

Stroke Prediction

Group 10

A large, dark blue, curved shape that starts from the bottom left and extends diagonally upwards towards the right, filling the lower half of the slide.

Why strokes?

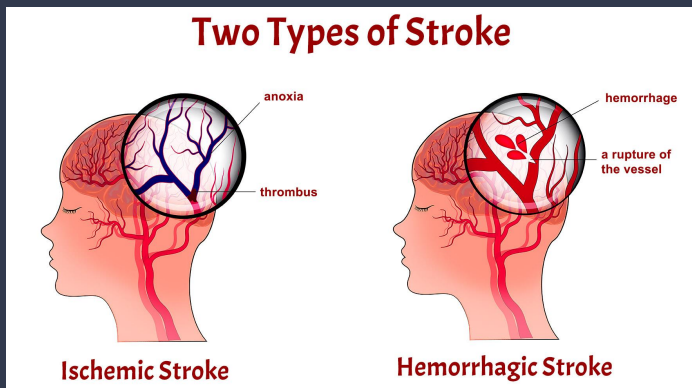
- Strokes are the 2nd leading cause of death in developed countries after heart disease (WHO)
- The #5 cause of death in the US
- One of the leading causes of disability
- Sometimes can be prevented



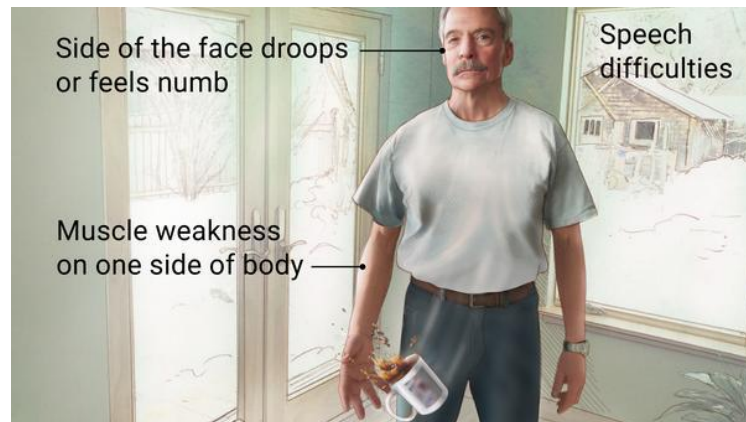
**American
Stroke
Association.**

*A division of the
American Heart Association.*

What is a stroke?



- Blood vessel that supplies nutrients to the brain is hindered by a clot (Ischemic -- 87 % of strokes)
- Blood vessel bursts (Hemorrhagic)
- Brain tissues can't receive necessary nutrients
- Leads to brain cells dying in only a few minutes
- Difficulty walking, speaking and understanding, paralysis of the face and or extremities
- Deadly if not treated immediately.



Stroke Risk Factors

- Age -- Especially for people > 65
- Gender -- women > men
- Hypertension
- High Cholesterol
- Smoking
- Diabetes
- Obesity - high BMI (over 30)
- Stress (work, marriage, location of residence)

Data Sources:

Kaggle



Analytics Vidhya



McKinsey Analytics
Online Hackathon
Healthcare Analytics



McKinsey Analytics Online Hackathon - Healthcare Analytics

Question we hope
to answer:

Can we reliably predict
a stroke based on
certain features of a
person's medical
history?

