Descriptive Analytics for Stroke

Surayyi Alqahtani

Health Informatics specialist

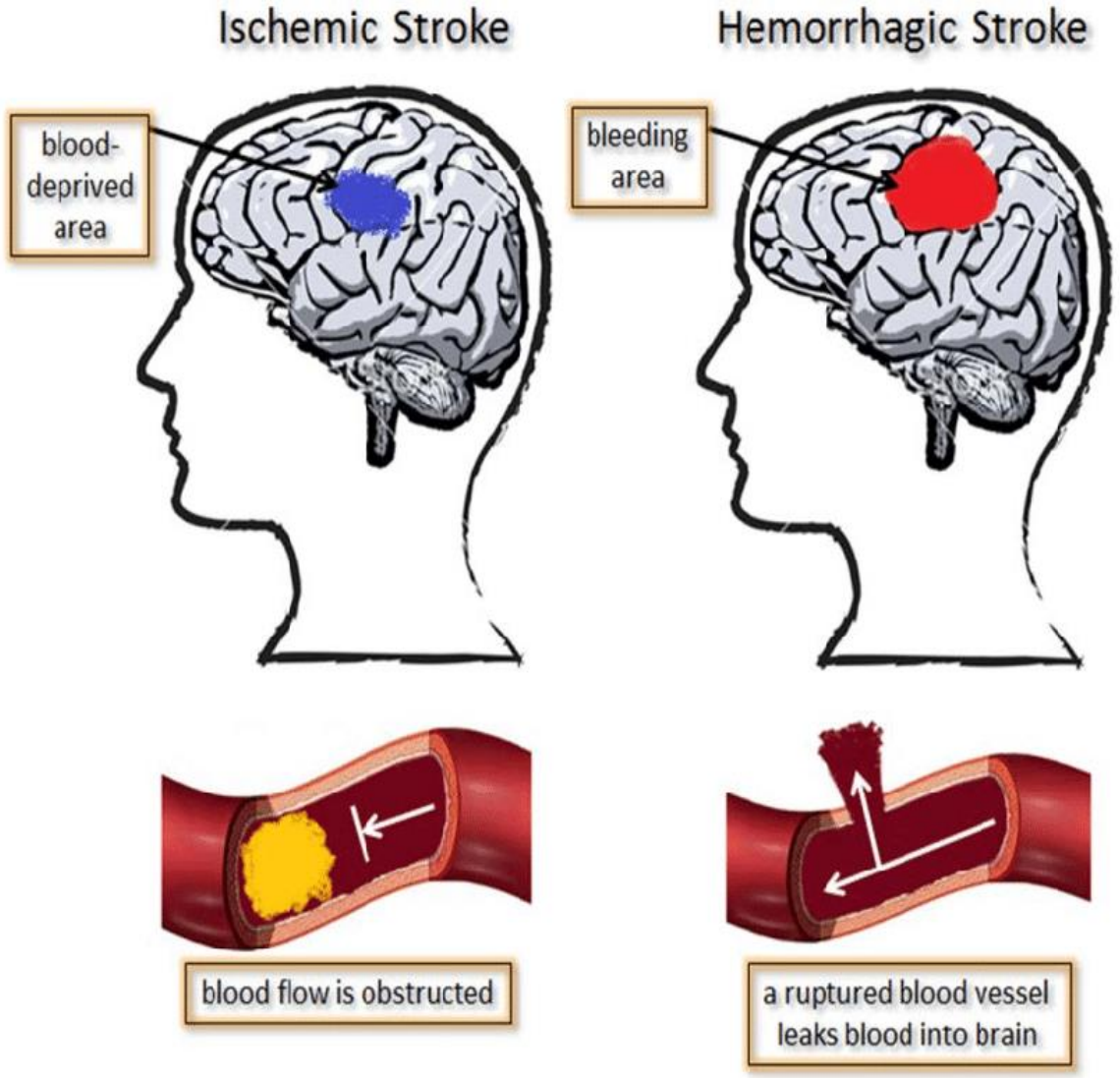
Outline:

- Introduction to Stroke.
- Question/Problem statement.
- Data Structure.
- Data Cleaning.
- Data Analysis.
- Conclusion.

Introduction to Stroke:

- **A stroke** occurs when the blood supply to part of brain is interrupted or reduced, preventing brain tissue from getting oxygen and nutrients.

