STROKE DISEASE PREDICTION

NAME:BRIAN KIMUTAI SIELE

INSTUCTOR :MWIKALI

Project Description (Introduction)

- Stroke is a serious medical condition that can lead to long-term disability or death. Early prediction of stroke risk can help save lives by enabling timely medical intervention. This project focuses on building a machine learning model to predict the likelihood of a stroke based on health data such as age, gender, blood pressure, and lifestyle habits. The goal is to assist healthcare providers in identifying high-risk individuals and improving stroke prevention efforts.

Project Overview

- The goal of this project is to develop a machine learning-based stroke prediction model that identifies individuals at high risk of stroke, providing insights to guide preventive healthcare strategies

Business Questions

1. Who are the individuals most at risk of experiencing a stroke based on their demographic, lifestyle, and clinical characteristics?
2. Which factors (e.g., age, smoking status, glucose levels, hypertension) are the most significant predictors of stroke?
3. How accurately can we predict the likelihood of a stroke using the given data?
4. What preventive actions or healthcare measures can be recommended for high-risk groups?
5. Can we design a model that is interpretable and can be used effectively by healthcare providers in both high- and low-resource settings?

Business Understanding

- Stroke is one of the leading causes of disability and death worldwide, and early prediction can significantly reduce its impact.
- This project focuses on building a predictive model for stroke risk, providing actionable insights for healthcare providers to improve early detection, prioritize high-risk individuals, and implement targeted interventions.

Data Understanding

- we explore the dataset to understand its structure, quality, and relevance to the project.The dataset typically contains various demographic, lifestyle, and clinical factors that influence the likelihood of stroke.

