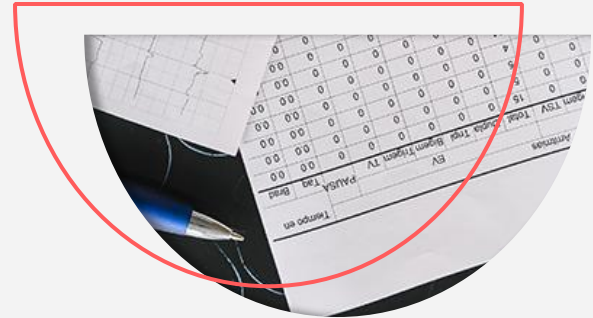




SAML

Medical Research

STROKE RESEARCH



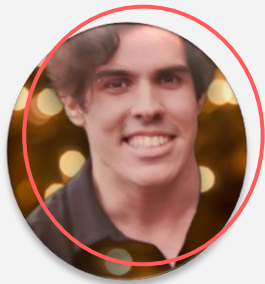


SAML TEAM



**ANGELA
PACATTE**

Data Analyst,
Communications
Director



**KELSEY
CORCORAN**

Data Analyst,
Machine Learning
Engineer



**KATIE
HOPKINS**

Data Analyst,
Chief Marketing
Officer



**JACK
BAUER**

Data Analyst,
Chief Information
Officer



**BOWEN
WILDER**

Data Analyst,
Project Manager

TABLE OF CONTENTS

01

INTRODUCTION

- What is a Stroke?
- Symptoms of a Stroke

02

PURPOSE

- Why Choose This Topic?
- Questions to Answer...

03

STROKE DATASET

- The Dataset
- Personal Indicators
- Medical Indicators
- EDA Findings

04

ML MODEL

- Data Preprocessing
- Analysis

05

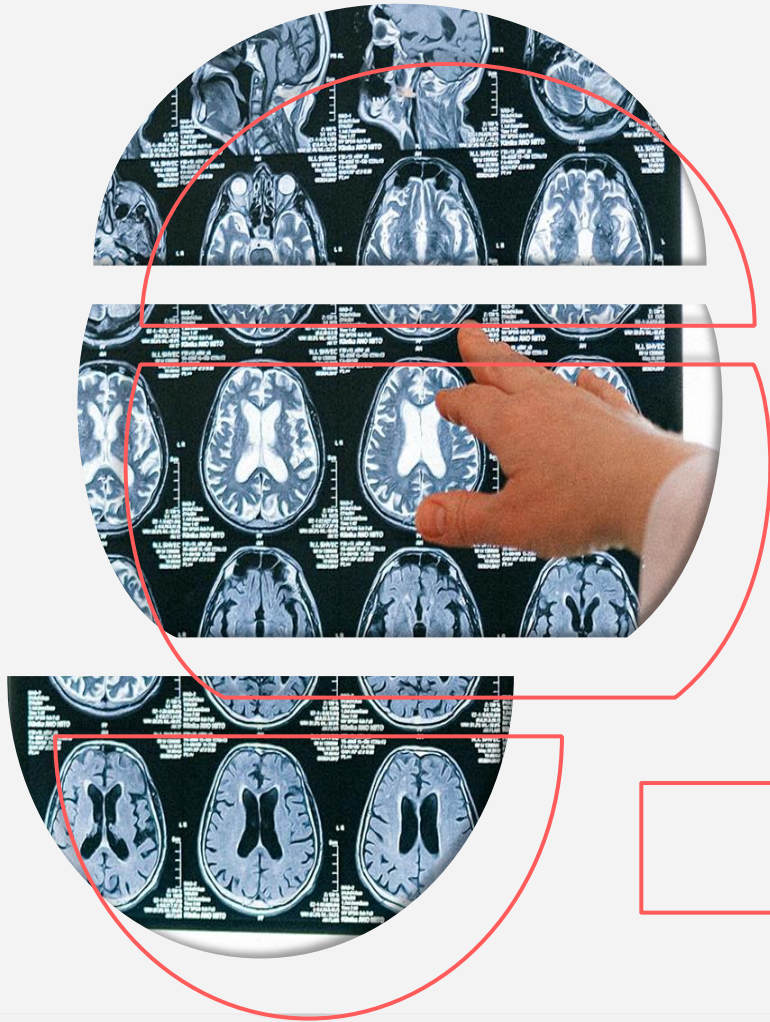
TOOLS

- Project Framework
- Tableau Dashboard

06

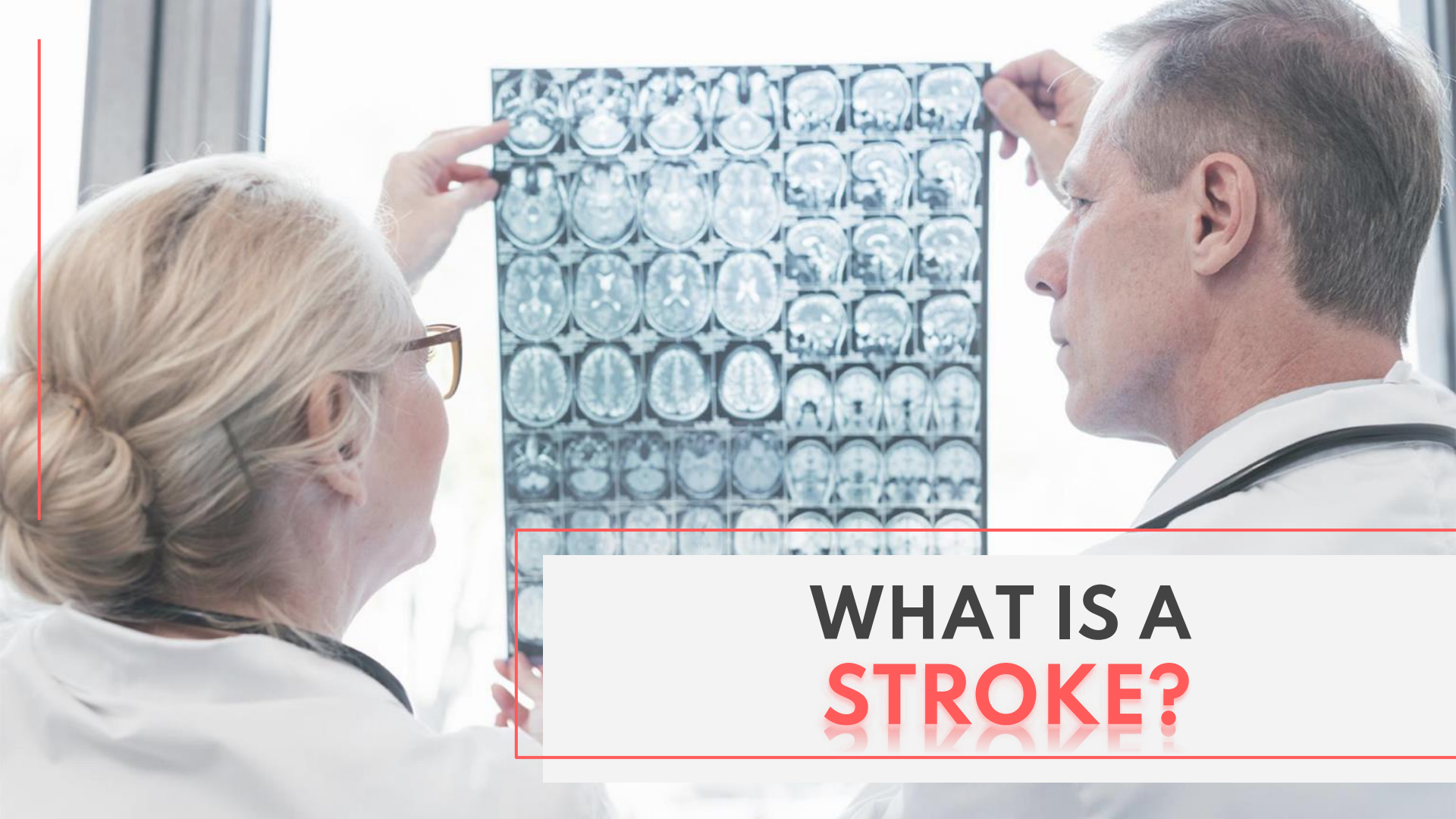
CONCLUSIONS

- ML Model Analysis
- What We Would Have Done Differently



01 INTRODUCTION

What is a stroke?
How do you detect a stroke?



