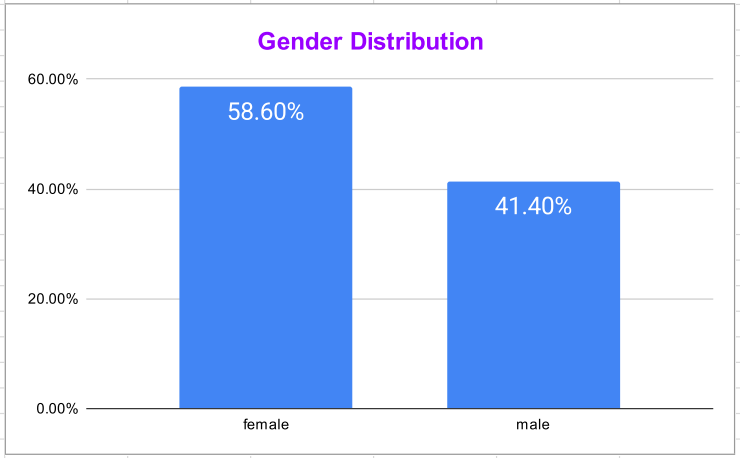
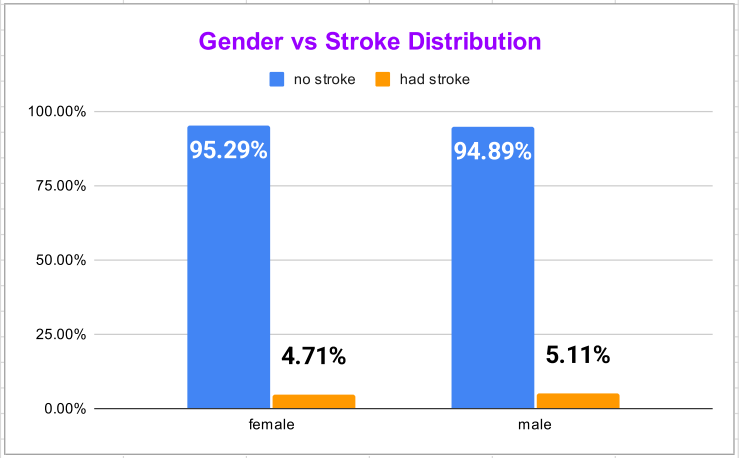
Gender Distribution Analysis

	stroke category	Values							
	no stroke			had stroke			Grand Total		
gender	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
female	2853	95.29%	58.70%	141	4.71%	56.63%	2994	100.00%	58.60%
male	2007	94.89%	41.30%	108	5.11%	43.37%	2115	100.00%	41.40%
Grand Total	**4860**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5109**	**100.00%**	**100.00%**



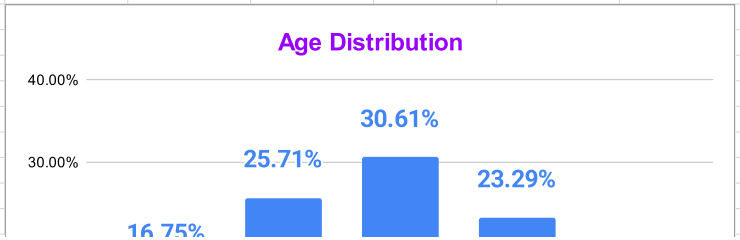


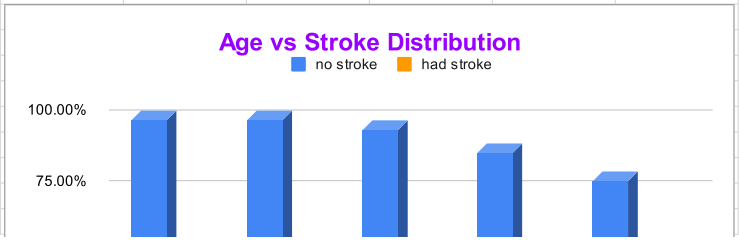
Insights

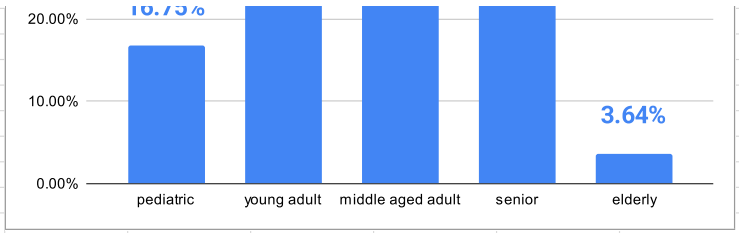
1. The female patients (58.60%) outnumber the male patients (41.40%) in this dataset.
2. The bar chart visualization shows that male patients' risk of having a stroke is 5.11% while the female patients have a 4.71% risk of having a stroke. The values are very similar when comparing the probability of having a stroke.

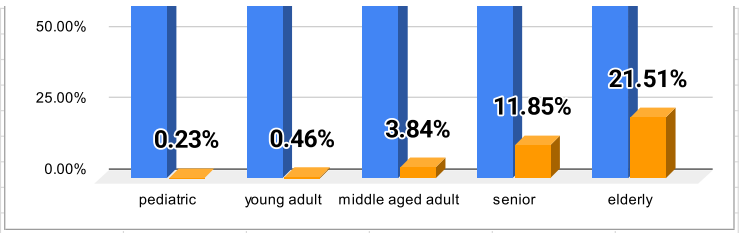
Age Distribution Analysis

		stroke category	Values							
		no stroke			had stroke			Grand Total		
age helper	age	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
1	pediatric	854	99.77%	17.57%	2	0.23%	0.80%	856	100.00%	16.75%
2	young adult	1308	99.54%	26.91%	6	0.46%	2.41%	1314	100.00%	25.71%
3	middle aged adu	1504	96.16%	30.94%	60	3.84%	24.10%	1564	100.00%	30.61%
4	senior	1049	88.15%	21.58%	141	11.85%	56.63%	1190	100.00%	23.29%
5	elderly	146	78.49%	3.00%	40	21.51%	16.06%	186	100.00%	3.64%
Grand Total		**4861**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5110**	**100.00%**	**100.00%**







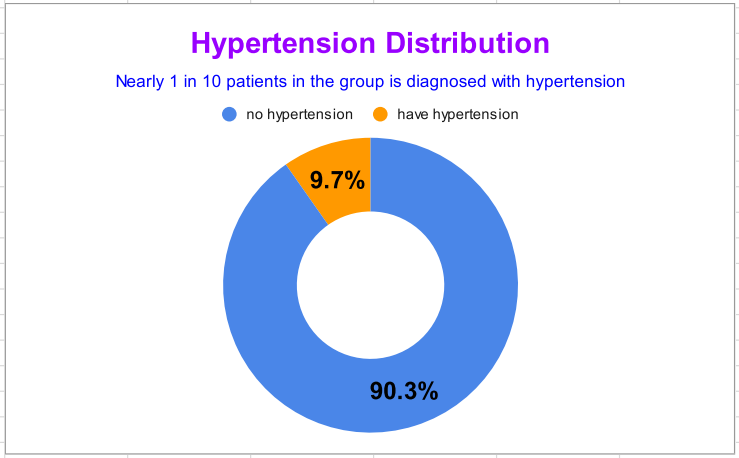


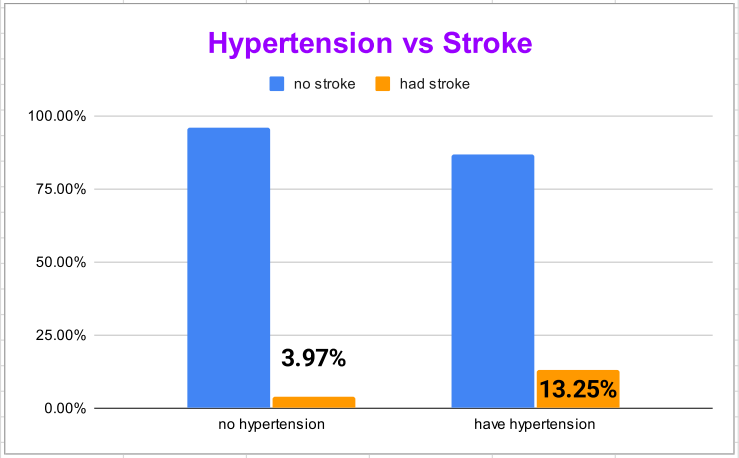
Insights

* The patient distribution according to age category follows a fairly normal distribution with the exception of elderly patients that make up 3.64% of the data.
* The highest number of patients belong to the middle aged adult bracket, which makes up 30.61% of the data.
* The bar chart visualization of the age vs stroke distribution shows that the risk of having a heart attack increases as the patient ages.
* Pediatric risk of stroke is 0.23% while elderly stroke risk is highest at 21.51%, which is almost twice as much as seniors.
* Note that the number of elderly patients are only 3.64% of the data, therefore the smallest movements in occurence of stroke can make significant changes.

Hypertension Distribution Analysis

	stroke category	*Values*							
	no stroke			had stroke			Grand Total		
hypertension	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
no hypertension	4429	96.03%	91.11%	183	3.97%	73.49%	4612	100.00%	90.25%
have hypertension	432	86.75%	8.89%	66	13.25%	26.51%	498	100.00%	9.75%
Grand Total	**4861**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5110**	**100.00%**	**100.00%**



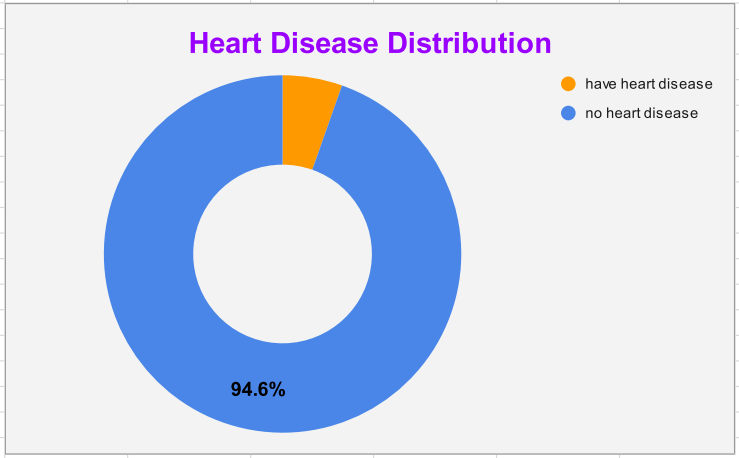


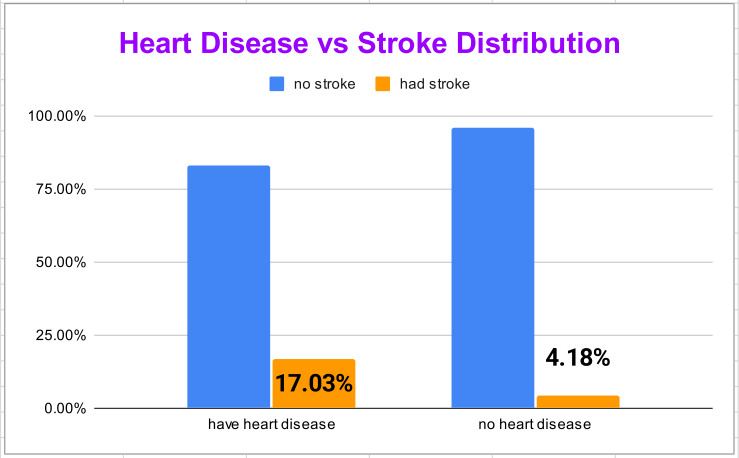
Insights

* Nearly 1 in 10 patients have been diagnosed with hypertension, this represents a heavily imbalanced data distribution.
* The risk of having a stroke for patients with hypertension is 13.25%, which is nearly four times as much as patients without hypertension (3.97%).

Heart Disease Distribution Analysis

	stroke category	Values							
	no stroke			had stroke			Grand Total		
heart disease	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
have heart disease	229	82.97%	4.71%	47	17.03%	18.88%	276	100.00%	5.40%
no heart disease	4632	95.82%	95.29%	202	4.18%	81.12%	4834	100.00%	94.60%
Grand Total	**4861**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5110**	**100.00%**	**100.00%**



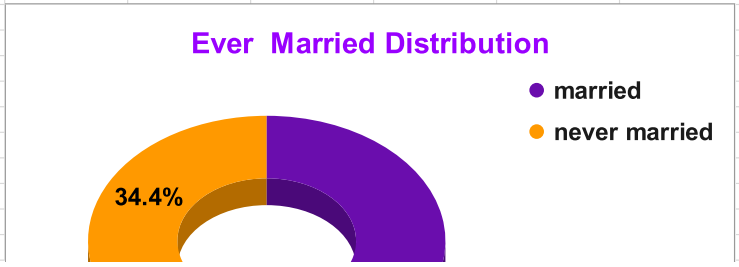


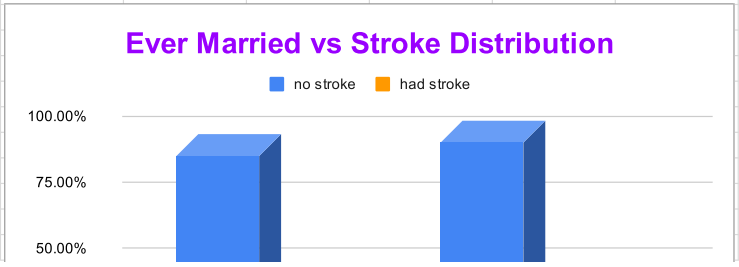
Insights

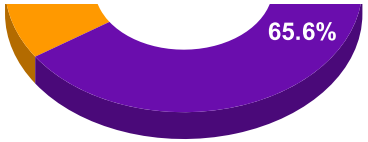
* Only 5.4% of the data represents patients with heart disease.
* Patients with heart disease posts a risk of 17.03% of having a stroke - this is the highest risk percentage that we have observed so far.
* Patients with heart disease is a little more than four times as likely to have a stroke than patients with no heart disease (4.18%)

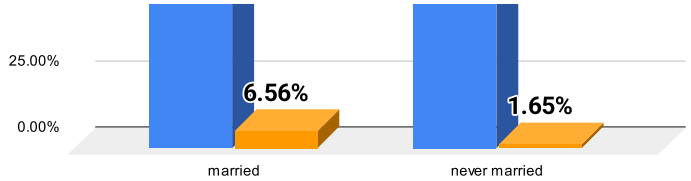
Ever Married Distribution Analysis

	stroke category	Values							
	no stroke			had stroke			Grand Total		
ever married	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
married	3133	93.44%	64.45%	220	6.56%	88.35%	3353	100.00%	65.62%
never married	1728	98.35%	35.55%	29	1.65%	11.65%	1757	100.00%	34.38%
Grand Total	**4861**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5110**	**100.00%**	**100.00%**







