ALZHEIMER'S DISEASE DATASET EDA PPT

Sakshi Andhale

TABLE OF CONTENTS

INTRODUCTION OF
ALZHEIMER'S DISEASE EDA

Alzheimer's Disease is a progressive neurodegenerative disorder that affects millions of people worldwide. Early diagnosis and understanding of various contributing factors can aid in managing the disease better. This exploratory data analysis (EDA) aims to delve into a dataset containing patient demographics, lifestyle factors, medical history, clinical measurements, and symptoms to uncover meaningful patterns that could assist in improving diagnosis and patient care.

01.

Dataset Overview

OVERVIEW

Purpose of Analysis:

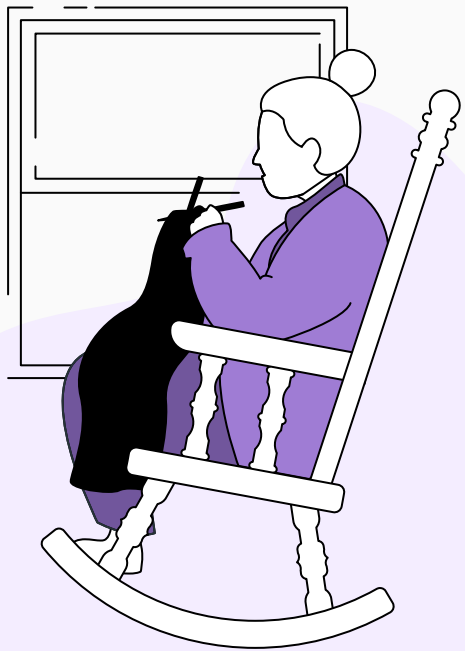
The purpose of this Exploratory Data Analysis (EDA) is to identify key patterns, trends, and risk factors associated with Alzheimer's Disease using a dataset that captures demographic details, lifestyle factors, clinical measurements, cognitive assessments, and symptoms.

Why this dataset?

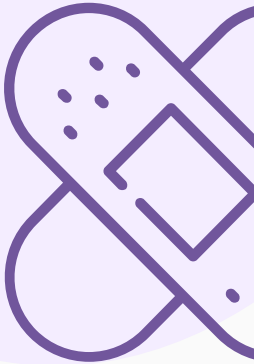
Alzheimer's Disease is a complex neurological disorder. Understanding the underlying patterns in demographic and medical data may help in early diagnosis and treatment strategies.

Tools Used

Python Libraries: pandas, numpy, seaborn, matplotlib



DATASET DESCRIPTION



The dataset contains detailed health information of 2,149 patients including:

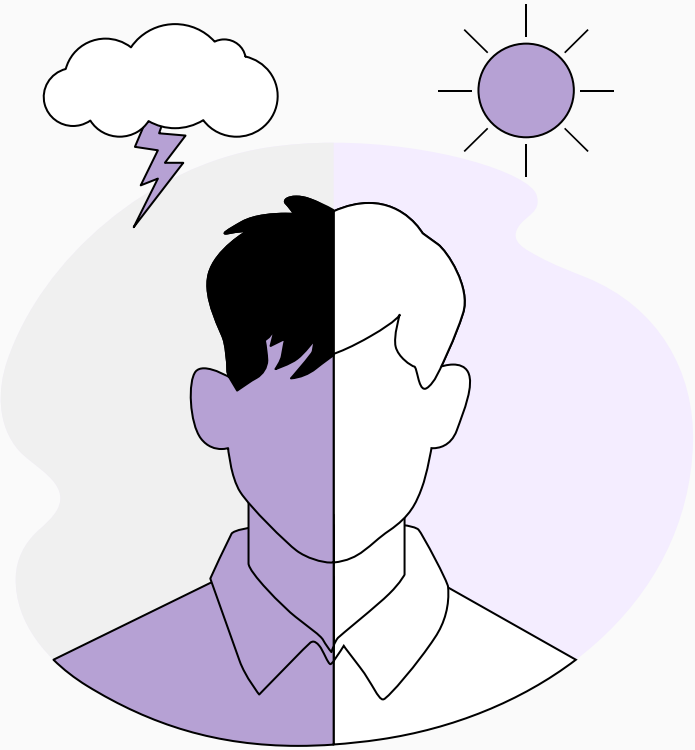
Demographic Data: Age, Gender, Ethnicity, and Education Level
Lifestyle Factors: Smoking, Alcohol Consumption, Diet, Physical Activity, etc.
Medical History: Cardiovascular Disease, Diabetes, Depression, and more
Clinical Measurements: Blood Pressure, Cholesterol Levels, BMI
Cognitive Assessments: MMSE Scores, Memory Complaints, Behavioral Problems
Symptoms: Confusion, Disorientation, Forgetfulness
Diagnosis: Whether the patient is diagnosed with Alzheimer's or not (Yes/No)

Rows and Columns:
Rows: 2,149
Columns: 20+

Questions Raised

1. What is the distribution of age in the dataset?
2. What is the gender distribution in the dataset?
3. What is the distribution of ethnicity among patients?
4. Does age vary across different education levels?
5. Is there a relationship between BMI and the presence of diabetes?
6. How does systolic and diastolic blood pressure correlate?
7. How many patients have been diagnosed with Alzheimer's disease?

And More..



02.

Data Cleaning

DATA CLEANING PROCESS

01.

Null Values

Checked and removed null values in critical columns like Diagnosis and Cognitive Scores to avoid skewed results.

02.

Outliers

Identified and treated outliers in numerical columns like BMI and Systolic BP using boxplots.

DATA CLEANING PROCESS

03.

Data Types

Encoded categorical columns (Gender, Ethnicity, Education Level) for analysis.

04.

Missing Values

Handled missing values in lifestyle columns such as Smoking and Alcohol Consumption through imputation.

05.

Duplicate

No duplicate entries were found in the dataset.

03.

Visualisation

Age Distribution of Diagnosed Patients

Data Presented:

The age range of patients (60-90 years old) with a specific focus on those diagnosed with Alzheimer's.

Questions Answered:

What is the most affected age group in Alzheimer's patients?

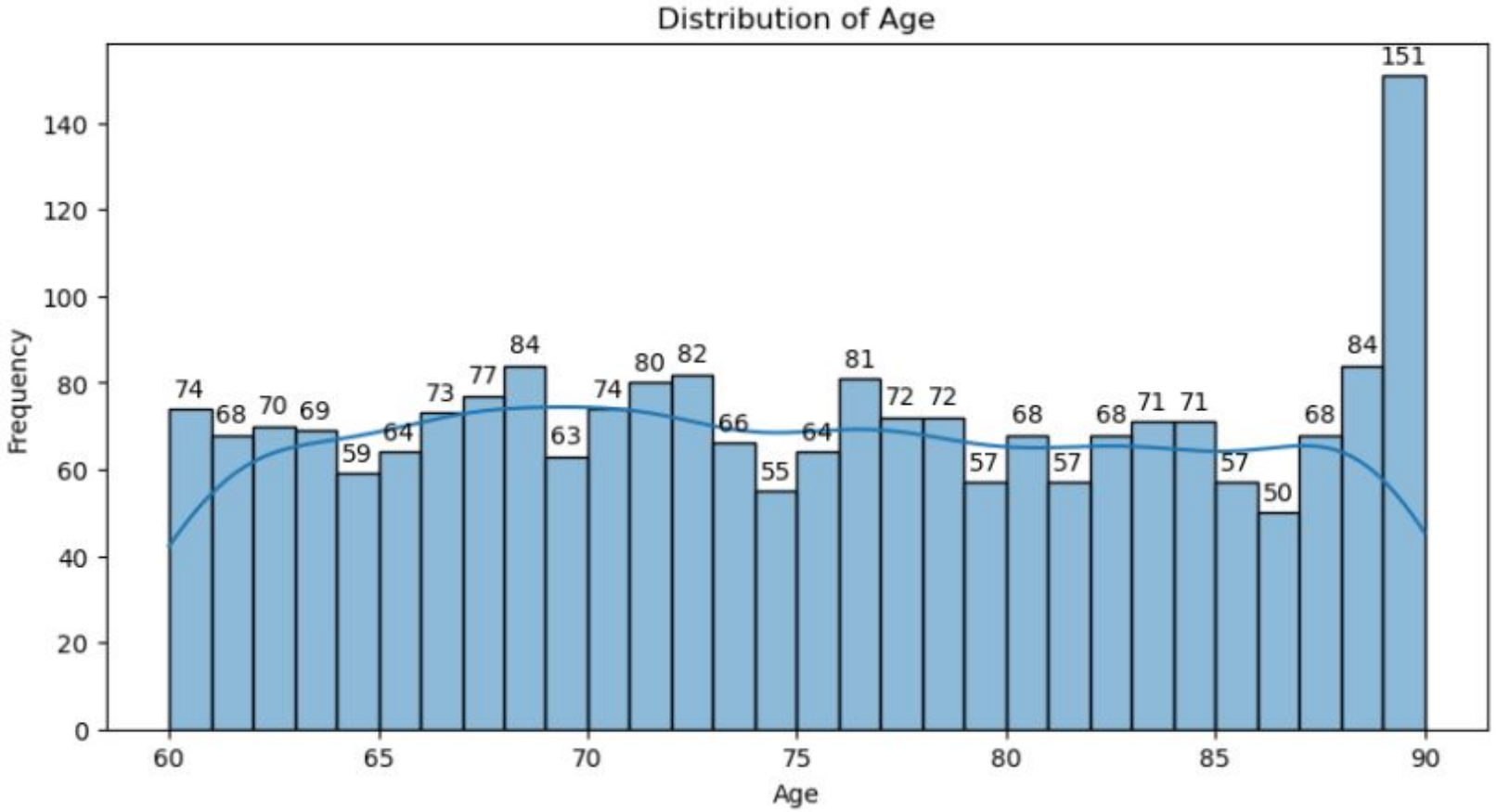
Key Insights:

we can gather that patients recorded in the dataset begin from age 60 with slightly peaking frequency at ages 68-69 and 72-73. The frequency of tested patients spike at age 89-90.

Conclusion:

Increased screening and early detection efforts should focus on individuals aged 70-80 to catch potential cases early.

Age Distribution of Diagnosed Patients



Gender Distribution

Data Presented:

Pie chart distributing male and female subjects.

Questions Answered:

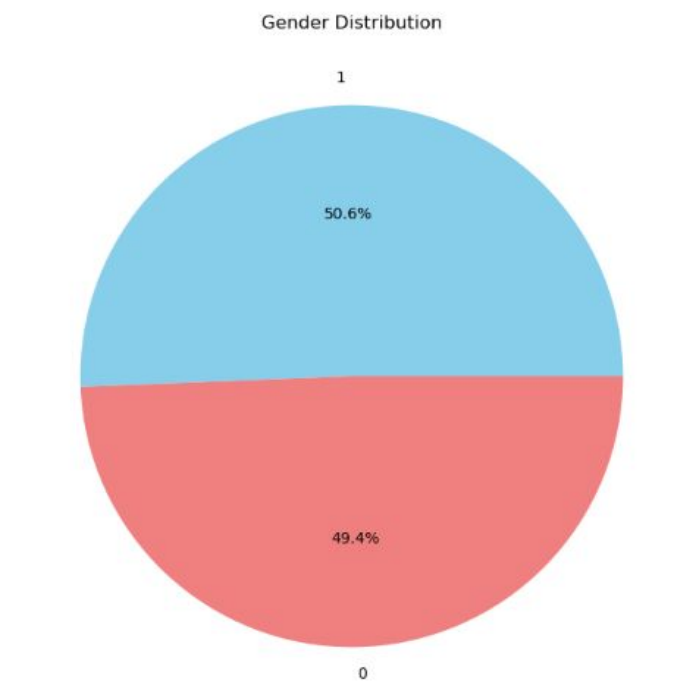
What is the gender distribution in the dataset?

Key Insights:

The pie chart reveals that 50.6% of the recorded patients are women and 49.4% are men.

Conclusion:

The distribution is nearly equal indicating that gender doesn't affect the outcome or frequency of contracting Alzheimer's Disease.



Ethnicity Distribution

Data Presented: Ethnicity counts.

Questions Answered:

What is the distribution of

Ethnicity among patients?

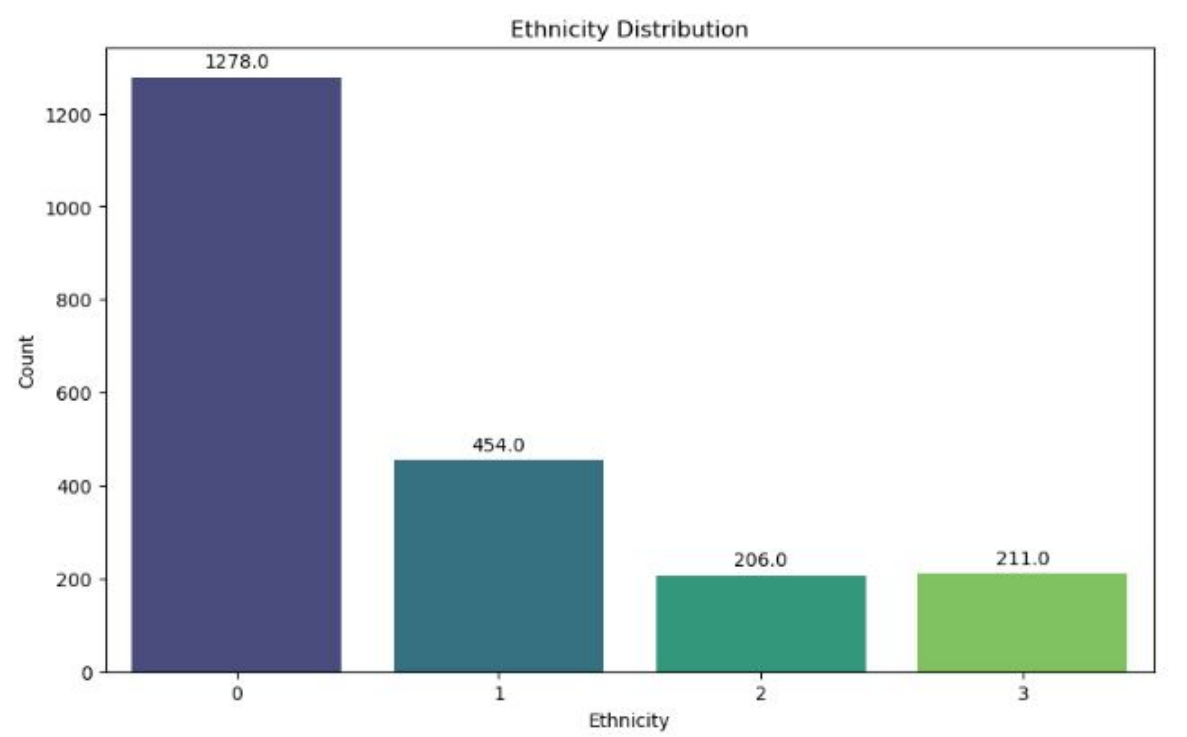
Key Insights:

Caucasian: 1278 patients (highest count)

African American: 454 patients

Asian: 206 patients

Other: 211 patients



Age Distribution by Education Level

Data Presented: Distribution of age for different education levels.

Questions Answered:

Does age vary across different education levels?

Key Insights:

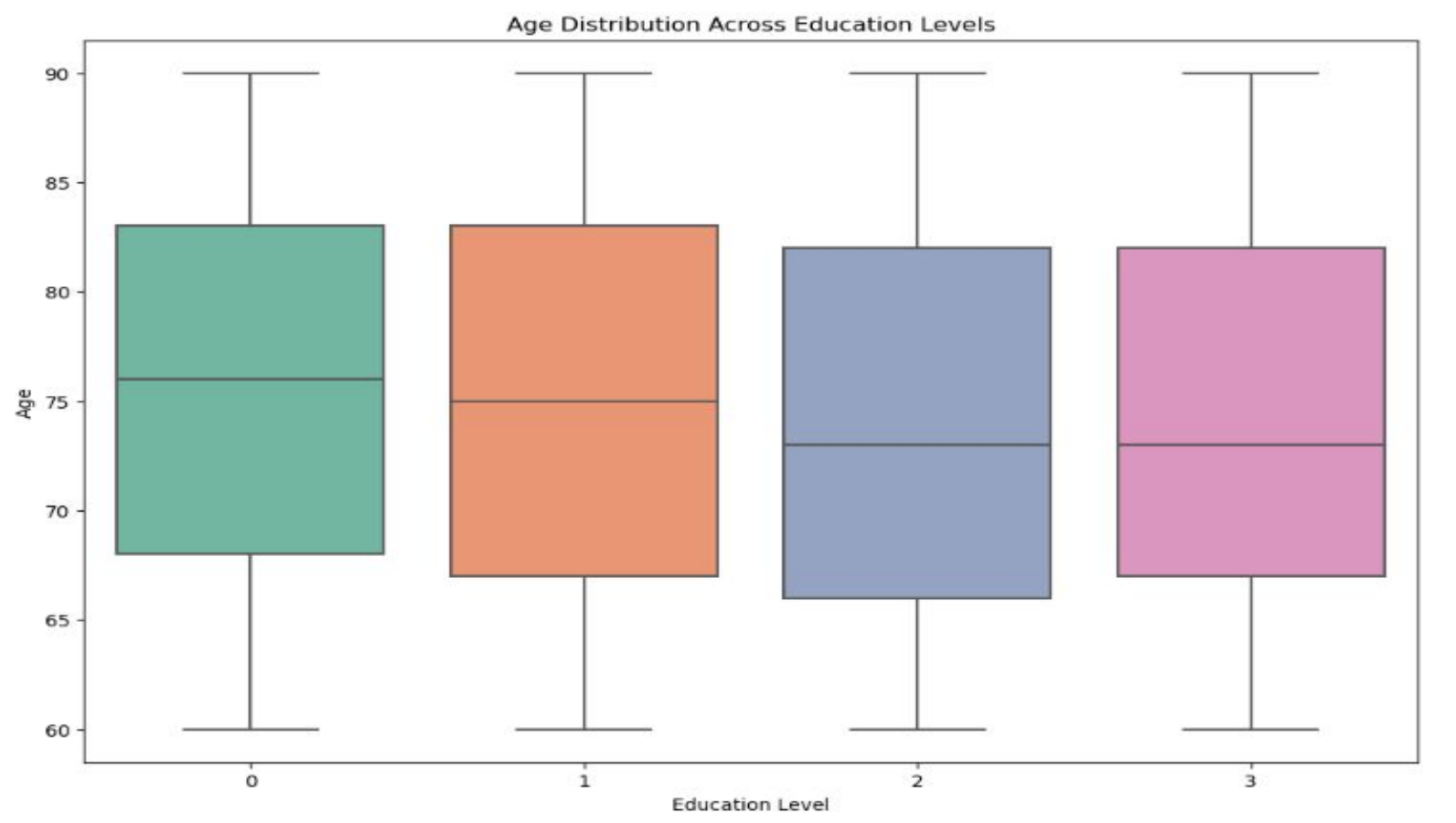
The median age is relatively consistent across the education levels, with slight variations.

Conclusion:

Higher education levels show a broader range in age, which could suggest that individuals with higher education levels may remain cognitively active for longer.

The data suggests that while education level does not have a strong direct impact on age, individuals with higher education tend to exhibit a broader range of ages.

Age Distribution by Education Level



BMI Distribution by Diabetes Status

Data Presented: Distribution of BMI values for patients with and without diabetes.

Questions Answered: Is there a relationship between BMI and the presence of diabetes?

Key Insights:

Diabetic patients tend to have a higher BMI, with most values concentrated between 25 and 40.

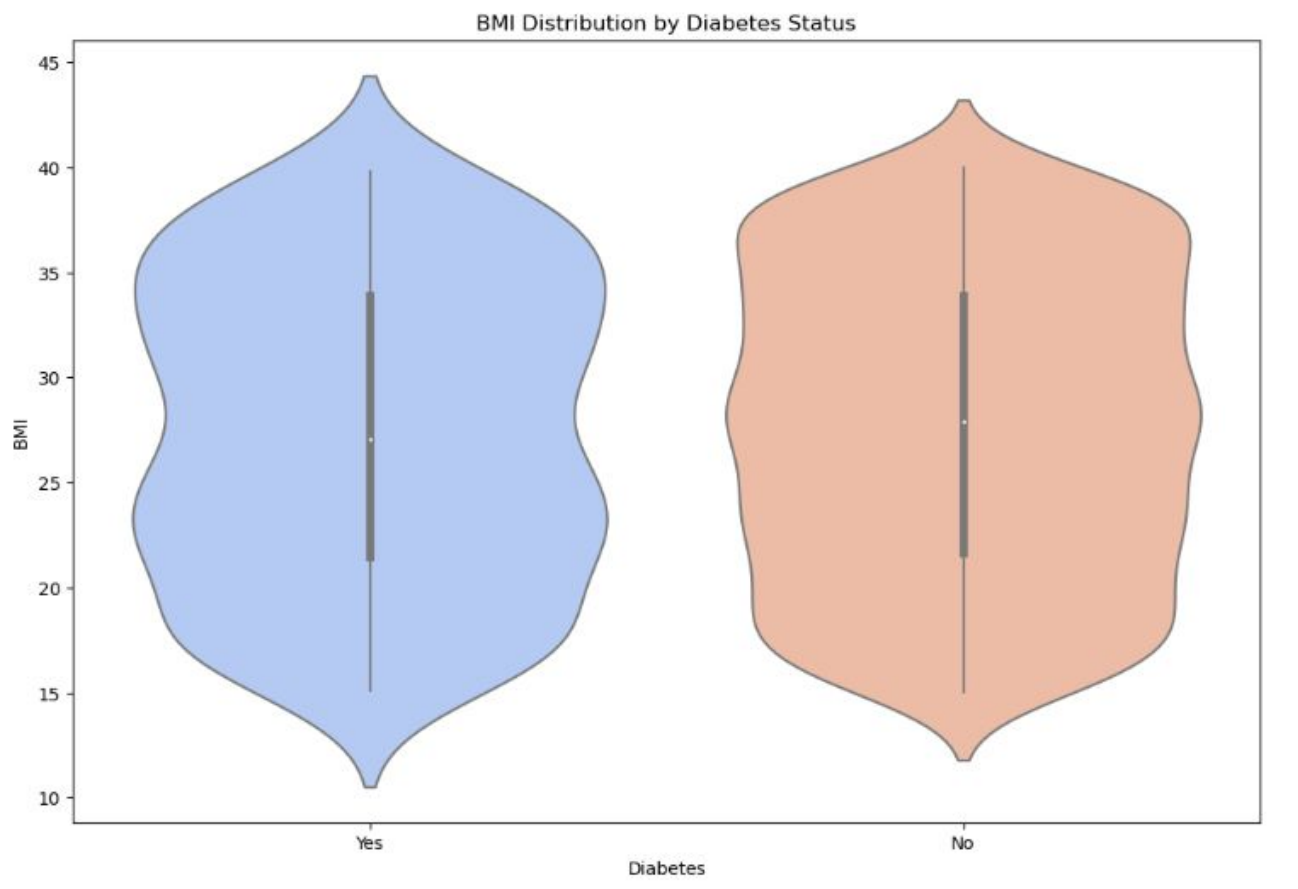
Non-diabetic patients have a more distributed BMI range, with lower median values around 20-25.

Conclusion:

Diabetic individuals generally have higher BMIs, indicating that obesity could be a significant factor in diabetes prevalence.

Highlighting the relationship between BMI and diabetes could guide healthcare providers to focus on lifestyle changes, such as diet and exercise programs, to reduce the risk of diabetes and related cognitive decline.

BMI Distribution by Diabetes Status



Systolic VS Diastolic Blood Pressure

Data Presented: The correlation between systolic and diastolic blood pressure.

Questions Answered: How does systolic and diastolic blood pressure correlate?

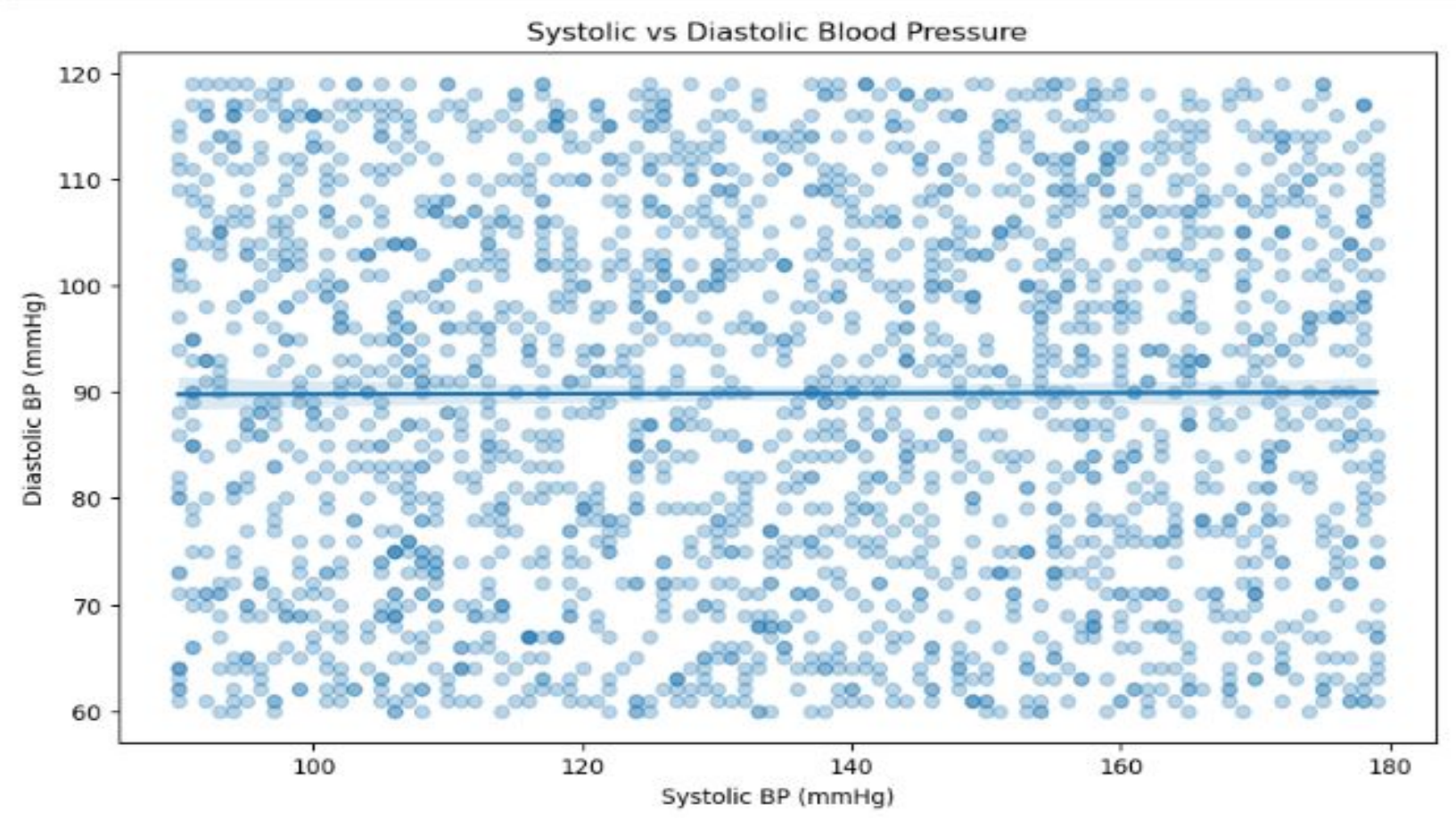
Key Insights:

The regression line indicates a weak correlation, suggesting that systolic and diastolic pressures do not vary in a tightly linear fashion.

Conclusion:

This insight suggests that changes in blood pressure, though important for general health, might not have a strong correlation in this dataset, possibly requiring other factors to be analyzed for a more comprehensive understanding of its impact on cognitive health.

Systolic VS Diastolic Blood Pressure



Alzheimer's Disease Diagnosis

Data Presented: Number of patients diagnosed with Alzheimer's disease.

Questions Answered: How many patients have been diagnosed with Alzheimer's disease?

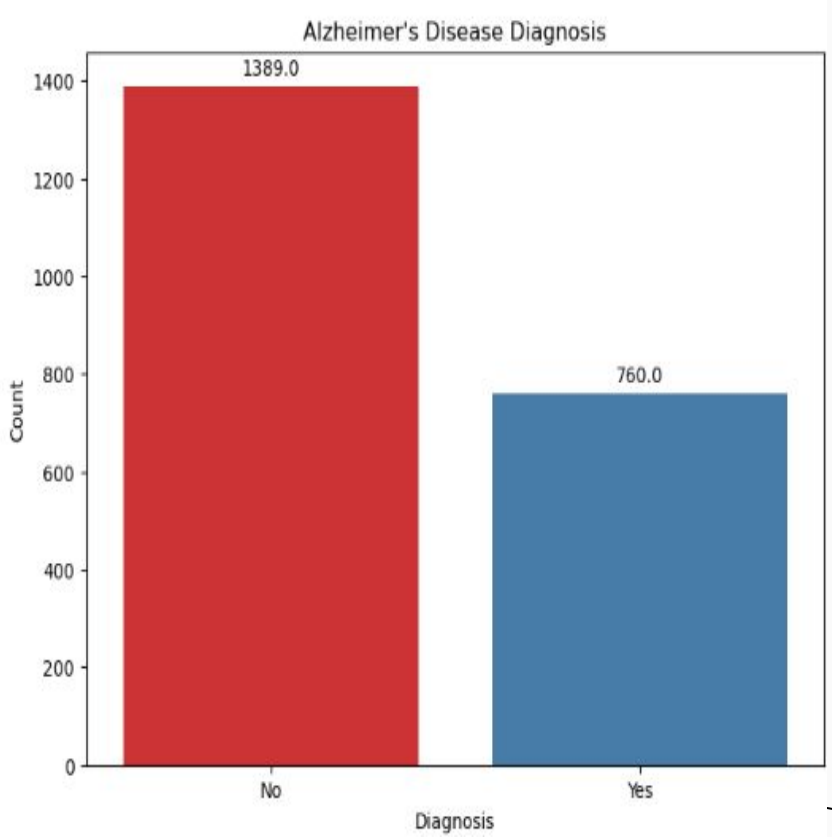
Key Insights:

Diagnosed: 760 patients (35.4%)

Not Diagnosed: 1389 patients (64.6%)

Conclusion:

Approximately two-thirds of the patients in the dataset have not been diagnosed with Alzheimer's



Age Distribution by Alzheimer’s Diagnosis

Data Presented: Age distribution among diagnosed and non-diagnosed patients.

Questions Answered: What is the age distribution for patients with and without Alzheimer’s?

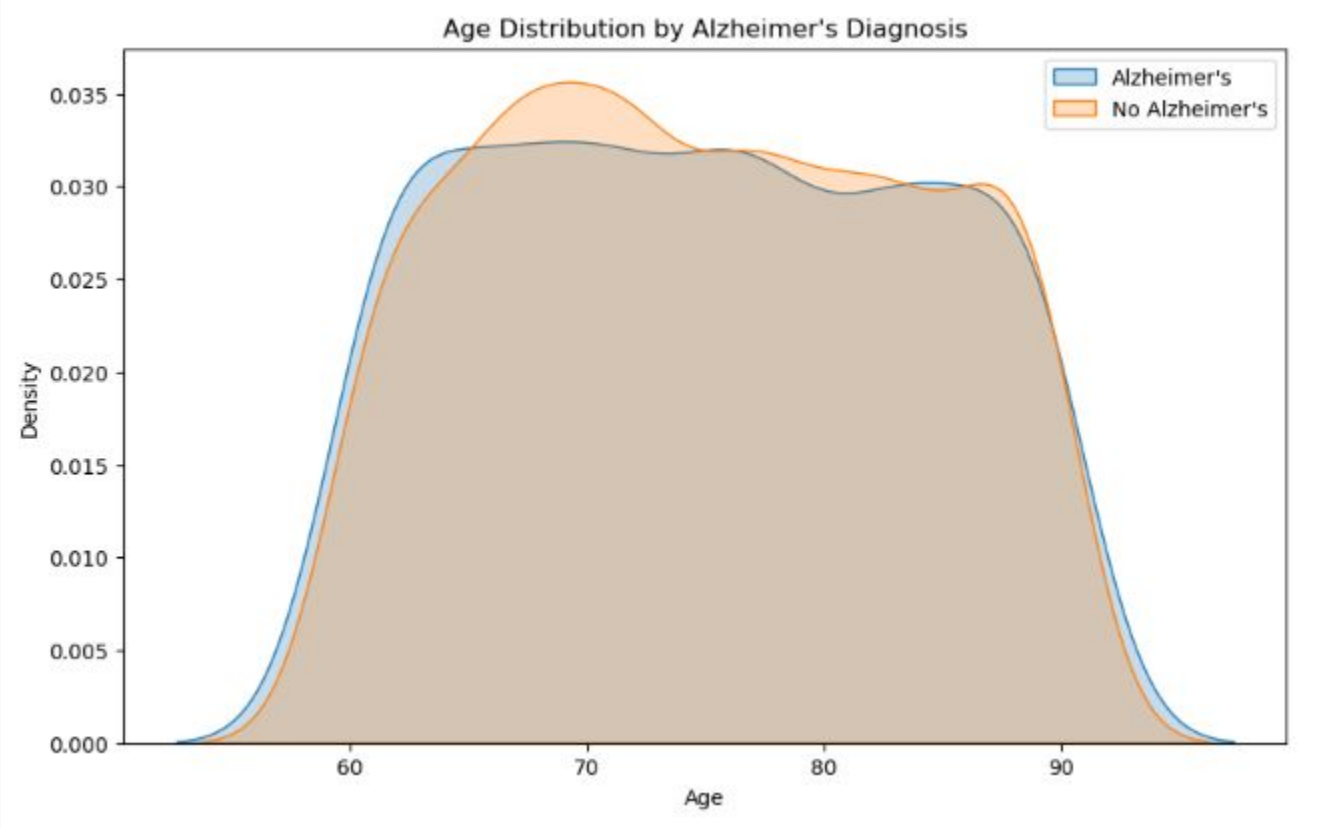
Key Insights:

Alzheimer’s patients tend to peak in the age range of 80, while non-Alzheimer’s patients peak around 70.

Conclusion:

The risk of Alzheimer's increases with age, particularly beyond 70 years. Screening and preventive measures should be targeted toward this age group for better early diagnosis.

Age Distribution by Alzheimer’s Diagnosis



Cholesterol Levels by Smoking Status

Data Presented: Cholesterol levels among smokers and non-smokers.

Questions Answered: How does cholesterol level vary with smoking habits?

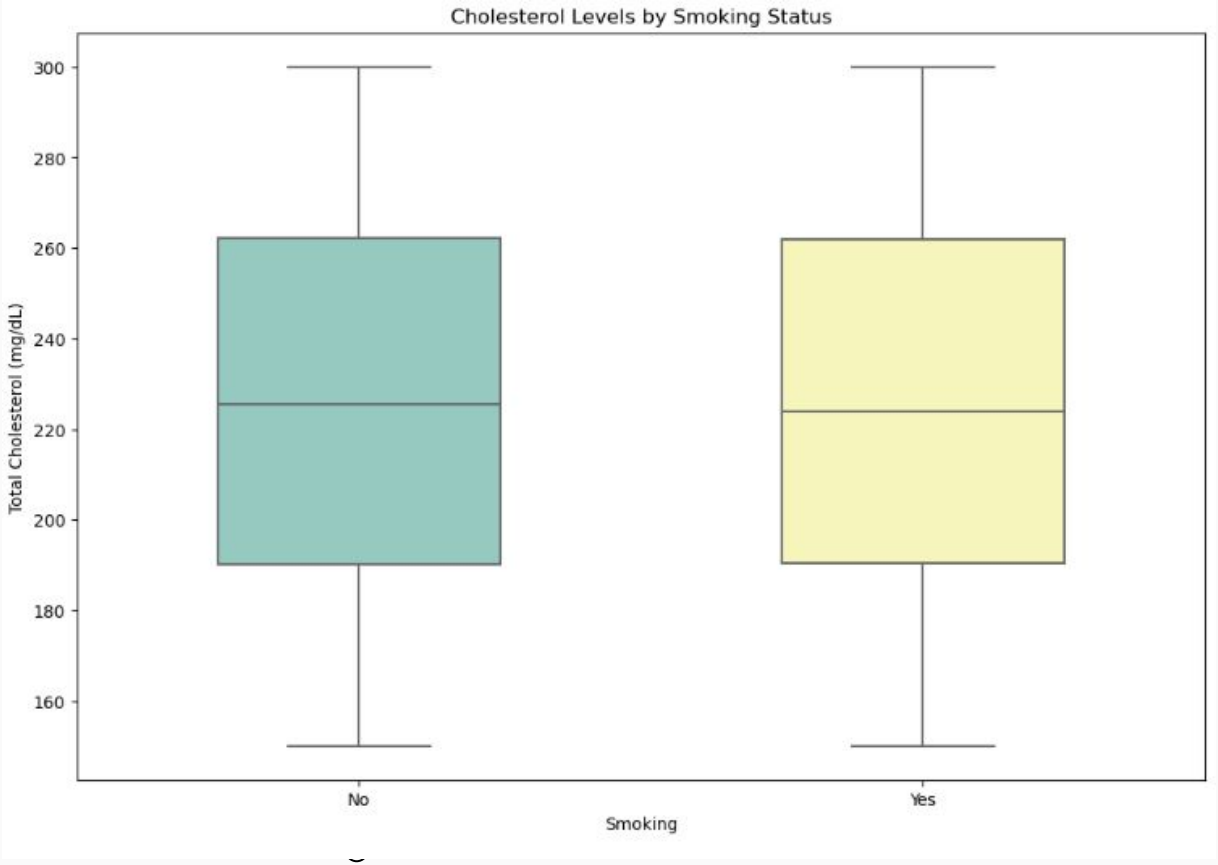
Key Insights:

Smokers tend to have slightly higher cholesterol levels, with more variability in their cholesterol levels compared to non-smokers.

Conclusion:

The higher cholesterol levels in smokers suggest that smoking cessation could reduce cardiovascular risks, which is essential in managing Alzheimer's patients and reducing health complications.

Cholesterol Levels by Smoking Status



Memory Complaints by Education Level

Data Presented: The proportion of memory complaints by different education levels.

Questions Answered: What is the relationship between education level and memory complaints?

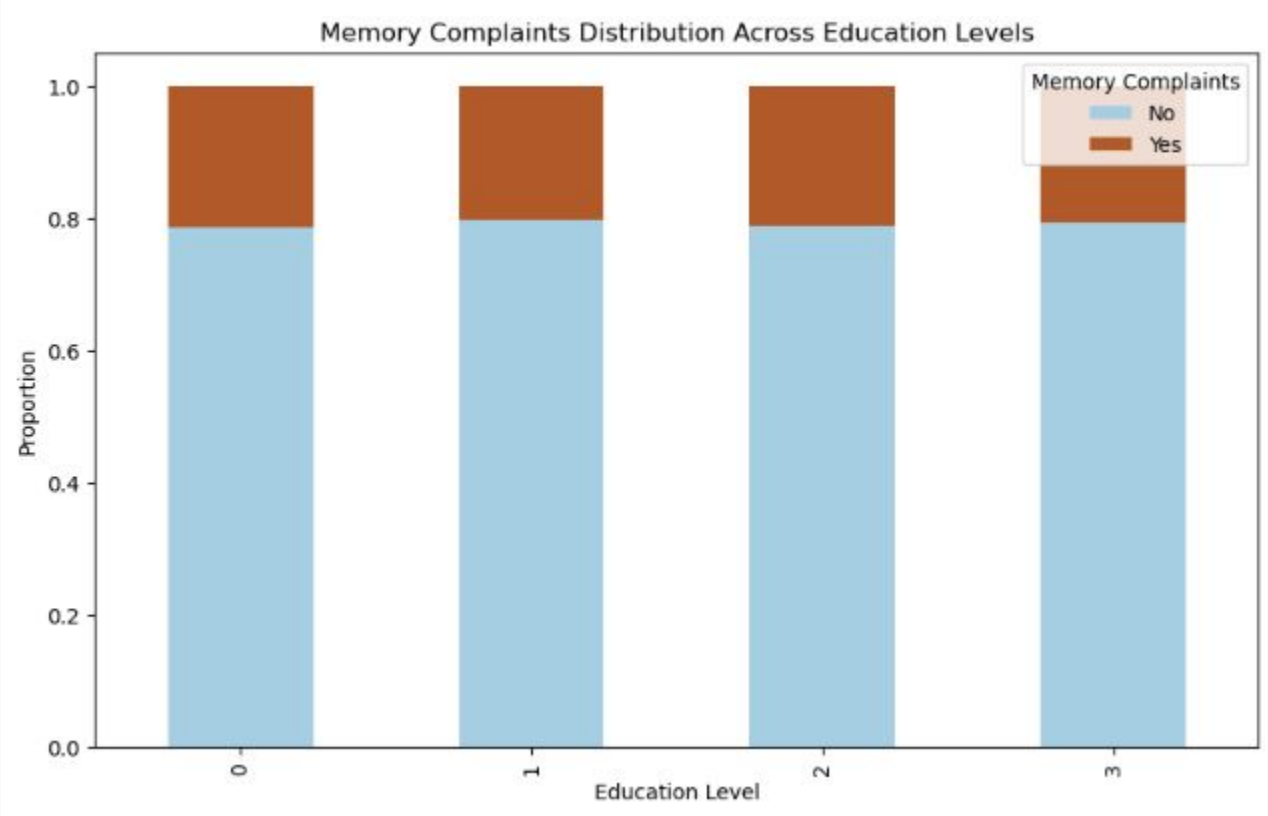
Key Insights:

Patients with lower education levels report more memory complaints compared to higher education levels.

Conclusion:

Lower education levels are associated with increased memory complaints. This suggests that individuals with less education may benefit from targeted cognitive health interventions and preventive measures to reduce the risk of Alzheimer’s.

Memory Complaints by Education Level



Alcohol Consumption by Cardiovascular Disease Status

Data Presented: Alcohol consumption among patients with and without cardiovascular disease.

Questions Answered: Is there a relationship between alcohol consumption and cardiovascular disease?

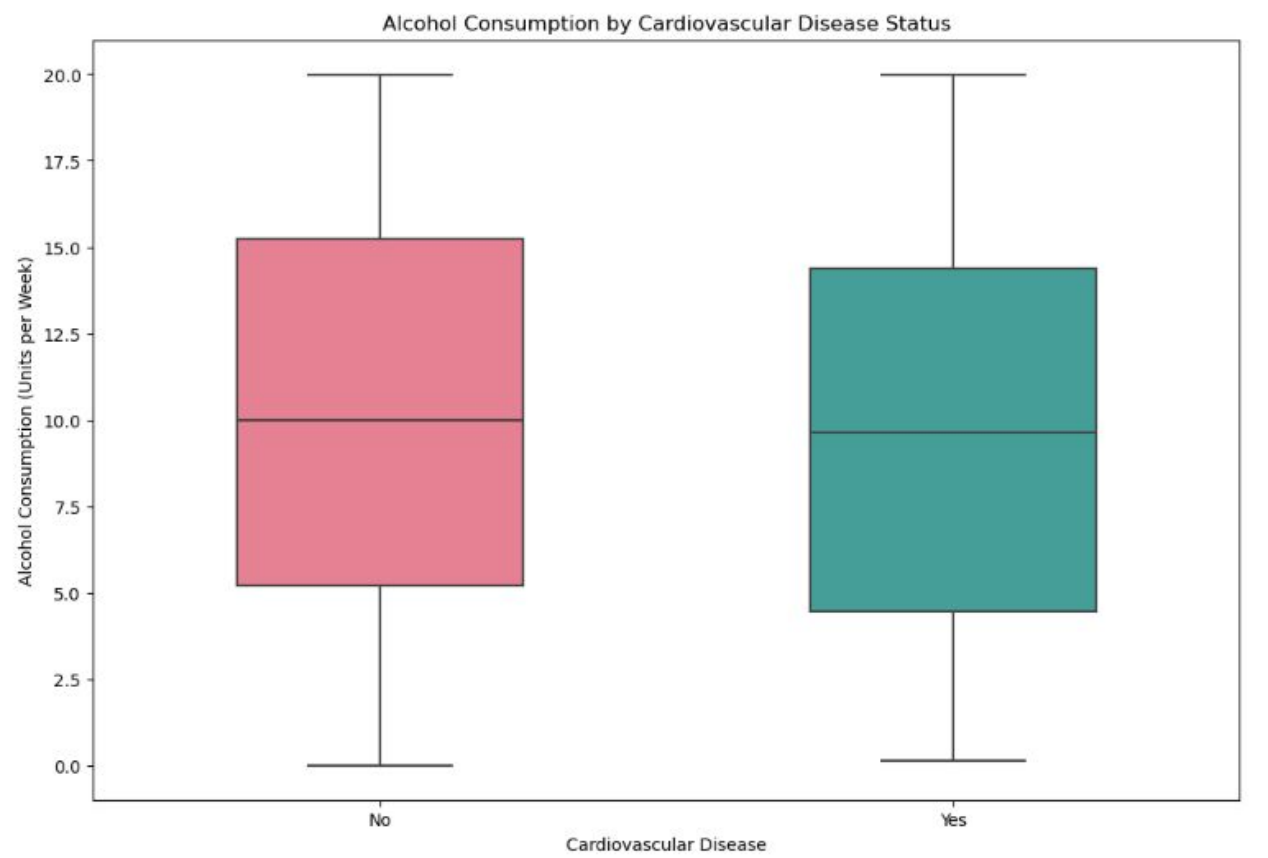
Key Insights:

The median alcohol consumption does not vary significantly between those with and without cardiovascular disease, although there is more variability among patients without cardiovascular disease.

Conclusion:

Alcohol consumption may not be a significant factor in cardiovascular disease within this dataset, but further exploration could involve assessing alcohol type or quantity in relation to Alzheimer's risk.

Alcohol Consumption by Cardiovascular Disease Status



Impact of Family History on Alzheimer’s

Data Presented: The proportion of Alzheimer's patients with and without a family history of the disease.

Questions Answered: How does family history affect Alzheimer’s diagnosis?

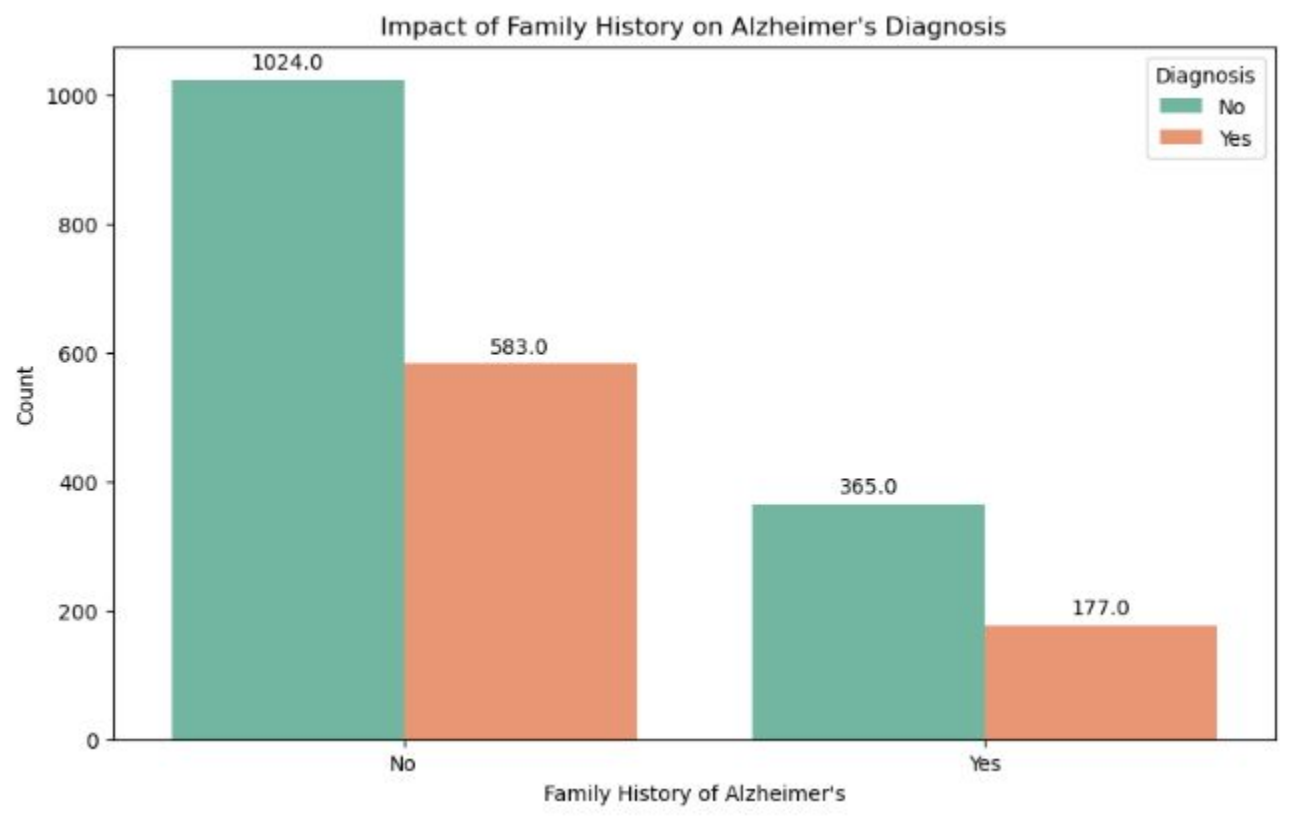
Key Insights:

Patients with a family history of Alzheimer’s have a higher likelihood of being diagnosed with the disease (33.9% compared to 17.8%).

Conclusion:

Family history is a strong predictor of Alzheimer's risk, reinforcing the need for early screening and genetic counseling for individuals with a familial history of the disease.

Impact of Family History on Alzheimer’s



Sleep Quality vs Physical Activity

Data Presented: The relationship between sleep quality and physical activity levels.

Questions Answered: How does physical activity impact sleep quality?

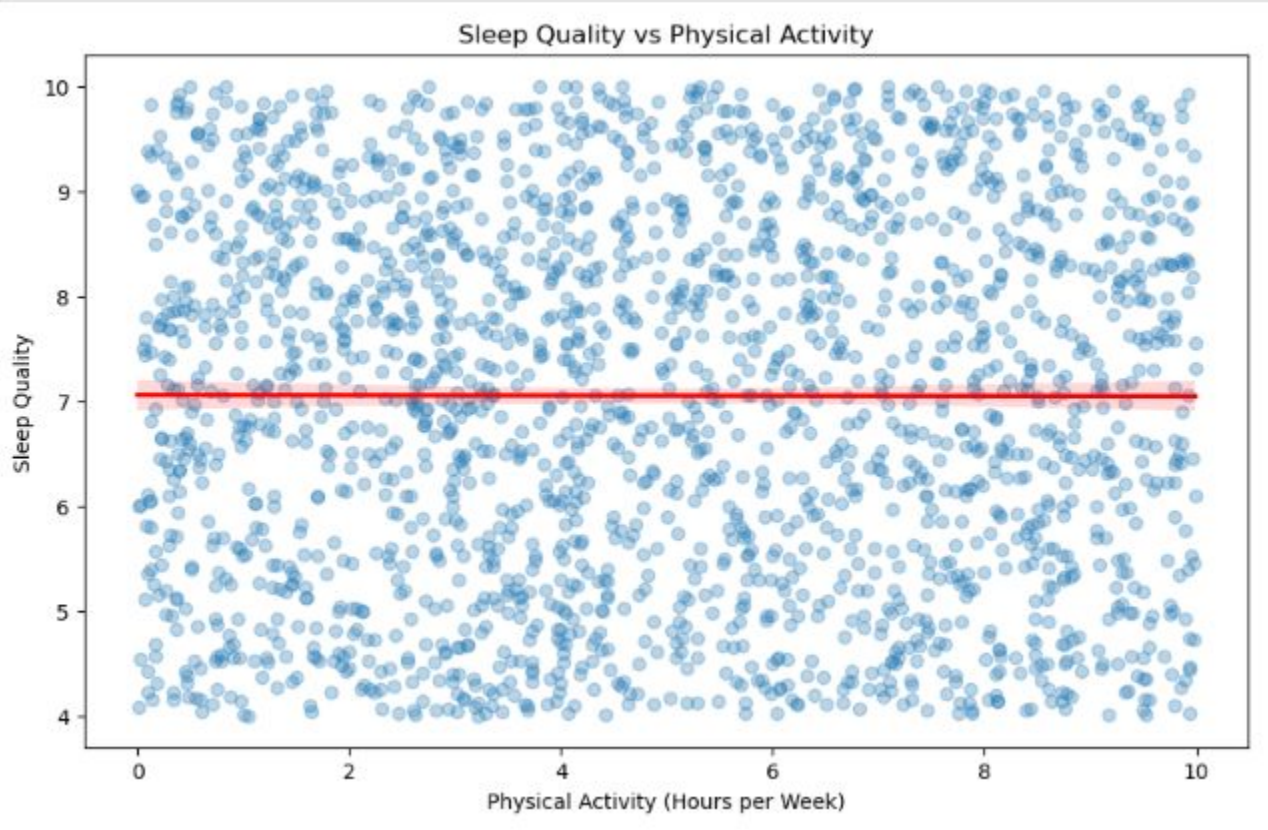
Key Insights:

Patients with higher physical activity levels tend to report better sleep quality, though there is variability across all activity levels.

Conclusion:

Increased physical activity is associated with improved sleep quality, which may benefit cognitive health in Alzheimer's patients. Physical exercise should be promoted as part of a holistic care plan.

Sleep Quality vs Physical Activity



Depression History and Alzheimer's Disease

Data Presented: The relationship between a history of depression and Alzheimer's diagnosis.

Questions Answered: What is the relationship between depression history and Alzheimer's diagnosis?

