Stroke Analysis: BMI and Health

This analysis focuses on exploring the association between a patient's likelihood of having a stroke and a range of risk factors: BMI, smoking status, hypertension, gender, age, work-type, heart disease and residence type. Understanding this correlation can deeply impact preventive measures and healthcare strategies.

Y **by Yongli Lyu**

Business Context

- The World Health Organization has identified stroke as a major global health concern.
- Stroke occurs when there is an interruption of blood flow to the brain, leading to brain damage and potentially severe health consequences.
- Understanding the factors that contribute to stroke incidence is crucial for healthcare professionals to develop effective prevention and treatment strategies.

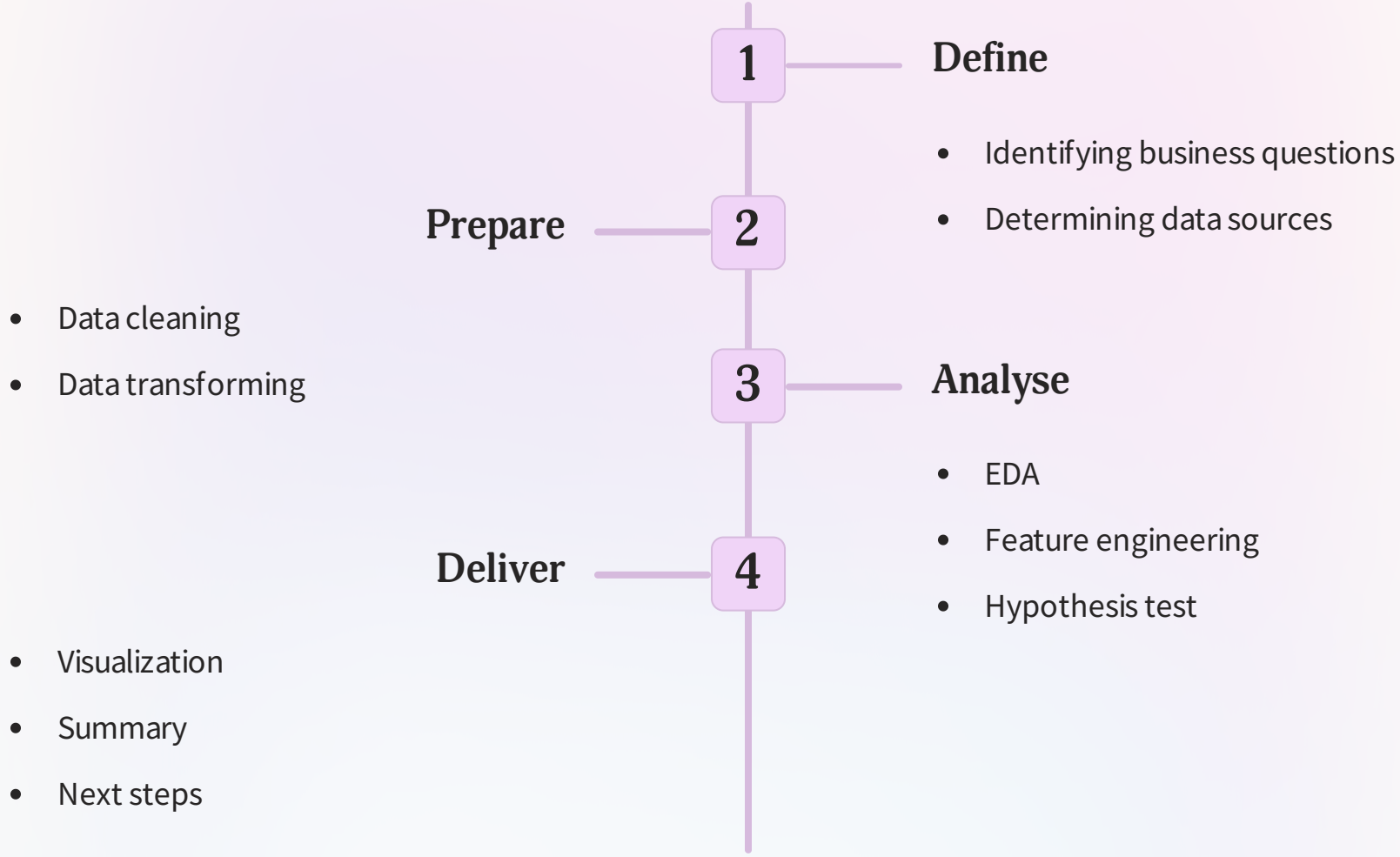
Business Question

Is there a significant distinction in the mean BMI between individuals who experienced a stroke and those who did not?

Explanation:

This question focuses on determining if there is a notable difference in the average BMI (Body Mass Index) between individuals who have had a stroke and those who have not. By exploring and analyzing the relevant data, we can gain valuable insights into the relationship between BMI and stroke.

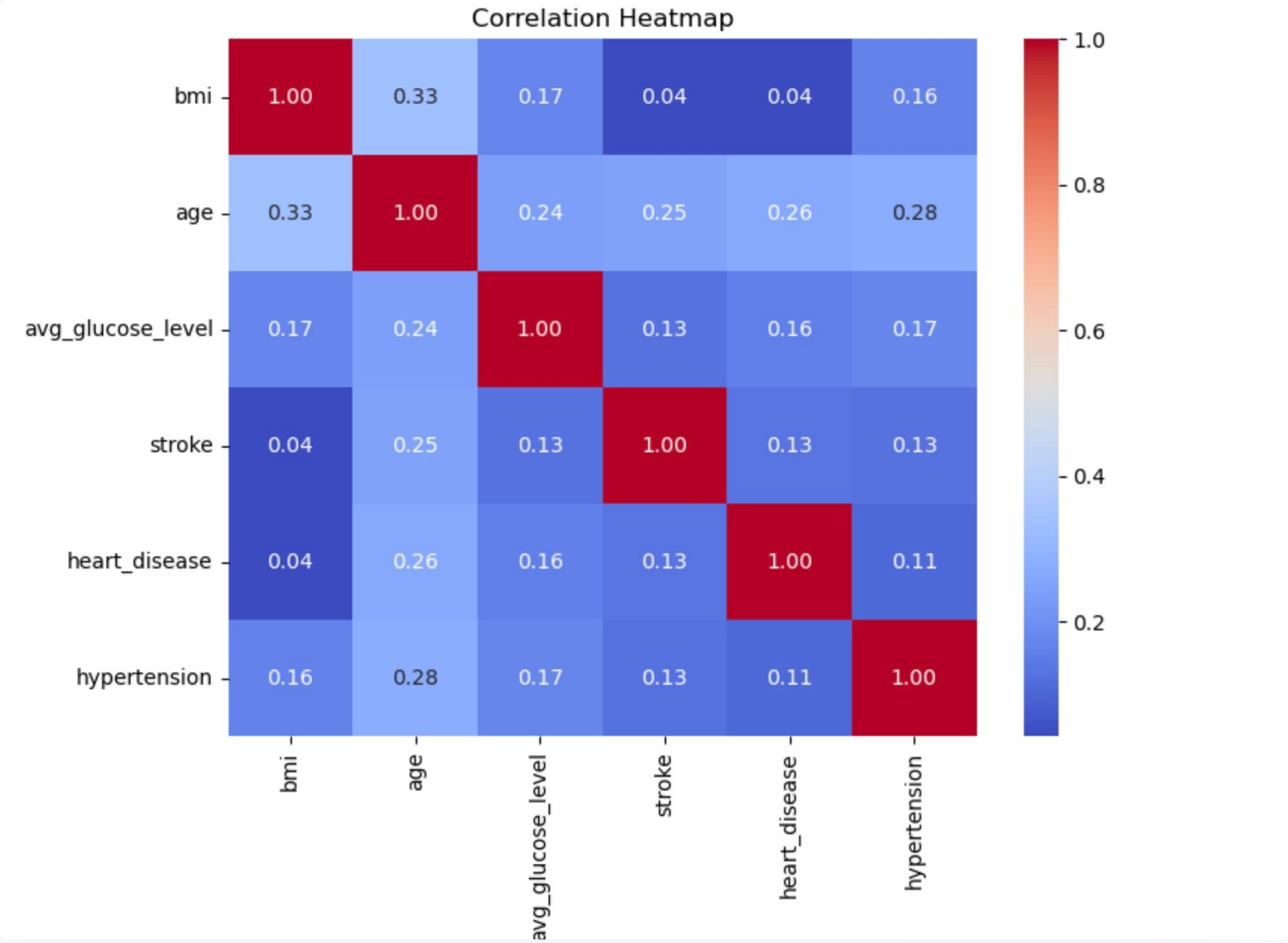
Data Pipeline



Data Summary

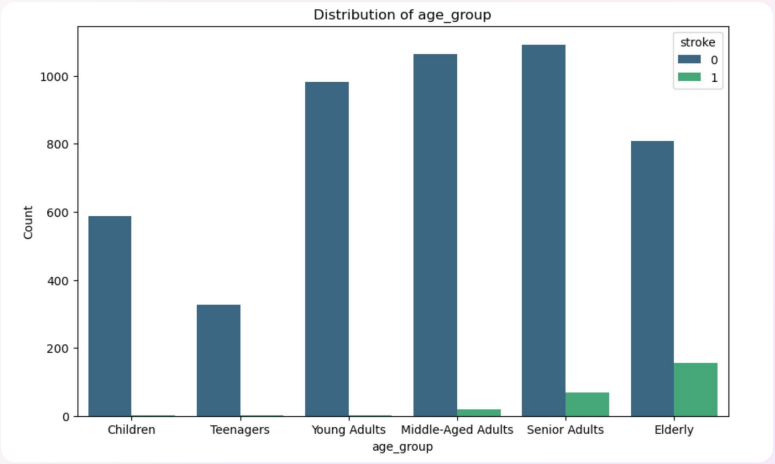
- The dataset includes 5110 observations with
- 4.9% of people who had stroke
- 9.7% of people who has hypertension
- 5.4% of people who has heart_disease
- 20% of people who has diabetes.

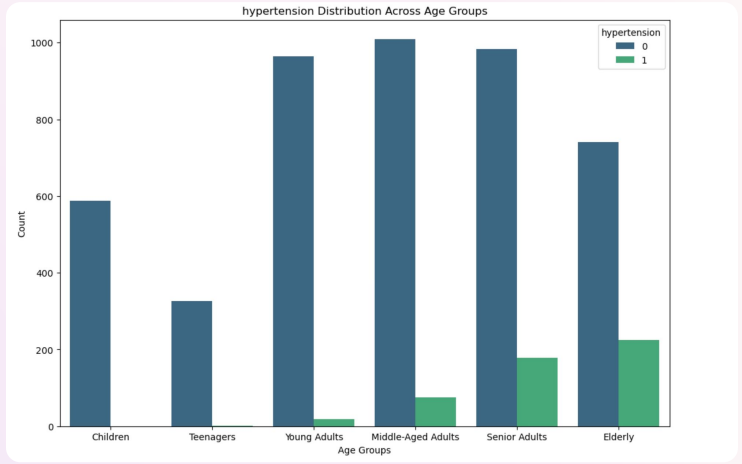
EDA - Heat map

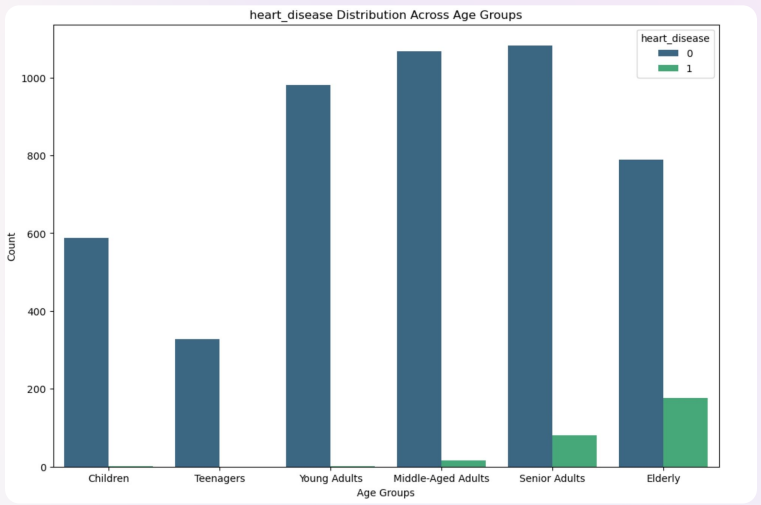


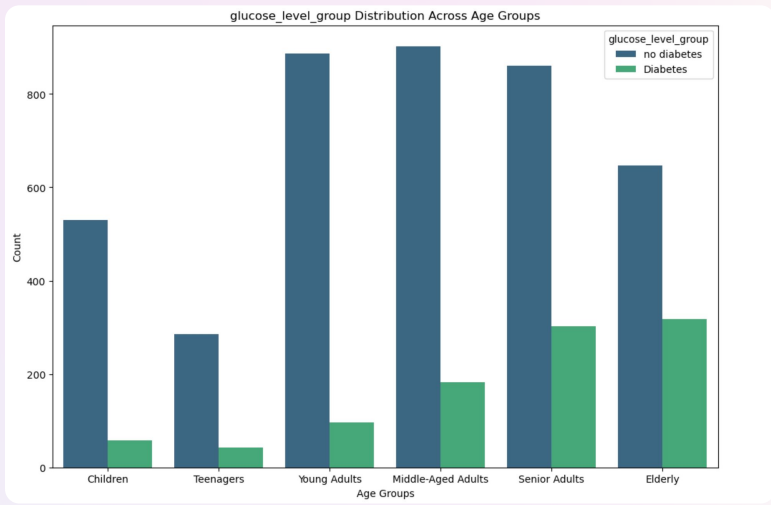
- The correlation heatmap reveals that the 'age' feature exhibits more noticeable correlations, consistently exceeding 0.2 with various other variables.
- 'BMI' exhibits a stronger correlation with 'age' than with 'stroke,' with a coefficient of only 0.04.
- In this project, my aim is to investigate the relationships between the features 'age' and 'BMI' in conjunction with four disease-related features.

Age and hypertension, heart disease, stroke and diabetes.





