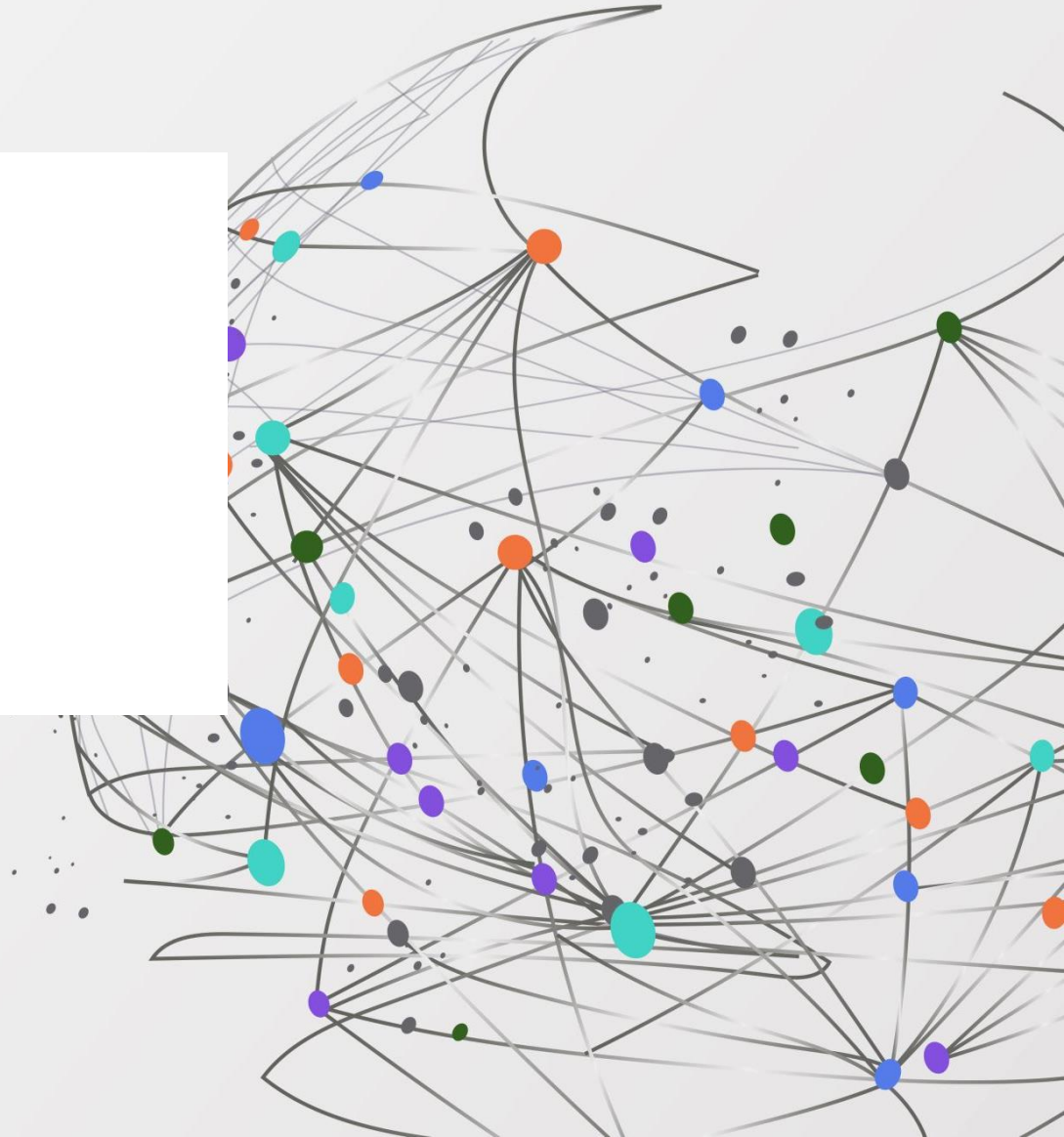# Datascience

STROKE
PREDICTION

Overview

- A health company is piloting a drug that cleans blood vessels, helping to prevent strokes. They have identified volunteers that are interested in joining the pilot group. The health company would like to ensure the drug is administered to the volunteers that are at risk of stroke.