**WHAT IS A
STROKE?**

SYMPTOMS OF A **STROKE**

EYES

Trouble *seeing* in one or both eyes

NUMBNESS

Sudden *weakness* in the face, arm, or leg, especially on one side of the body

BRAIN

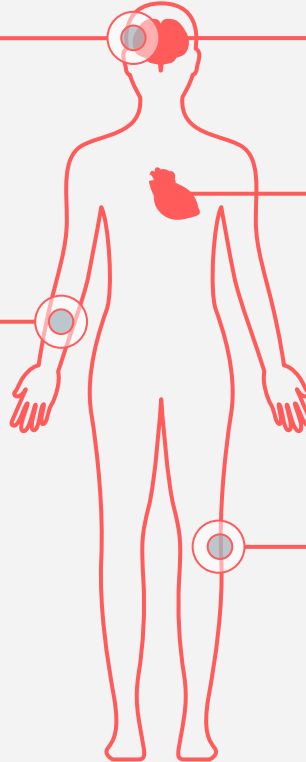
Sudden *confusion*, trouble speaking, or difficulty understanding speech.
Severe headache

HEART

Having a stroke increases a person's risk for *cardiac trouble*

MOBILITY

Sudden *trouble walking*, dizziness, loss of balance, or lack of coordination



02

PURPOSE



Why is predicting strokes important?
What do we hope to conclude?

**WHY CHOOSE
THIS TOPIC?**



QUESTIONS TO ANSWER...



SUCCESS

Can our Machine Learning model be used to predict stroke risk?



ACCURACY

Which aspect is more accurate to predict risk:
Medical or Personal data?

03 STROKE DATA

What does the dataset include?
What were our initial findings?



STROKE DATASET

ID#	Gender	Age	Hypertension	Heart_Dise...	Avg_Glucose_Lvl	BMI	Ever_Married	Work_Type	Residence_Type	Smoker	Stroke
int	str	int	int	int	float	float	str	str	str	str	int
1	Male	67	0	1	228.69	36.6	Yes	Private	Urban	Former	1
3	Male	80	0	1	105.92	32.5	Yes	Private	Rural	Never	1
4	Female	49	0	0	171.23	34.4	Yes	Private	Urban	Current	1
5	Female	79	1	0	174.12	24.0	Yes	Self-employed	Rural	Never	1
6	Male	81	0	0	186.21	29.0	Yes	Private	Urban	Former	1
7	Male	74	1	1	70.09	27.4	Yes	Private	Rural	Never	1
8	Female	69	0	0	94.39	22.8	No	Private	Urban	Never	1
10	Female	78	0	0	58.57	24.2	Yes	Private	Urban	Unknown	1
11	Female	81	1	0	80.43	29.7	Yes	Private	Rural	Never	1
12	Female	61	0	1	120.46	36.8	Yes	Govt_job	Rural	Current	1

Shape

(5,109, 12)

Features

Categorical: 8

Numerical: 3

Missing Values

201 NaNs in "BMI" column

	All Criteria
	Medical Criteria
	Personal Criteria

PERSONAL CRITERIA



AGE

From birth to 82 years of age



GENDER

Male or Female



EVER MARRIED

Yes or No



WORK TYPE

Government, Private, Self-Employed,
& Raise Children



RESIDENCE TYPE

Rural or Urban



SMOKING STATUS

Current, Former, Never,
& Unknown

MEDICAL CRITERIA



AGE

From birth to 82 years of age



GENDER

Male or Female



HYPERTENSION

Yes or No



HEART DISEASE

Yes or No



AVG GLUCOSE LVL

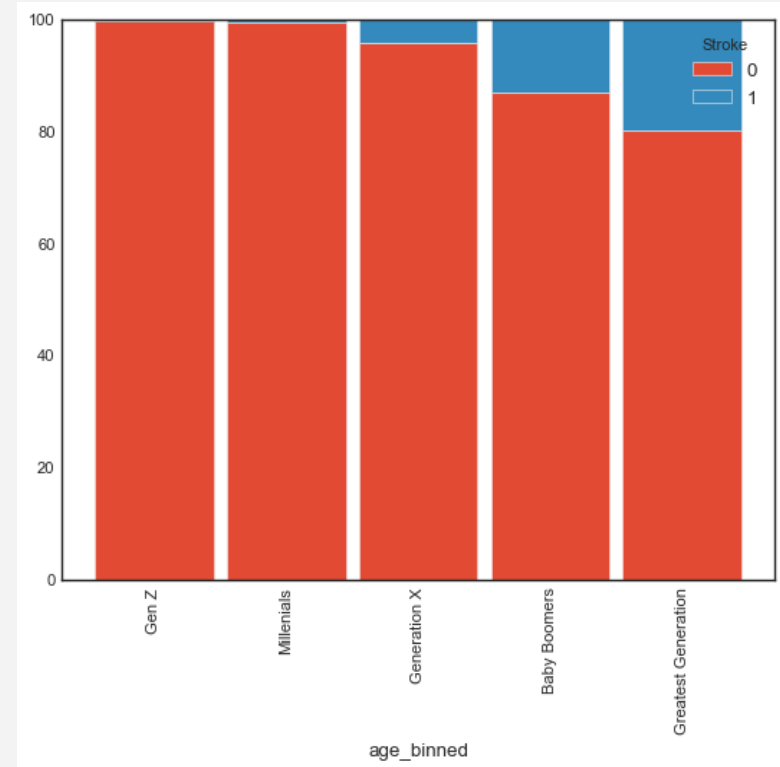
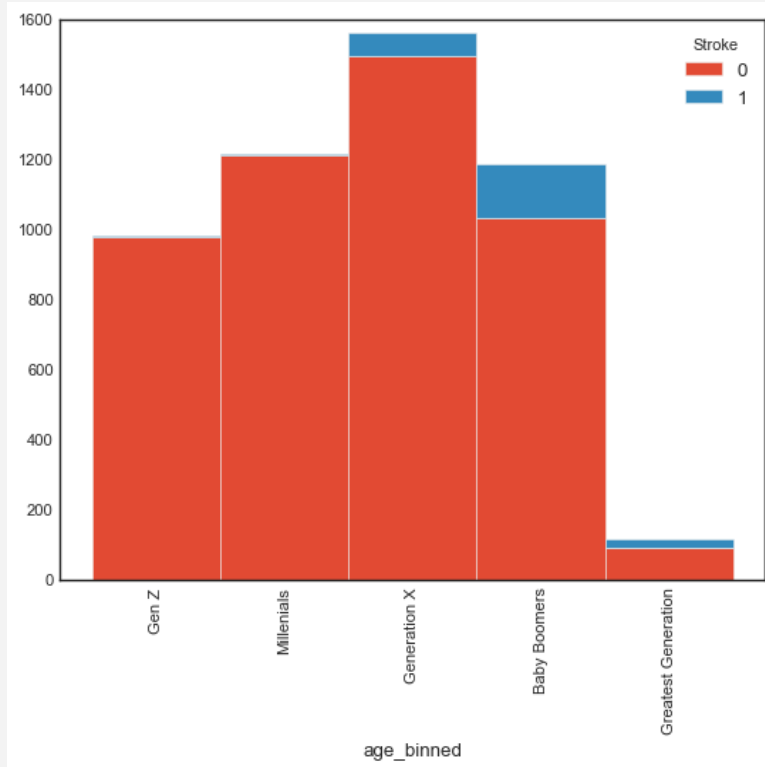
From 55 to 268



