The Chances of a Person Having a Stroke

What is a Stroke?

A Stroke is defined as 'a sudden disabling attack or loss of consciousness caused by an interruption in the flow of blood to the brain, especially through thrombosis.'

Stroke is a leading cause of death and disability, causing around 38,000 deaths each year in the UK.

There are around 100,000 strokes every year in the UK

126,000 hospital admissions in England per year

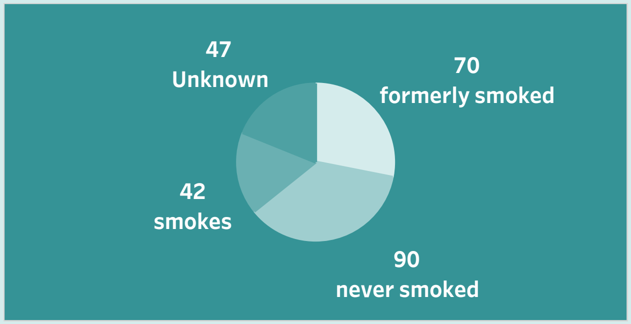
In the UK there are approximately 1.3 million people living with stroke

Presented by -
Hanna, Issa, Vera and Violetta

The Chances of a Person Having a Stroke

Stroke Statistics

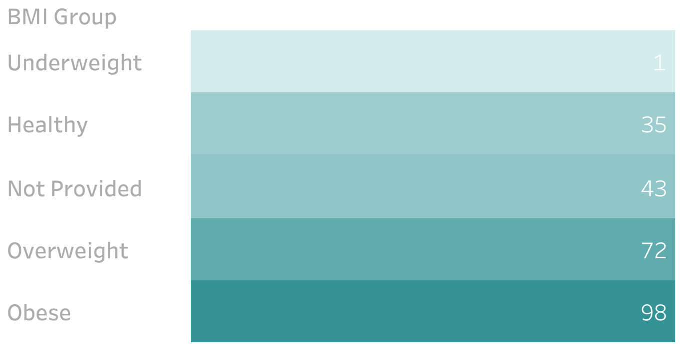
Has the victim ever smoked?



