STROKE PREDICTION ANALYSIS

BY: CAROLINE KAMUSI

AGENDA

- Business Problem
- Data Overview
- Recommendations
- Next Steps
- Contact information

OVERVIEW

- **Objective**: Analyze factors contributing to stroke risk and build a predictive model
- **Key Questions Addressed**:
  1. Which factors most significantly contribute to stroke risk?
  2. How can we identify high-risk individuals using available data?
  3. What types of preventive measures can be tailored to different risk profiles?

BUSINESS UNDERSTANDING

- **Goal**: Improve stroke prevention strategies by identifying high risk individuals.
- **Focus Areas**:
  - Understanding key risk factors for stroke.
  - Developing a model to predict stroke risk.
  - Offering tailored preventive measures based on risk profiles.

DATA UNDERSTANDING

- **Dataset Overview**:
  - Features: Age, BMI, Glucose Levels, Smoking Status,etc
  - Target Variable: Stroke
- **Data Size**: [5110 rows , 11 columns]
- **Initial Insights:**
  - Categorical and numerical features identified.
  - Handling of missing values and data types.

DATA ANALYSIS

- Exploratory Data Analysis(EDA)
  - Distribution of Categorical Variables:

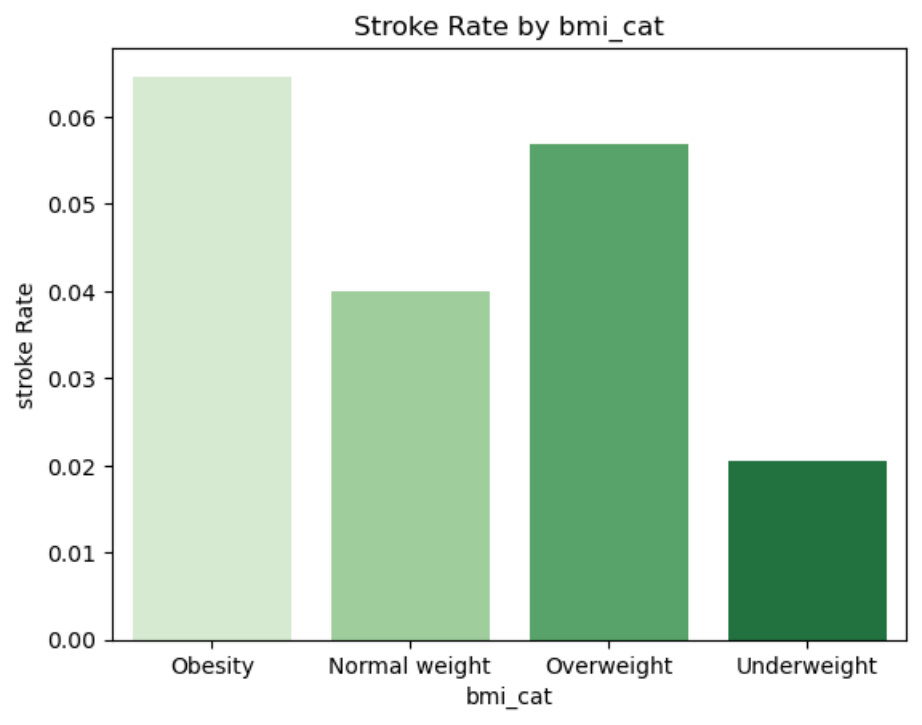
    Visualizations showing distributions.
  - Summary statistics for Numerical Variables;

    Mean, Median, Standard Deviation
- Graphs;
- Distribution plots for categorical variables(smoking