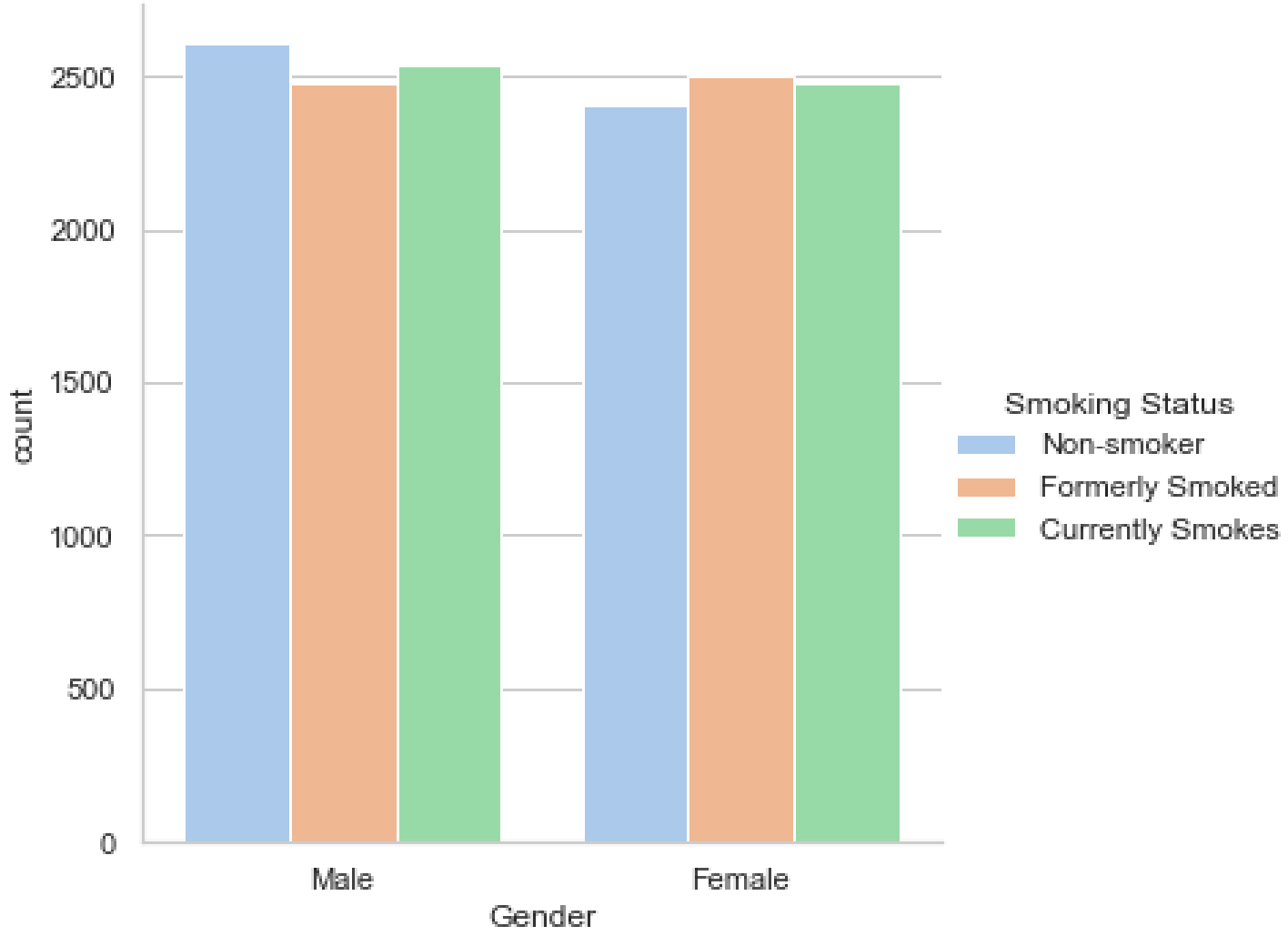
Data Preparation

- Data Cleaning:
- Handling missing values
- Removing unwanted images
- Checking outliers