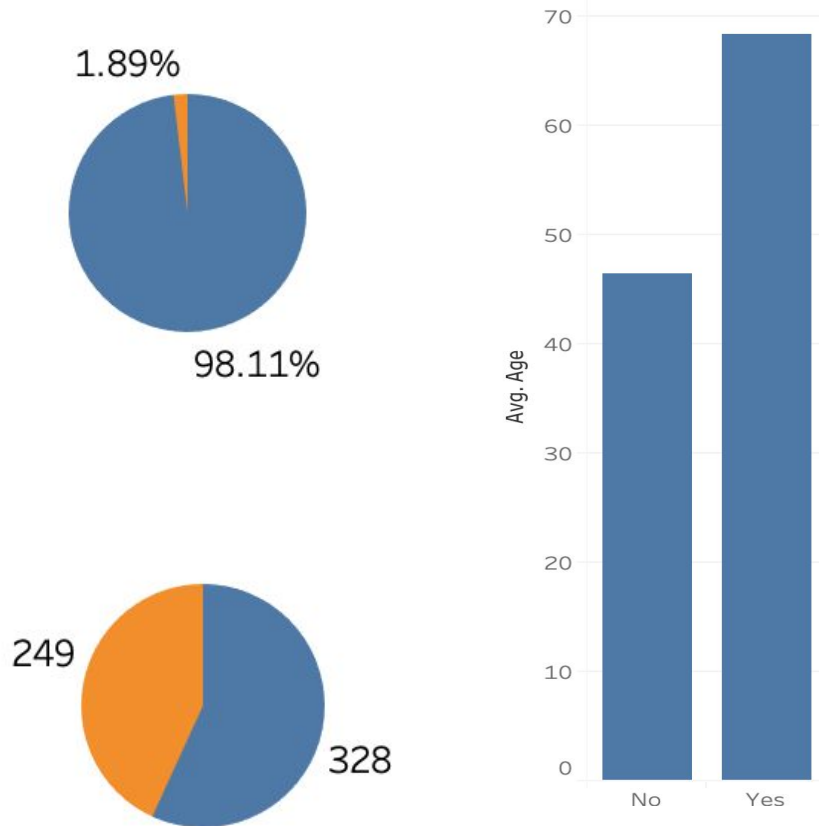
Features

- Gender (Male, Female, Unknown)
- Age
- Hypertension (high blood pressure)
- Heart Disease (yes or no)
- Has ever been married (yes or no)
- Work type (Private, self-employed, government job, etc.)
- Residence type (Urban or Rural)
- Avg glucose level
- BMI (>30 considered high risk)
- Smoking status
- Has patient ever experienced a stroke

Exploring the data:

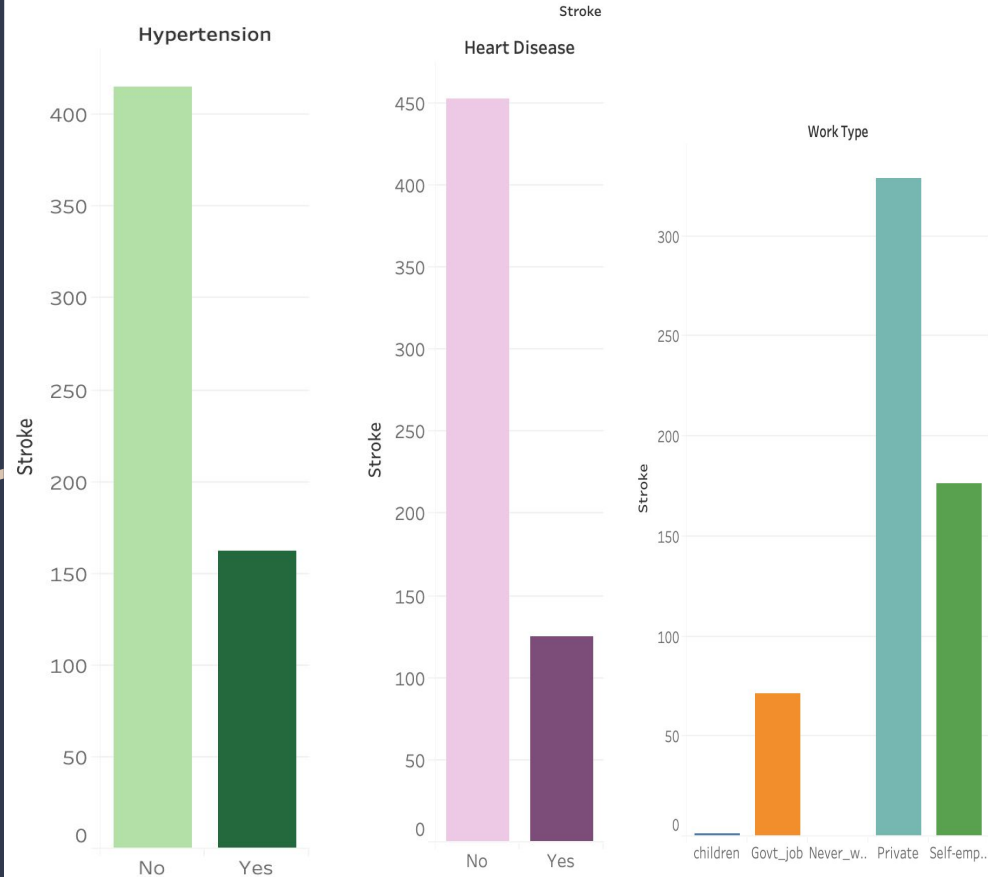
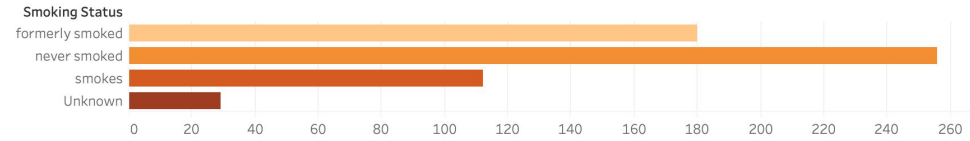
- Nearly 2% of people in the dataset had a stroke
- Slight gender imbalance toward women
- Average age is 68.35 years

