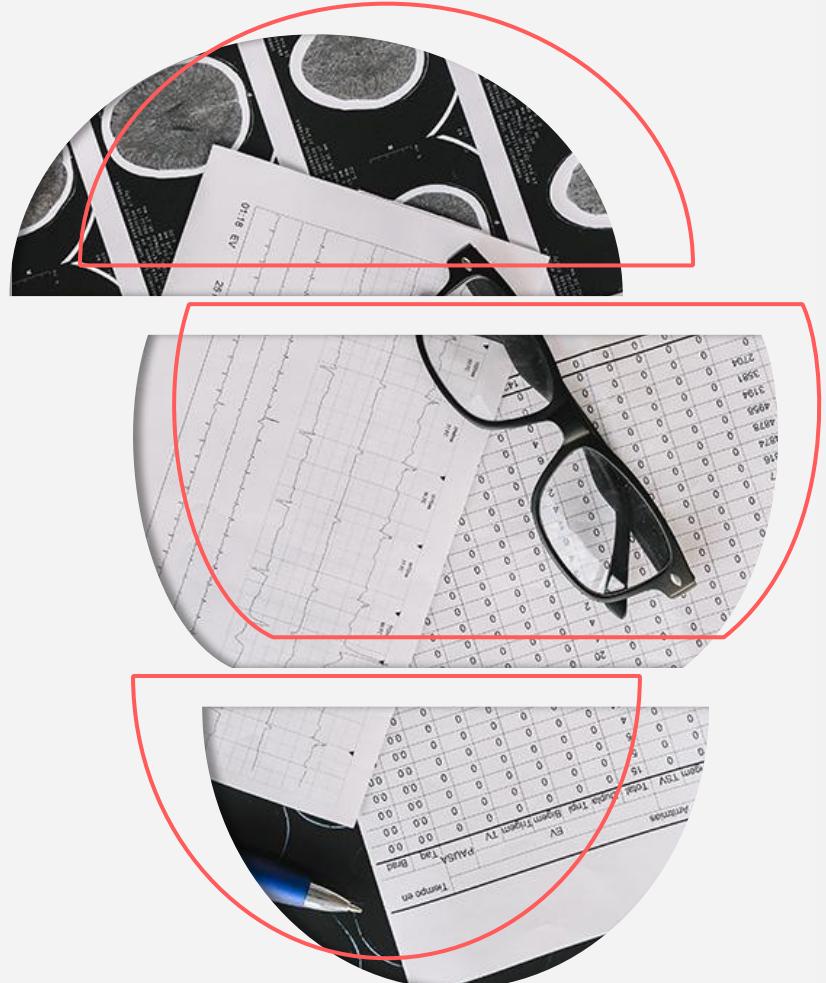




SAML

*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

- Analysis

05

## TOOLS

- Project Framework
- Tableau Dashboard

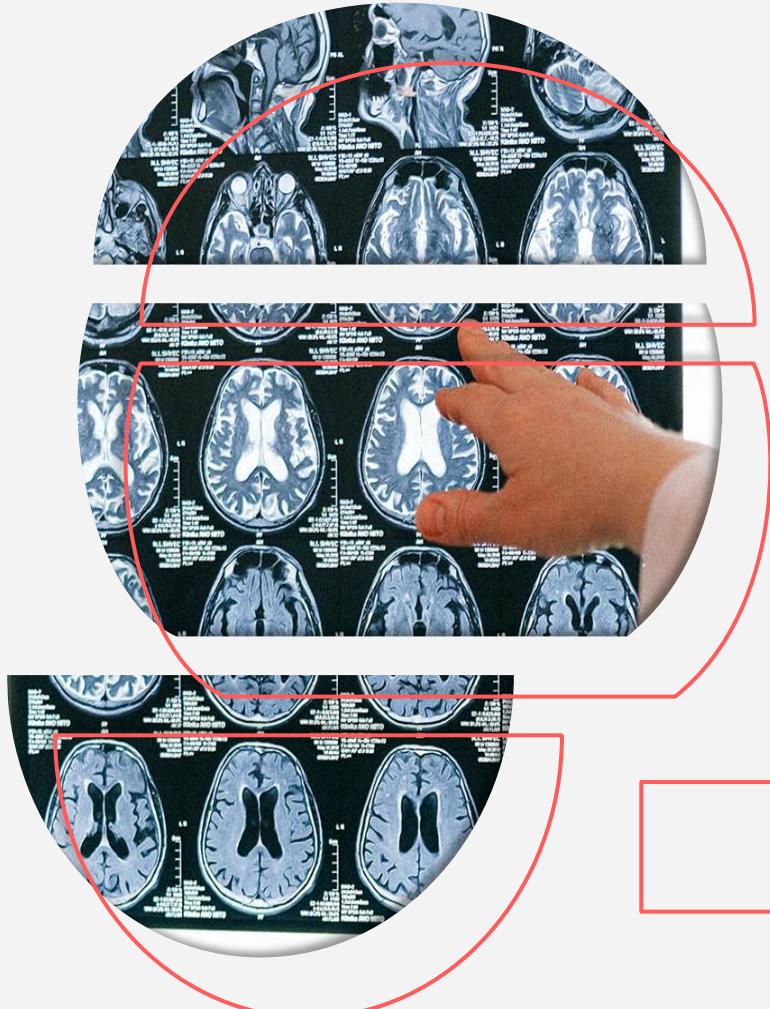
06

## CONCLUSIONS

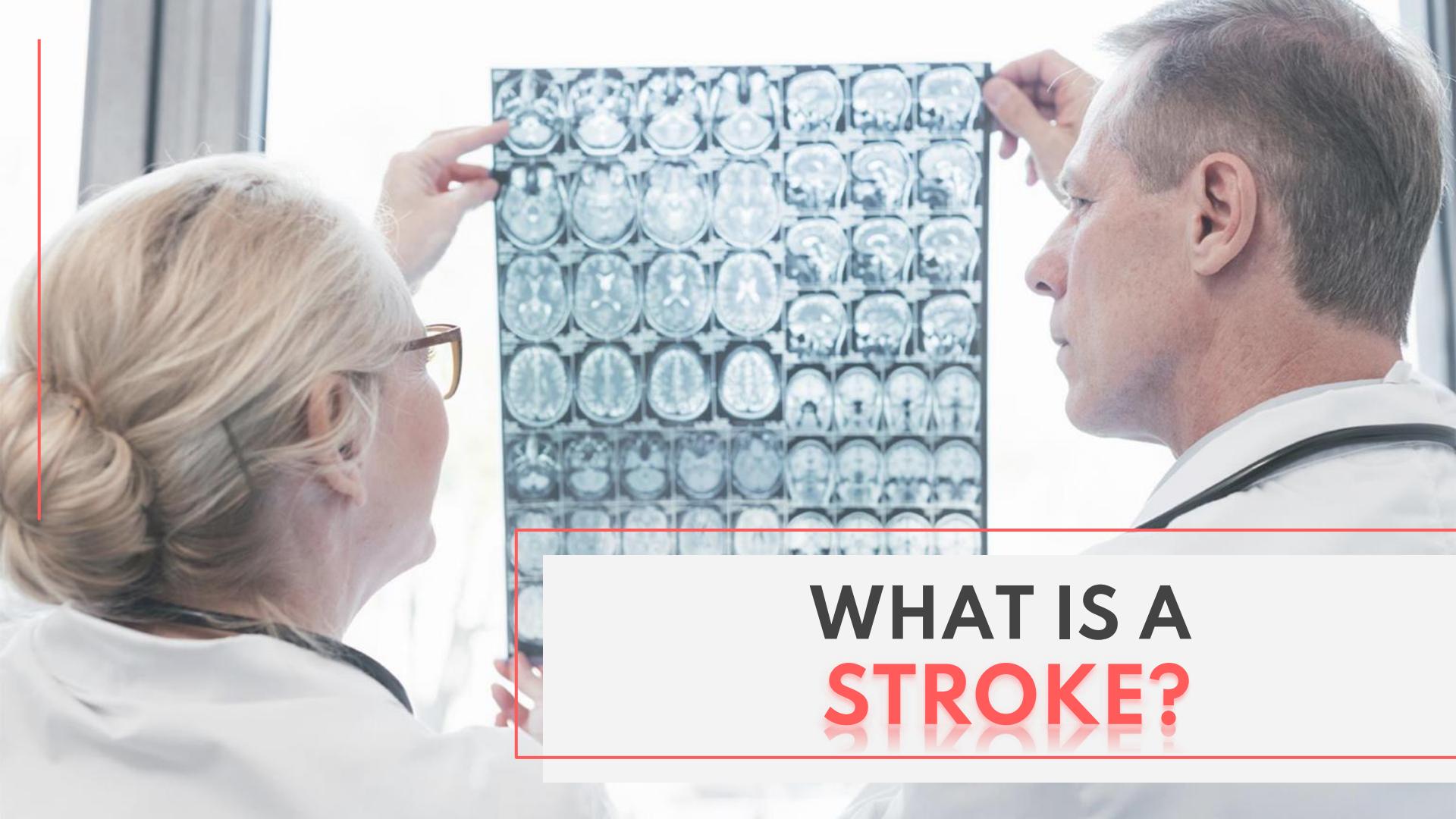
- Results
- Recomendations