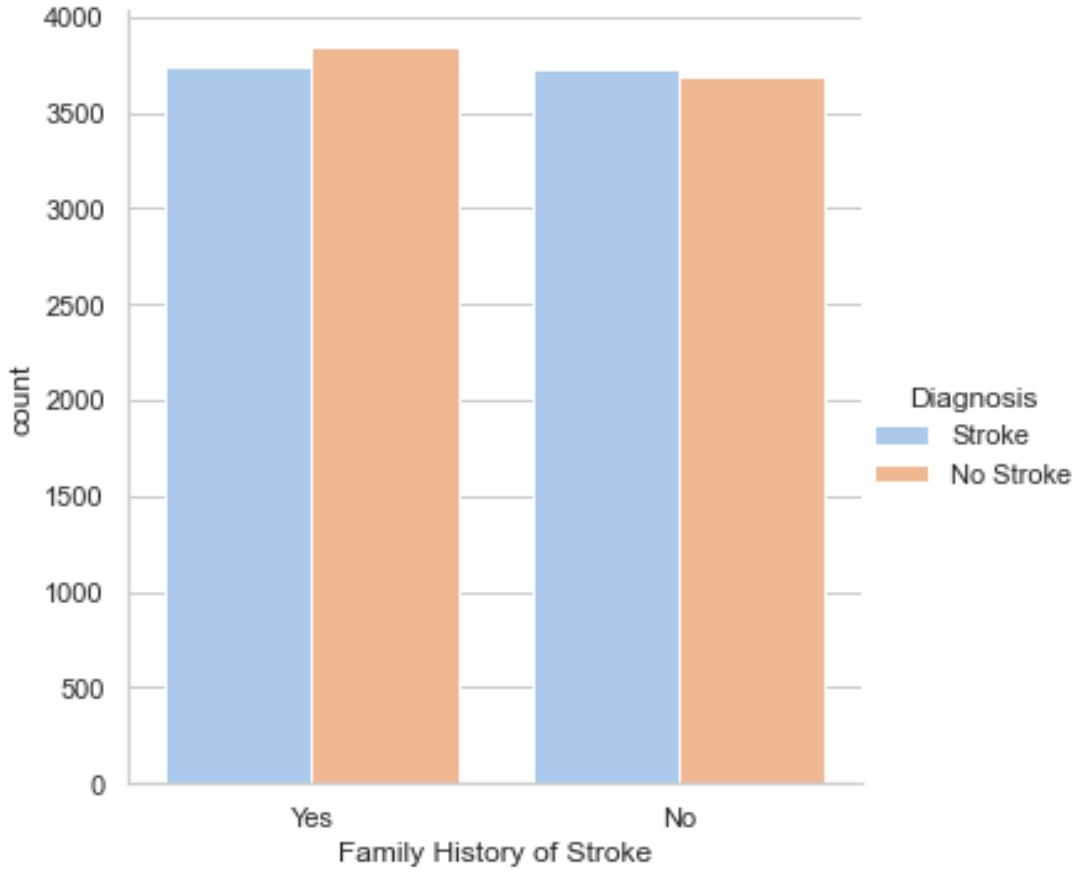
Data Analysis

- The data analysis process for stroke disease prediction involves exploring and preparing the dataset to uncover patterns and relationships that can help build a predictive model

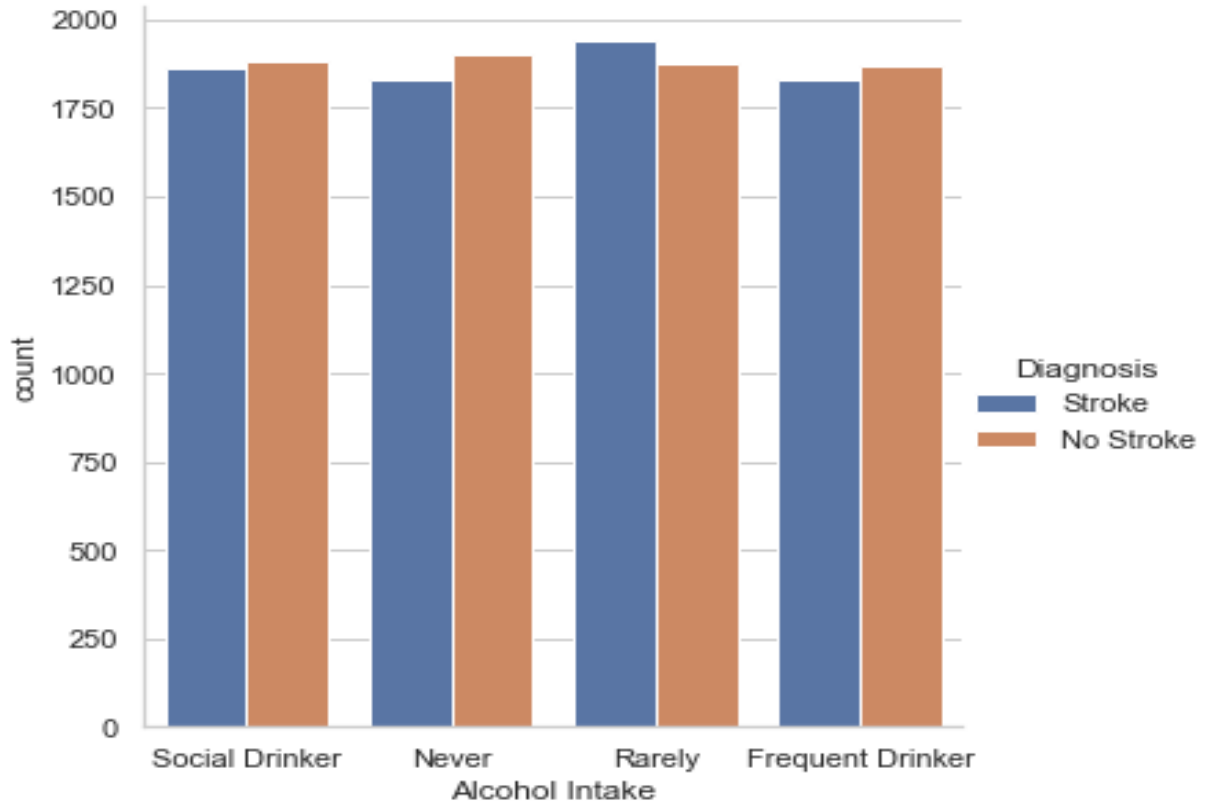
Data Visualization
Plot of gender against smoking status