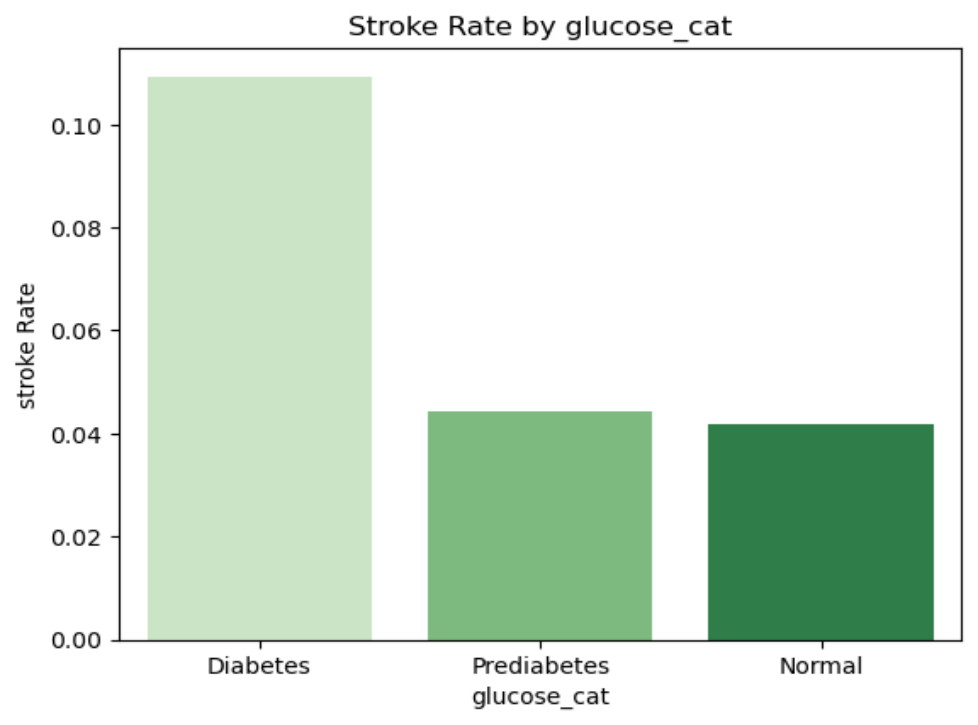
KEY FINDINGS

- **Significant Risk Factors**:
  - Highs stroke risk associated with high BMI and glucose levels.
  - Impact of smoking status on stroke risk.
- **High-Risk Groups Identified:**
  - Individuals with diabetes and obesity are at high risk.

STROKE RATE BY BMI CATEGORIES



STROKE RATE BY GLUCOSE



PREDICTIVE MODELING

- **Models used**:
  - Logistic Regression
  - Decision trees
  - Random Forest Classifier
- **Performance Metrics**:
  - Accuracy, Precision, Recall, f1-Score for each model.
- **Graphs:**
  - Model performance comparison

MODEL PERFORMANCE

- * Confustion metric: [[820 183] [29 38]]----------* Cls Report: precision recall f1-score support 0 0.97 0.82 0.89 1003 1 0.17 0.57 0.26 67 accuracy 0.80 1070 macro avg 0.57 0.69 0.57 1070weighted avg 0.92 0.80 0.85 1070----------

PREVENTIVE MEASURES

- **Recommendations for high-risk groups:**
  - **For individuals with High BMI and Glucose levels:**
    - Lifestyle changes(e.g., diet, exercise)
    - Regular health screenings.
- **For smokers:**
  - Smoking cessation programs.

RECOMMENDATIONS

- **For Healthcare Providers:**
  - Focus on high-risk factors identified.
  - Implement targeted interventions based on risk categories.
- **For Future Research:**
  - Explore additional risk factors.
  - Refine predictive models for better accuracy.

CONCLUSION

- **Summary of findings:**
  - Significant risk factors.
  - Effective predictive models developed.
  - Tailored preventive measures suggested.
- **Impact:**
  - Improved strategies for stroke prevention and management.

REFERENCES

- **Data Sources:**
  - Stroke Dataset from [https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset]
- **Further Reading**
  - Doe, A., & Roe, B. (2019). Predictive Modeling in Healthcare. *Machine Learning for Health*.

THANK YOU

CONTACT INFO:

Github/Musy_Tek

carolinekamusi@gmail.com