BMI

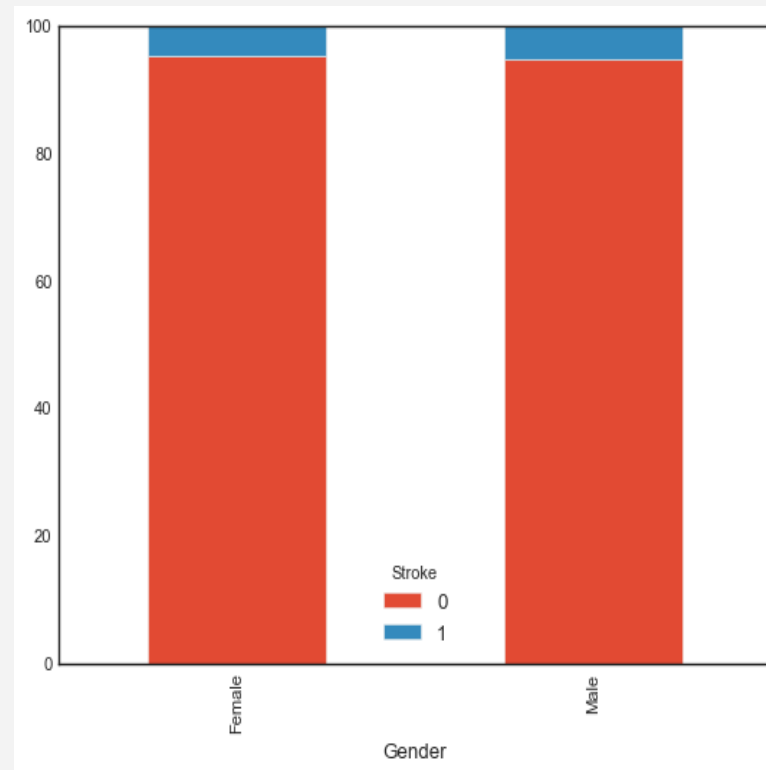
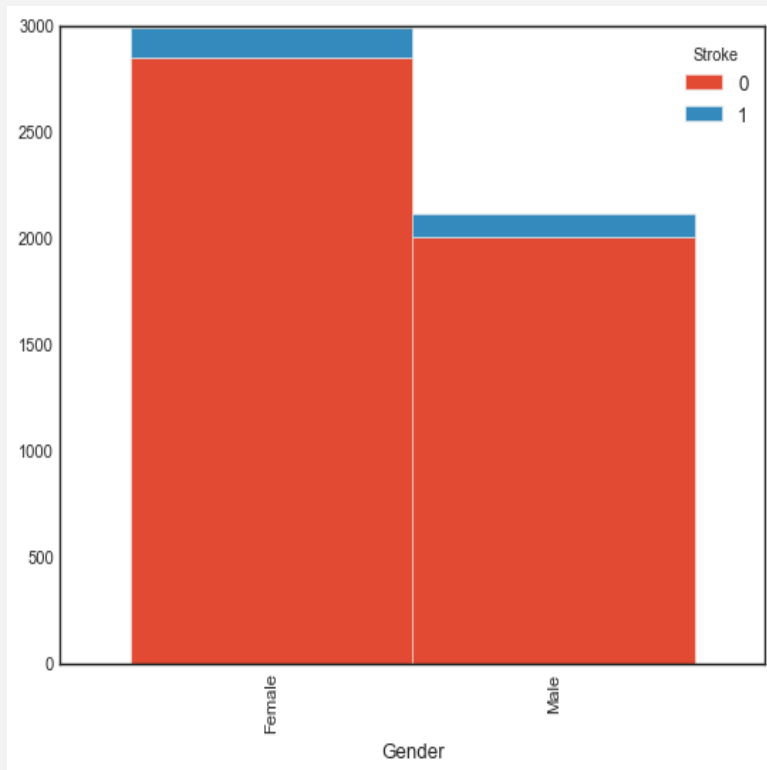
From 10 to 98

DEMOGRAPHIC ANALYSIS (AGE)

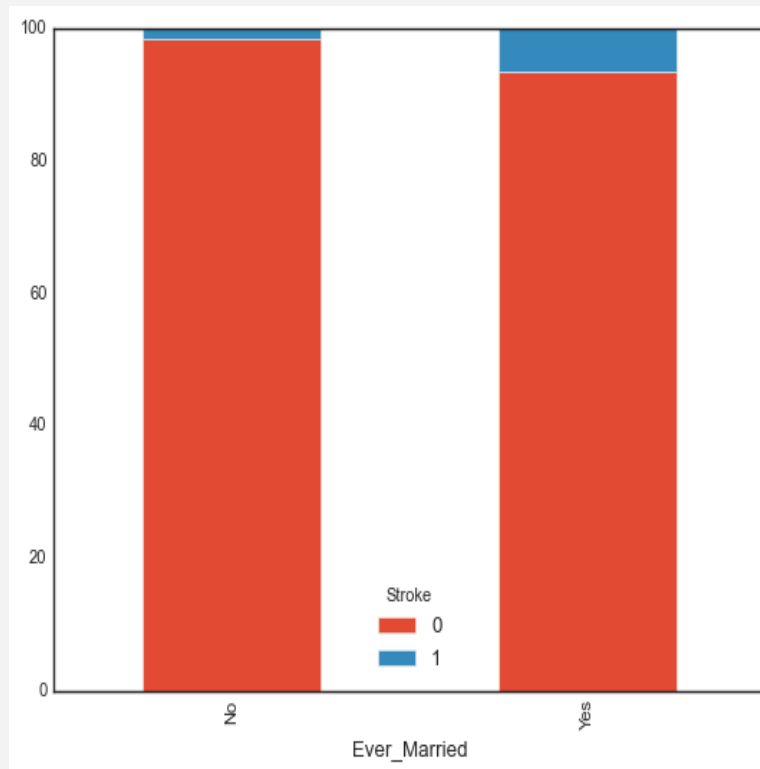


DEMOGRAPHIC ANALYSIS

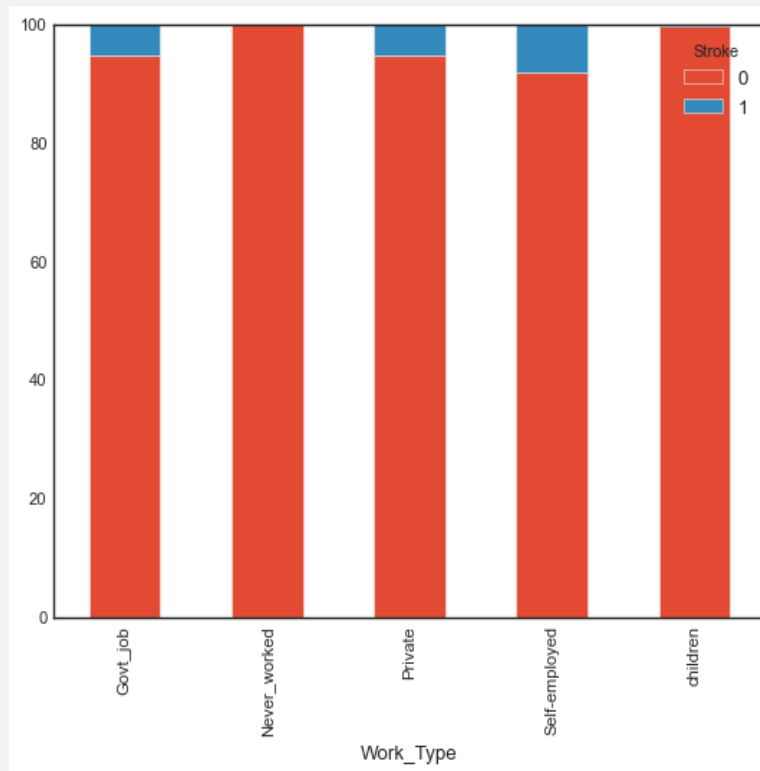
(GENDER)



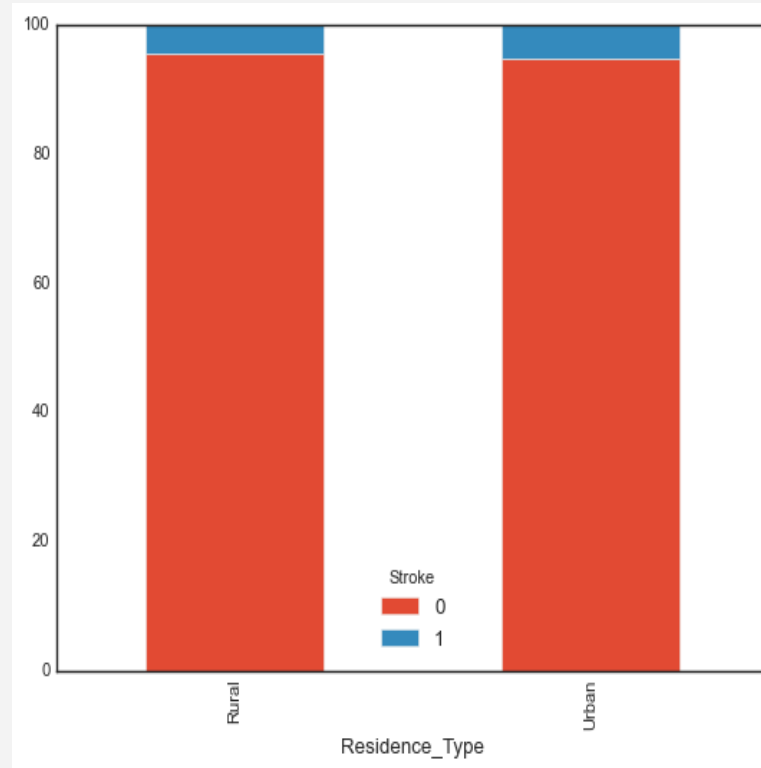
PERSONAL ANALYSIS (EVER MARRIED)



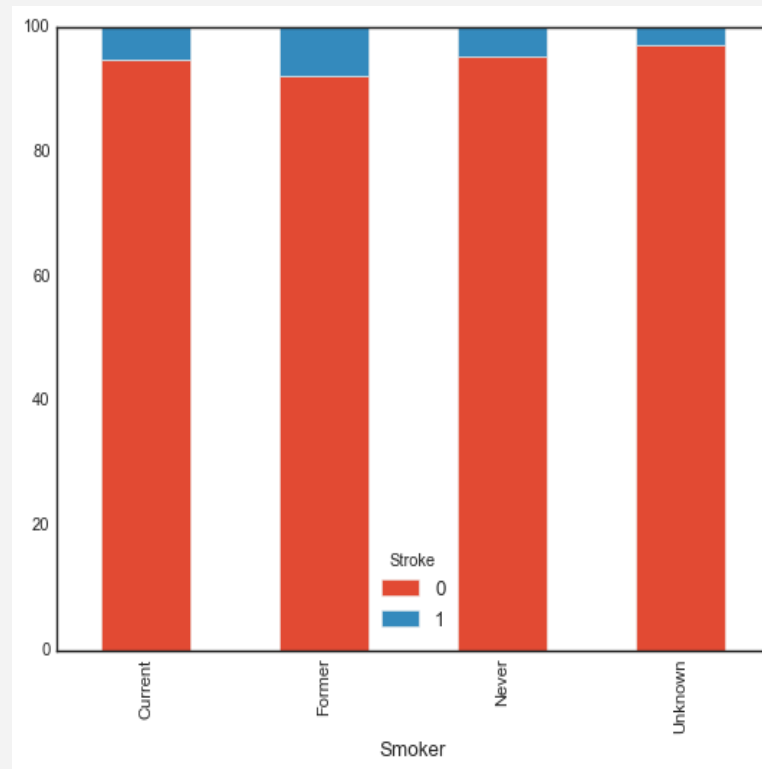
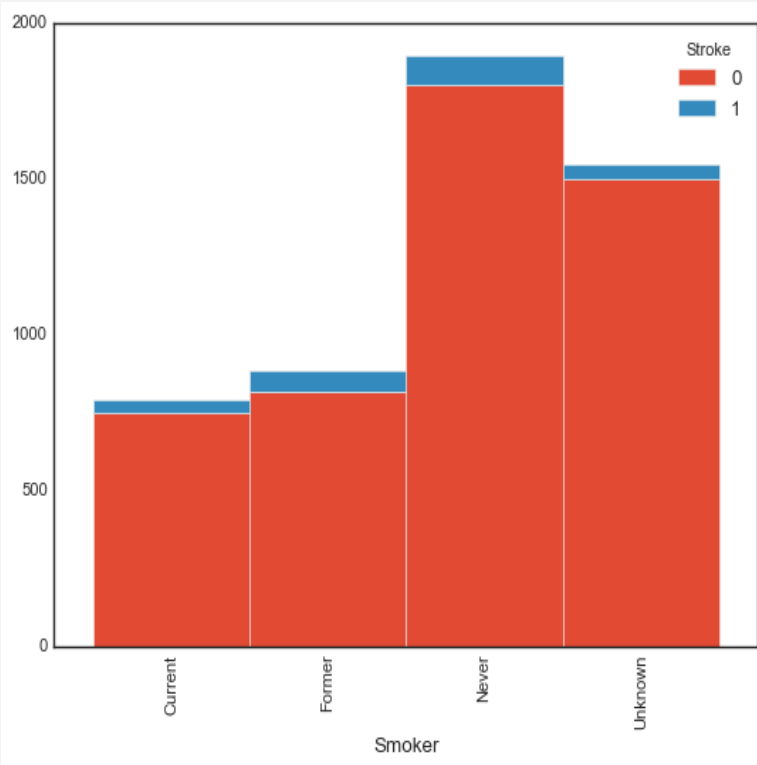
PERSONAL ANALYSIS (WORK TYPE)



PERSONAL ANALYSIS (RESIDENCE TYPE)