# 01

# INTRODUCTION



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



# WHAT IS A **STROKE?**

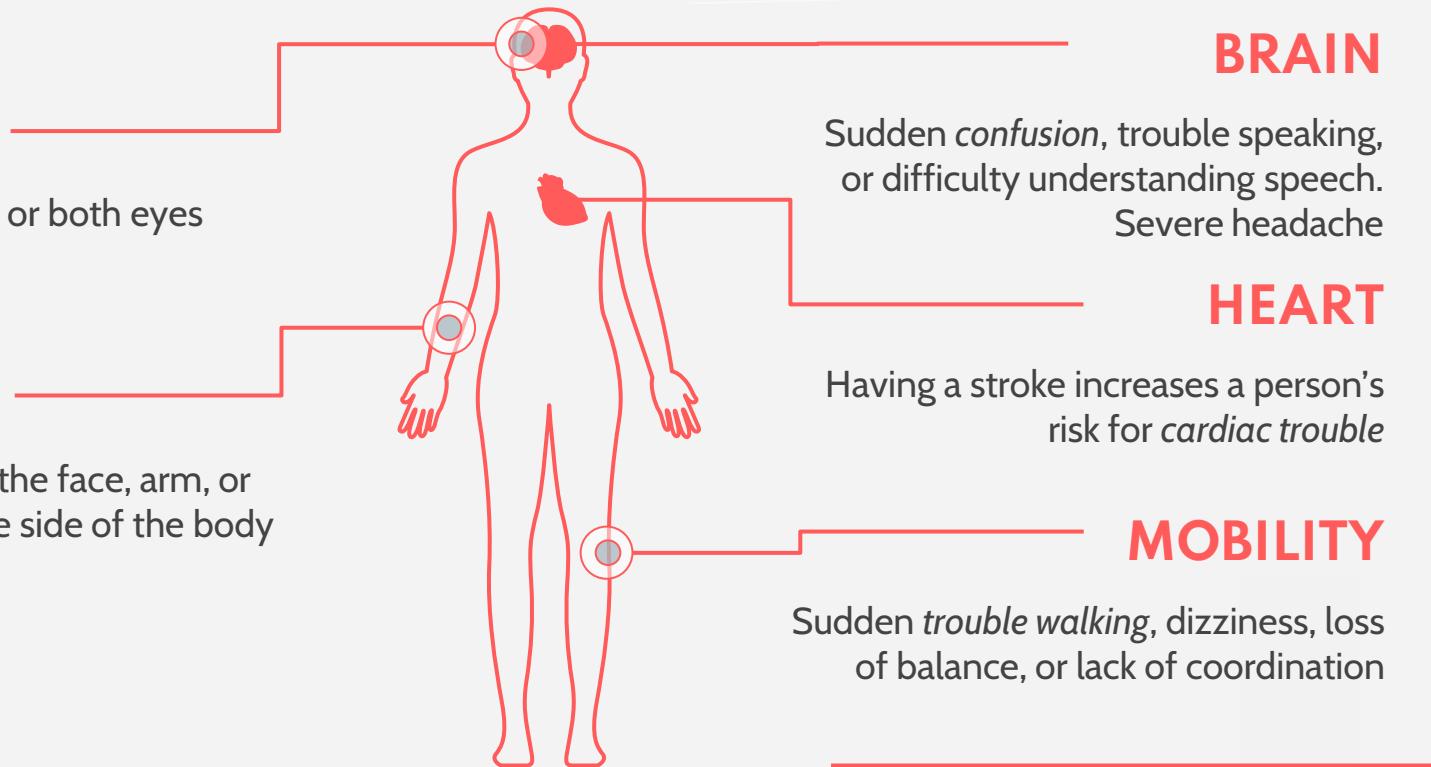
# SYMPTOMS OF A STROKE

## 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
	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

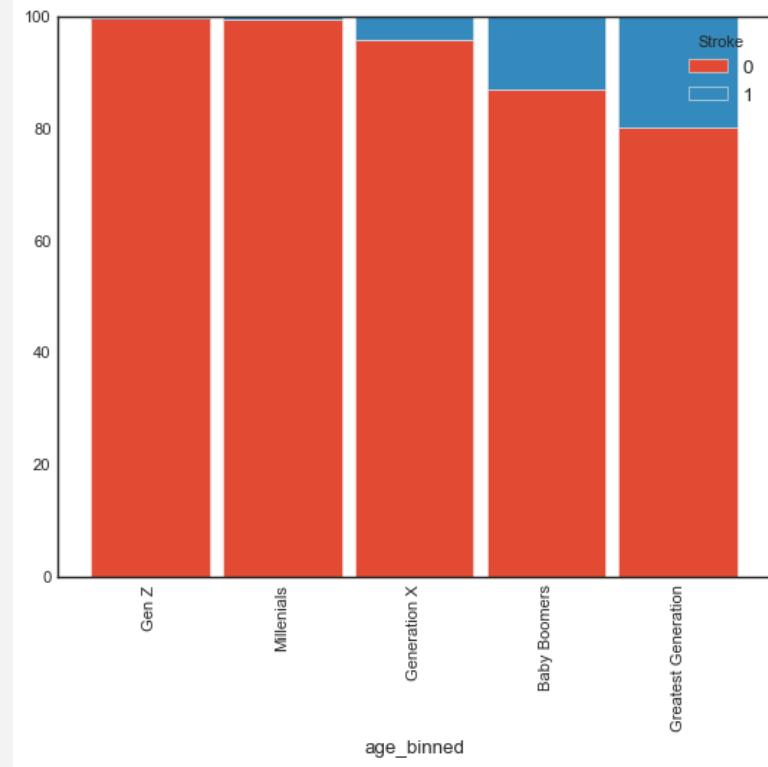
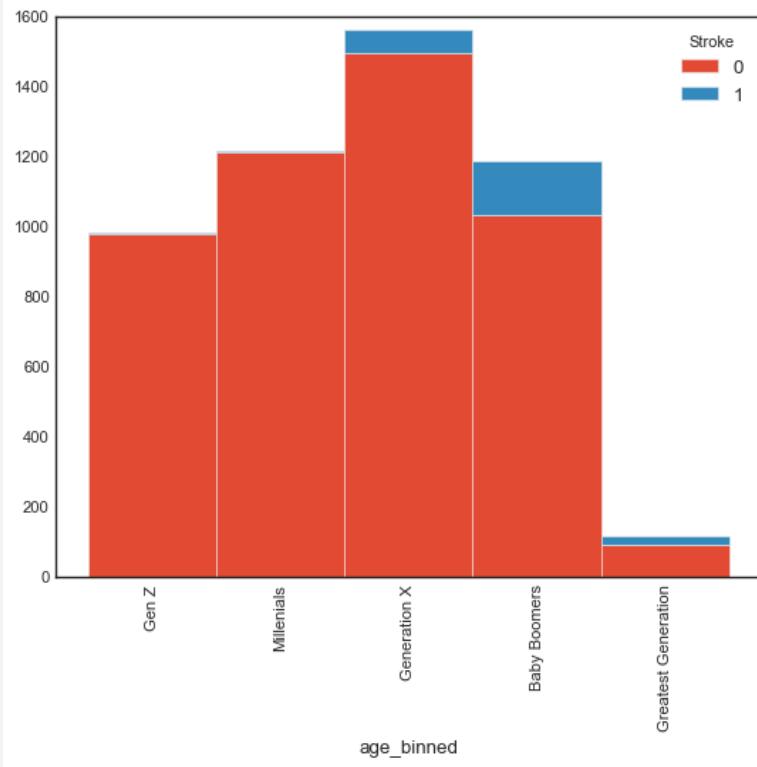