- **What Are the Types of Strokes?**
  - - Ischemic Stroke.
  - - Hemorrhagic Stroke.



Question/Problem statement:

- **Scenario:** The Ministry of Health intends to establish an awareness campaign to raise awareness about the risk of stroke and wants to choose the right clinic that contributes to reaching the largest possible number of patients likely to have a stroke.

- **Q\What are the most common diseases that associated with stroke?**

Data Structure:

- The Dataset is form Kaggle (Stroke Prediction Dataset) .
- It is consisted of 12 columns and 5111 rows.

- df.head(3)

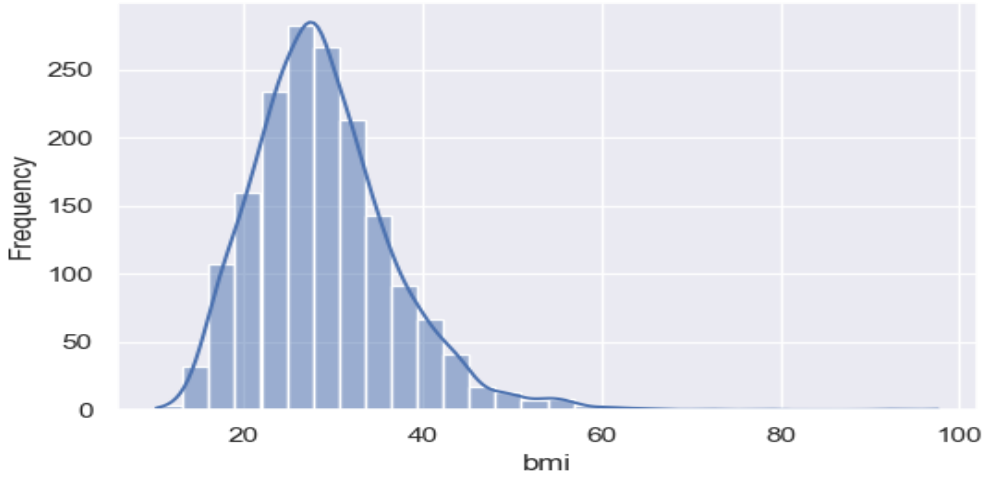
	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1

Data Cleaning:

- We have around 200 missing values in BMI column.

```
9   bmi                 4909 non-null   float64
```

- First, We want to see the distribution of BMI to decide by which we will Replace the NaN values. mean or median ?



- It is Skewed distribution , so the median is the best.
- The median = 28.1
- After cleaning: df.head(3)

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke	BMI_type
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1	Extreme Obesity
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	28.1	never smoked	1	Overweight
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1	Obesity

Data Analysis:

- Correlation is not the proper way to our project because most of data is categorical data not a numerical .
- The best way is counting patients by each disease with stroke using mask method.
- For example, when we want to count the number of patients who have hypertension and stroke:

```
mask_hypertension_Stroke = (df.hypertension == 1) & (df.stroke == 1)

df[mask_hypertension_Stroke].shape[0]
print(f"The Total Number of patient that have hypertension and stroke in same time is {df[mask_hypertension_Stroke].shape[0]}")
```

```
The Total Number of patient that have hypertension and stroke in same time is 66
```

- We will apply this method to all diseases to count the total number of patients in each.

Data Analysis (part2):

- In BMI , we will generate new column based on BMI type as following picture:

- df.sample(3)

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke	BMI_type
3778	63729	Female	19.0	0	0	No	Private	Urban	65.79	28.6	smokes	0	Overweight
4363	43138	Male	15.0	0	0	No	Private	Urban	55.79	21.3	never smoked	0	Normal weight
428	59906	Female	40.0	0	0	Yes	Private	Rural	139.90	31.7	smokes	0	Obesity

- After that we are focusing on obesity and extreme obesity and do mask also.