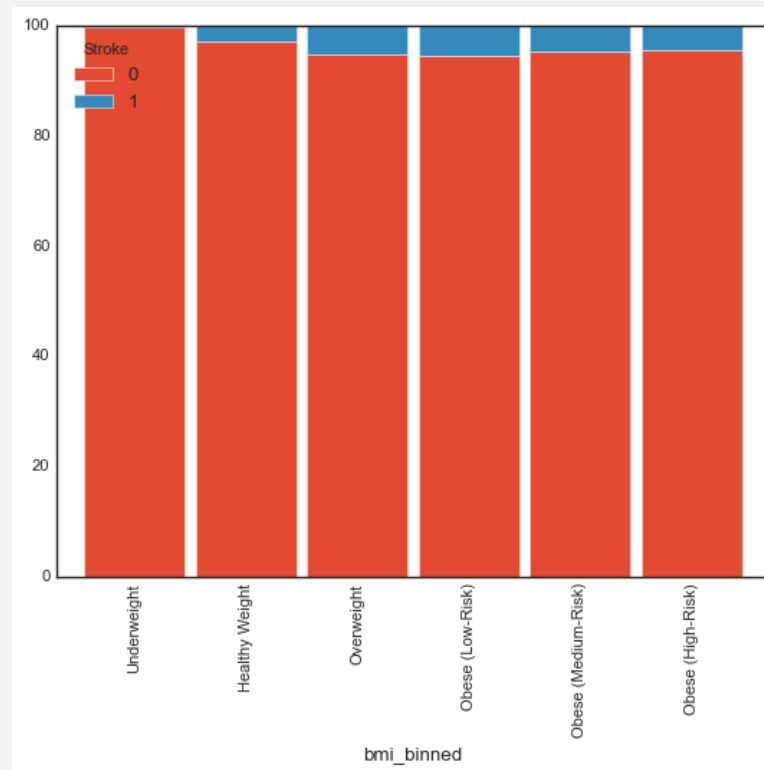
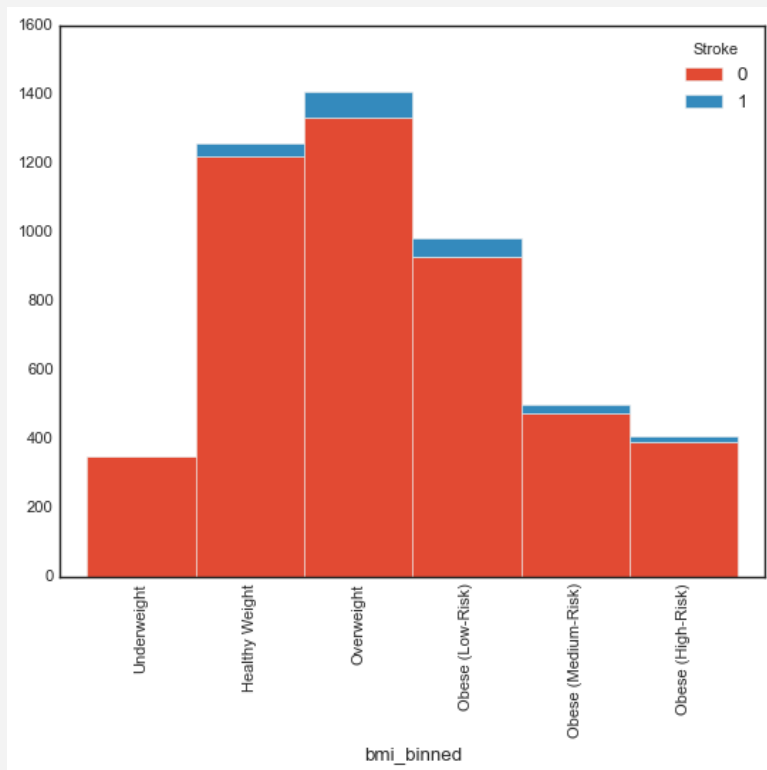


PERSONAL ANALYSIS (SMOKING STATUS)



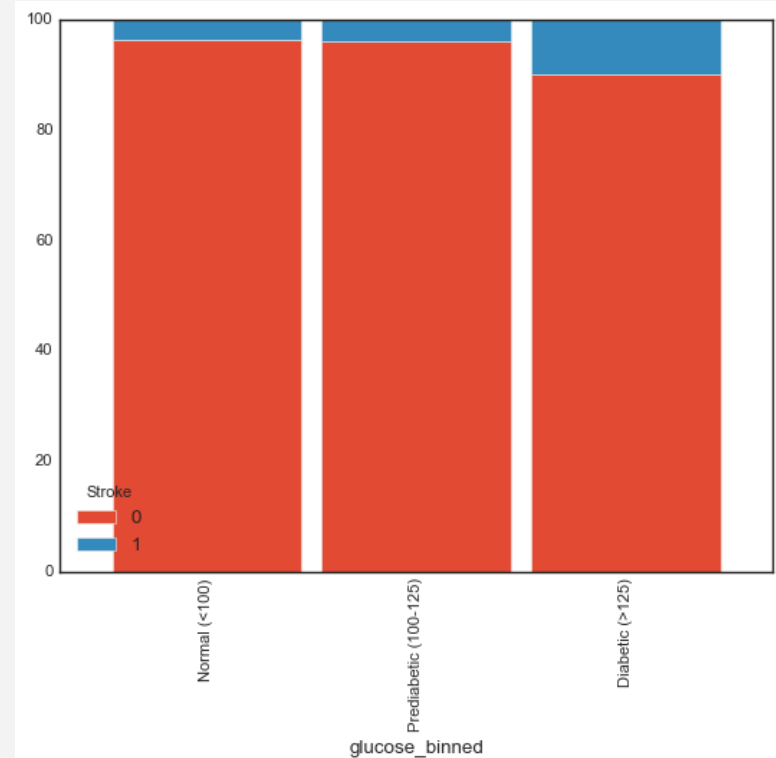
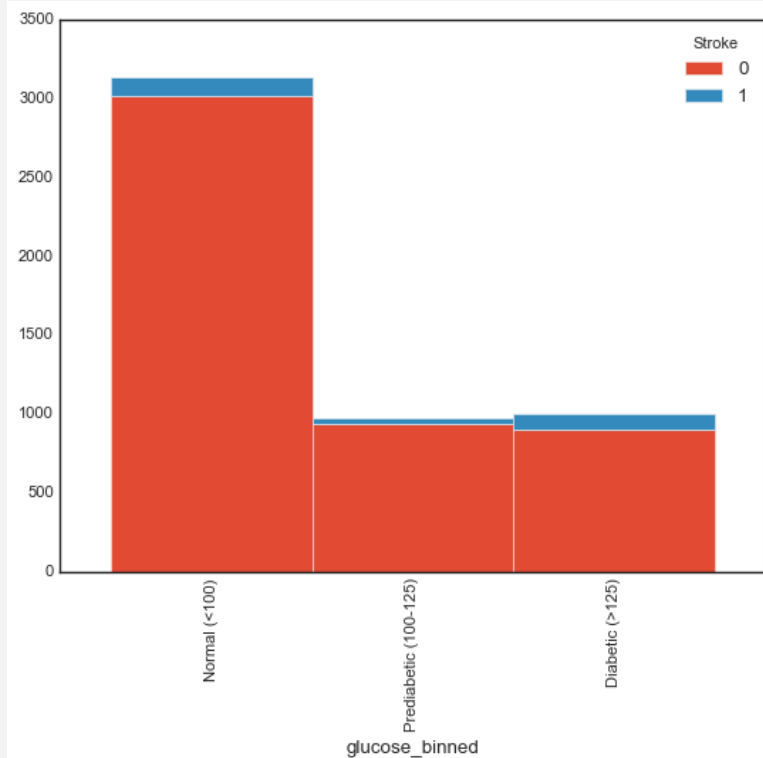
MEDICAL ANALYSIS

(BMI)



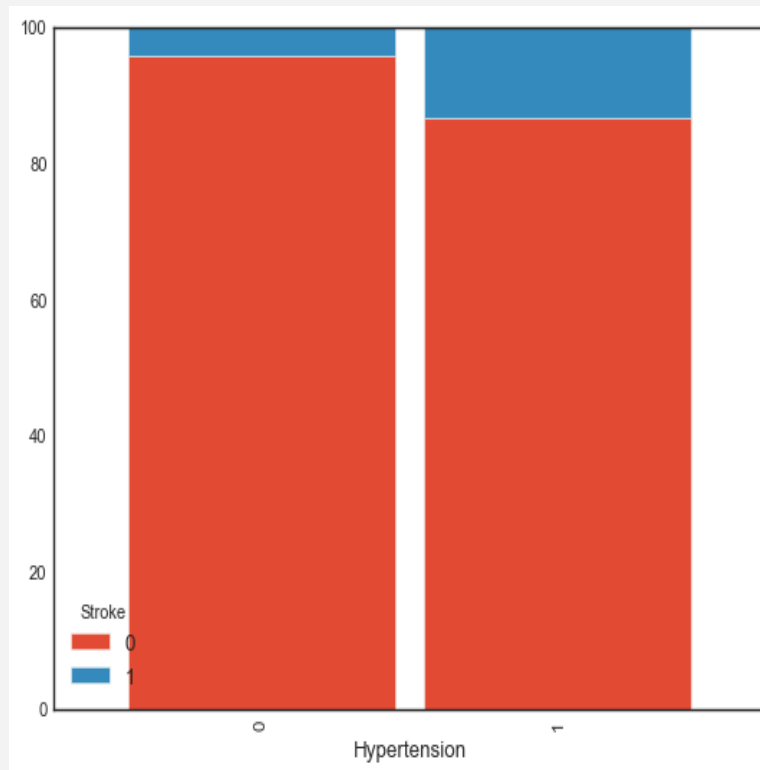
MEDICAL ANALYSIS

(AVG GLUCOSE)