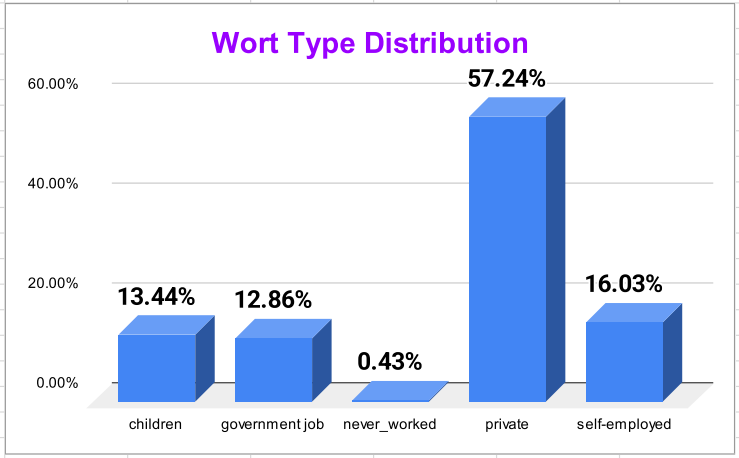


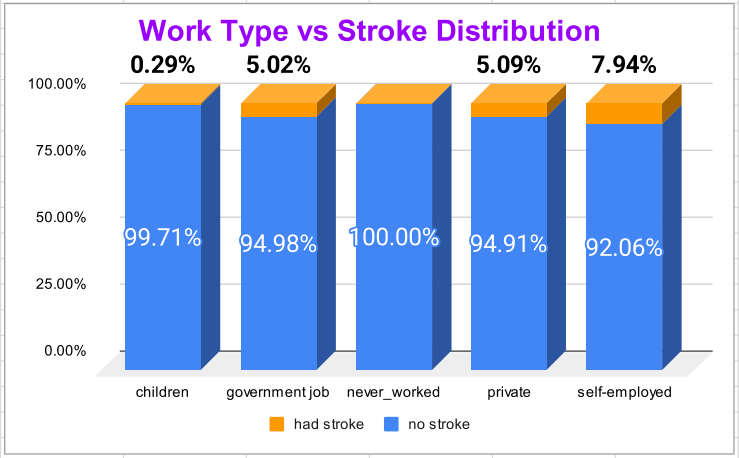
Insights

* Married and once married patients make up **65.6%** of the data, which is **almost under twice** as much as the never married patients (**34.4%**)
* Married and once married patients have a **6.56%** likelihood of having a stroke while only patients that were never married only have a **1.65%** risk of having a stroke - around four times less likely to have a stroke.

Work Type Distribution Analysis

	stroke category	*Values*							
	no stroke			had stroke			Grand Total		
work type	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
children	685	99.71%	14.09%	2	0.29%	0.80%	687	100.00%	13.44%
government job	624	94.98%	12.84%	33	5.02%	13.25%	657	100.00%	12.86%
never_worked	22	100.00%	0.45%				22	100.00%	0.43%
private	2776	94.91%	57.11%	149	5.09%	59.84%	2925	100.00%	57.24%
self-employed	754	92.06%	15.51%	65	7.94%	26.10%	819	100.00%	16.03%
Grand Total	**4861**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5110**	**100.00%**	**100.00%**