Data Analysis (part3):

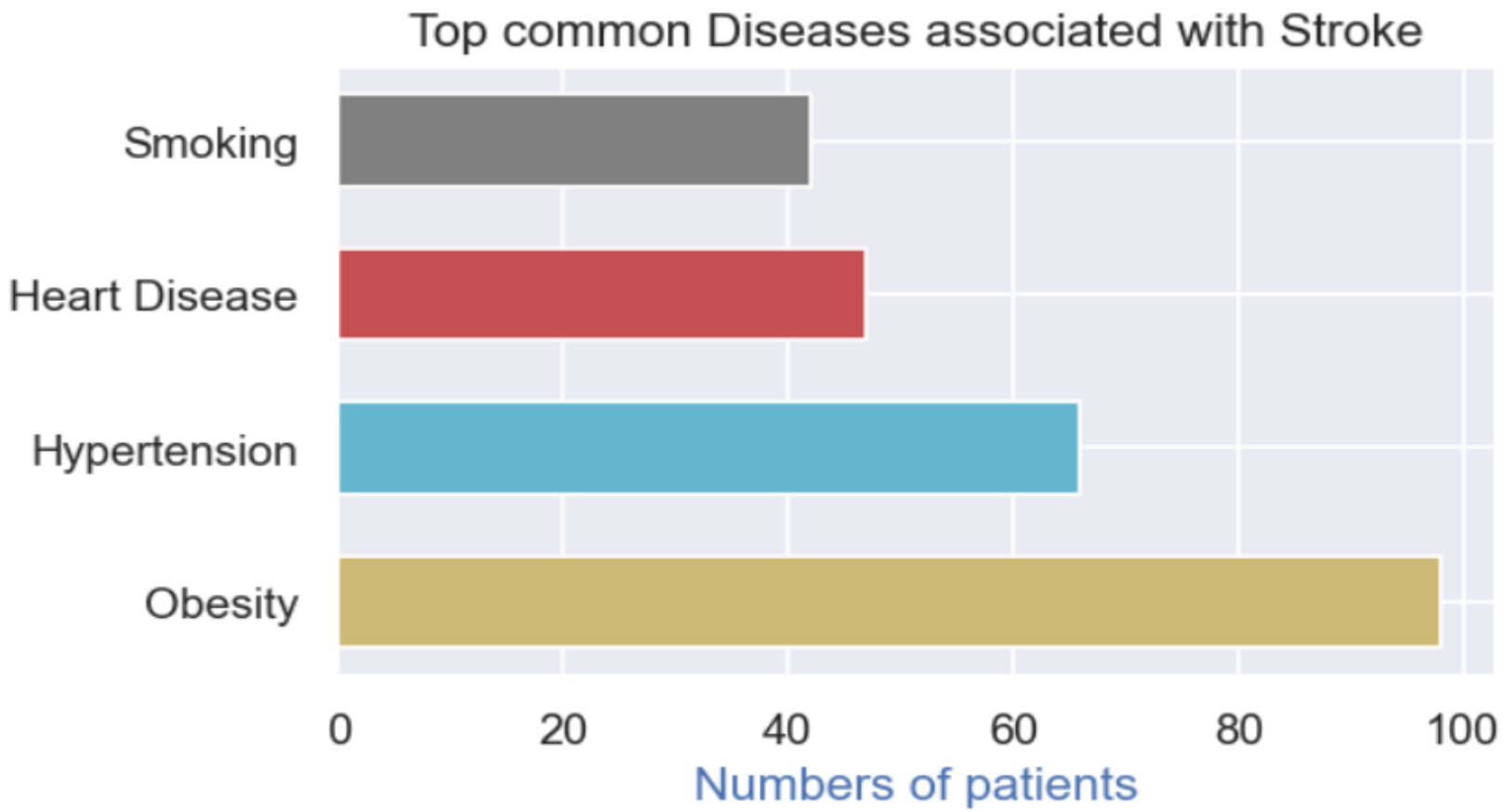
- In heart disease, we follow same technique as hypertension

- In smoking status , we are focusing just on smokers' patients, without patients who was formerly smoked .

- Last, average glucose level we can't decide if the patient is diabetic or not. why? there is important missing information which is when the test was taken and is the patient fasting or after meal ?

The result:

- This is bar graph showing the top common diseases associated with stroke



Conclusion:

- Finally, Stroke is a life-threatening condition that is caused by a lack of blood and oxygen flow to the brain cells.

- We can order the diseases that associated with stroke based on analysis phase as follow:
- 1- Obesity.
- 2- Hypertension.
- 3- Heart Disease.
- 4- Smoking.

- In conclusion, the most disease that can cause the stroke is obesity.