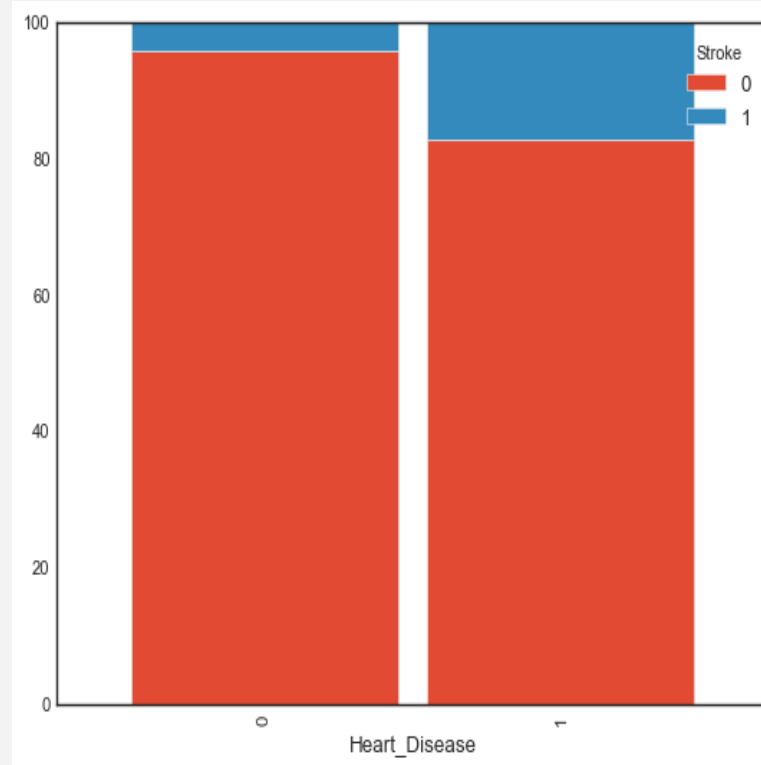


MEDICAL ANALYSIS

(HYPERTENSION)



MEDICAL ANALYSIS (HEART DISEASE)



EDA SUMMARY

Demographics

Stroke risk increases among patients that:

- Are greater in “Age”
- * Stroke risk between “Genders” is undetermined from the dataset

Personal

Stroke risk increases among patients that:

- Are or were married at one point in time
- Are “Self-Employed”, “Private” and “Government” employed workers
- Were “Current” or “Former” smokers
- * Stroke risk between “Residence Types” is undetermined from the dataset

Medical

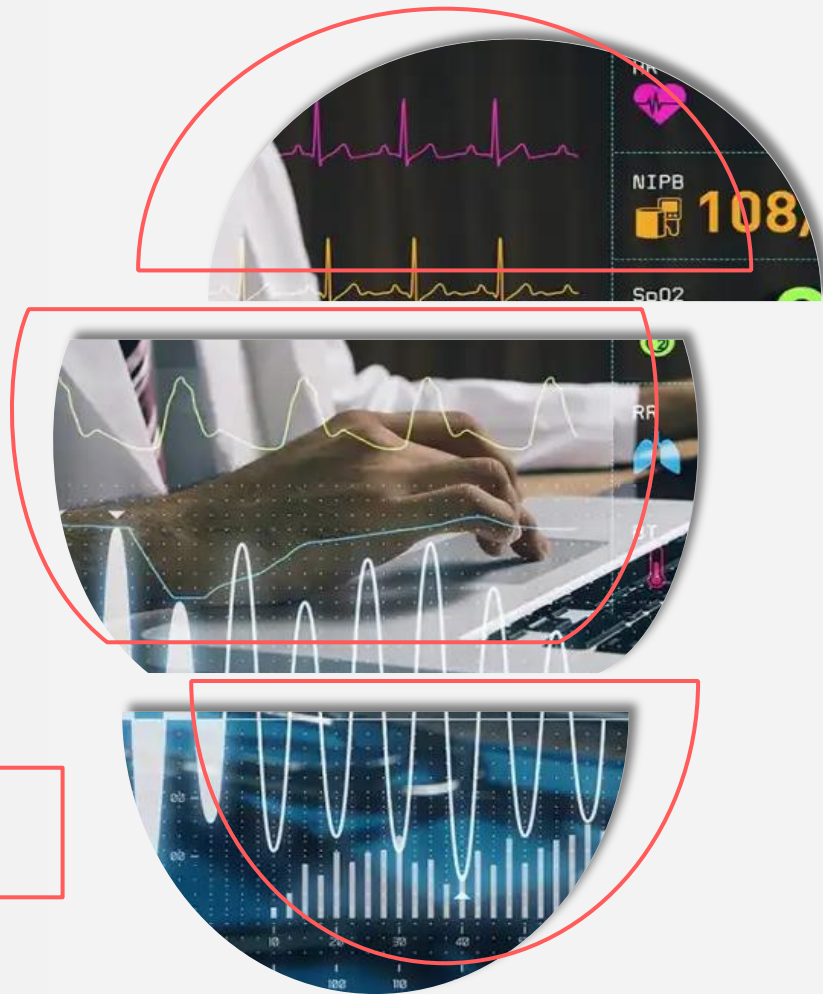
Stroke risk increases among patients that:

- Are “Overweight” or “Obese”
- Have “Diabetes”
- Have “Hypertension”
- Have “Heart Disease”