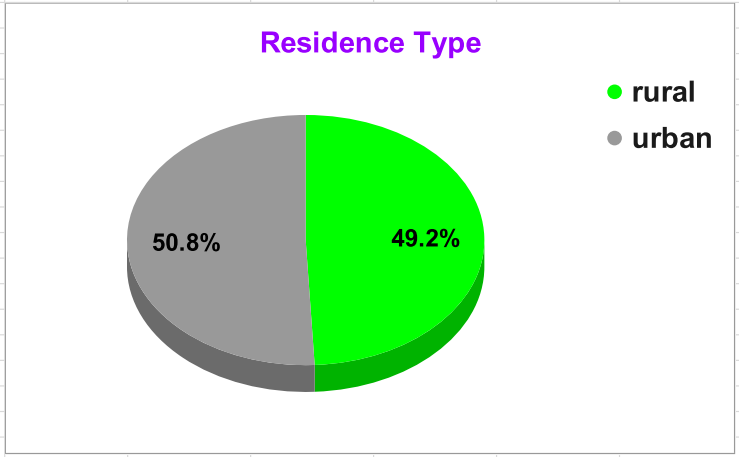


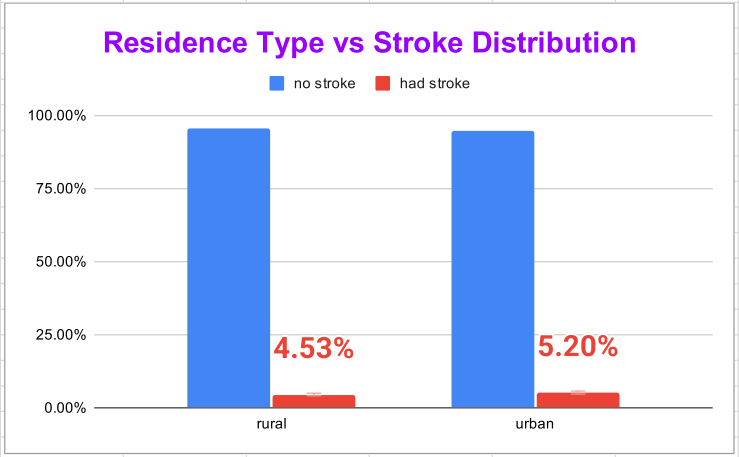
Insights

* 57.24% of the data is composed of patients that worked in the private sector.
* The patients that never worked makeup 0.43% of the dataset, this is not significant enough to make affirmative conclusions about the risk factors of this category in the group.
* Patients that are self-employed carries the highest risk factor of 7.94%, but this is close to the risk factor of patients taht work government jobs (5.02%) and private company jobs (5.09%) -- which suggests that work type may not be a strong determining factor of stroke risk. Furtehr analysis required to confirm its statistical significance.

Residence Type Distribution Analysis

	stroke category	Values							
	no stroke			had stroke			Grand Total		
residence type	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
rural	2400	95.47%	49.37%	114	4.53%	45.78%	2514	100.00%	49.20%
urban	2461	94.80%	50.63%	135	5.20%	54.22%	2596	100.00%	50.80%
Grand Total	**4861**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5110**	**100.00%**	**100.00%**



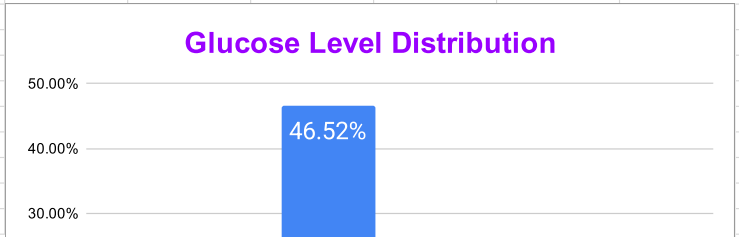


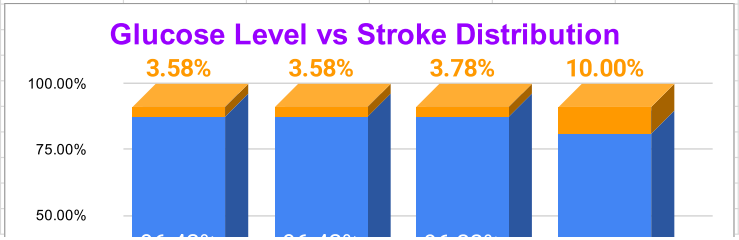
Insights

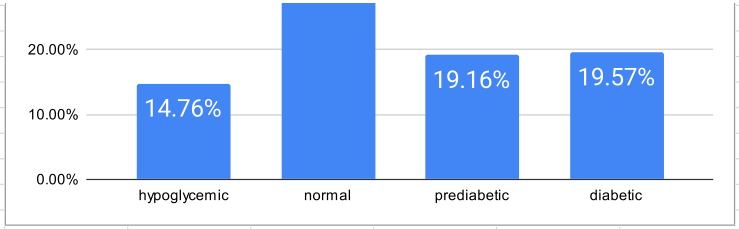
* The dataset is about evenly distributed (50.8% urban vs 49.2% rural) in terms of residence type.
* Patients that lives in urban areas have a slightly higher risk of having a stroke at 5.20% while patients that live in rural areas have a 4.53% chance of having a stroke.