## BMI

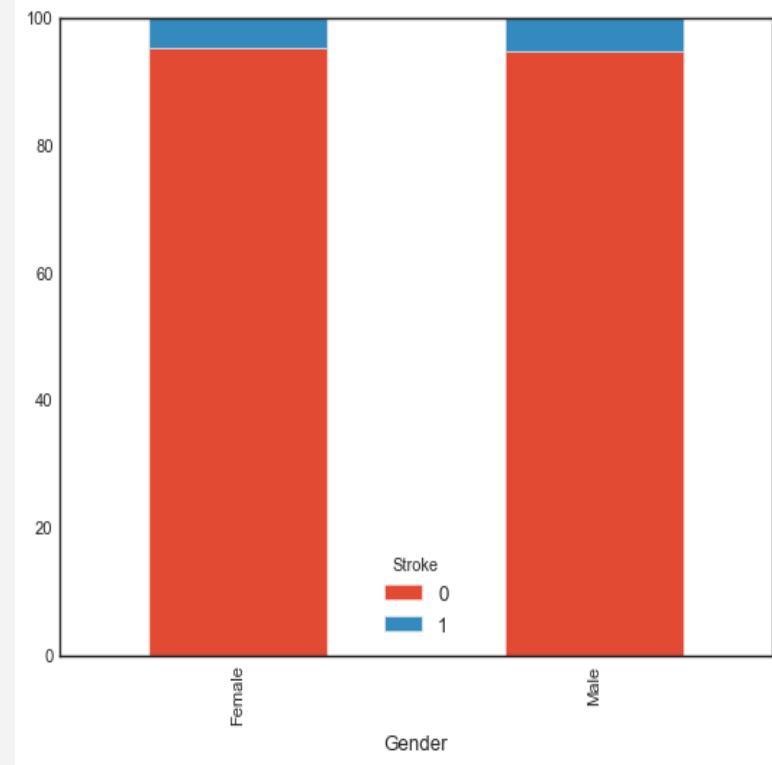
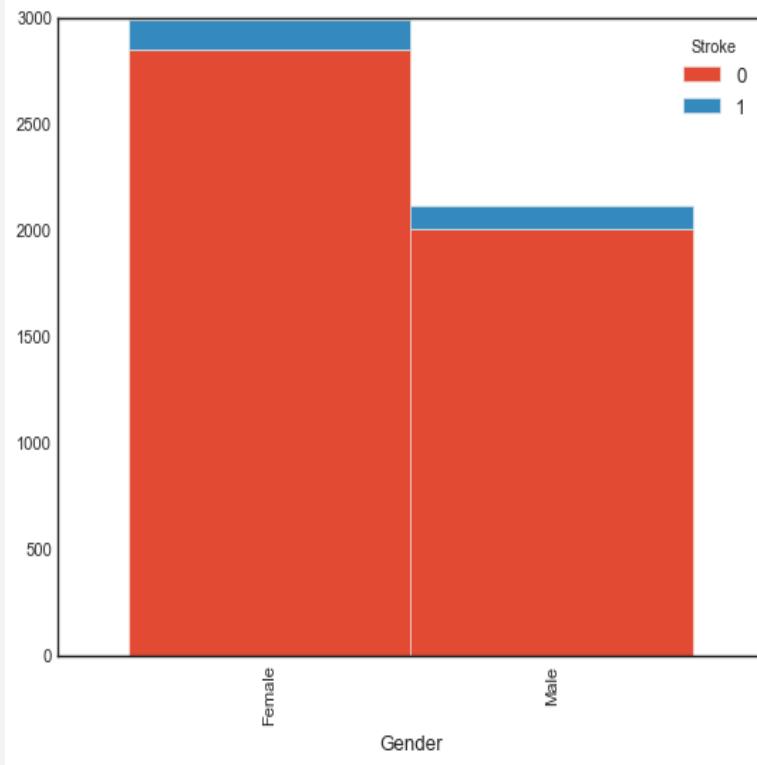
From 10 to 98

# DEMOGRAPHIC ANALYSIS

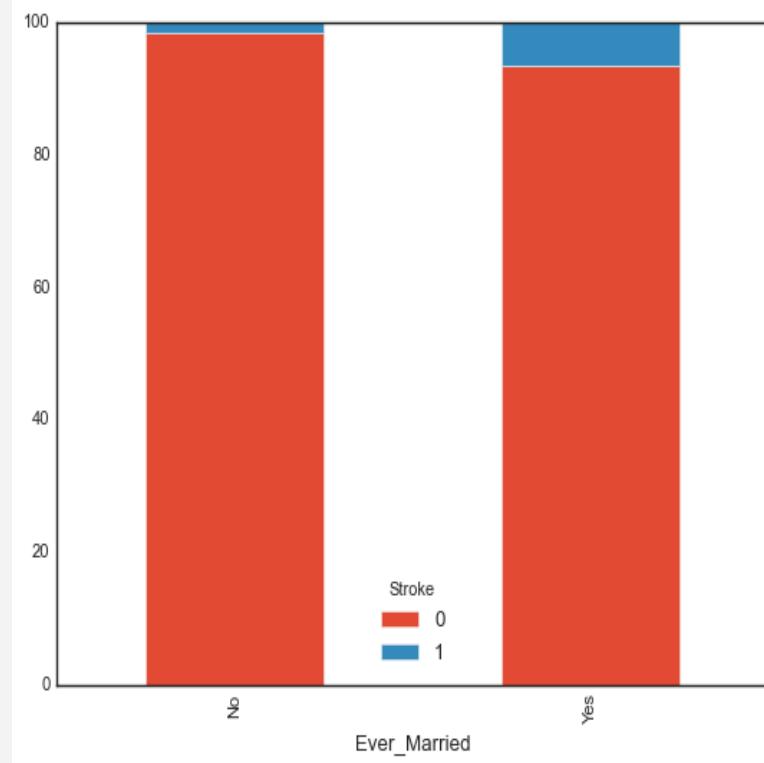
## (AGE)



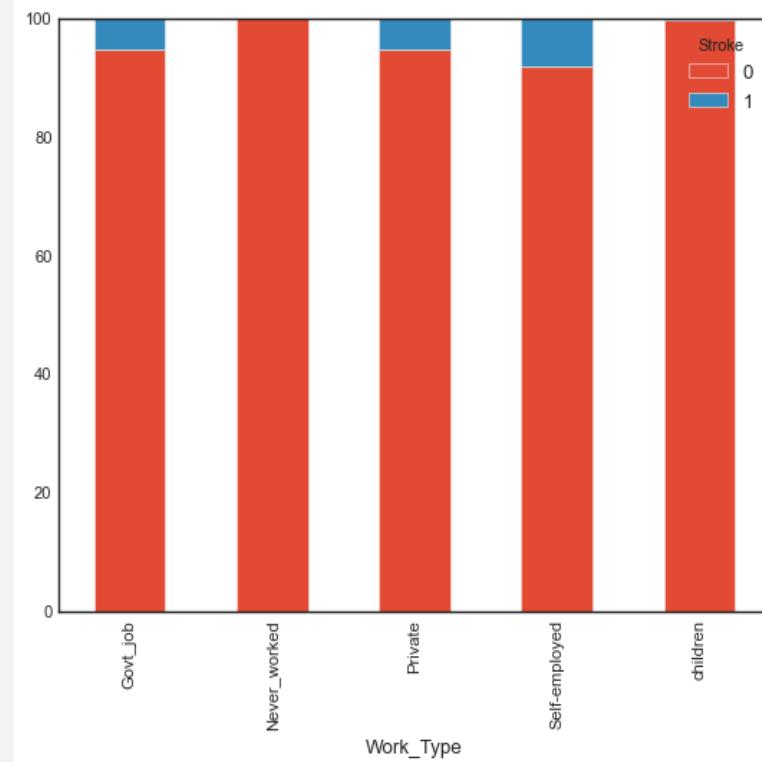
# DEMOGRAPHIC ANALYSIS (GENDER)



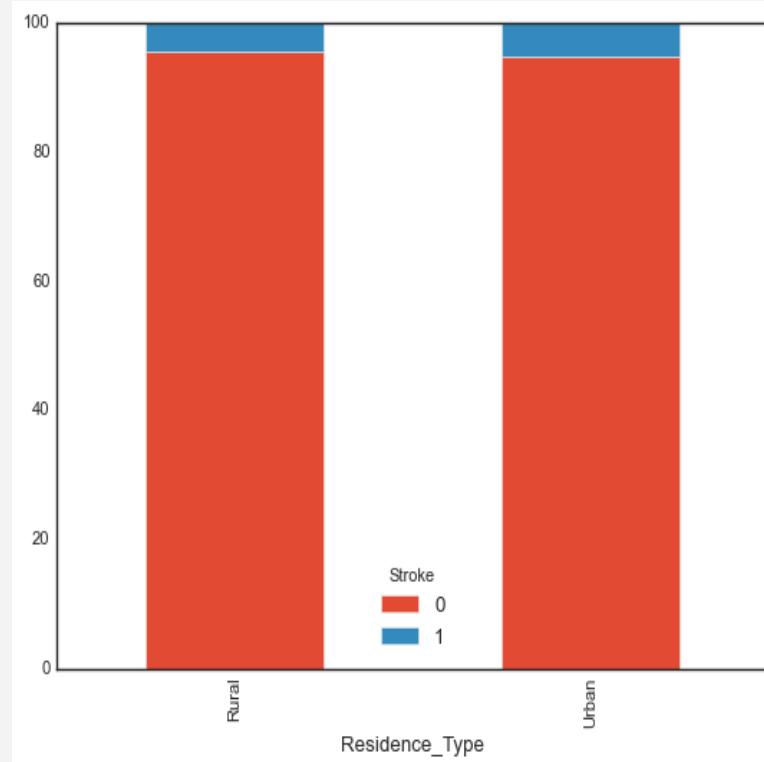
# PERSONAL ANALYSIS (EVER MARRIED)