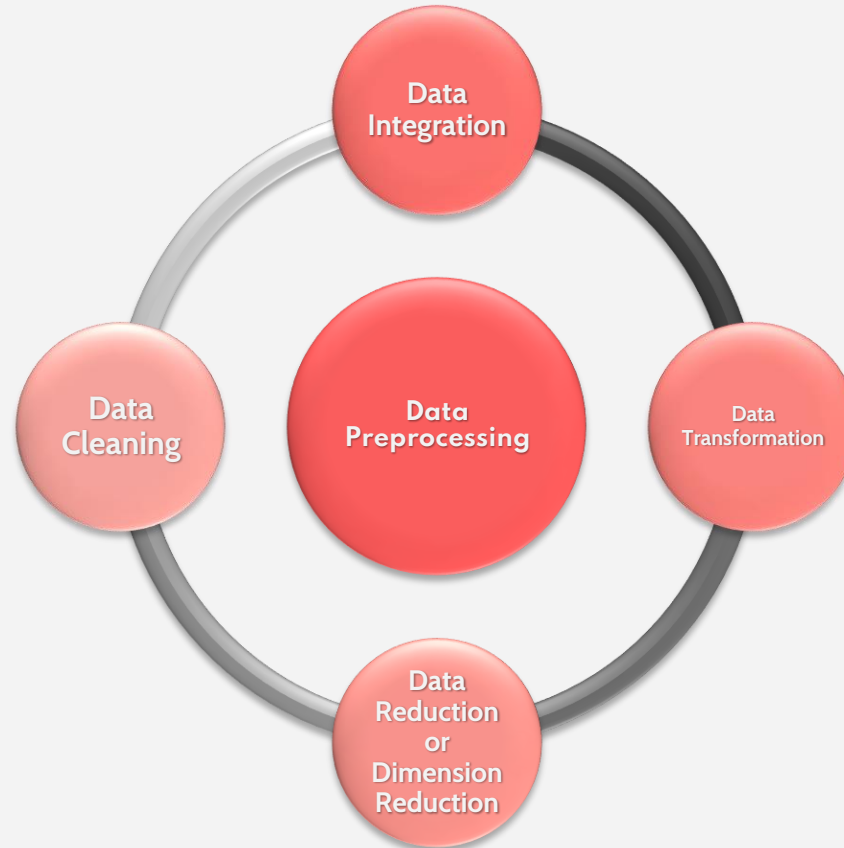
04

ML MODEL

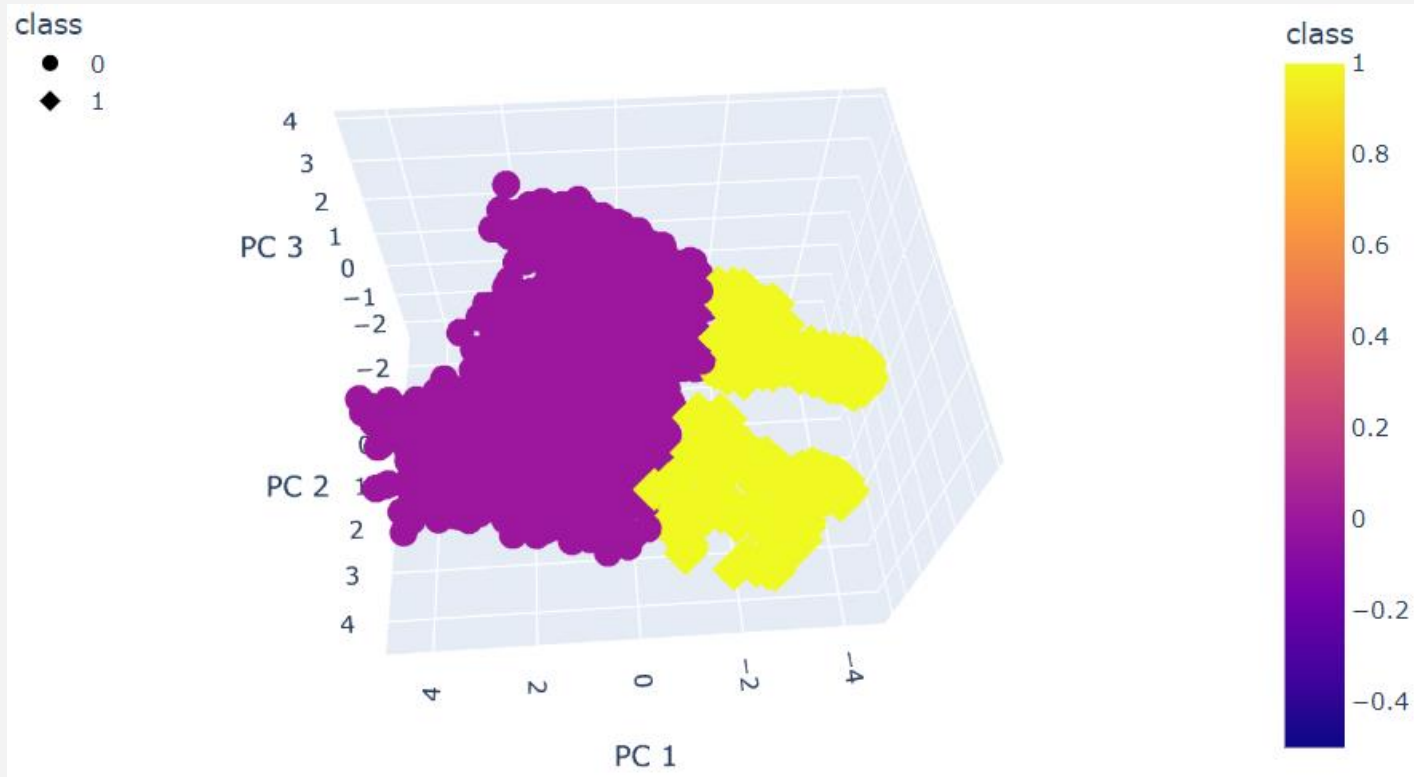
What does the dataset include?
What were our initial findings?



DATA PREPROCESSING



ANALYSIS



05

TOOLS



What applications were needed?
How can we sufficiently display our findings?