Total Stroke Percentage

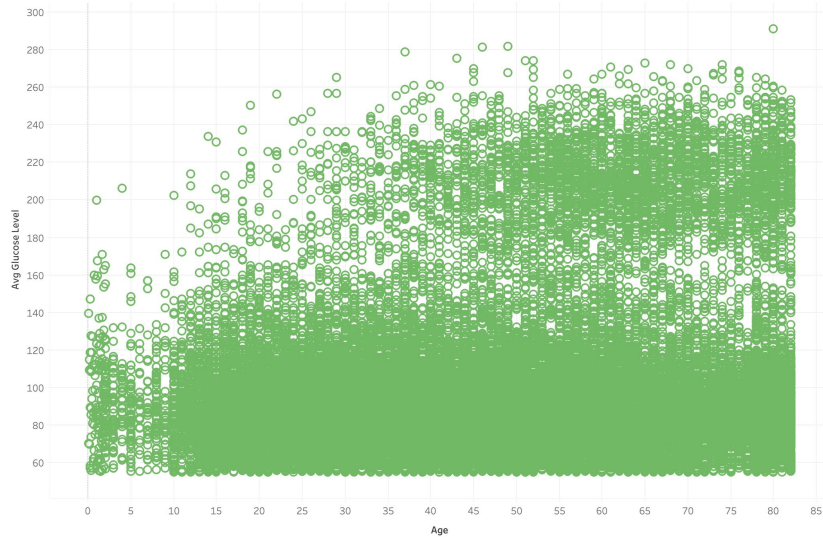


- Higher number of strokes for former smokers compared to current smokers (180 : 112)
- 28% or 162 people had hypertension
- 22% or 125 people had heart disease
- 329 (57%) Private company employers and 176 self-employed (31%) had strokes

