www.ijircce.com

Scan to save the contact details