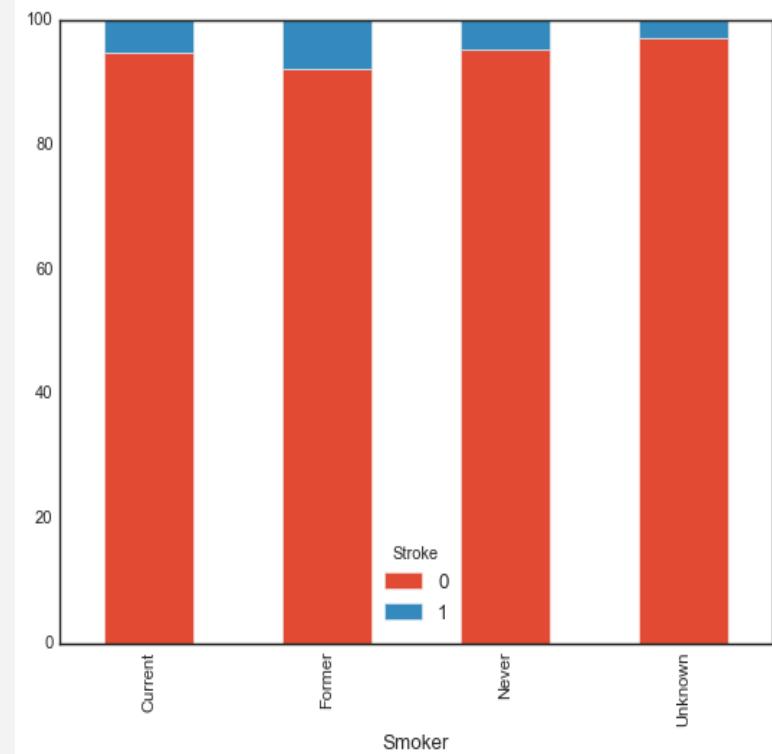
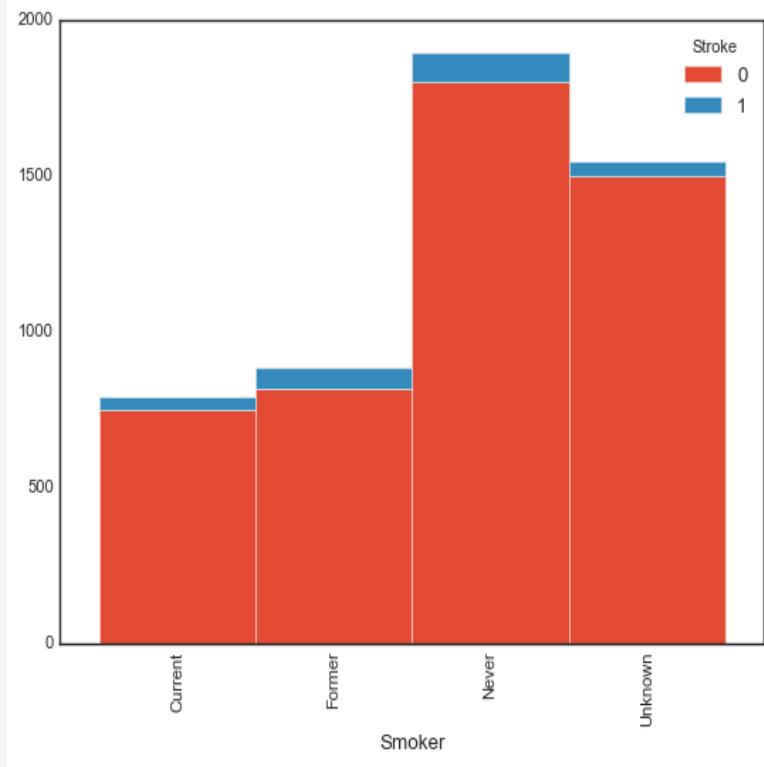
# PERSONAL ANALYSIS (WORK TYPE)



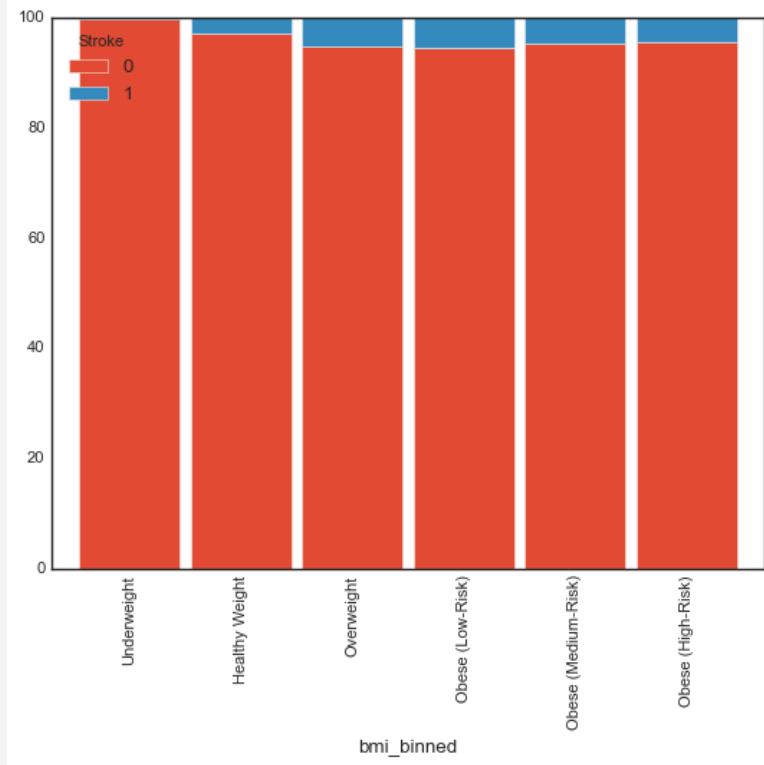
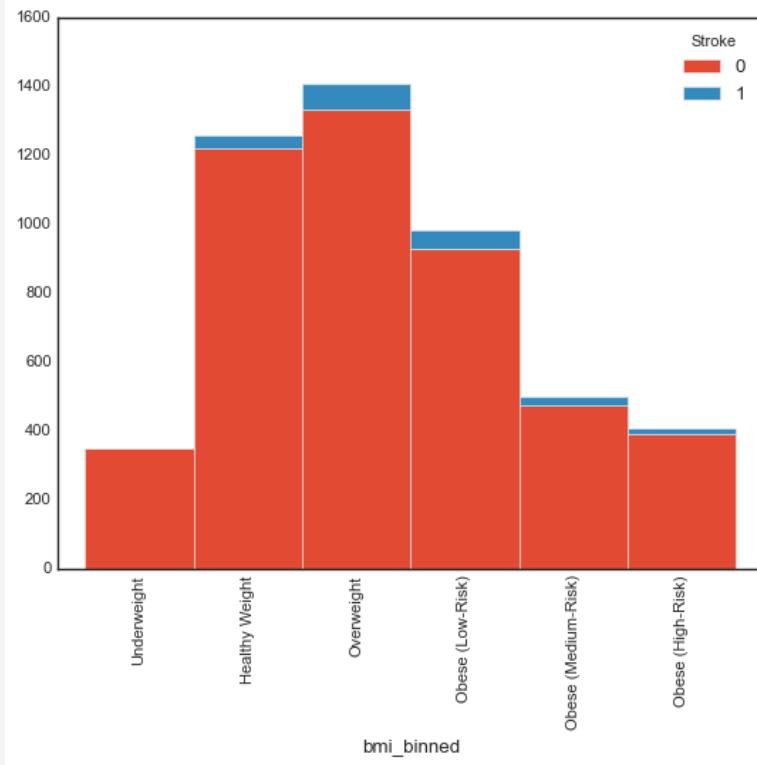
# PERSONAL ANALYSIS (RESIDENCE TYPE)



