

# Table of CONTENTS

01

## Project Overview

**Key user attributes :** Gender, Age, Hypertension, Heart disease, Ever married, Work type, Residence type, Average glucose level, BMI, Smoking status, Stroke.

02

## Libraries and Data Handling

**Libraries used :** Pandas, NumPy, Matplotlib, Seaborn.

**Data Loading and preprocessing :** Read CSV File, Understand the Dataset, Handle Missing Values, Encode Categorical Variables, Feature Scaling, Feature Selection, Split the Dataset.

03

## Data Analysis Technique

**Descriptive statistics :** Summary Statistics, Distribution Analysis, Categorical Data Analysis, Correlation Analysis.

**Inferential Statistics :** Hypothesis Testing, Confidence Interval.

**Predictive Modeling :** Logistic Regression, Decision Trees, Random Forests and GRM, Model Evaluation.

04

## Visual Insights

**Descriptive Statistics Visualizations :** Histograms, Box Plots, Bar Charts, Pie Charts.

**Correlation and Relationship Analysis :** Correlation Heatmap, Scatter Plots.

**Predictive Modeling and Evaluation :** ROC Curves, Confusion Matrices, Feature Importance Plots.

05

## Key Findings

**Major Findings from the Analysis :** Age as a Major Predictor, Impact of Health Conditions, BMI and Glucose Levels, Gender Differences, Stroke Effects of Marriage, Smoking Status, Work Type.



# Table of CONTENTS

06

## Advanced Analysis

**Model Selection and Evaluation :** Multiple Models, Model Evaluation Metrics, Cross-Validation.

**Hyperparameter Tuning :** Random Search

**Model Interpretation :** SHAP Values

**Contribution to Understanding Broader Market**

**Dynamics or Seasonal Patterns:** Trend Analysis, Seasonal Patterns, Predictive Insights, Policy Implications.

07

## Conclusion

Summary of key insights derived from the analysis and their implications for future strategic healthcare decisions.

## Appendix

**Code Snippets:** Provided Python code used for loading, cleaning, transforming data, and generating visualizations.

**Datasets :** Sample dataset of Healthcare Strokes for data analysis.

**Additional References:** Referenced any external datasets or tools used during the analysis process.