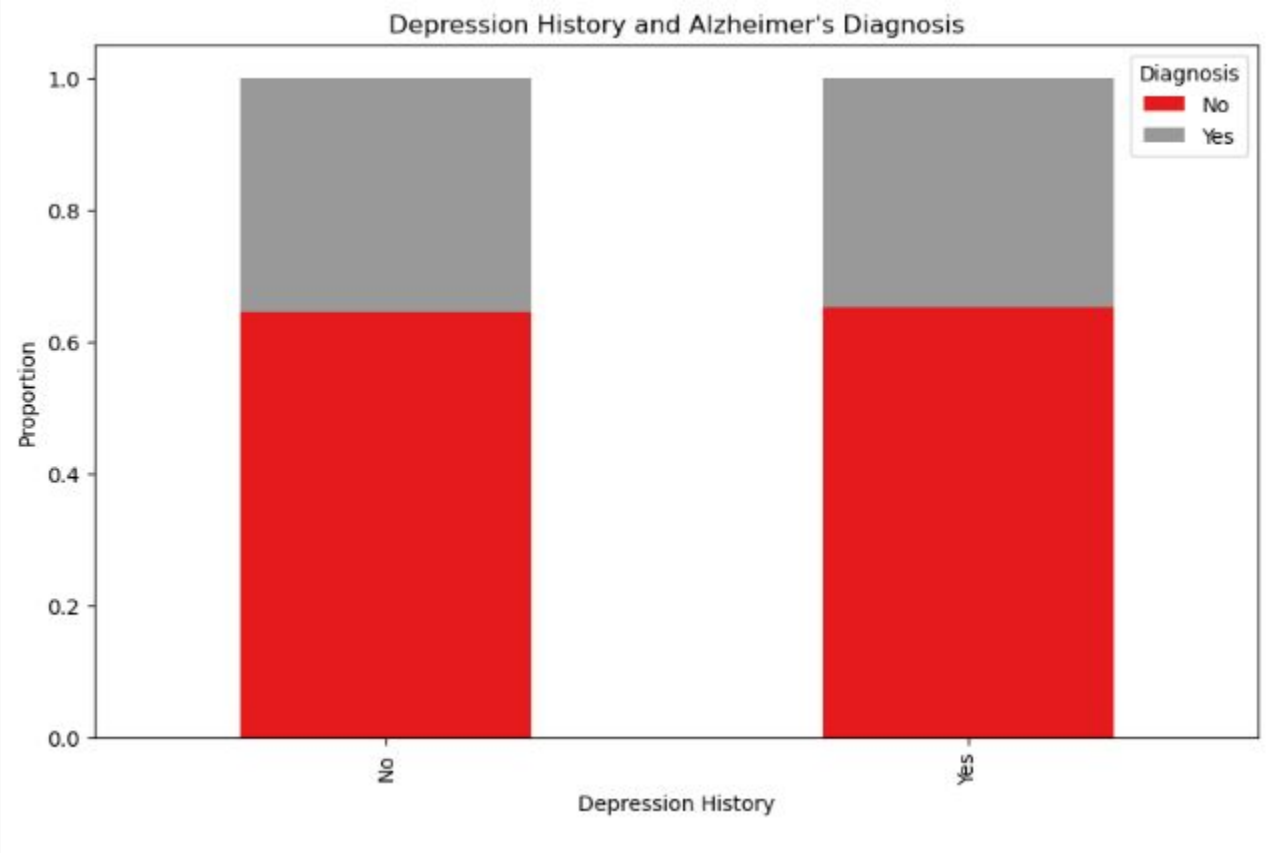
Key Insights:

Approximately 70% of patients with a history of depression are also diagnosed with Alzheimer's, compared to 40% for those without a history of depression.

Conclusion:

A history of depression is strongly linked to Alzheimer's diagnosis. This underscores the importance of mental health management in reducing the risk or delaying the onset of Alzheimer's in at-risk populations.

Depression History and Alzheimer’s Disease



BMI Distribution by Behavioral Problems

Data Presented: The relationship between BMI and the presence of behavioral problems.

Questions Answered: Is there a relationship between BMI and behavioral problems in Alzheimer's patients?

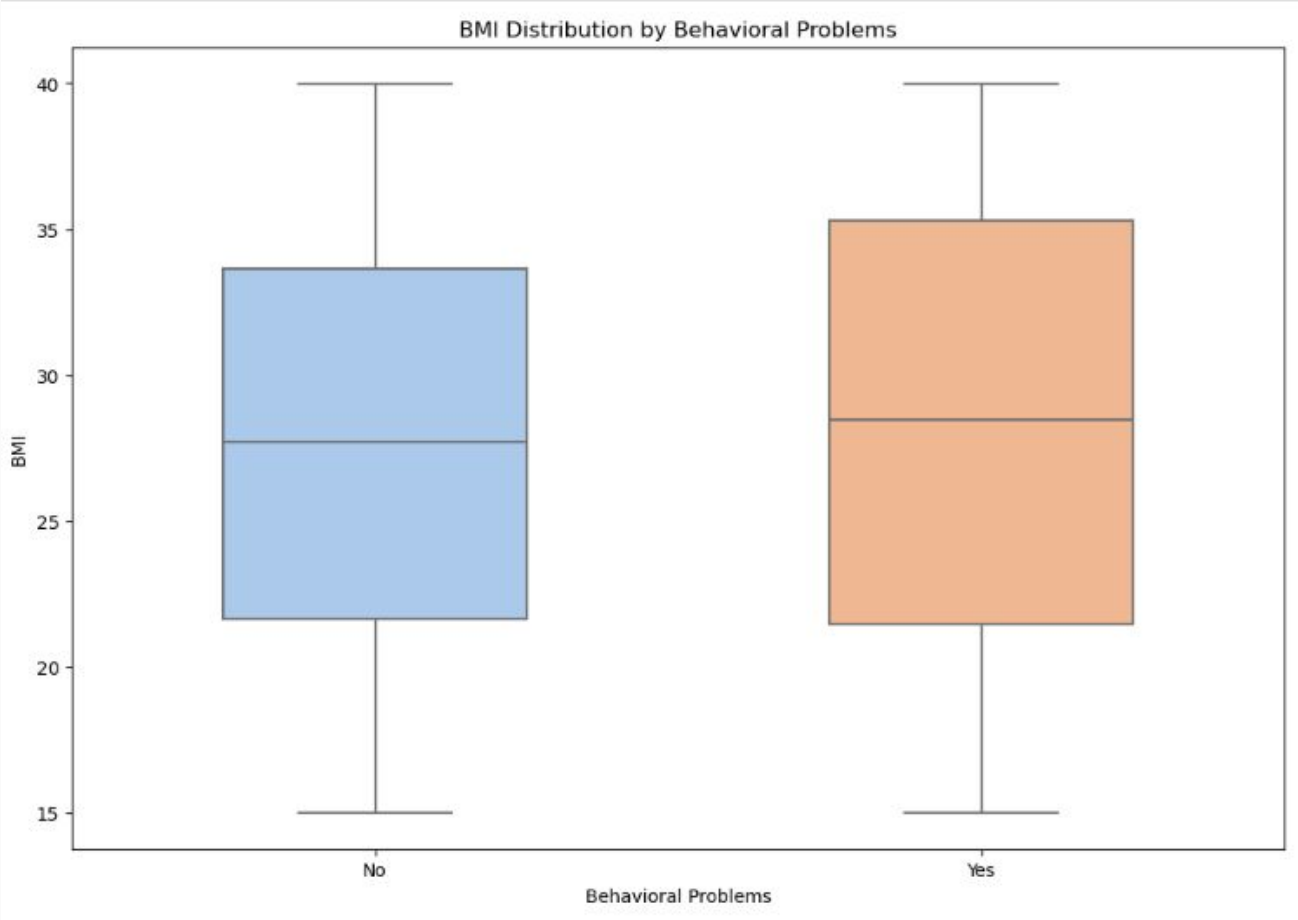
Key Insights:

Patients with behavioral problems tend to have higher BMI, with more variability in BMI compared to those without behavioral issues.

Conclusion:

The association between higher BMI and behavioral problems suggests that managing weight could play a role in mitigating behavioral issues in Alzheimer’s patients. Behavioral health strategies should also consider physical health factors such as BMI.

BMI Distribution by Behavioral Problems



04.

Conclusion

CONCLUSIONS

The EDA reveals significant correlations between demographic factors, lifestyle choices, medical history, and Alzheimer's Disease, offering numerous points of intervention. Focused attention on age, physical health (BMI, cardiovascular health, and diabetes), and mental health (depression history) will be critical in early diagnosis and management. Furthermore, promoting physical activity, ensuring proper sleep, and addressing lifestyle factors like smoking and alcohol consumption may aid in slowing disease progression and improving patient outcomes.

These findings provide a foundation for healthcare providers to develop more personalized care strategies, emphasizing the need for early diagnosis, lifestyle interventions, and mental health support. By identifying and understanding key patterns in this dataset, researchers can continue advancing knowledge on Alzheimer's Disease, potentially guiding future medical research and public health policies.