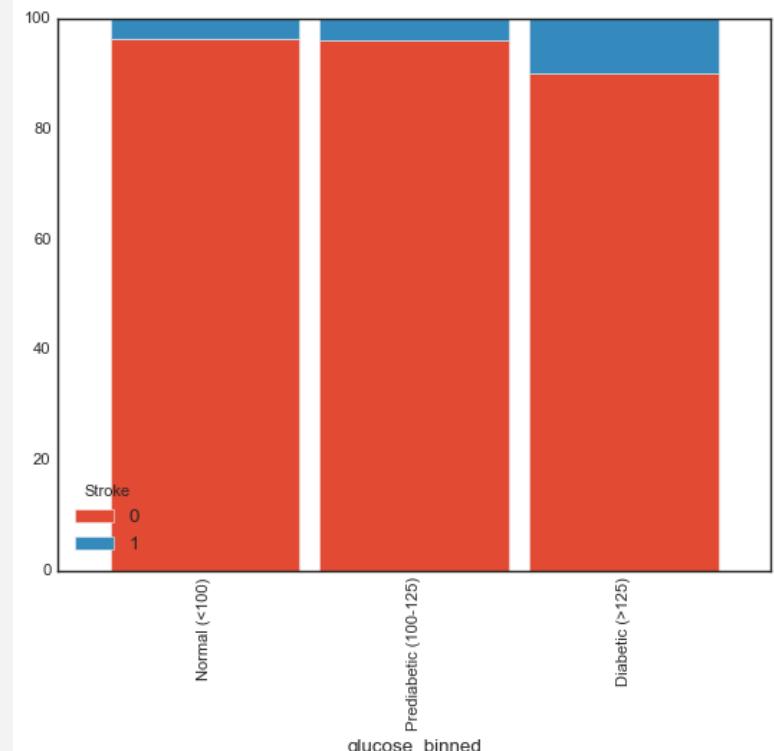
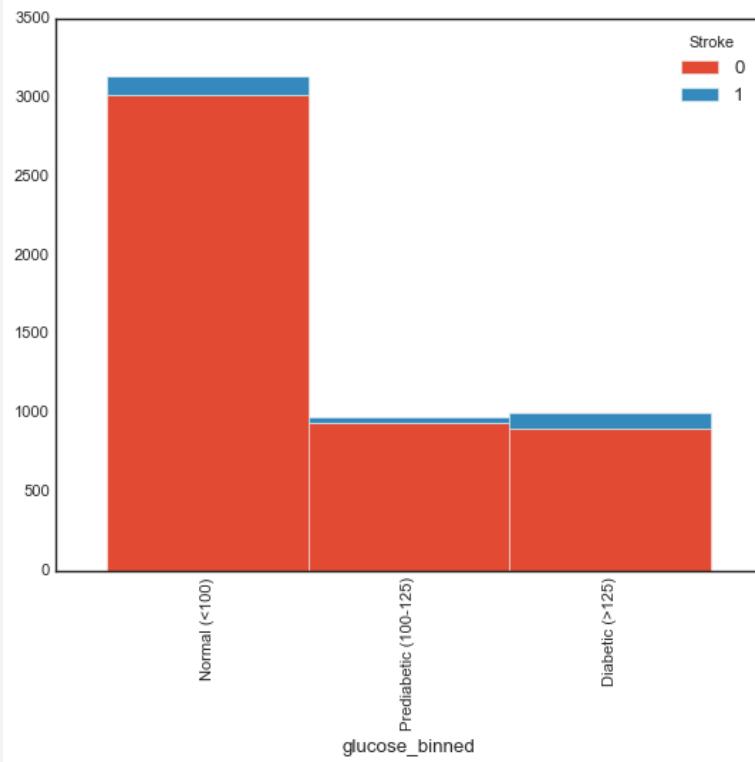
# PERSONAL ANALYSIS (SMOKING STATUS)



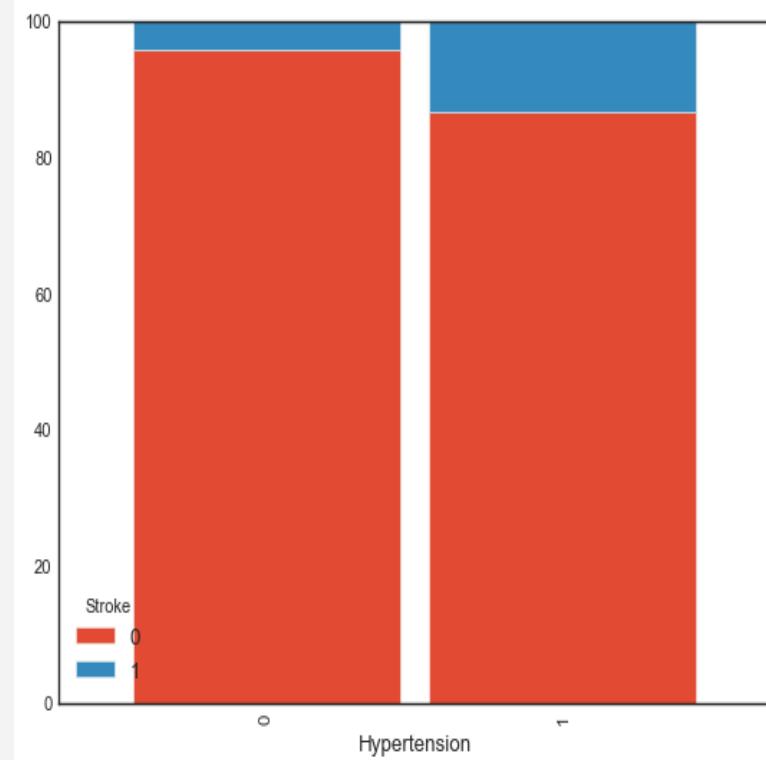
# MEDICAL ANALYSIS (BMI)



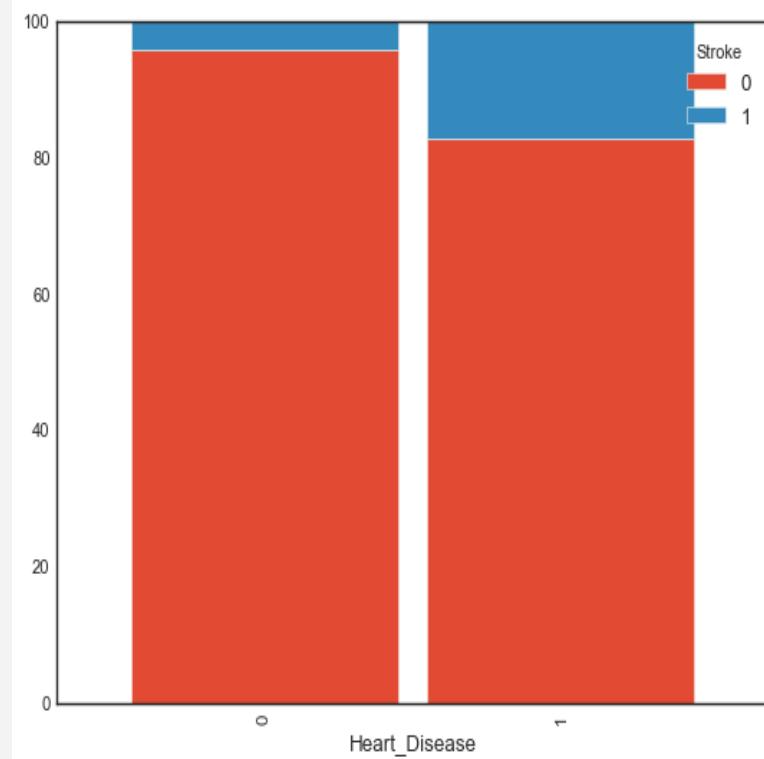
# MEDICAL ANALYSIS (AVG GLUCOSE)



# 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:

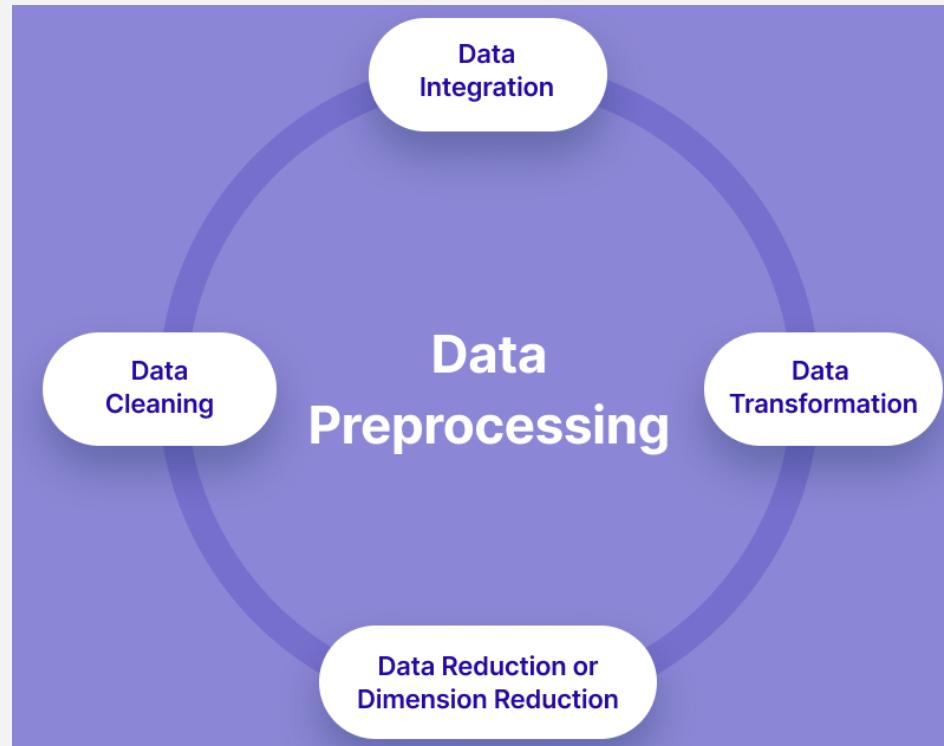
- Are “Overweight” or “Obese”
- Have “Diabetes”
- Have “Hypertension”
- Have “Heart Disease”