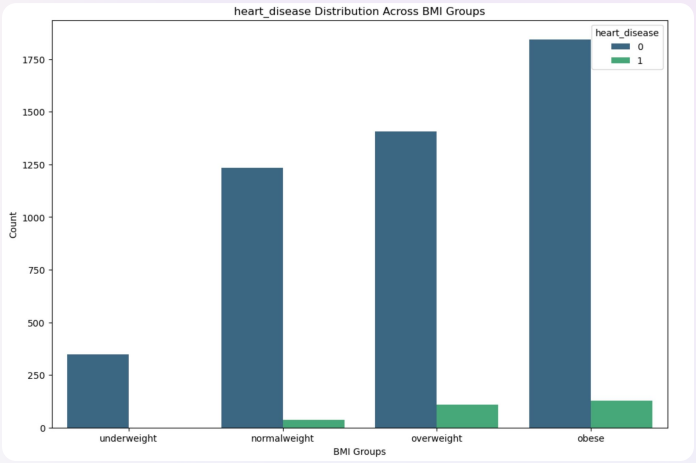


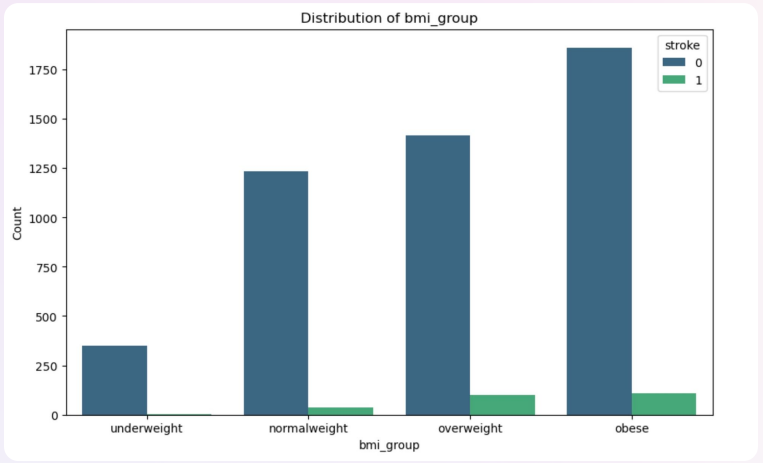


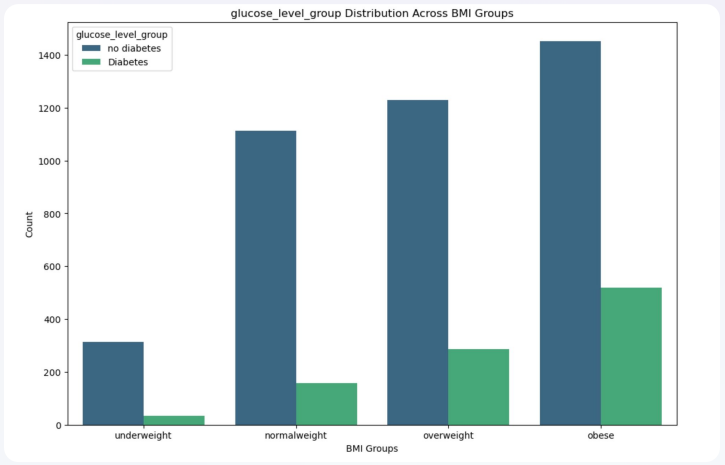
- **General Trends:**
  - Across all three diseases (hypertension, heart disease, stroke, diabetes), there is a consistent pattern of increased prevalence with advancing age.
  - The age distribution for individuals with these diseases tends to shift towards older age groups.
- **Interpretation:**
  - Age appears to be a common factor influencing the likelihood of developing these health conditions.
  - Understanding age-related patterns is crucial for targeted healthcare interventions and preventive measures.

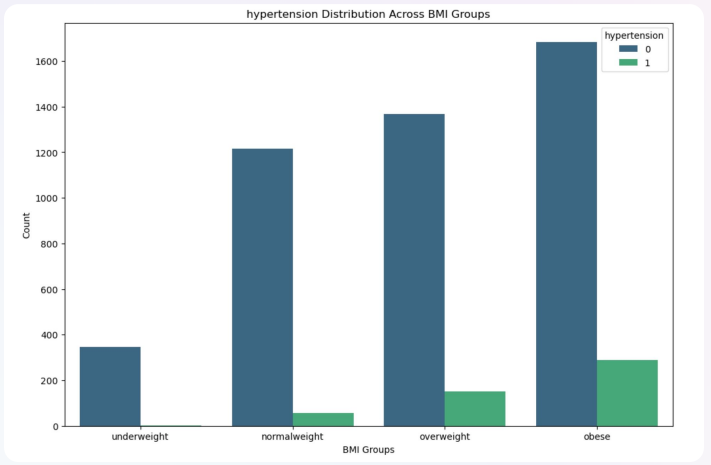
BMI

- **General Trends:**
  - Across all three diseases (hypertension, heart disease, stroke, disbetes), there is a trend of higher prevalence with increasing BMI.
  - The BMI distribution for individuals with these diseases tends to shift towards higher BMI values.
- **Interpretation:**
  - BMI emerges as a potential common factor associated with the likelihood of developing these health conditions.
  - Integrating BMI assessments into disease prevention and management plans may be beneficial.

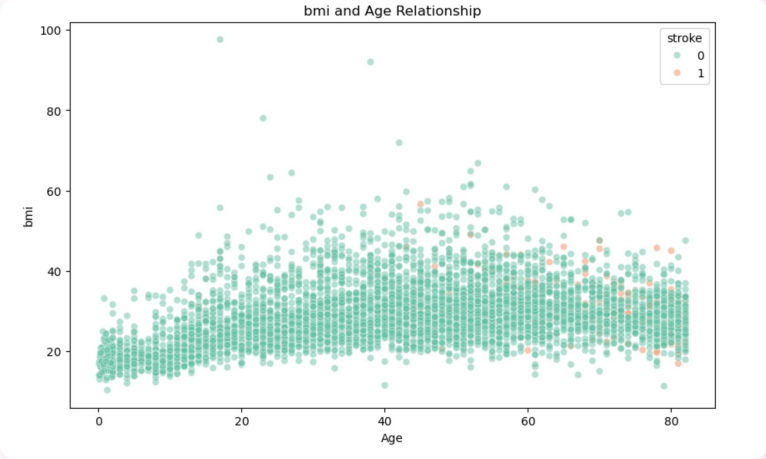


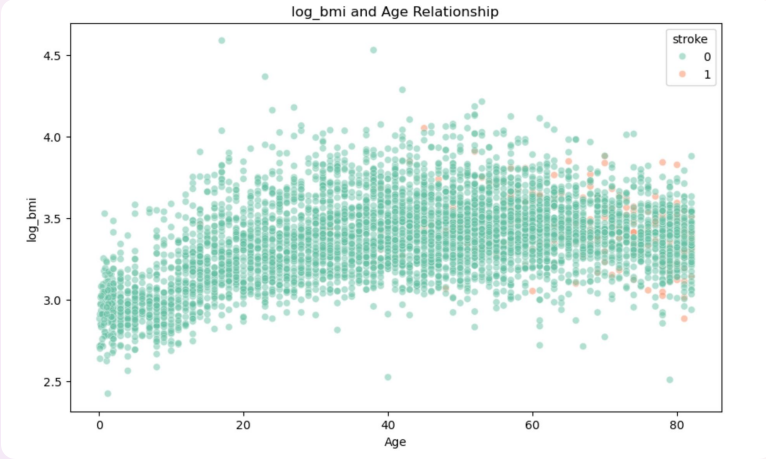






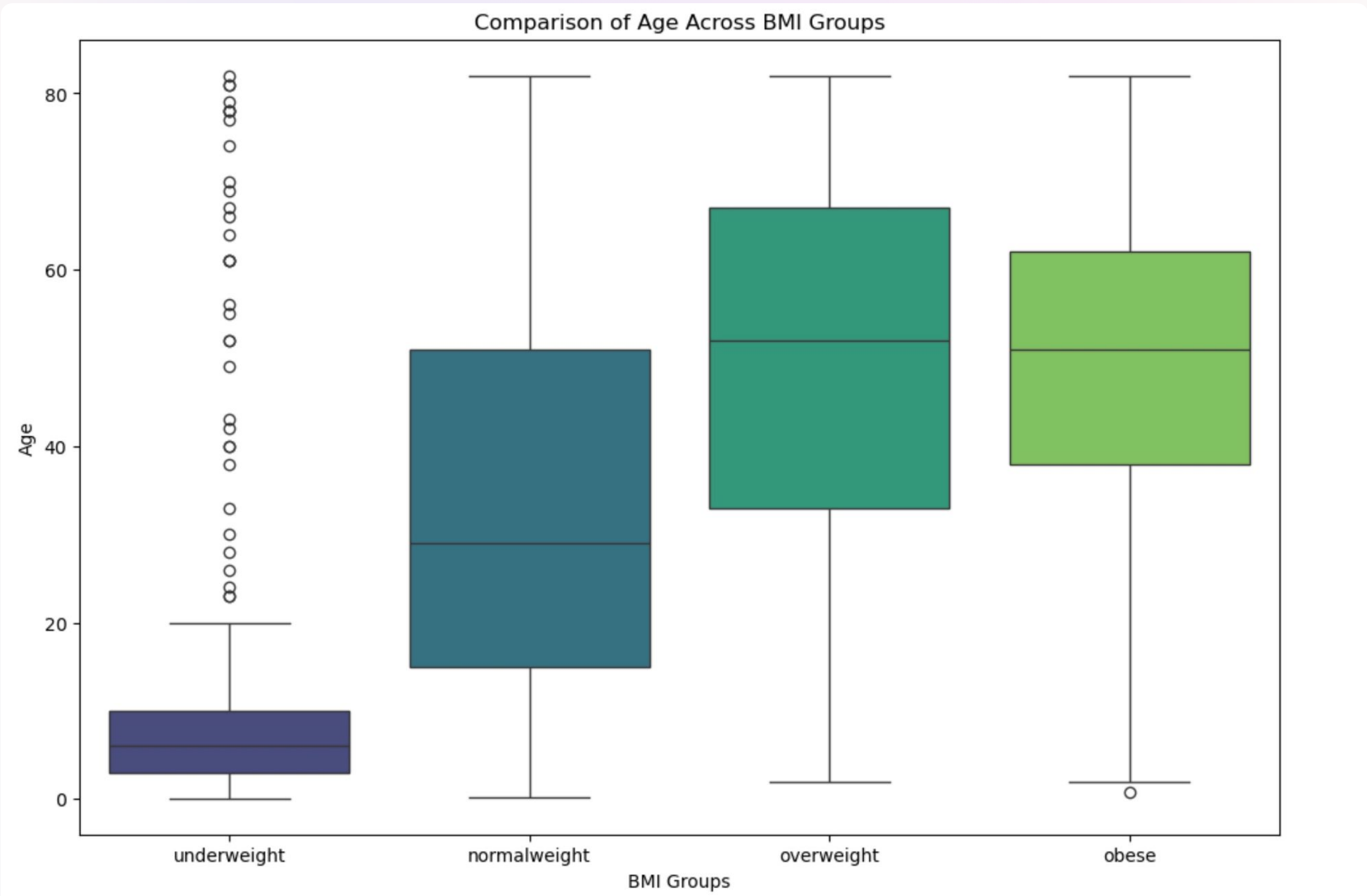
BMI and Age





As age increases, there is a slight linear relationship with BMI, and this association becomes more evident with log-transformed BMI

The box plot similarly depicts this trend, where the mean age progressively increases, accompanied by a rise in BMI until reaching the obese group.



Hypothesis Testing

The null and alternative hypotheses will be rigorously tested to determine the significance of the difference in mean BMI for patients who have experienced a stroke versus those who have not. Alpha is set to 0.05

1

Null Hypothesis (H0)

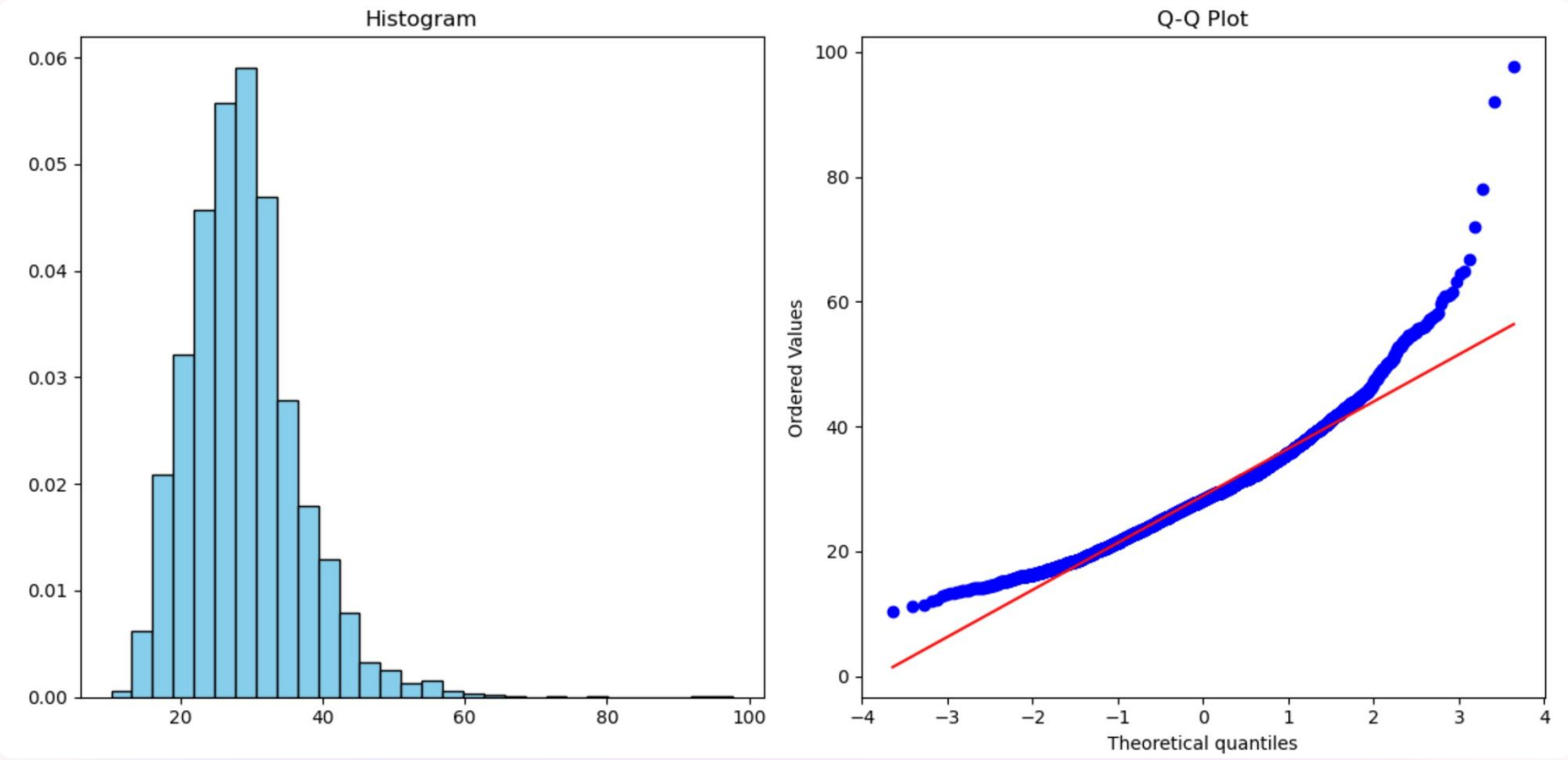
The mean BMI is the same for patients who had a stroke and those who did not.