PROJECT FRAMEWORK

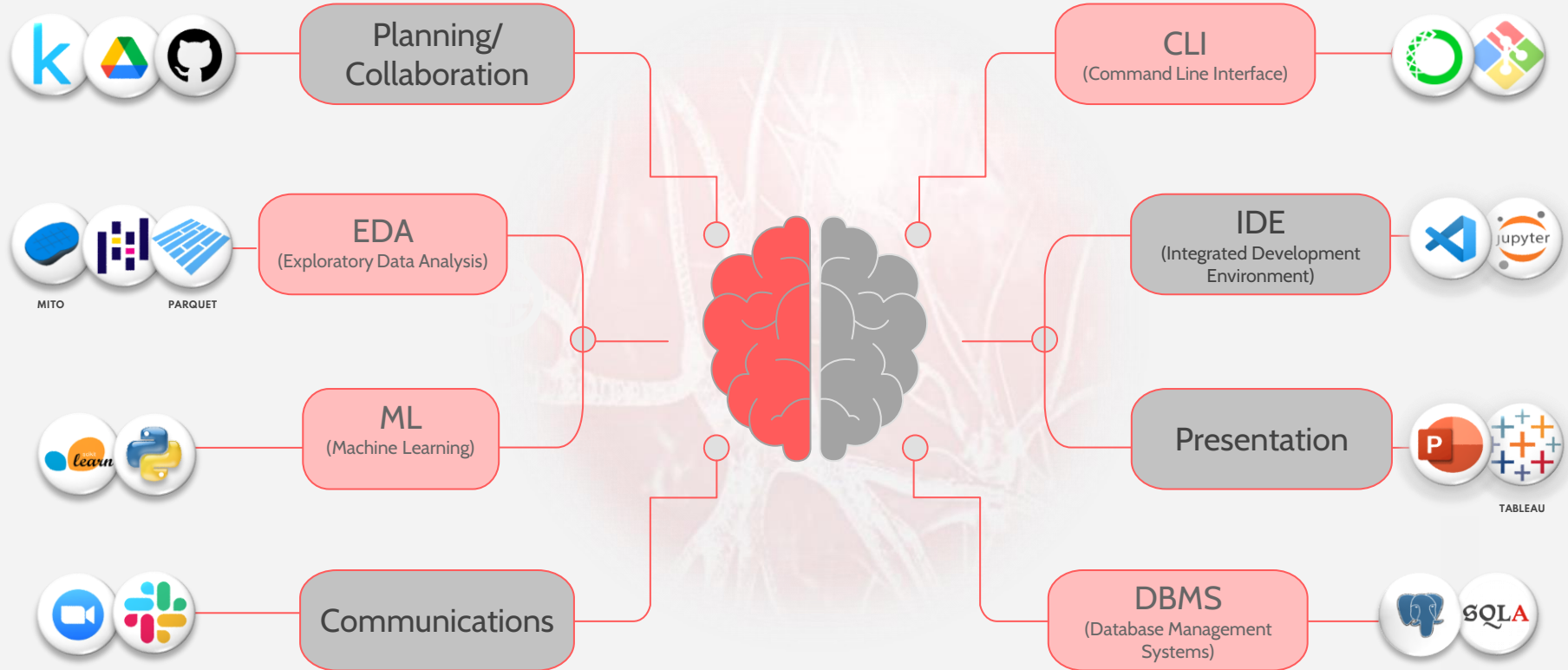




TABLEAU DASHBOARD

06

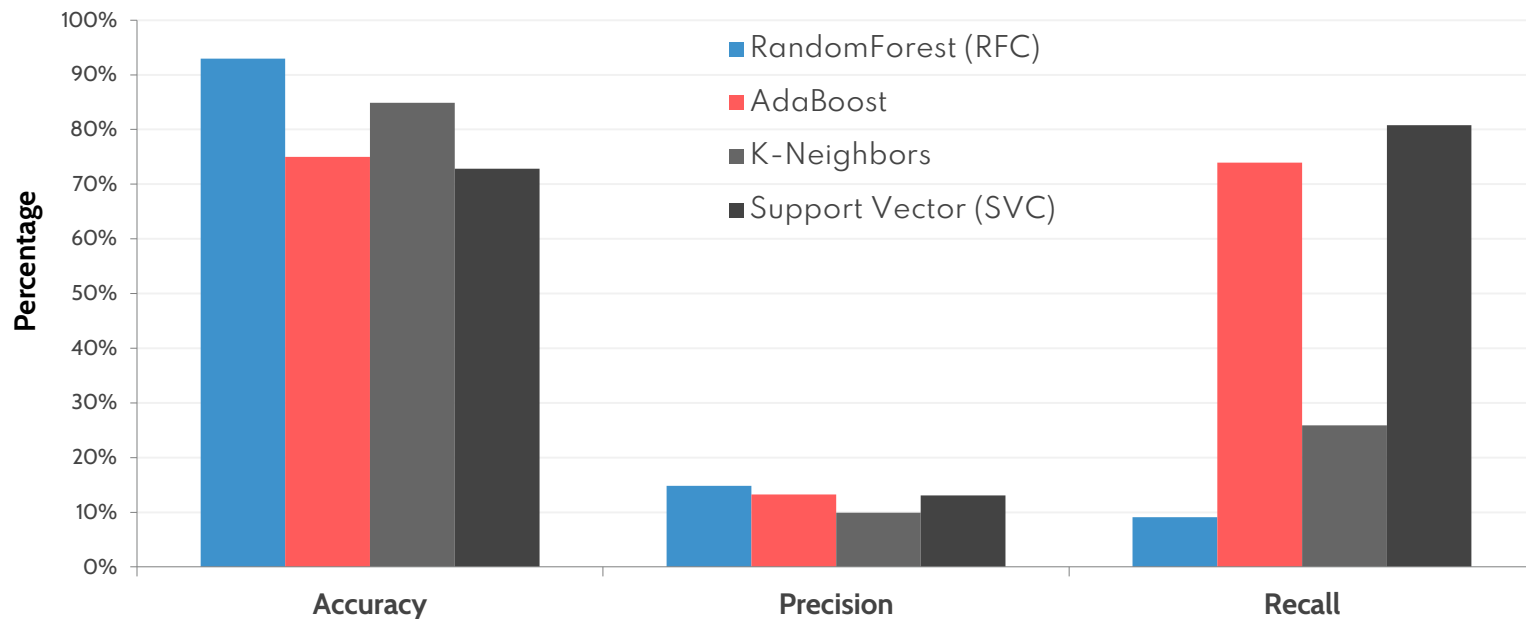
CONCLUSIONS

Which model was most suited to predicting stroke risk?
What changes could have increased the project's utility?



ML MODEL ANALYSIS

Machine Learning Model Comparison



Source: SAML

© 2022 SAML Medical Research

WHAT WOULD WE HAVE DONE DIFFERENTLY?

- ❑ Larger Dataset
- ❑ More ML Models
- ❑ Deeper Correlations
- ❑ Supervised ML
- ❑ Deep Learning
- ❑ Javascript

THANKS!



SAML

DO YOU HAVE ANY QUESTIONS?

info@SAML.com

+1 512 555 SAML

SAML Medical Research (Github)



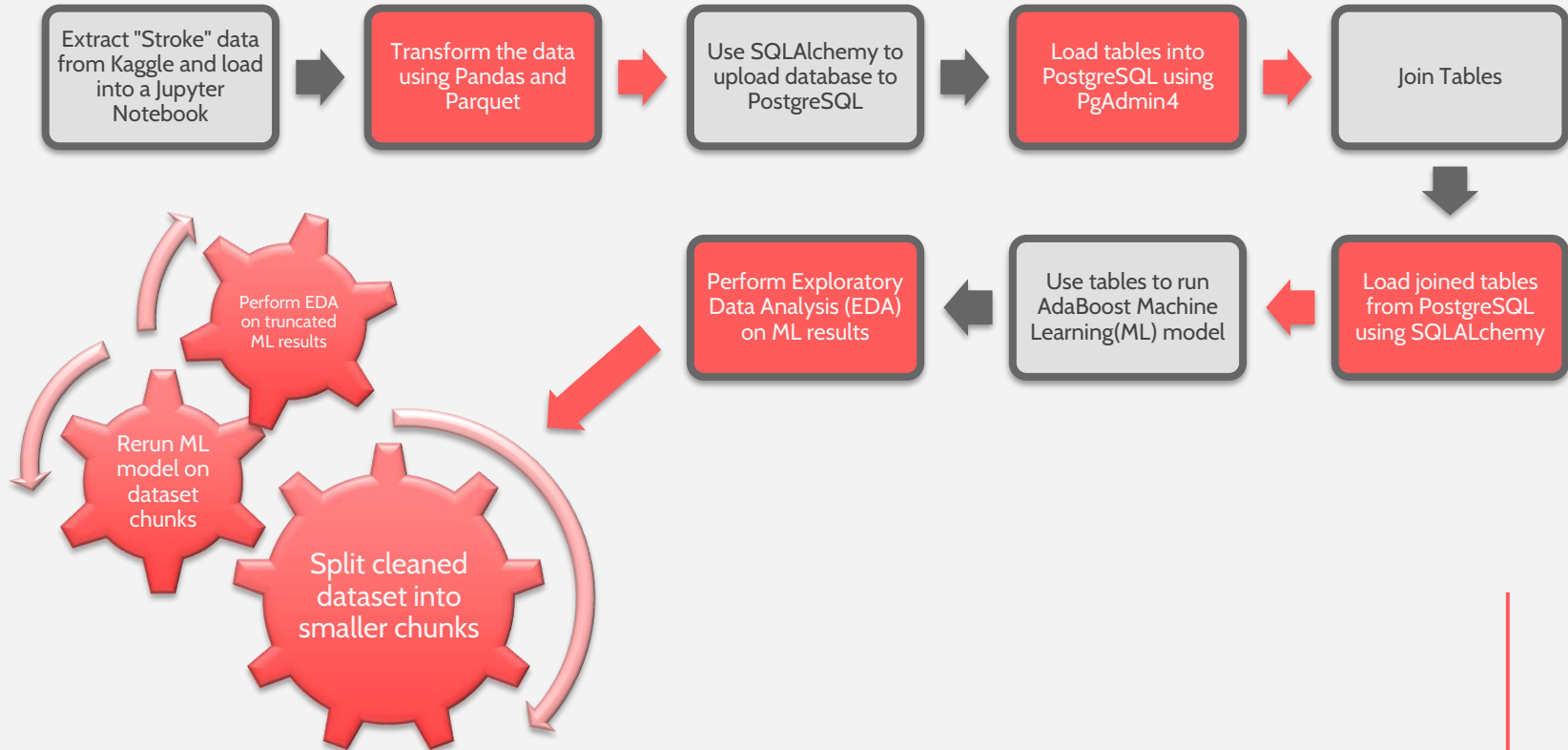


**END OF
PRESENTATION**



PROJECT OUTLINE





















PROJECT OUTLINE





TECHNOLOGIES

TECHNOLOGIES

Languages	Tools	Collaboration & Storage	Communications	CLIs	IDEs	Presentation/ Dashboard
 PYTHON	 PANDAS	 GOOGLE DRIVE	 SLACK	 ANACONDA	 JUPYTER NOTEBOOK	 TABLEAU
 POSTGRESQL	 MITO	 GITHUB	 ZOOM	 GITBASH	 VS CODE	 POWER POINT
	 PARQUET	 KAGGLE			 PgADMIN	
	 SCIKIT-LEARN	 GIT				
	 SQLALCHEMY					

A decorative red line runs vertically down the left side of the slide. In the top right corner, there are several red geometric shapes, including a semi-circle and a rectangle with rounded corners.

COLOR & FONTS

Fonts & colors used

This presentation has been made using the following fonts:

Spartan

(<https://fonts.google.com/specimen/Spartan>)

Cabin

(<https://fonts.google.com/specimen/Cabin>)

#434343

#f3f3f3

#ff5b5b

#666666