# 04

# ML MODEL

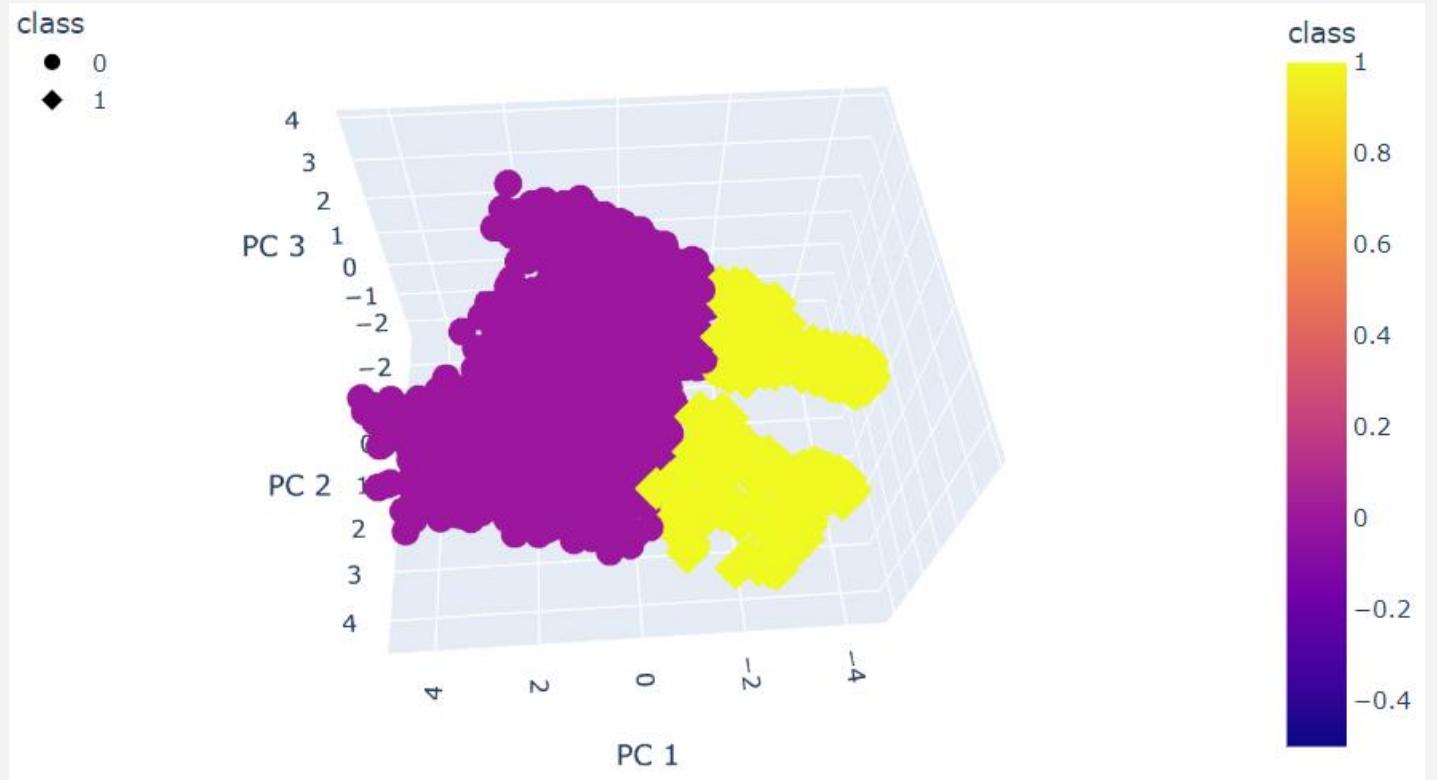


# ANALYSIS





# ANALYSIS





---

# SUPERVISED ANALYSIS

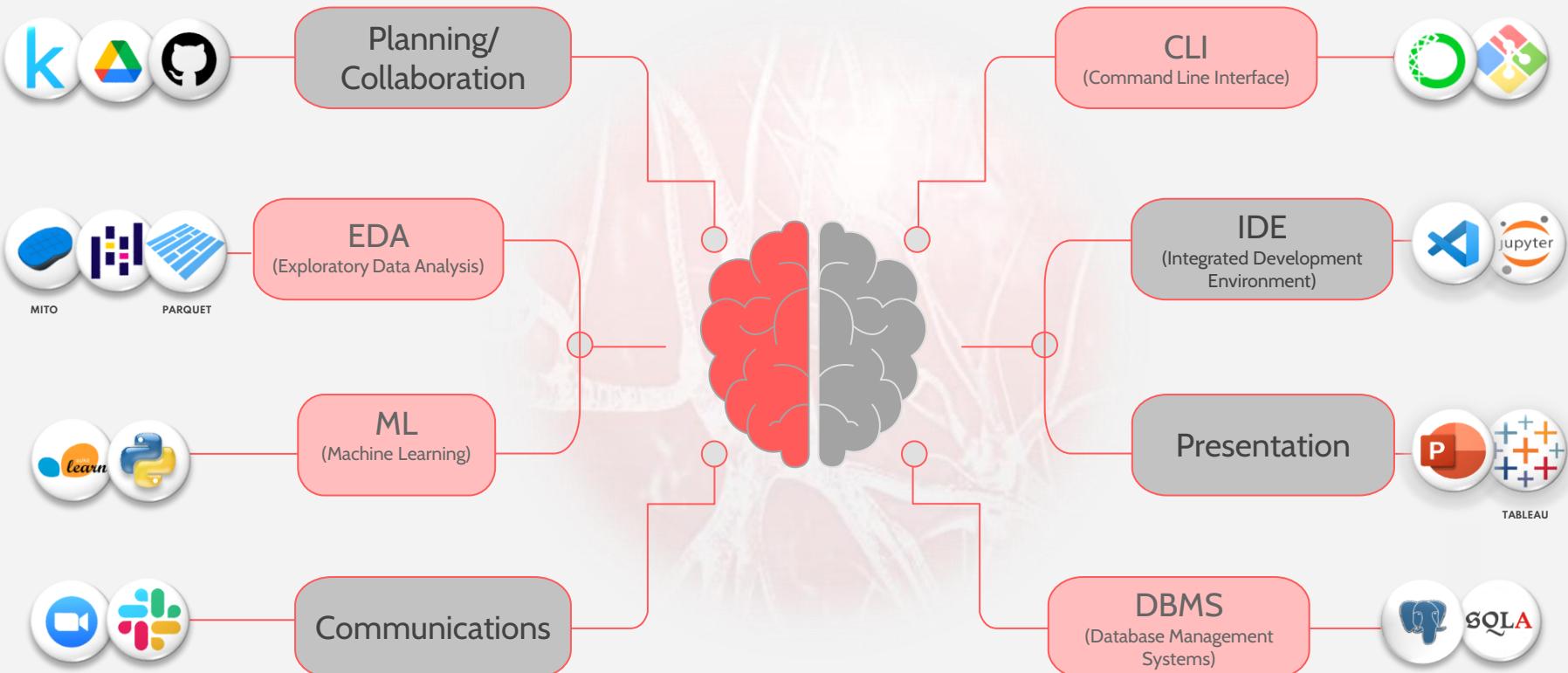
---

05

# TOOLS



# PROJECT FRAMEWORK

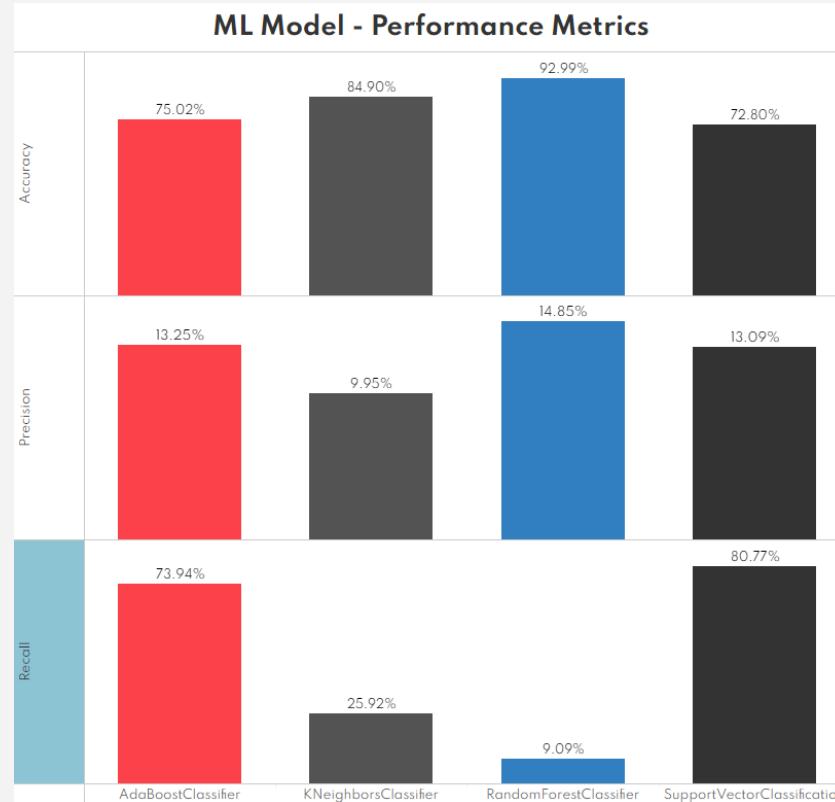


06

# CONCLUSIONS



# ANALYSIS



# WHAT WOULD WE HAVE DONE DIFFERENTLY?

1. Find a dataset with more datapoints (specifically one with more positive stroke cases)
2. Exhaust options for ML Models
3. Go deeper into finding correlations between features/combination-of-features with target
4. Attempt a Supervised ML model first and not focus so intently on supervised models
5. Attempt a Deep Learning model as we believe it would be well suited for this dataset
6. Develop a webpage using Javascript that allows users to input their information and assess their personal stroke risk

