Healthcare Analysis

Problem: The World Health Organization (WHO) is committed to preventing disease and promoting better global health. The organization seeks to improve its understanding of stroke risk factors and trends.

Approach: To meet the World Health organization goal, I will use SQL to extract insights to predict whether a person is likely to get stroke based on gender, age, various diseases, and smoking status.

Attribute Information 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

1. High risk groups

```
SELECT gender, age, work_type, COUNT(*) as stroke_count

FROM stroke_prediction

WHERE stroke = 1

GROUP BY gender, age, work_type
```

```
ORDER BY stroke_count DESC;
```

	gender	age	work_type	stroke_count
0	Female	79.0	Private	9
1	Female	78.0	Private	7
2	Male	78.0	Self-employed	7
3	Female	80.0	Self-employed	5
4	Male	58.0	Private	5
...	...	...	...	...
128	Male	73.0	Govt_job	1
129	Male	73.0	Self-employed	1
130	Male	74.0	Self-employed	1
131	Male	76.0	Self-employed	1
132	Male	77.0	Self-employed	1

133 rows × 4 columns

2. Geographical

```
SELECT Residence_type, COUNT(*) as stroke_count

FROM stroke_prediction

WHERE stroke = 1

GROUP BY Residence_type;
```

Residence_type	stroke_count
Rural	114
Urban	135

3. **Stroke based on gender.**

```
SELECT gender,
COUNT(*) as gender_strokes
FROM stroke_prediction
WHERE stroke = 1
GROUP BY gender;
```

gender	gender_strokes
Female	141
Male	108

4. **Stroke based on work type.**

```
SELECT work_type,
COUNT(*) as work_type_strokes
FROM stroke_prediction
WHERE stroke = 1
GROUP BY work_type;
```

work_type	work_type_strokes
Govt_job	33
Private	149
Self-employed	65
children	2

5. **Average BMI of individuals who had a stroke based on gender.**

```
SELECT gender, AVG(bmi) as avg_bmi_stroke

FROM stroke_prediction

WHERE stroke = 1

GROUP BY gender;
```

gender	avg_bmi_stroke
Female	30.218333
Male	30.812360

6. **Percentage of Individuals with Hypertension who had a stroke based on gender.**

```
SELECT gender, hypertension, COUNT(*) as hypertension_strokes,
```

```
Round((COUNT(*) * 100.0 / SUM(COUNT(*)) OVER (PARTITION BY gender)),2) as percentage

FROM stroke_prediction

WHERE stroke = 1

GROUP BY gender, hypertension;
```

gender	hypertension	hypertension_strokes	percentage
Female	0	102	72.34
Female	1	39	27.66
Male	0	81	75.00
Male	1	27	25.00

7. **Average glucose level and BMI of individuals who had a stroke based on smoking status.**

```
SELECT smoking_status,

Round(AVG(avg_glucose_level),1) as avg_glucose_stroke,

AVG(bmi) as avg_bmi_stroke

FROM stroke_prediction

WHERE stroke = 1

GROUP BY smoking_status;
```

smoking_status	avg_glucose_stroke	avg_bmi_stroke
Unknown	124.8	29.351724
formerly smoked	132.8	31.014035
never smoked	132.9	30.439286
smokes	140.0	30.579487

8. **Stroke percentage for each age group**

```
SELECT
 CASE
  WHEN age BETWEEN 0 AND 9 THEN '0-9'
  WHEN age BETWEEN 10 AND 19 THEN '10-19'
  WHEN age BETWEEN 20 AND 29 THEN '20-29'
  WHEN age BETWEEN 30 AND 39 THEN '30-39'
  WHEN age BETWEEN 40 AND 49 THEN '40-49'
  WHEN age BETWEEN 50 AND 59 THEN '50-59'
  WHEN age BETWEEN 60 AND 69 THEN '60-69'
  WHEN age BETWEEN 70 AND 79 THEN '70-79'
  WHEN age BETWEEN 80 AND 89 THEN '80-89'        ELSE '90+'    END AS age_group,
round((COUNT(*) * 100.0 / (SELECT COUNT(*) FROM stroke_prediction WHERE stroke = 1)),1) as
percentage_strokes  FROM stroke_prediction
```

```
WHERE stroke = 1

GROUP BY age_group

ORDER BY age_group;
```

age_group	percentage_strokes
0-9	0.4
10-19	0.4
30-39	2.4
40-49	4.8
50-59	19.3
60-69	18.9
70-79	37.8
80-89	16.1

9. **Distribution of Stroke Cases Based on Marital Status**

```
SELECT   ever_married,

 COUNT(*) as stroke_count

FROM stroke_prediction

WHERE stroke = 1 GROUP BY ever_married order by ever_married desc;
```

ever_married	stroke_count
Yes	220
No	29

10. **Glucose Level and BMI for Individuals Who Had a Stroke based on residence and gender.**

```
SELECT

 Residence_type,gender,

 Round(AVG(avg_glucose_level),2) as avg_glucose_stroke,

 Round(AVG(bmi),2) as avg_bmi_stroke

FROM stroke_prediction

WHERE stroke = 1

GROUP BY Residence_type, gender;
```

Residence_type	gender	avg_glucose_stroke	avg_bmi_stroke
Rural	Female	119.90	29.19
Rural	Male	143.28	31.24
Urban	Female	128.16	31.15
Urban	Male	143.06	30.41