- The data set provided helps us understand the profile of the potential customer in terms of:
  1. Health history
  2. Lifestyle
  3. Marital status
  4. Stroke history

Data Insights

Ever Married vs stroke

Type of work vs stroke
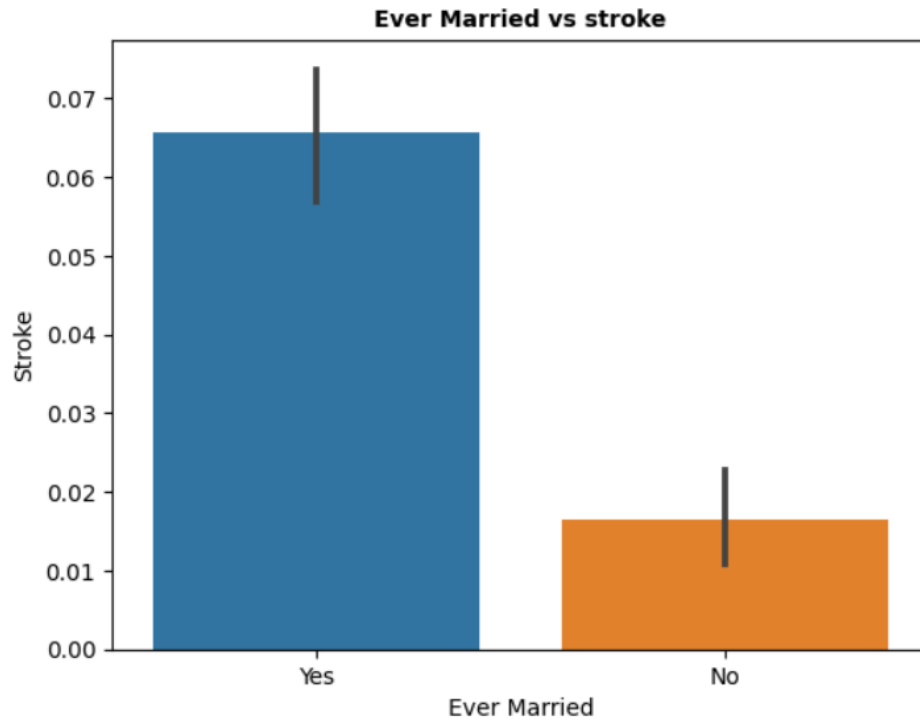
- Married individuals are more likely to suffer a stroke, indicating your key target marketing profile.
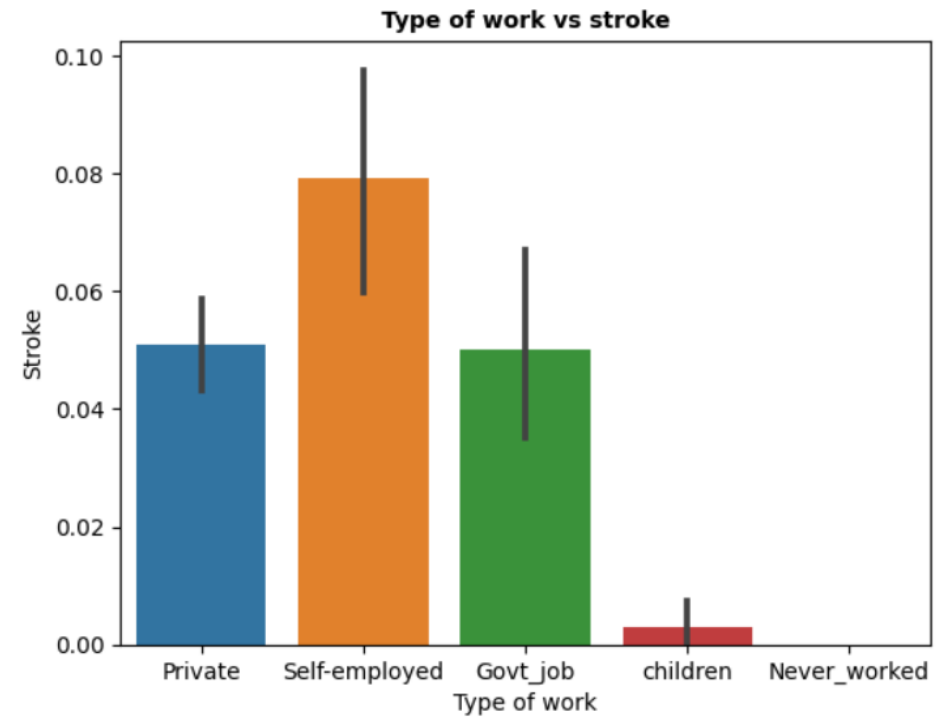
- Individuals who are self-employed are more likely to suffer a stroke.
- Children are an additional market

Model Results

- Multiple models were applied to the dataset
  - KNN model
  - Random Forest Tree model

- The ideal model was selected and tuned for further accuracy.
  The tuned Random Forest model produced:
  - A precision of 95%, therefore predictions are 95% correct.
  - 0.04% (FN) accuracy in predicting the classification of the individual to the wrong pilot group, predicting no stroke but it was a stroke individual.
  - 4% (FP) accuracy in predicting the classification of the individual to the wrong pilot group, predicting a stroke individual but it was a not stroke individual.

  - The balance of the dataset is unbalanced as 95% has not had a stroke and 4.8% has had a stroke, which is why the FP is high.

Recommendations

- A balanced dataset would ensure the model would be more accurate.

- Target your sales campaign at married individuals.
- Target your sales campaign towards individuals who own their own business and work in private and government sector.
- Consider additional markets like children