# THANKS!

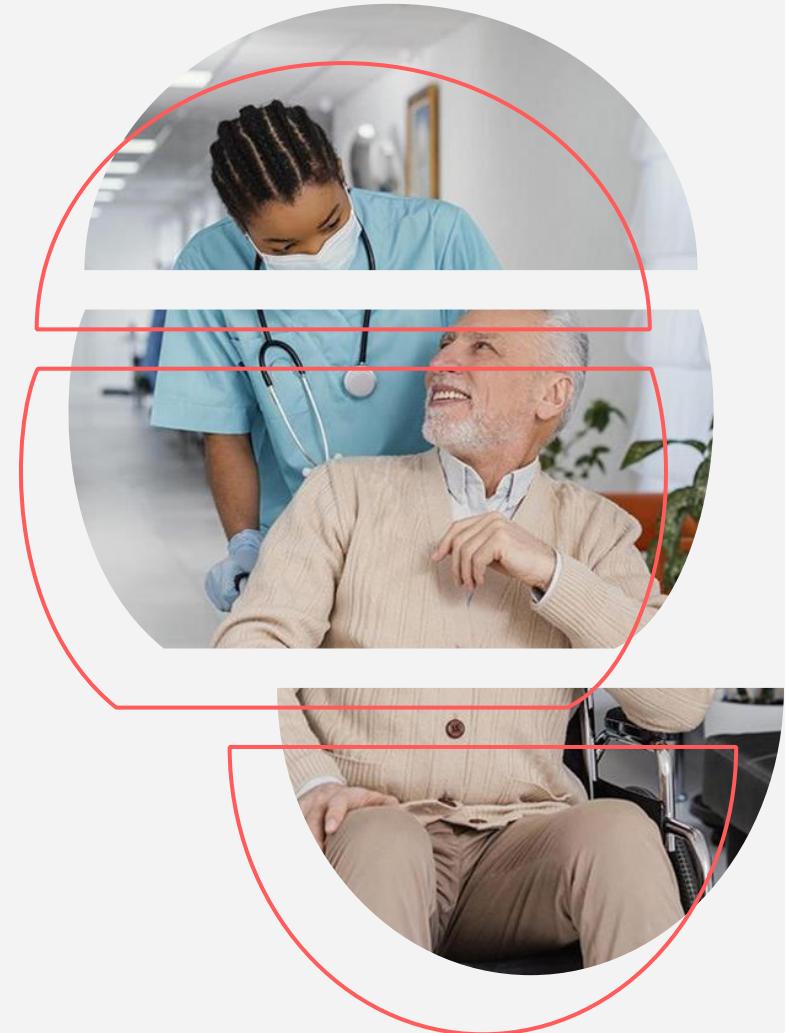


**DO YOU HAVE ANY QUESTIONS?**

info@SAML.com

+1 512 555 SAML

[https://github.com/boborodono/San\\_Antonio](https://github.com/boborodono/San_Antonio)

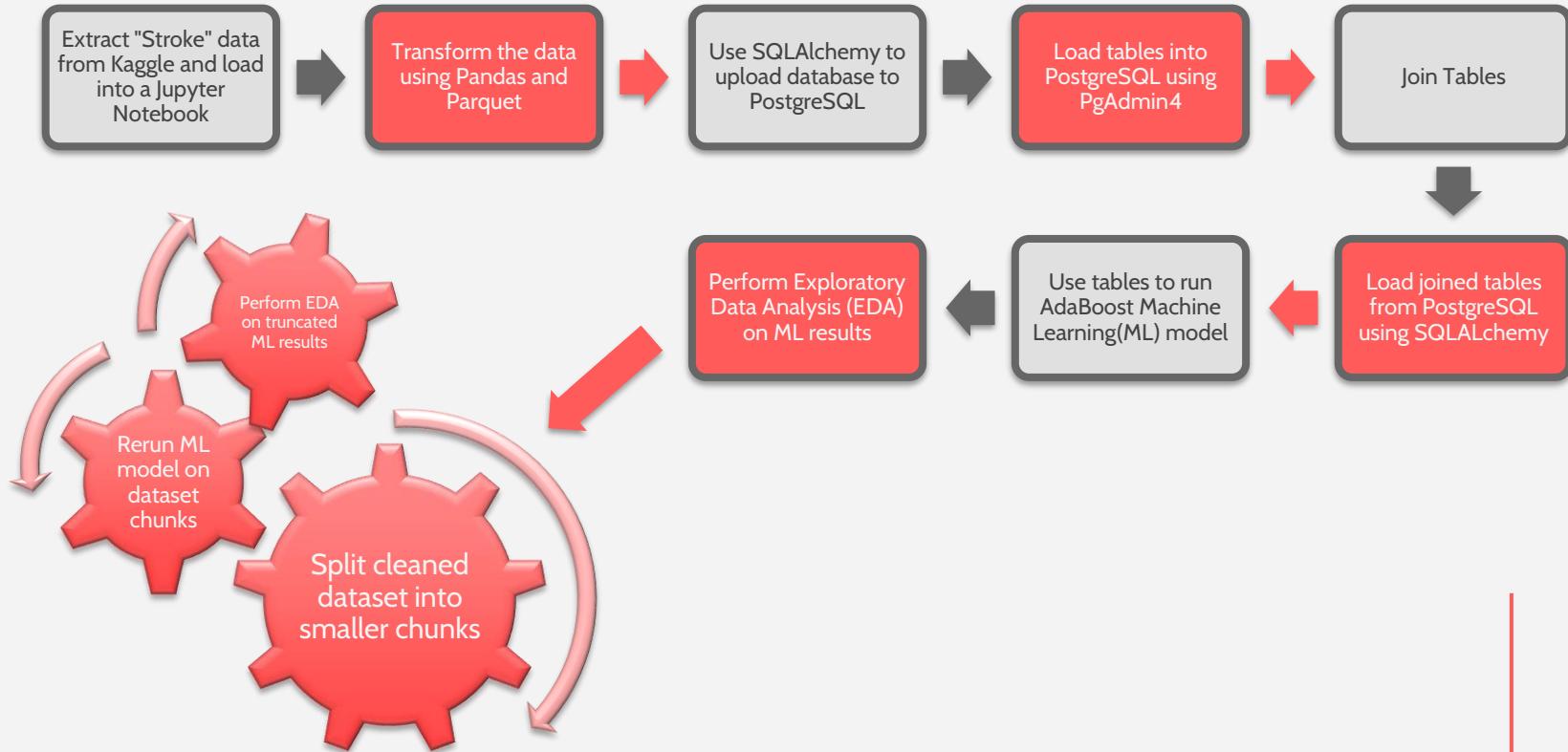




**END OF  
SLIDE DECK**

# PROJECT OUTLINE

# PROJECT OUTLINE



# 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	 POWERPOINT
	 PARQUET	 KAGGLE			 PgADMIN	
	 SCIKIT-LEARN	 GIT				
 SQLALCHEMY						



# 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

**END OF  
FONTS &  
COLOR**

---

# **ABOUT THE DISEASE...**

---



# RECOMMENDATIONS

## NEPTUNE

Mercury is the closest planet to the Sun and the smallest one



## JUPITER

Venus has a beautiful name and is the second planet from the Sun



Despite being red, Mars is actually a cold place. It's full of iron oxide dust



Saturn is a gas giant and has several rings. It's composed of hydrogen



# RECOMMENDATIONS

## STEP 1

Venus is the second planet from the Sun



## STEP 3

Despite being red, Mars is a cold place



## STEP 2

Mercury is the smallest planet

## STEP 4

Neptune is very far from the Sun

# CONCLUSIONS

Do you know what helps you make your point clear?

Lists like this one:

- They're simple
- You can organize your ideas clearly
- You'll never forget to buy milk!