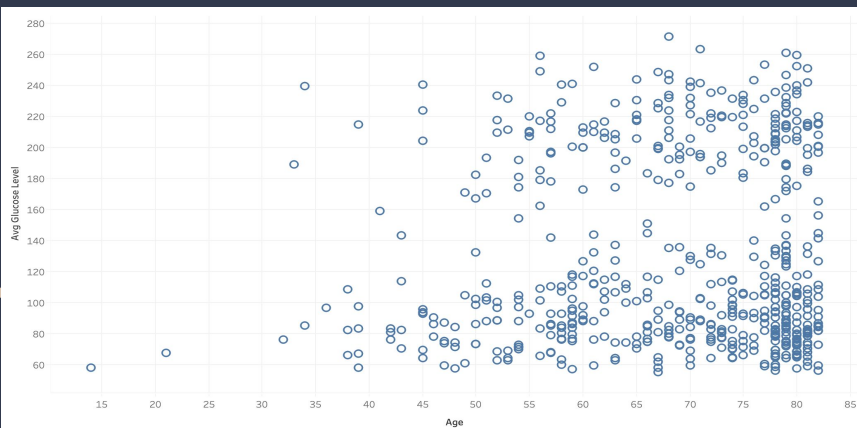
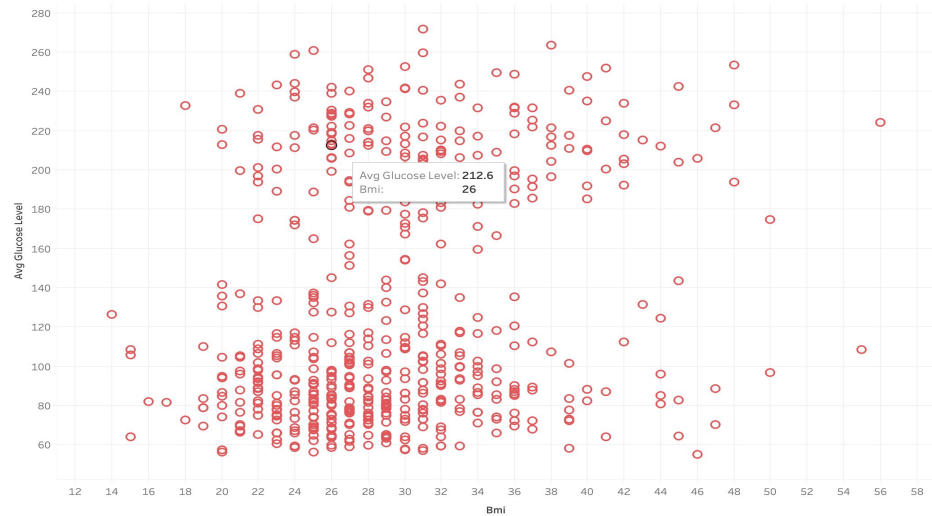
Stroke/Smoking Data



Age vs avg_glucose_level



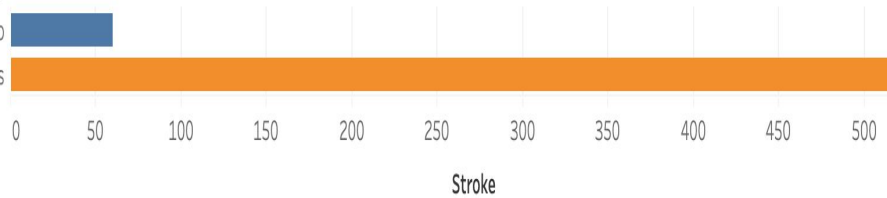
bmi vs avg_glucose_level



Ever Married

No

Yes



Analysis

- Random Forest Classifier
- Logistic Regression

Random Forest Classifier

```
forest = RandomForestClassifier(n_estimators = 100)
forest.fit(X_train, y_train)

forest_score = forest.score(X_train, y_train)
forest_test = forest.score(X_test, y_test)

y_pred = forest.predict(X_test)

print('Training Score', forest_score)
print('Testing Score \n', forest_test)
print(cm)
print(classification_report(y_test, y_pred))
```

Training Score 0.9993777777777778
Testing Score
0.8173333333333334
[[5848 0]
 [1652 0]]

		precision	recall	f1-score	support
	0	0.84	0.94	0.89	5848
	1	0.65	0.37	0.47	1652
<hr/>					
	accuracy			0.82	7500
	macro avg	0.75	0.66	0.68	7500
	weighted avg	0.80	0.82	0.80	7500

Logistic Regression

```
model = LogisticRegression(solver="lbfgs", max_iter=200)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

print('Testing Score \n', score)

print(classification_report(y_test, y_pred))

cm = confusion_matrix(y_test, y_pred)
print(cm)
```

Testing Score
0.7797333333333333

		precision	recall	f1-score	support
	0	0.78	1.00	0.88	5848
	1	0.00	0.00	0.00	1652
<hr/>					
	accuracy			0.78	7500
	macro avg	0.39	0.50	0.44	7500
	weighted avg	0.61	0.78	0.68	7500

[[5848 0]
 [1652 0]]

Neural Network

239/239 - 1s - loss: 0.0930 - accuracy: 0.9811
Loss: 0.09303482621908188, Accuracy: 0.9811444282531738

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 80)	880
dense_1 (Dense)	(None, 30)	2430
dense_2 (Dense)	(None, 1)	31

Total params: 3,341

Trainable params: 3,341

Non-trainable params: 0

Analysis Tools Used

- GitHub
- Python
- Jupyter Notebook
- PostgreSQL
- Scikit learn library
- Keras library
- Tensorflow
- Tableau
- SQLAlchemy
- Visual Studio Code