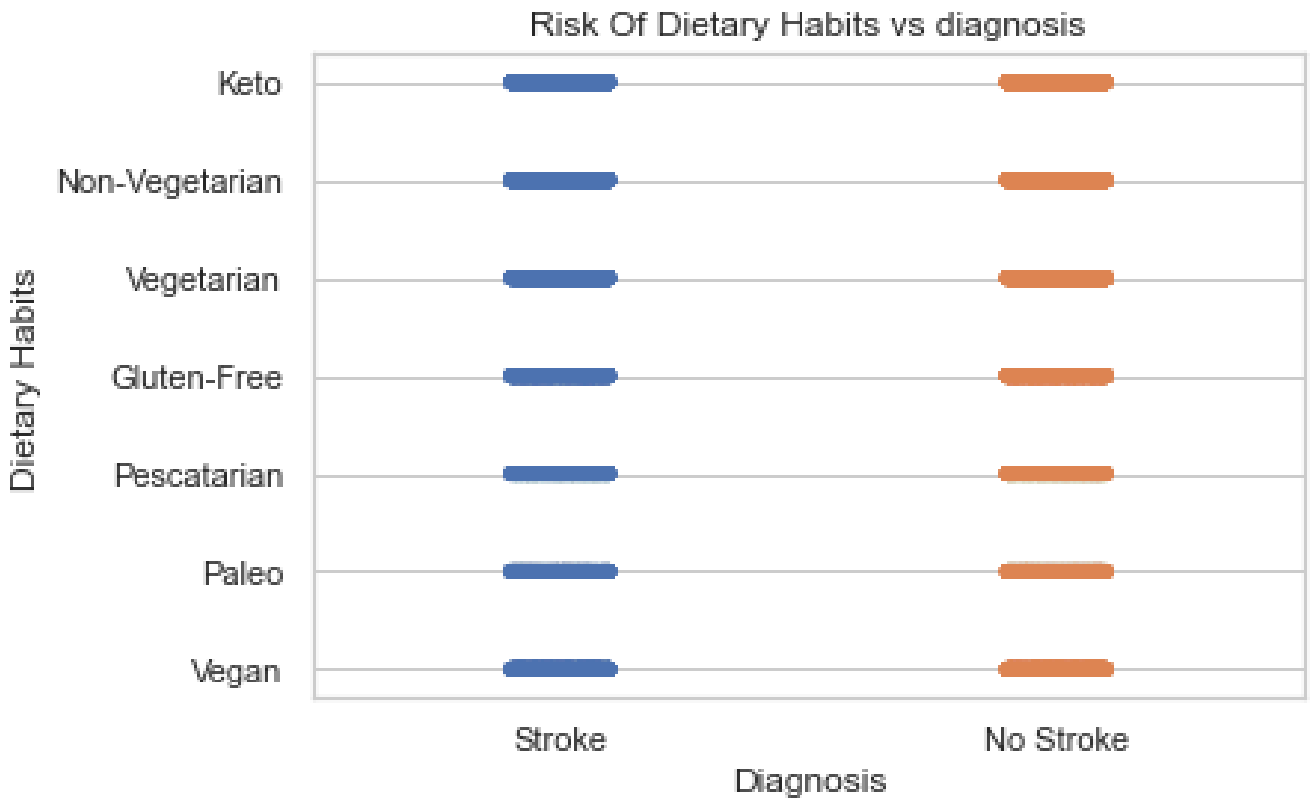
Plot of Family History of Stroke against Diagnosis



Plot for Alcohol intake against Diagnosis of stroke



PLot for relations between Diagnosis with Dietary habits



Conclusion

Key Risk Factors:

- - Age: Older individuals are at higher risk for strokes.
- - Smoking: Smokers, especially heavy smokers, are more prone to strokes.
- - Hypertension: Prevalent hypertension significantly increases stroke risk.
- - Diabetes: Individuals with diabetes are at higher risk.
- - Obesity (High BMI): Obesity is a contributing factor to stroke.
- - High Cholesterol and Glucose Levels: Elevated levels correlate with a higher stroke risk.
- - Lifestyle Factors: Lack of exercise and poor diet can exacerbate these risks

Recommendation

- 1. Public Health Interventions:
- 2. Healthcare Strategies:
- 3. Personalized Prevention Plans:
- 4. Model Utilization:
- 5. Awareness Campaigns: