GROUP 6

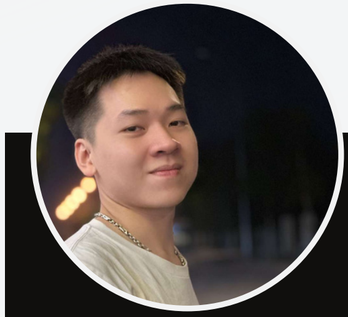
FINAL PROJECT

CONTENT

OUR TEAM

NGUYEN HUU HOANG PHONG

Member of group 6

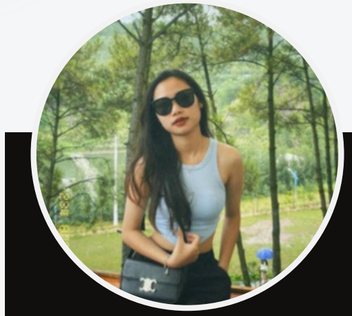


DAO DUC MANH

Leader of Group 6

BUI ANH QUAN

Member of group 6



BUI NGOC HA

Member of group 6

NGUYEN DUC PHUONG

Member of group 6

NGUYEN MANH QUANG

Member of group 6

INTRODUCE

According to the World Health Organization (WHO) stroke is the 2nd cause of death globally, responsible for approximately 11% of total deaths.

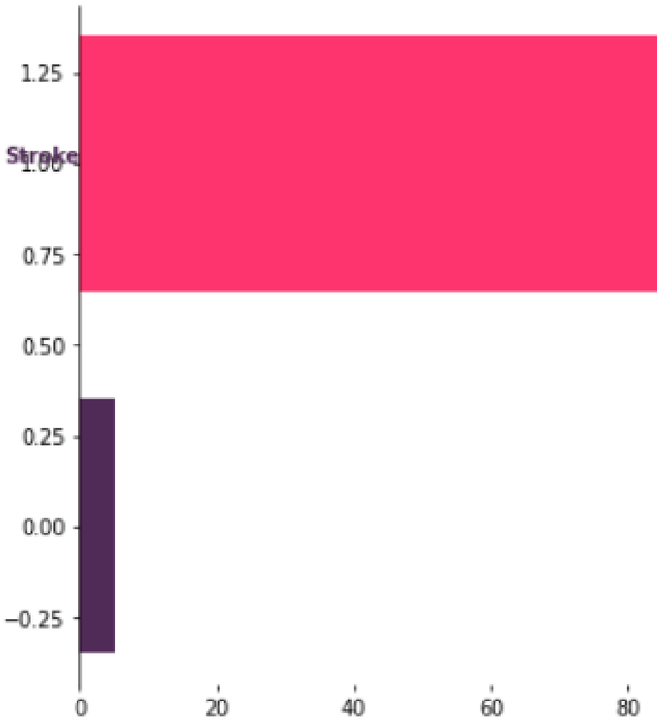
From all the analysis -> Suggest some methods to reduce the incidence of stroke

ABOUT THE DATASET

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

ANALYSIS

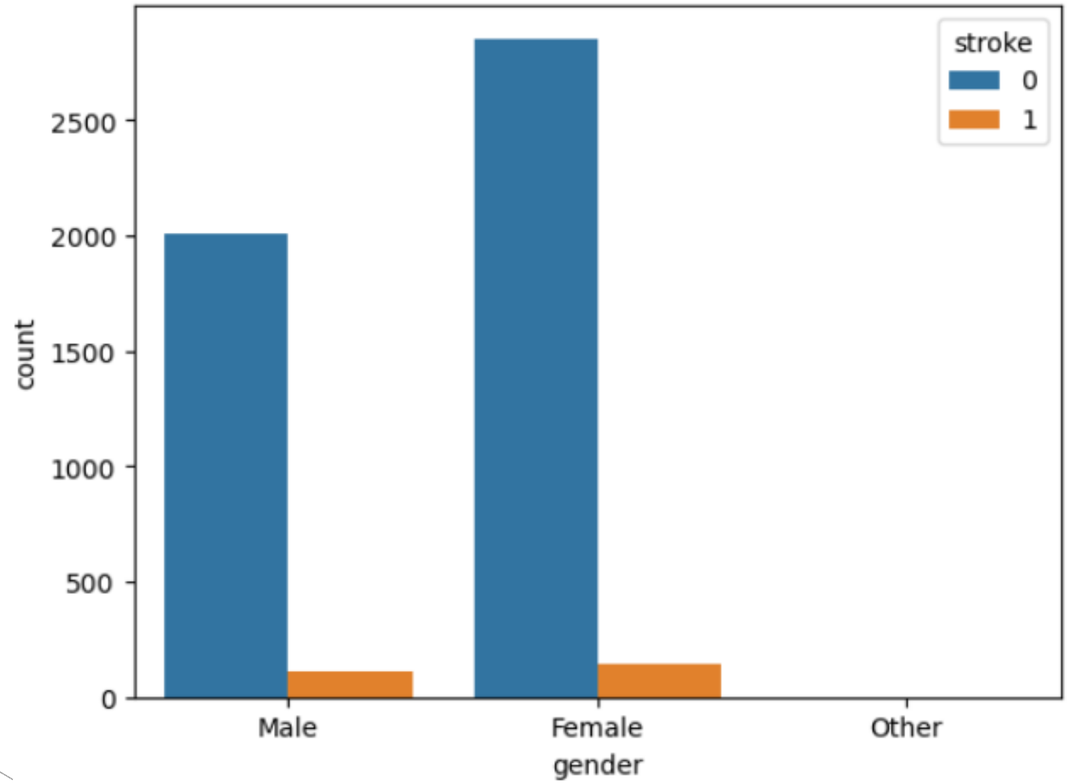
HEALTHY/STROKE

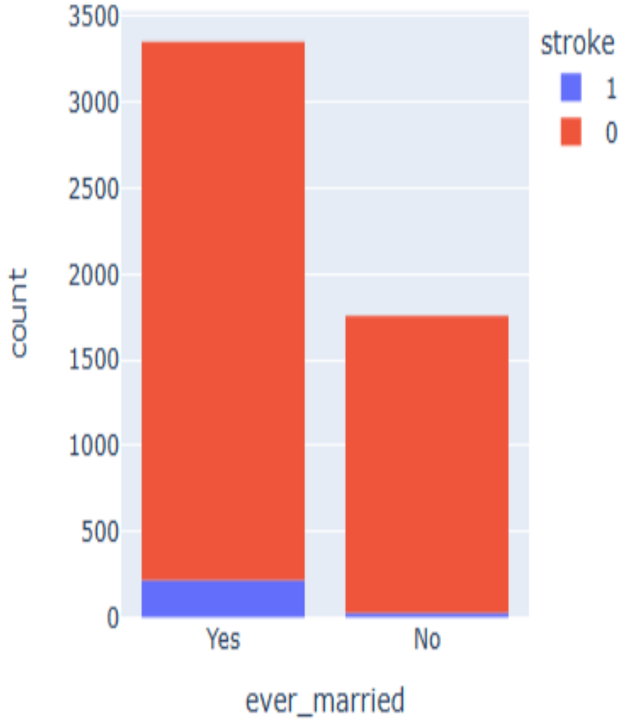


PERCENTAGE OF PEOPLE HAVING STROKES

-> Highly unbalanced distribution, 5 in 100 people susceptible to heart stroke

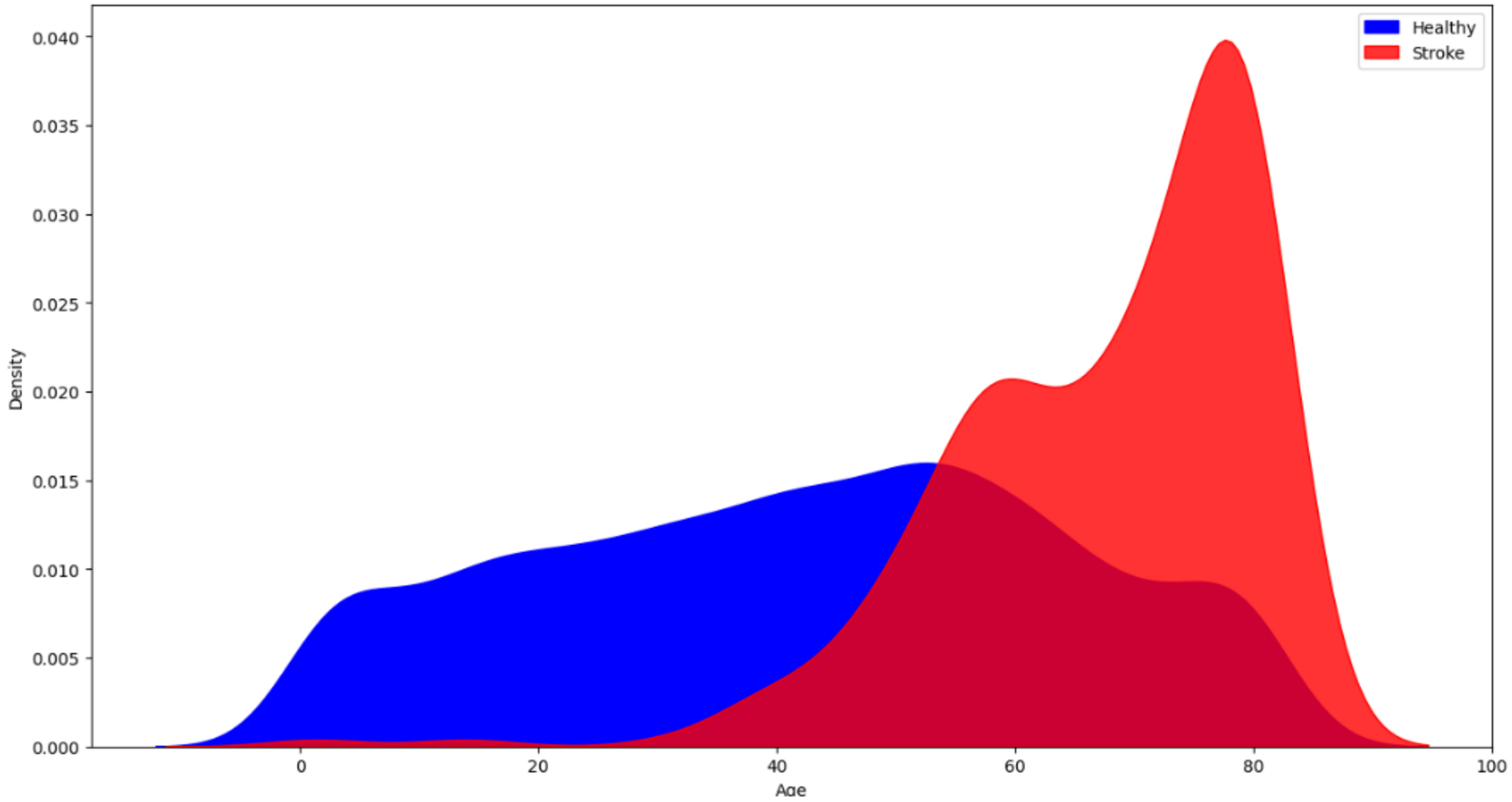
ANALYSIS





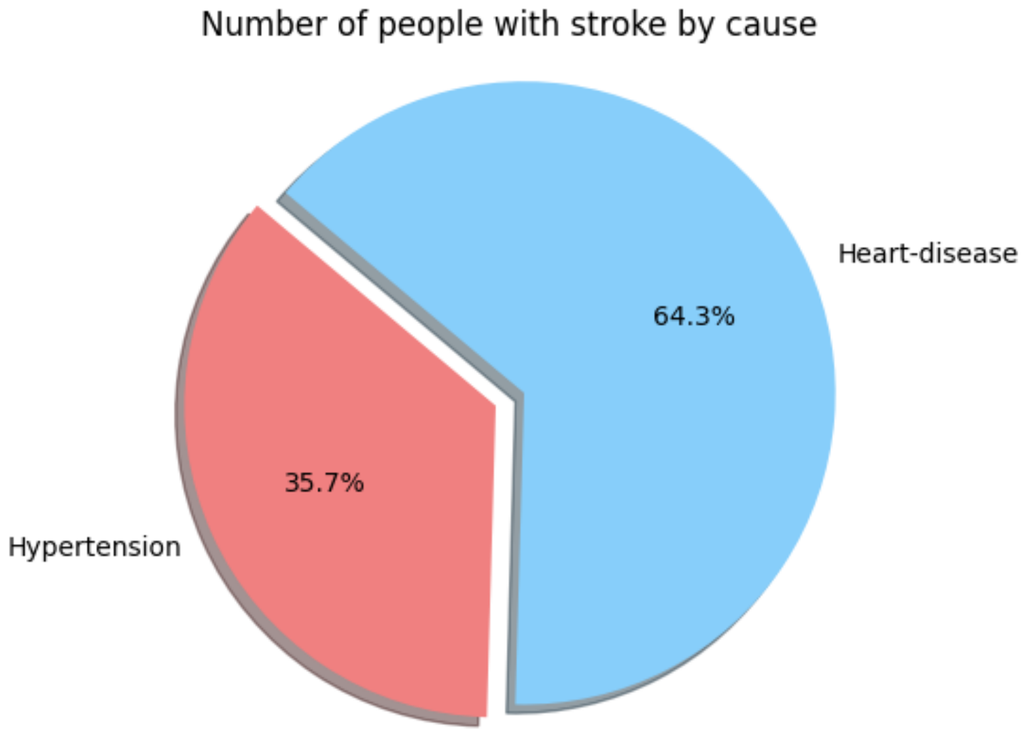
PEOPLE WITH STROKE BASED ON GENDER AND MARRIED

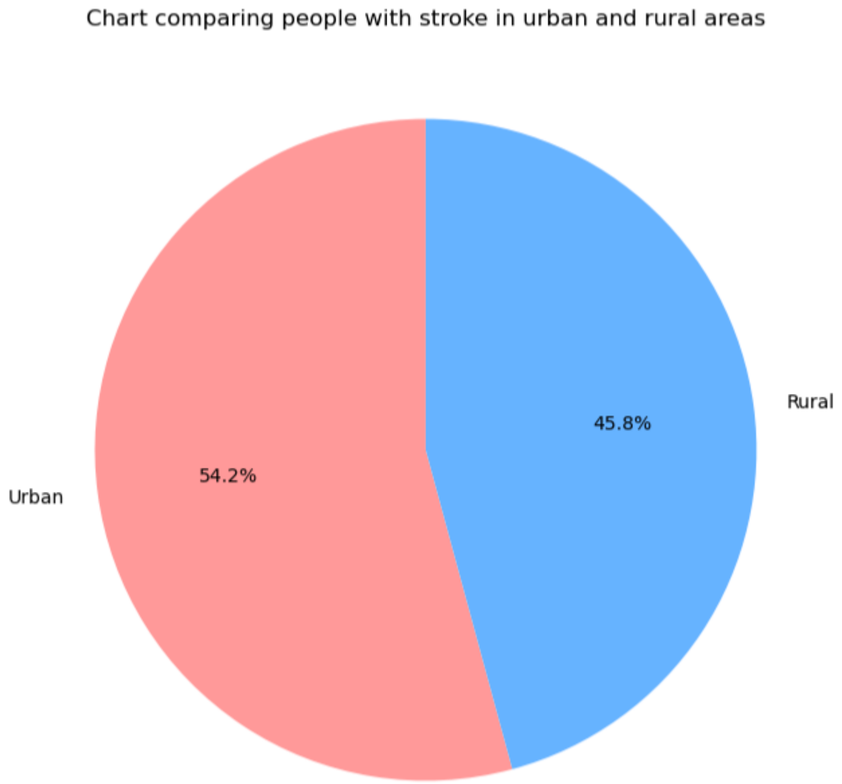
ANALYSIS



AGE DISTRIBUTION OF STROKE AND HEALTHY PATIENTS

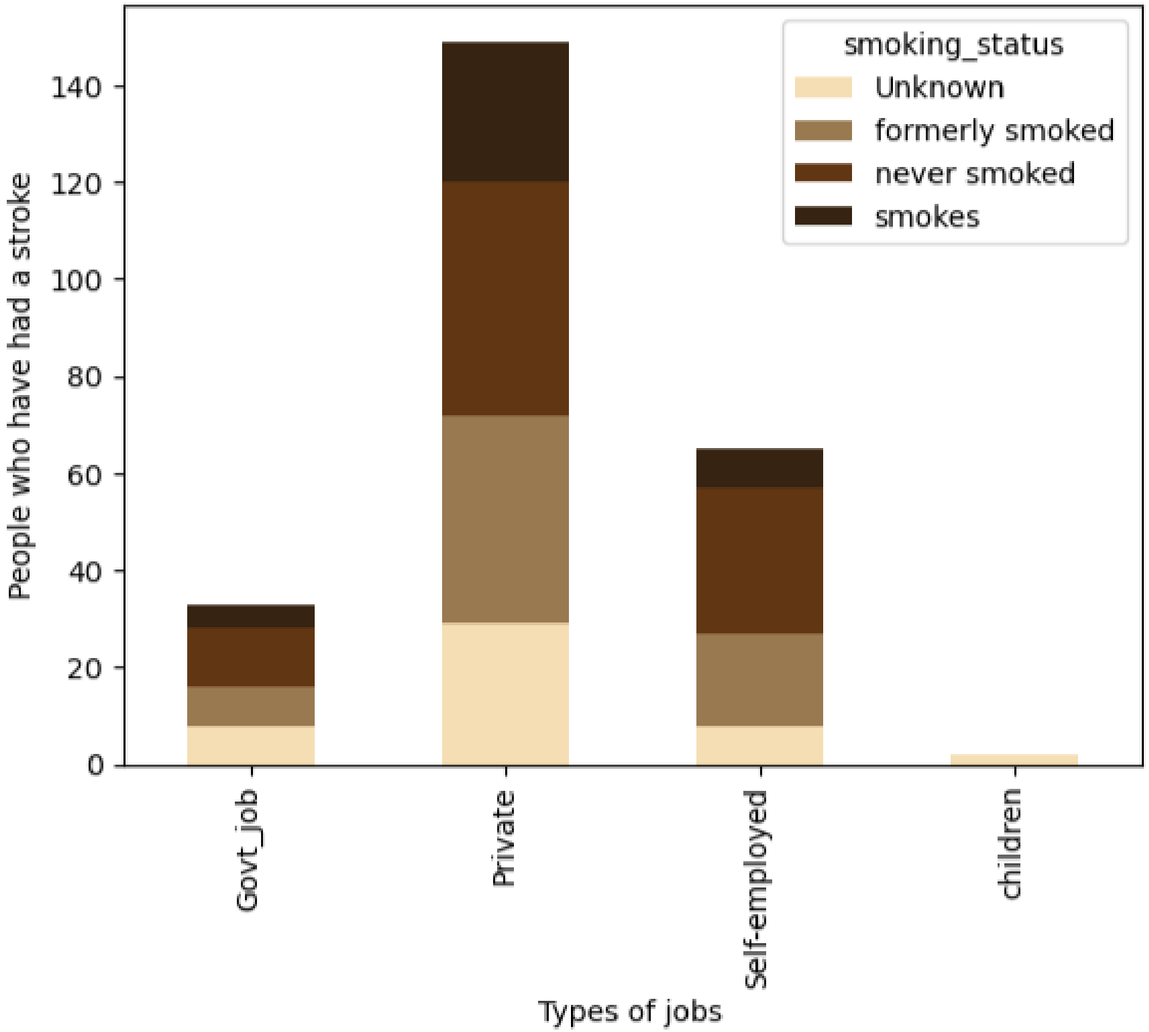
ANALYSIS





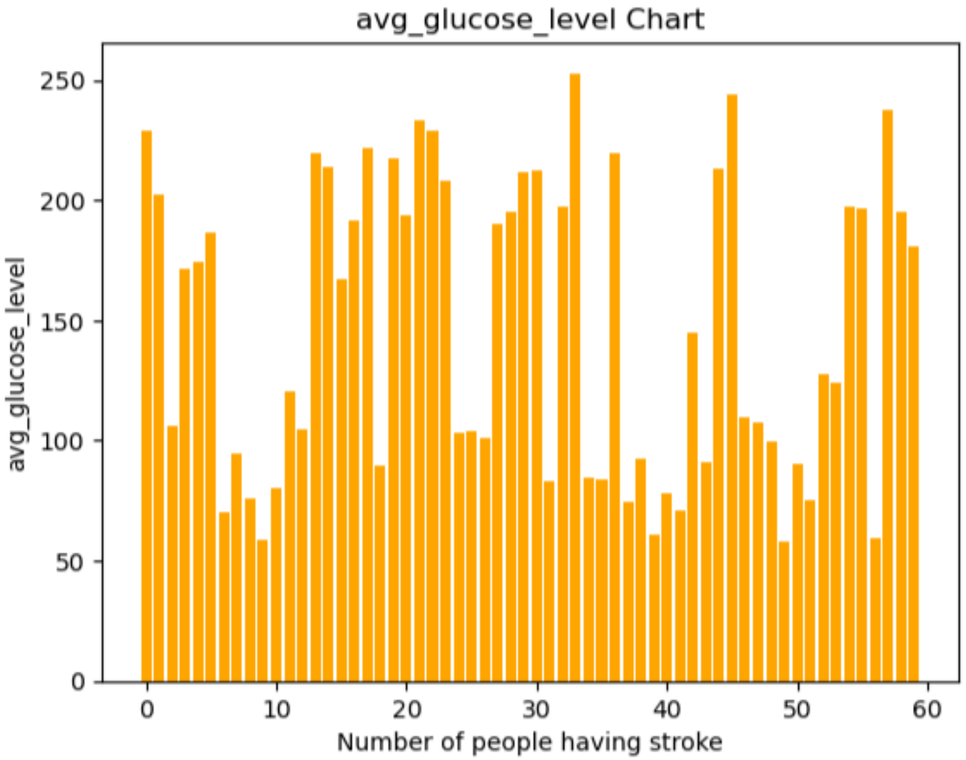
PEOPLE WITH STROKE BASED ON HYPERTENSION AND HEART DISEASE

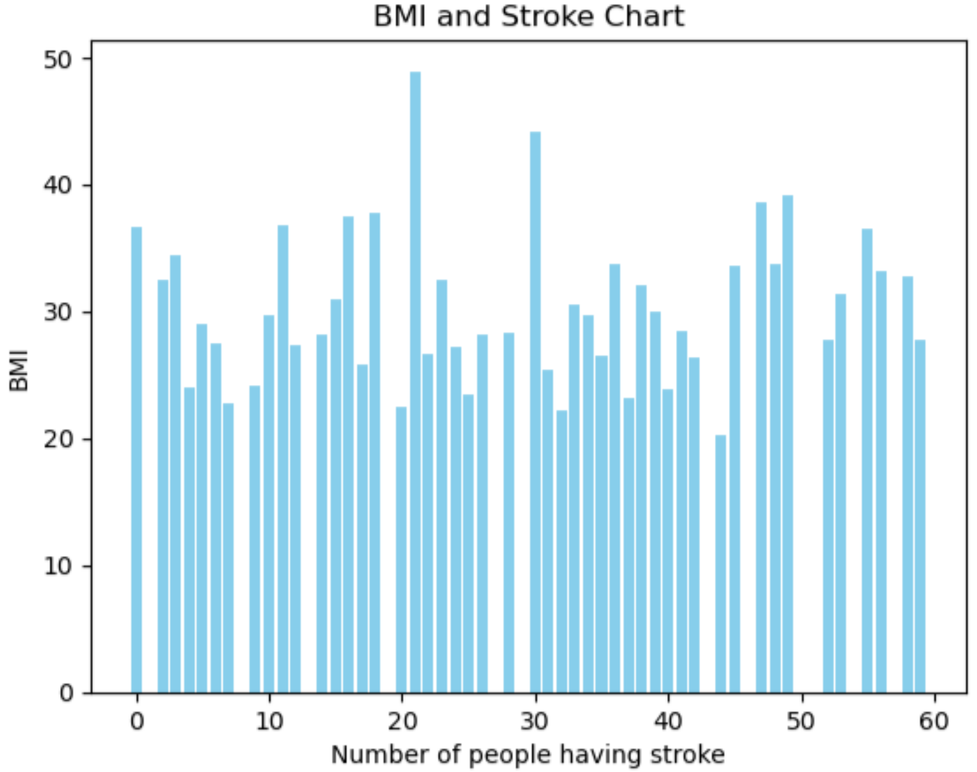
ANALYSIS



'COMPARE THE NUMBER OF PEOPLE WHO HAVE HAD A STROKE BY WORK TYPE AND SMOKING STATUS

ANALYSIS





PEOPLE WITH STROKE BASED ON AVG GLUCOSE AND BMI

KNN ALGORITHM

```
con_cols=['age','avg_glucose_level','bmi']
X=df.drop(['stroke'],axis=1)
y=df['stroke']

from sklearn.model_selection import train_test_split

testperc=0.2
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = testperc, random_state = 0)

from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
X_train[con_cols]=scaler.fit_transform(X_train[con_cols])
X_test[con_cols]=scaler.transform(X_test[con_cols])

from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
from sklearn.model_selection import GridSearchCV

modelx_knn = KNeighborsClassifier()

k_range = list(range(1, 11))
param_grid = dict(n_neighbors=k_range)
param_grid.update({"metric":["minkowski","cityblock","euclidean"]})


grid = GridSearchCV(modelx_knn, param_grid, cv=2, scoring='accuracy', return_train_score=False,verbose=1)

grid_search=grid.fit(X_train[con_cols], y_train)

print(grid_search.best_params_)
print("The best accuracy achieved is",grid_search.best_score_ *100 )
```

=> **THE RESULT IS:**

FITTING 2 FOLDS FOR EACH OF 30 CANDIDATES, TOTALLING 60 FITS

{'METRIC': 'MINKOWSKI', 'N_NEIGHBORS': 9}

THE BEST ACCURACY ACHIEVED IS 95.77286729132537

KNN ALGORITHM

This result indicates that when using the KNN Algorithm with 'metric' parameters of 'minkowski' and 'n_neighbors' of 9 to classify data points, the model achieved an accuracy of 95.77.

This may show that this KNN model is capable of classifying data points quite well.

95.77%

RECOMMENDATION

- Diet

An unhealthy diet can increase your chances of having a stroke because it may lead to an increase in your blood pressure and cholesterol levels.

- Do Exercise

Combining a healthy diet with regular exercise is the best way to maintain a healthy weight.

Regular exercise can also help lower your cholesterol and keep your blood pressure healthy.

– Stop Smoking

Smoking significantly increases your risk of having a stroke. This is because it narrows your arteries and makes your blood more likely to clot.

– Cut down on alcohol

Excessive alcohol consumption can lead to high blood pressure and trigger an irregular heartbeat (atrial fibrillation), both of which can increase your risk of having a stroke.

THANK'S FOR LISTENING

GROUP 6