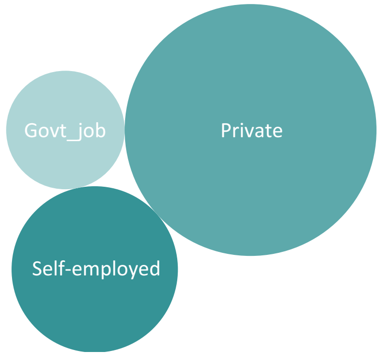
4.87%
% Stroke Victims

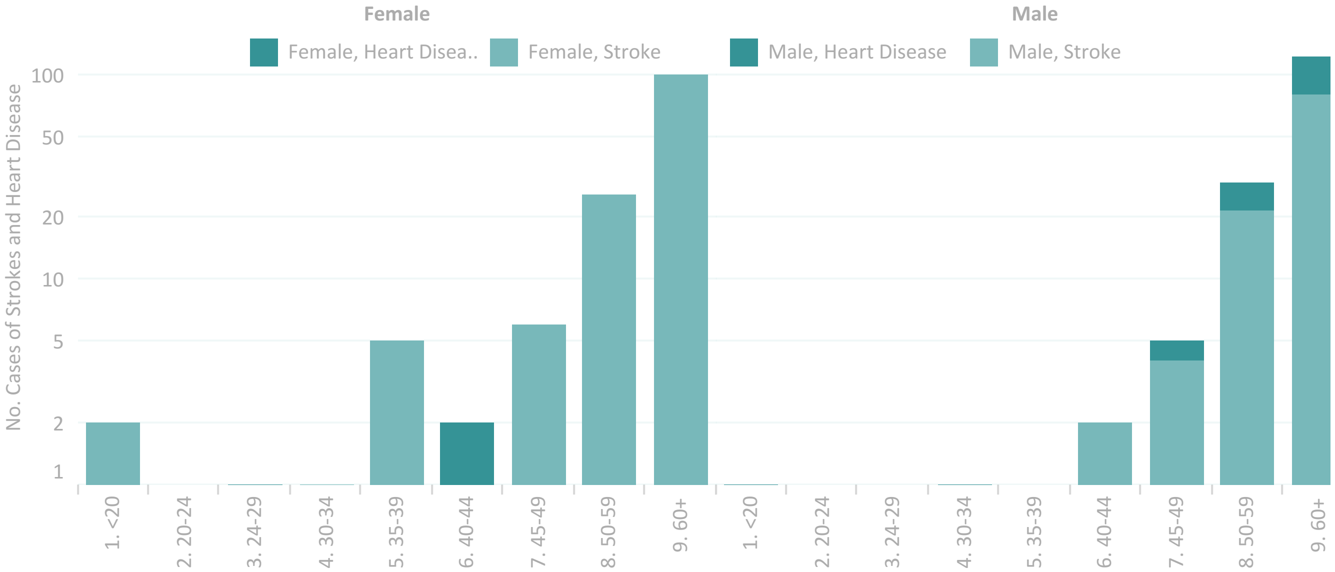
249
No. Strokes Victims

67.73
Avg. Age for a Stroke



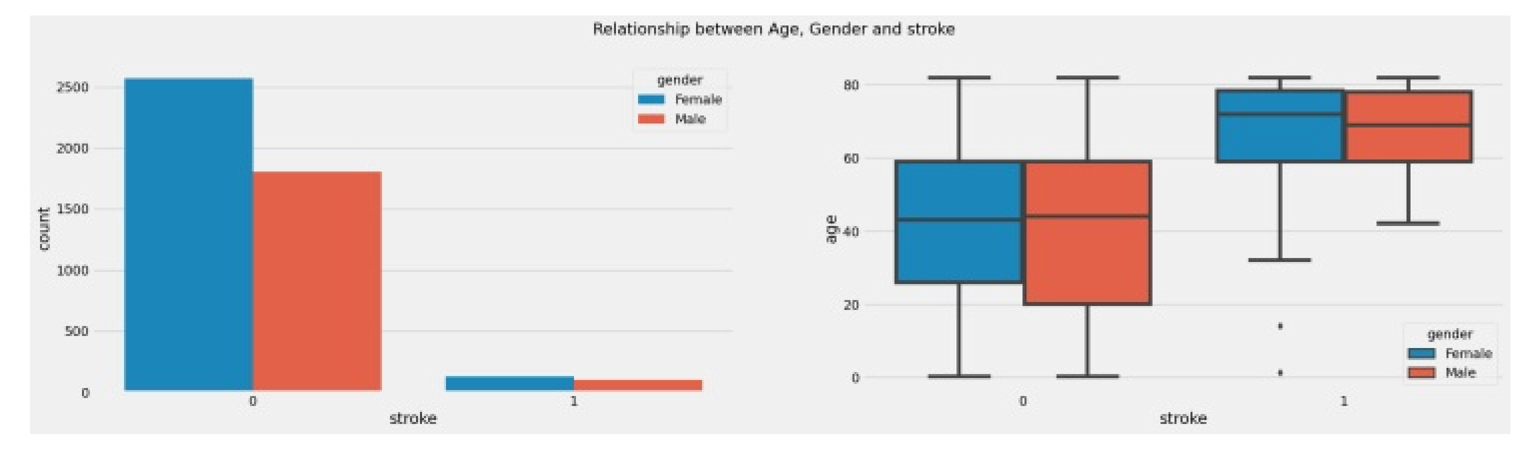
Work Type of Stroke Victims

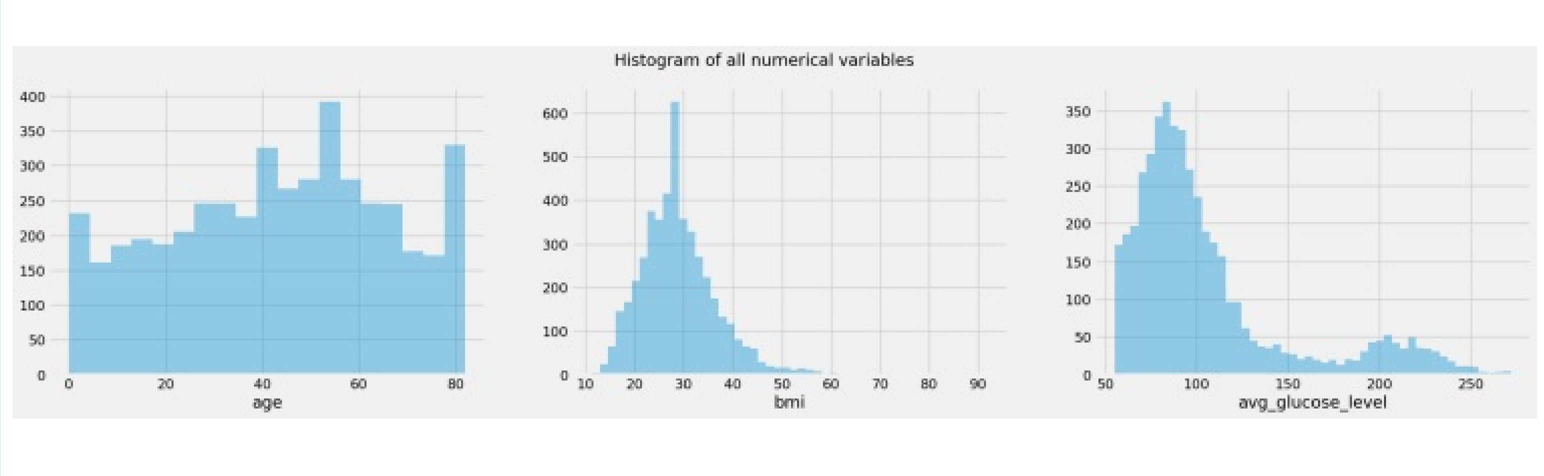




The Chances of a Person Having a Stroke

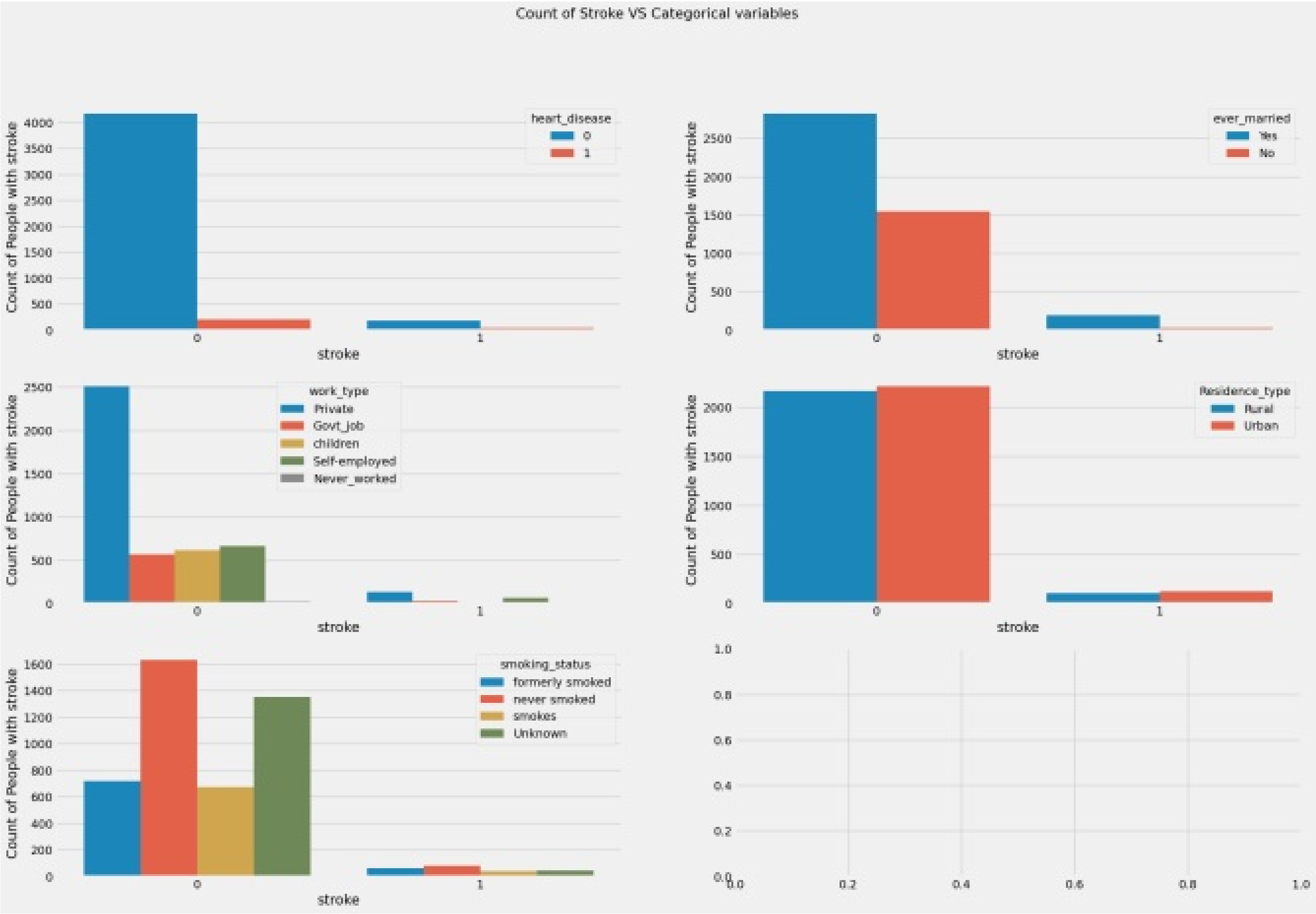
Understanding the Data - Part 1





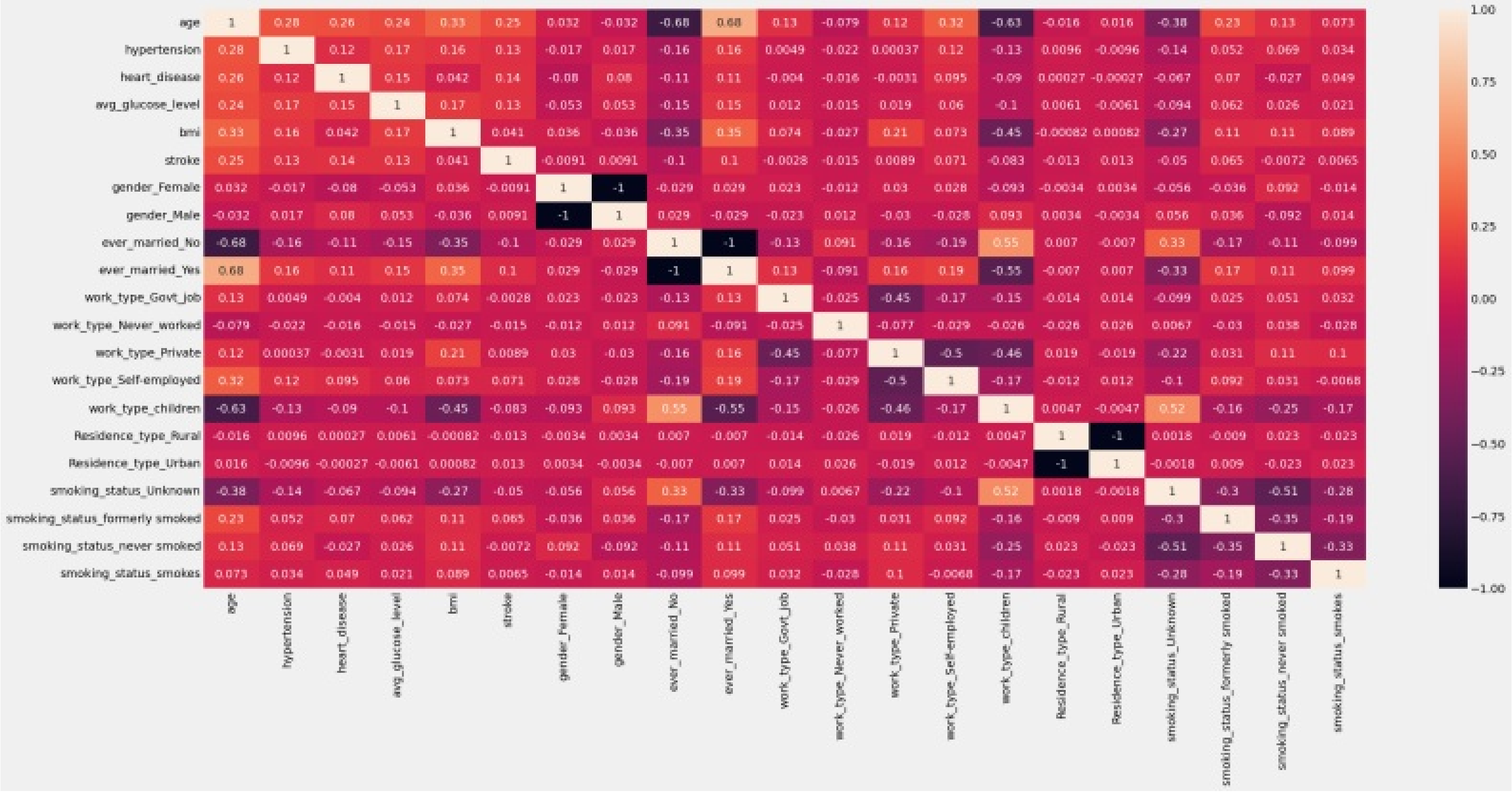
The Chances of a Person Having a Stroke

Understanding the Data - Part 2



The Chances of a Person Having a Stroke

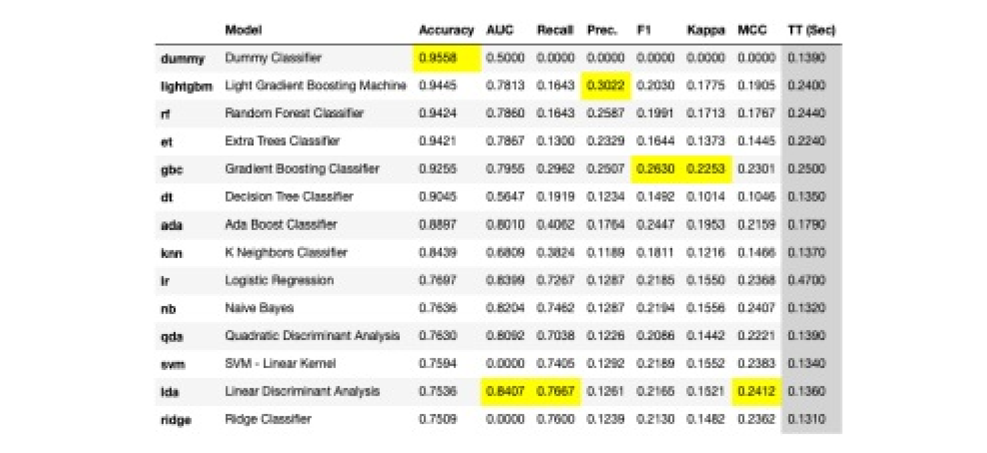
Understanding the Data - Part 3



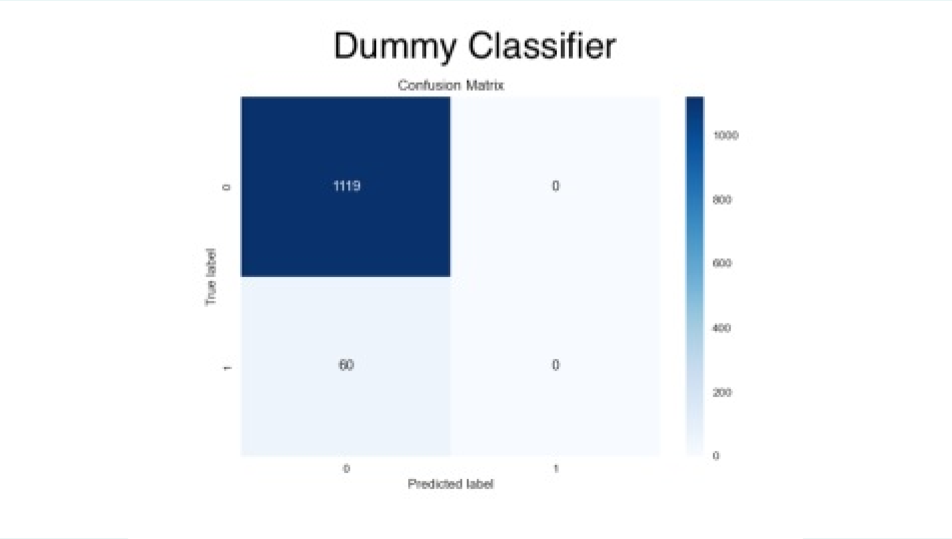
	age	hypertension	heart_disease	avg_glucose_level	bmi	stroke	gender_Female	gender_Male	ever_married_No	ever_married_Yes	work_type_Govt_job	