2

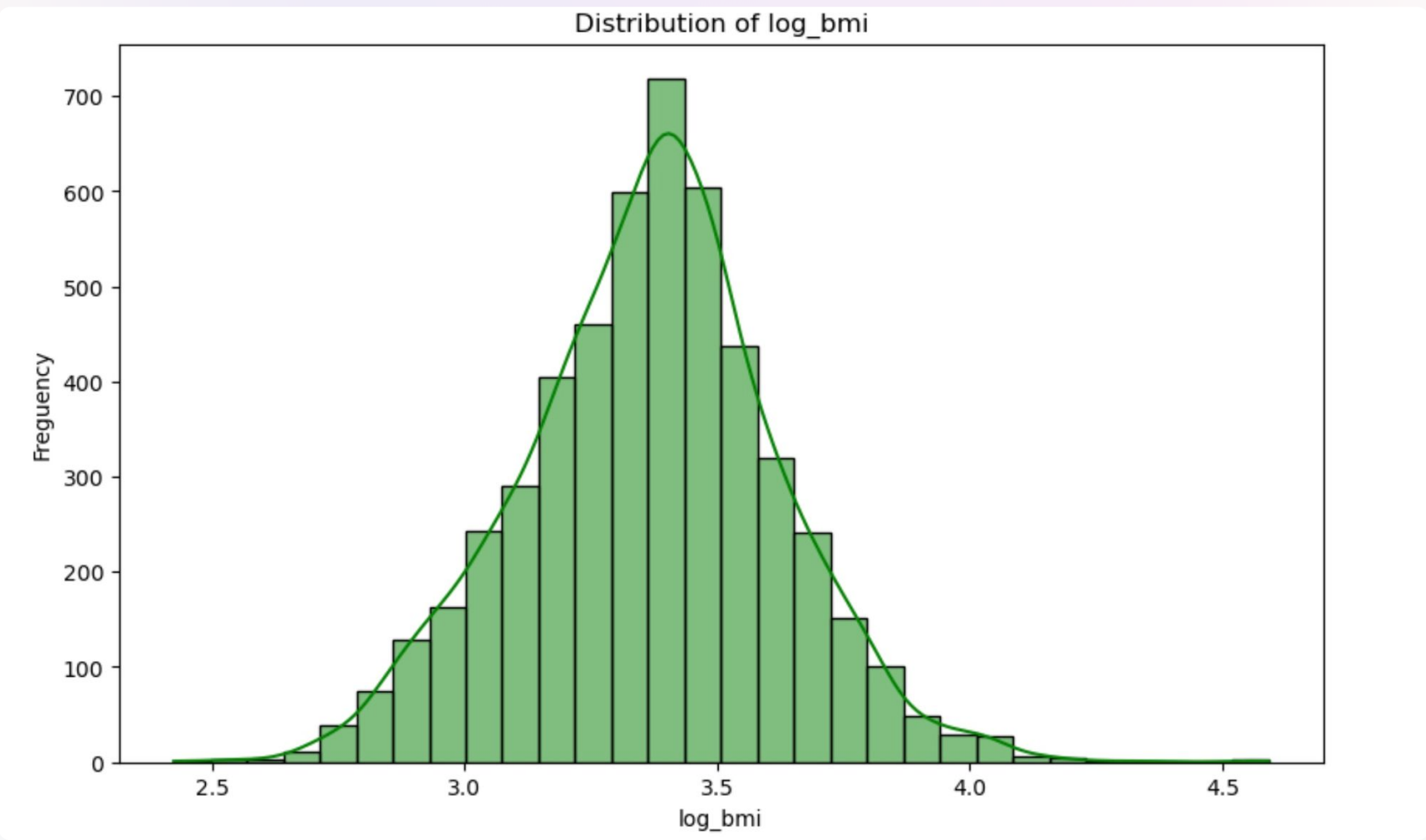
Alternative Hypothesis (H1)

The mean BMI is different for patients who had a stroke compared to those who did not.

Bmi Distribution



The original BMI distribution is right-skewed (skewness = 1). After log transformation, skewness is significantly reduced, approaching 0. This transformation normalizes the distribution, making it more symmetric.



Summary

- The t-test yielded significant results ($t = 3.62$, $p = 0.00038$).
- The small p-value ($p = 0.00038$) leads to the rejection of the null hypothesis.
- The findings suggest a meaningful distinction between the compared groups in the hypothesis test.

Next Steps

- Next steps involve conducting further analysis to build a predictive model for diabetes using existing features.
- This process will entail selecting relevant variables, exploring feature engineering techniques, and applying suitable modeling approaches to enhance our ability to predict diabetes outcomes accurately.

Appendix

- The data source is from Kaggle: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset/data.
- https://www.aihw.gov.au/reports/heart-stroke-vascular-disease/hsvd-facts/contents/all-heart-stroke-and-vascular-disease/stroke