And the most important thing: the audience won't miss the point of your presentation



# WRLD. CANCER AWRNSS. DAY PACK

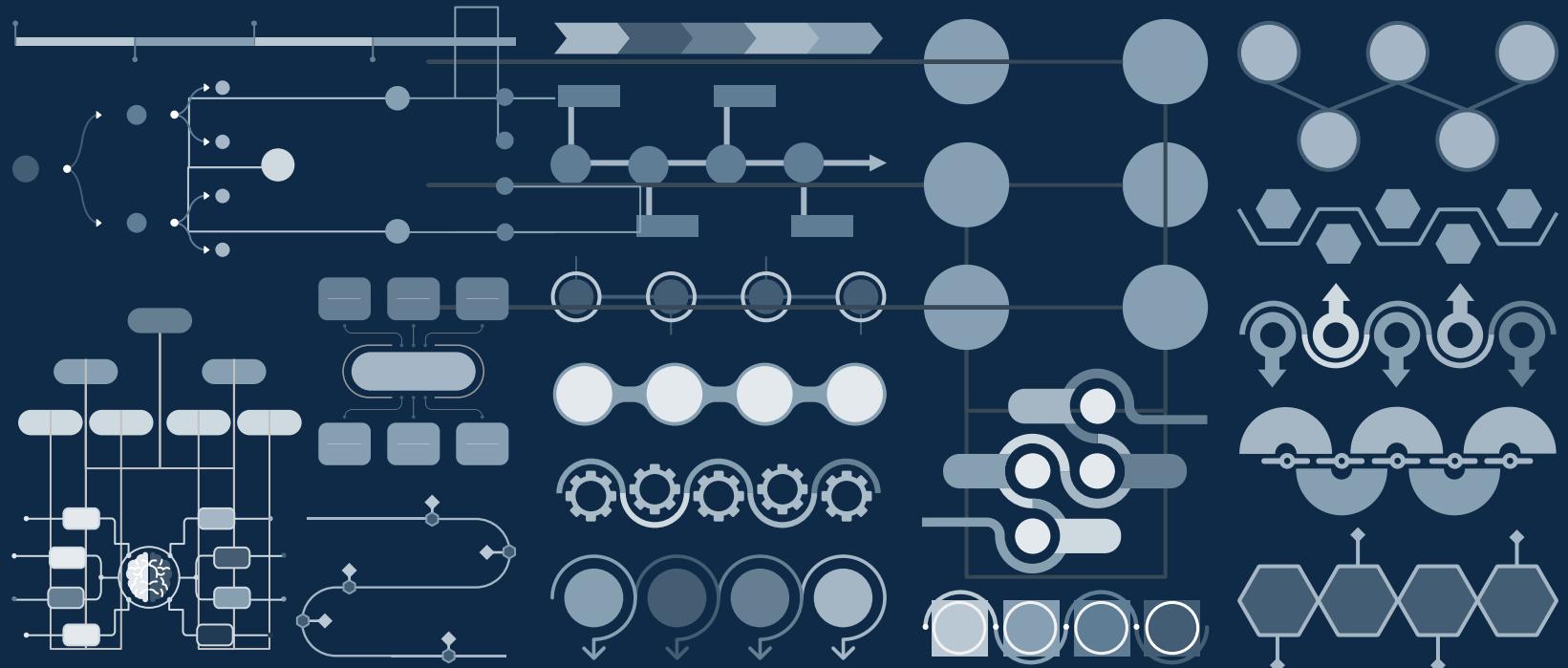


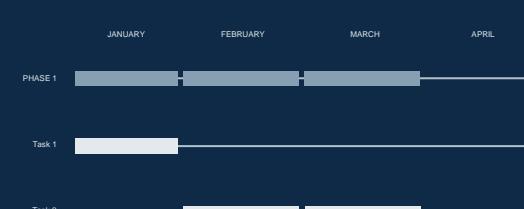
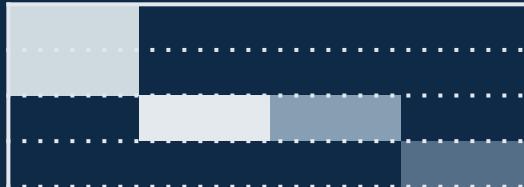
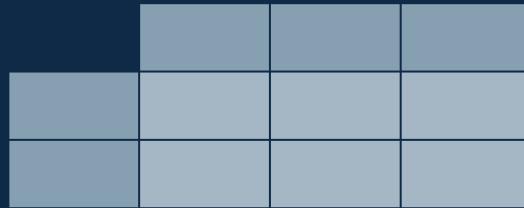
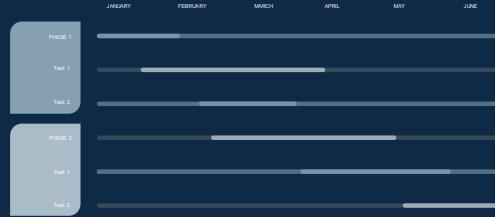
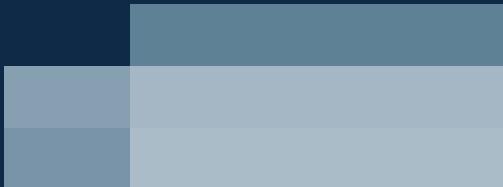
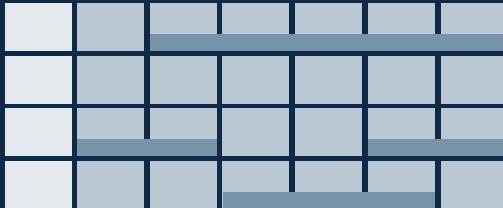
**Use our editable graphic resources...**

You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more infographics on Slidesgo.

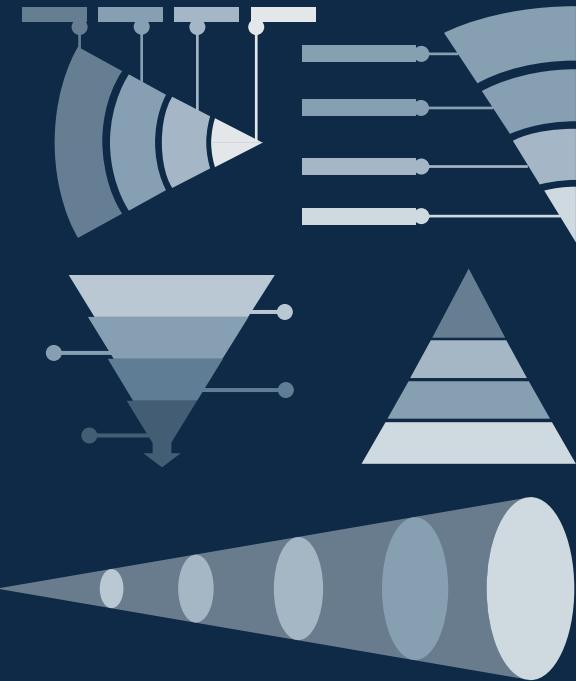
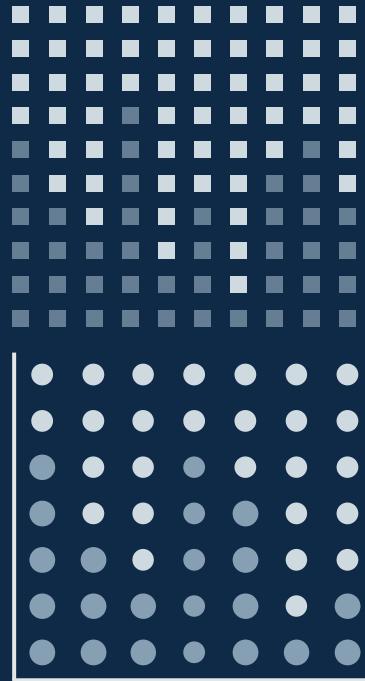












# ...and our sets of editable icons

You can resize these icons without losing quality.

You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.

In Google Slides, you can also use Flaticon's extension, allowing you to customize and add even more icons.



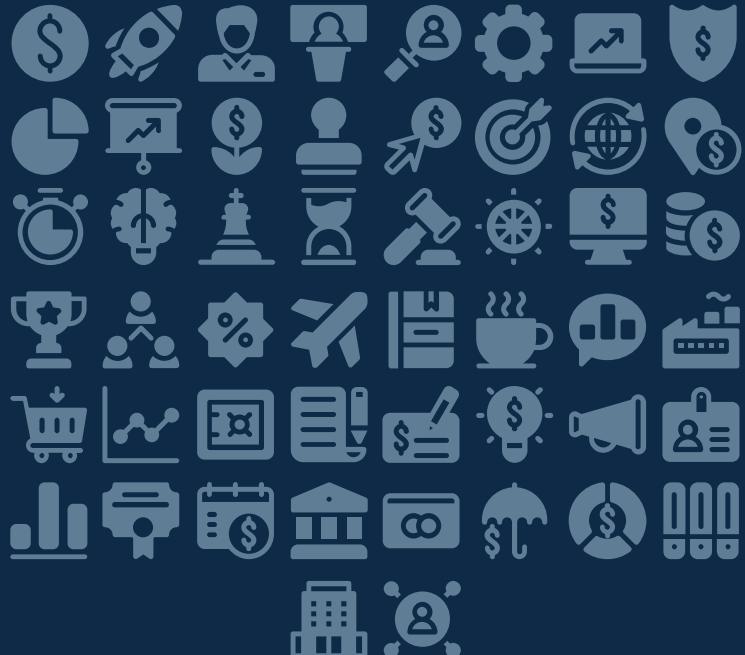
## Educational Icons



## Medical Icons



# Business Icons



# Teamwork Icons



## Help & Support Icons



## Avatar Icons



## Creative Process Icons



## Performing Arts Icons



# Nature Icons



# SEO & Marketing Icons