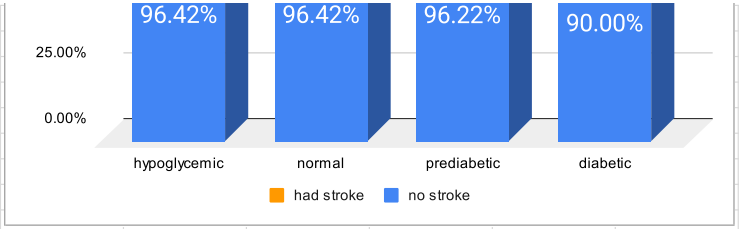
Glucose Level Distribution Analysis

		stroke category	Values							
		no stroke			had stroke			Grand Total		
glucose helper	glucose category	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
1	hypoglycemic	727	96.42%	14.96%	27	3.58%	10.84%	754	100.00%	14.76%
2	normal	2292	96.42%	47.15%	85	3.58%	34.14%	2377	100.00%	46.52%
3	prediabetic	942	96.22%	19.38%	37	3.78%	14.86%	979	100.00%	19.16%
4	diabetic	900	90.00%	18.51%	100	10.00%	40.16%	1000	100.00%	19.57%
Grand Total		**4861**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5110**	**100.00%**	**100.00%**







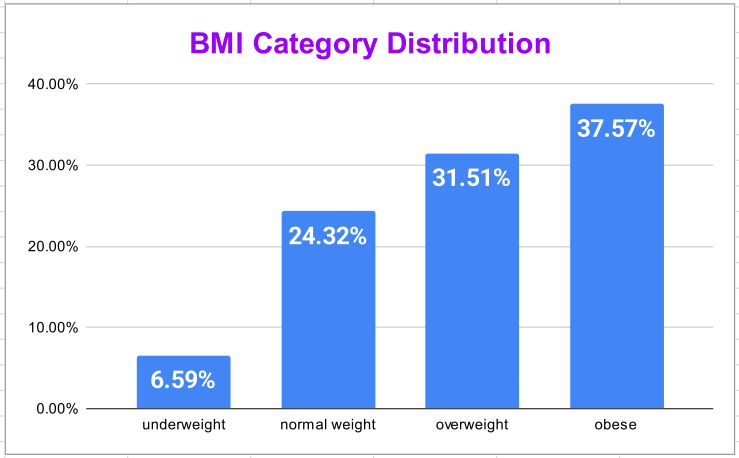


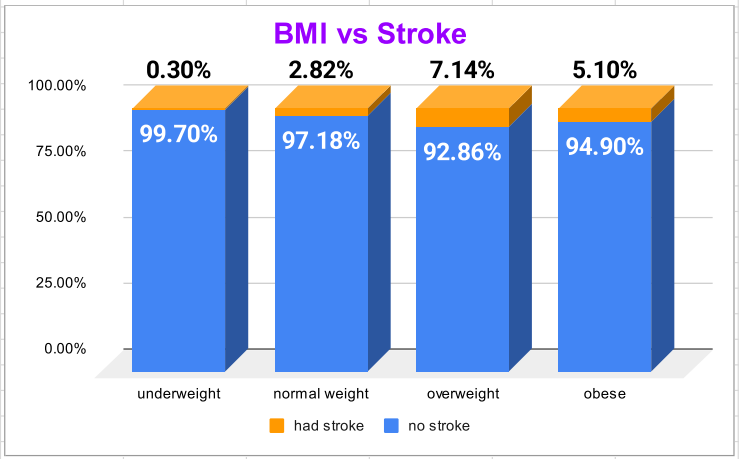
Insights

* Patients with normal glucose level has the highest composition at 46.52% of the data, while the other three categories are nearly at the same level.
* The most at risk patients of having a stroke are patients with diabetes (10%) which is just under three times the risk of hypoglycemic (3.58%), normal (3.58%), and prediabetic (3.78) patients.

BMI Category Distribution Analysis

		stroke category	*Values*							
		no stroke			had stroke			Grand Total		
bmi helper	*bmi category*	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
1	underweight	336	99.70%	6.91%	1	0.30%	0.40%	337	100.00%	6.59%
2	normal weight	1208	97.18%	24.85%	35	2.82%	14.06%	1243	100.00%	24.32%
3	overweight	1495	92.86%	30.75%	115	7.14%	46.18%	1610	100.00%	31.51%
4	obese	1822	94.90%	37.48%	98	5.10%	39.36%	1920	100.00%	37.57%
Grand Total		**4861**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5110**	**100.00%**	**100.00%**