work_type_Never_worked	work_type_Private	work_type_Self-employed	work_type_children	Residence_type_Rural	Residence_type_Urban	smoking_status_Unknown	smoking_status_formerly smoked	smoking_status_never smoked	smoking_status_smokes
age	1	0.28	0.26	0.24	0.33	0.25	0.032	-0.032	-0.68	0.68	0.13	-0.079	0.12	0.32	-0.63	-0.016	0.016	-0.38	0.23	0.13	0.073
hypertension	0.28	1	0.12	0.17	0.16	0.13	-0.017	0.017	-0.16	0.16	0.0049	-0.022	0.00037	0.12	-0.13	0.0096	-0.0096	-0.14	0.052	0.069	0.034
heart_disease	0.26	0.12	1	0.15	0.042	0.14	-0.08	0.08	-0.11	0.11	-0.004	-0.016	-0.0031	0.095	-0.09	0.00027	-0.00027	-0.067	0.07	-0.027	0.049
avg_glucose_level	0.24	0.17	0.15	1	0.17	0.13	-0.053	0.053	-0.15	0.15	0.012	-0.015	0.019	0.06	-0.1	0.0061	-0.0061	-0.094	0.062	0.026	0.021
bmi	0.33	0.16	0.042	0.17	1	0.041	0.036	-0.036	-0.35	0.35	0.074	-0.027	0.21	0.073	-0.45	-0.00082	0.00082	-0.27	0.11	0.11	0.089
stroke	0.25	0.13	0.14	0.13	0.041	1	-0.0091	0.0091	-0.1	0.1	-0.0028	-0.015	0.0089	0.071	-0.083	-0.013	0.013	-0.05	0.065	-0.0072	0.0065
gender_Female	0.032	-0.017	-0.08	-0.053	0.036	-0.0091	1	-1	-0.029	0.029	0.023	-0.012	0.03	0.028	-0.093	-0.0034	0.0034	-0.056	-0.036	0.092	-0.014
gender_Male	-0.032	0.017	0.08	0.053	-0.036	0.0091	-1	1	0.029	-0.029	-0.023	0.012	-0.03	-0.028	0.093	0.0034	-0.0034	0.056	0.036	-0.092	0.014
ever_married_No	-0.68	-0.16	-0.11	-0.15	-0.35	-0.1	-0.029	0.029	1	-1	-0.13	0.091	-0.16	-0.19	0.55	0.007	-0.007	0.33	-0.17	-0.11	-0.099
ever_married_Yes	0.68	0.16	0.11	0.15	0.35	0.1	0.029	-0.029	-1	1	0.13	-0.091	0.16	0.19	-0.55	-0.007	0.007	-0.33	0.17	0.11	0.099
work_type_Govt_job	0.13	0.0049	-0.004	0.012	0.074	-0.0028	0.023	-0.023	-0.13	0.13	1	-0.025	-0.45	-0.17	-0.15	-0.014	0.014	-0.099	0.025	0.051	0.032
work_type_Never_worked	-0.079	-0.022	-0.016	-0.015	-0.027	-0.015	-0.012	0.012	0.091	-0.091	-0.025	1	-0.077	-0.029	-0.026	-0.026	0.026	0.0067	-0.03	0.038	-0.028
work_type_Private	0.12	0.00037	-0.0031	0.019	0.21	0.0089	0.03	-0.03	-0.16	0.16	-0.45	-0.077	1	-0.5	-0.46	0.019	-0.019	-0.22	0.031	0.11	0.1
work_type_Self-employed	0.32	0.12	0.095	0.06	0.073	0.071	0.028	-0.028	-0.19	0.19	-0.17	-0.029	-0.5	1	-0.17	-0.012	0.012	-0.1	0.092	0.031	-0.0068
work_type_children	-0.63	-0.13	-0.09	-0.1	-0.45	-0.083	-0.093	0.093	0.55	-0.55	-0.15	-0.026	-0.46	-0.17	1	0.0047	-0.0047	0.52	-0.16	-0.25	-0.17
Residence_type_Rural	-0.016	0.0096	0.00027	0.0061	-0.00082	-0.013	-0.0034	0.0034	0.007	-0.007	-0.014	-0.026	0.019	-0.012	0.0047	1	-1	0.0018	-0.009	0.023	-0.023
Residence_type_Urban	0.016	-0.0096	-0.00027	-0.0061	0.00082	0.013	0.0034	-0.0034	-0.007	0.007	0.014	0.026	-0.019	0.012	-0.0047	-1	1	-0.0018	0.009	-0.023	0.023
smoking_status_Unknown	-0.38	-0.14	-0.067	-0.094	-0.27	-0.05	-0.056	0.056	0.33	-0.33	-0.099	0.0067	-0.22	-0.1	0.52	0.0018	-0.0018	1	-0.3	-0.51	-0.28
smoking_status_formerly smoked	0.23	0.052	0.07	0.062	0.11	0.065	-0.036	0.036	-0.17	0.17	0.025	-0.03	0.031	0.092	-0.16	-0.009	0.009	-0.3	1	-0.35	-0.19
smoking_status_never smoked	0.13	0.069	-0.027	0.026	0.11	-0.0072	0.092	-0.092	-0.11	0.11	0.051	0.038	0.11	0.031	-0.25	0.023	-0.023	-0.51	-0.35	1	-0.33
smoking_status_smokes	0.073	0.034	0.049	0.021	0.089	0.0065	-0.014	0.014	-0.099	0.099	0.032	-0.028	0.1	-0.0068	-0.17	-0.023	0.023	-0.28	-0.19	-0.33	1

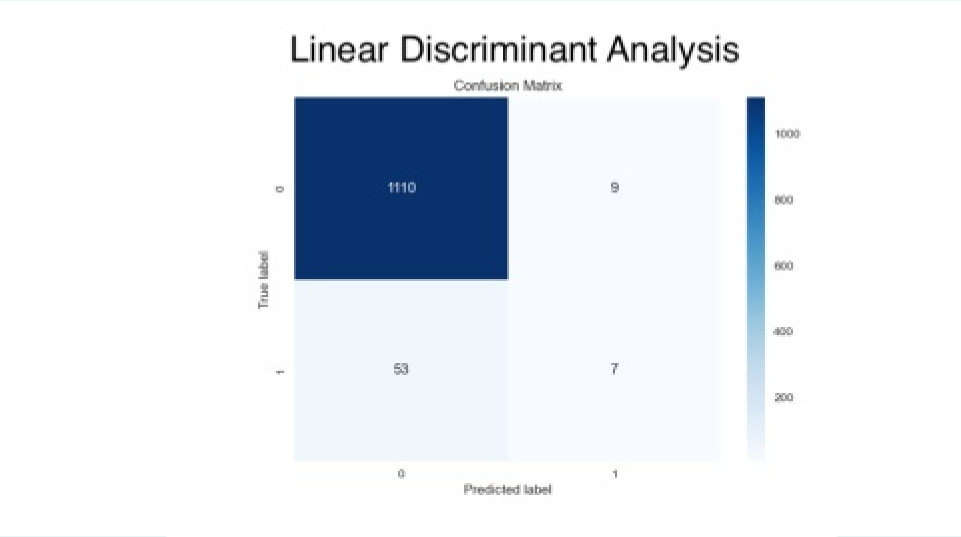
The Chances of a Person Having a Stroke

Machine Learning



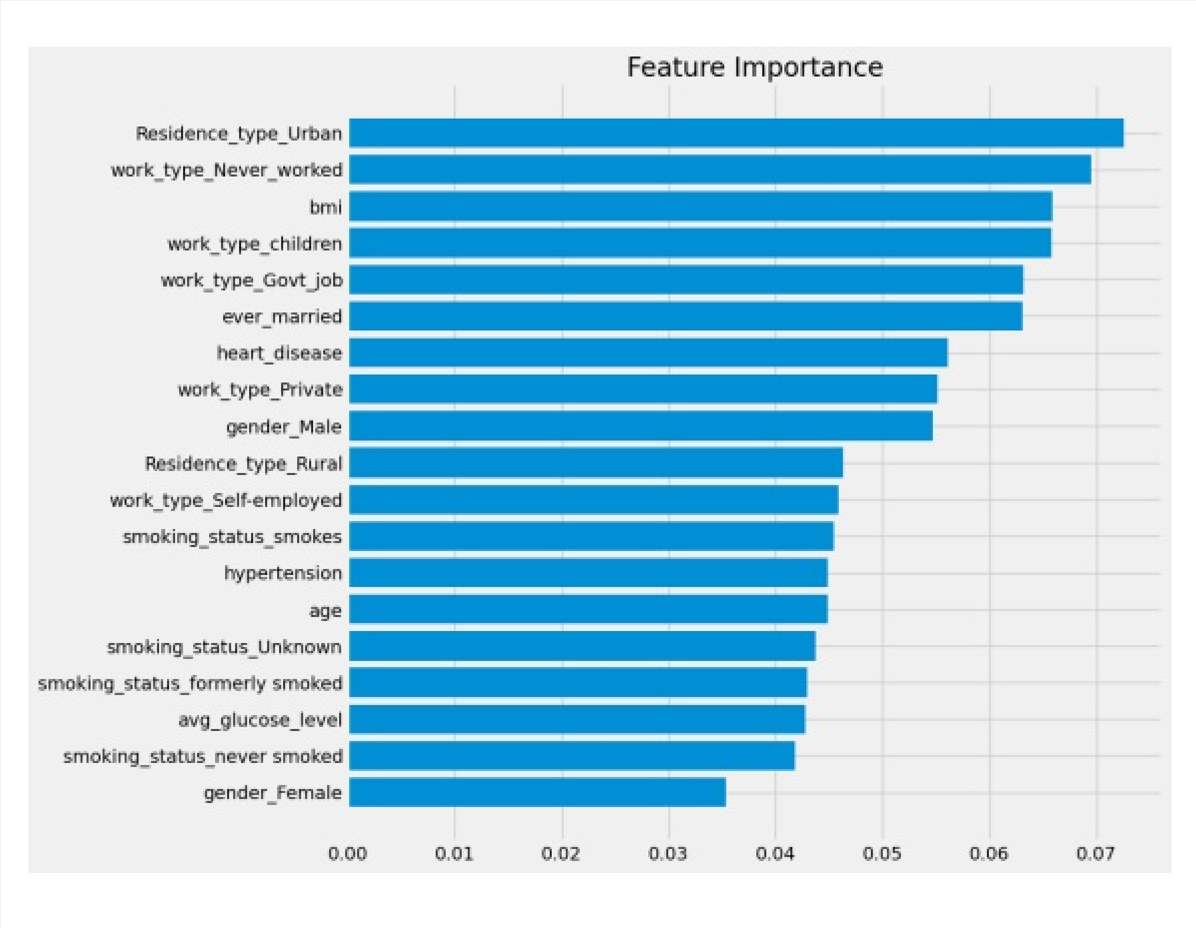
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
dummy	Dummy Classifier	0.9558	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1390
lightgbm	Light Gradient Boosting Machine	0.9445	0.7813	0.1643	0.3022	0.2030	0.1775	0.1905	0.2400
rf	Random Forest Classifier	0.9424	0.7860	0.1643	0.2587	0.1991	0.1713	0.1767	0.2440
et	Extra Trees Classifier	0.9421	0.7867	0.1300	0.2329	0.1644	0.1373	0.1445	0.2240
gbc	Gradient Boosting Classifier	0.9255	0.7955	0.2962	0.2507	0.2630	0.2253	0.2301	0.2500
dt	Decision Tree Classifier	0.9045	0.5647	0.1919	0.1234	0.1492	0.1014	0.1046	0.1350
ada	Ada Boost Classifier	0.8897	0.8010	0.4062	0.1764	0.2447	0.1953	0.2159	0.1790
knn	K Neighbors Classifier	0.8439	0.6809	0.3824	0.1189	0.1811	0.1216	0.1466	0.1370
lr	Logistic Regression	0.7697	0.8399	0.7267	0.1287	0.2185	0.1550	0.2368	0.4700
nb	Naive Bayes	0.7636	0.8204	0.7462	0.1287	0.2194	0.1556	0.2407	0.1320
qda	Quadratic Discriminant Analysis	0.7630	0.8092	0.7038	0.1226	0.2086	0.1442	0.2221	0.1390
svm	SVM - Linear Kernel	0.7594	0.0000	0.7405	0.1292	0.2189	0.1552	0.2383	0.1340
lda	Linear Discriminant Analysis	0.7536	0.8407	0.7667	0.1261	0.2165	0.1521	0.2412	0.1360
ridge	Ridge Classifier	0.7509	0.0000	0.7600	0.1239	0.2130	0.1482	0.2362	0.1310





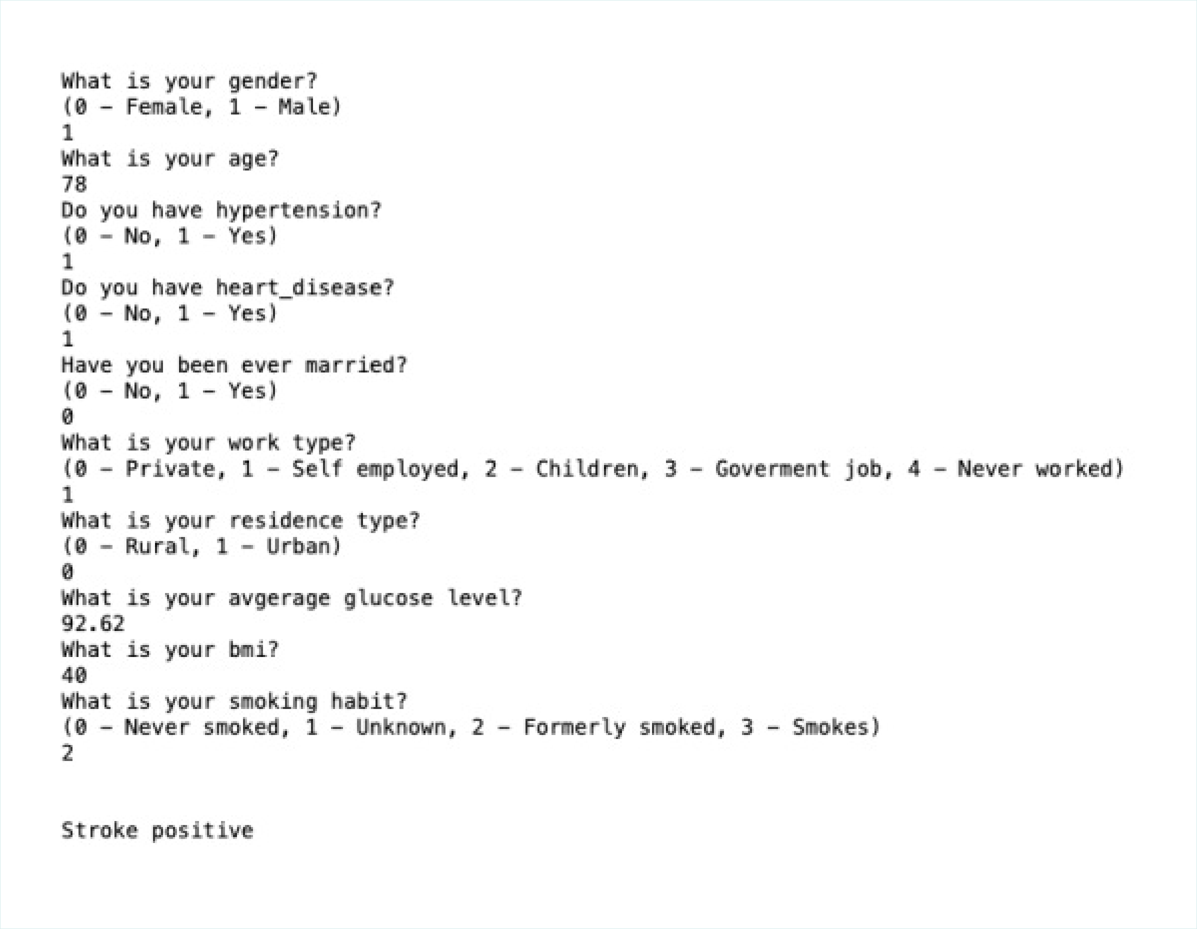
The Chances of a Person Having a Stroke

Feature Importance



The Chances of a Person Having a Stroke

Play area



```
What is your gender?
(0 - Female, 1 - Male)
1
What is your age?
78
Do you have hypertension?
(0 - No, 1 - Yes)
1
Do you have heart_disease?
(0 - No, 1 - Yes)
1
Have you been ever married?
(0 - No, 1 - Yes)
0
What is your work type?
(0 - Private, 1 - Self employed, 2 - Children, 3 - Goverment job, 4 - Never worked)
1
What is your residence type?
(0 - Rural, 1 - Urban)
0
What is your avgerage glucose level?
92.62
What is your bmi?
40
What is your smoking habit?
(0 - Never smoked, 1 - Unknown, 2 - Formerly smoked, 3 - Smokes)
2


Stroke positive
```

The Chances of a Person Having a Stroke

Conclusion

To determine the best model, we need to consider both accuracy and the confusion matrix. Here are a few observations:

> The first two models (Dummy Classifier and LDA) have similar accuracies, with the Dummy Classifier having a slightly higher accuracy. However, both models have a significant number of false negatives (FN), indicating a relatively high misclassification rate for positive instances (stroke cases).

> The XGBClassifier (with mean BMI) achieved the highest accuracy (95.1%) among all the models. However, it also had a higher number of false negatives (FN) compared to the first two models.

> The fourth model, which used the resampled dataset with XGBClassifier (with mean BMI), had a lower accuracy (92.4%) compared to the others. It also had a considerable number of false negatives (FN) and false positives (FP)