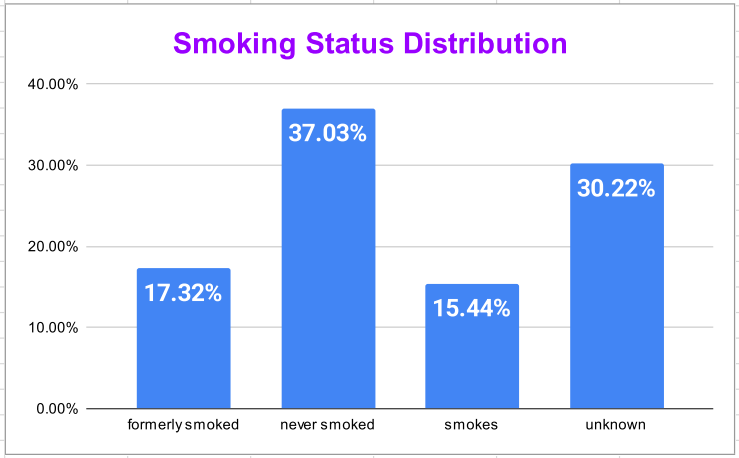


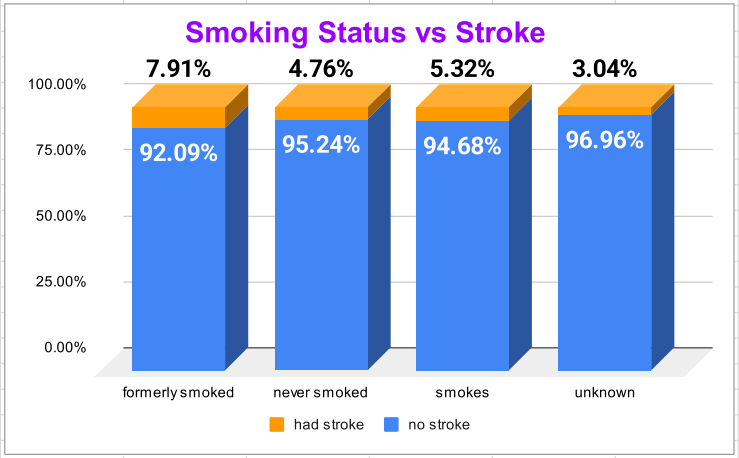
Insights

* The highest distribution for bmi category is pateints that are in the obese level with 37.57% risk of having a stroke. The distirbution is left skewed with underweight patiens representing 6.59%
* The stacked bar graph shows that patients in the overweight category (7.14%) have the highest risk of having a stroke, while patients i nthe obese level have a 5.10% chance of having a stroke - this is interesting as one would expect it to be the
* Further data gathering and study is necessary to uncover the reasons why overweight people have a higher risk of stroke compared to people in the overweight category.

Smoking Status Distribution Analysis

	stroke category	Values							
	no stroke			had stroke			Grand Total		
smoking status	Count	Row %	Column %	Count	Row %	Column %	Count	Row %	Column %
formerly smoked	815	92.09%	16.77%	70	7.91%	28.11%	885	100.00%	17.32%
never smoked	1802	95.24%	37.07%	90	4.76%	36.14%	1892	100.00%	37.03%
smokes	747	94.68%	15.37%	42	5.32%	16.87%	789	100.00%	15.44%
unknown	1497	96.96%	30.80%	47	3.04%	18.88%	1544	100.00%	30.22%
Grand Total	**4861**	**95.13%**	**100.00%**	**249**	**4.87%**	**100.00%**	**5110**	**100.00%**	**100.00%**





Insights

* Patienst that never smoked (37.03%) have the highest distribution in the dataset.
* Patients with the unknown smoking status comprise of 30.22% of the dataset. This is a significant portion of the data and should not simply be removed from the dataset.
* People that formerly smoked have the highest risk of having a stroke at 7.91% while people that smokes have a 5.32% chance of having a stroke.
* The risk of having a stroke for people that never smoked (4.76%) and people that smokes (5.32%) is very similar.