Alzheimer's Disease: A Study of Predictors and Risks

Aishwarya Movva

Table of contents
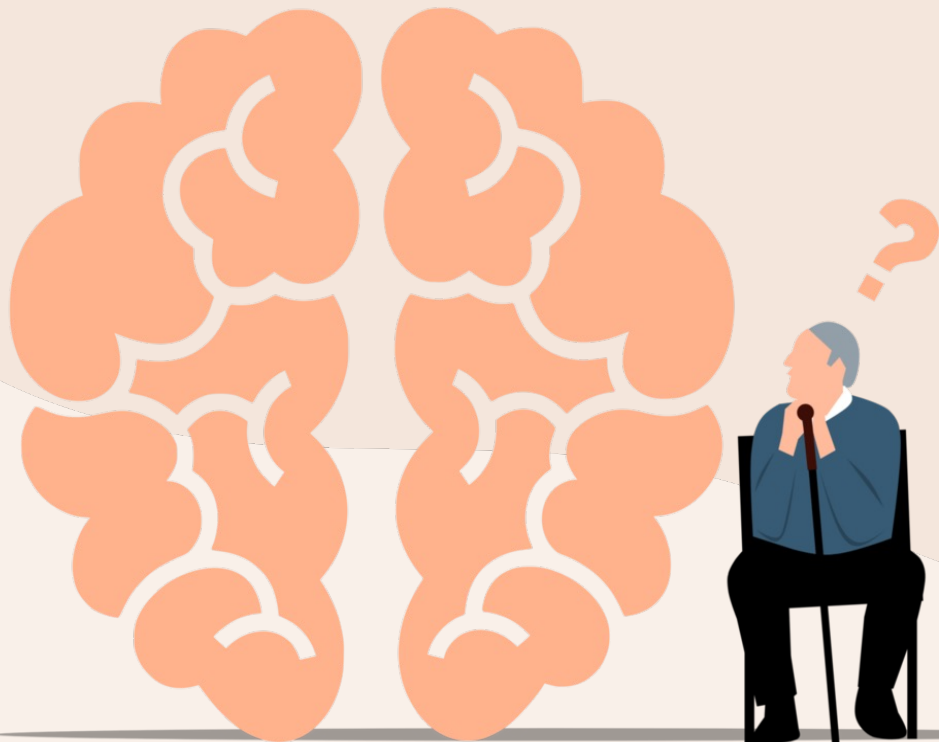
Overview of what we will be covering

INTRODUCTION

* **Background, Research Focus**

Background

Alzheimer's disease is a progressive neurodegenerative disorder and the most common cause of dementia. It primarily affects memory, cognition, and behavior, leading to significant functional decline.

Why is it important?

* **Growing prevalence:** Rising cases, especially in aging populations.

  Critical need for early detection: Identifying predictors can improve interventions and outcomes.

How this study helps:

* **Investigates** relationships between **demographics, lifestyle factors, and cognitive metrics** with Alzheimer's diagnosis.

  Aims to support **early detection** strategies for better patient care.

Why I'm interested in this topic?

* *I chose this topic because of my interest in understanding how data science can address critical healthcare challenges.*

  Alzheimer's disease affects millions globally, and I wanted to explore how data can provide insights for early detection and intervention.

Research Focus

This study aims to explore three key research questions:

QUESTION 1

Which demographic factors are most strongly associated with Alzheimer's diagnosis?

QUESTION 2

How do lifestyle and health metrics relate to diagnosis?

QUESTION 3

Are behavioral and cognitive assessments significant predictors of Alzheimer's diagnosis?

Understanding these factors supports early detection and improves patient outcomes

Data Overview

✳ **Data Summary**

Data Summary

Understanding the dataset

Dataset Source:

Sourced from Kaggle, designed to study Alzheimer's research

Structure:

- 2,143 patient records
- **18 key variables** selected from an original dataset of 36 variables.

Key Variable Categories:

- **Demographics**: Age, gender, ethnicity, education level.

- **Lifestyle**: Physical activity, diet quality, smoking.

- **Health Metrics**: BMI, cholesterol levels, blood pressure, family history of Alzheimer's.

- **Cognitive**: MMSE scores, memory complaints, behavioral changes.

Not Diagnosed: 65% | Diagnosed: 35%

Distribution of Alzheimer's Diagnosis
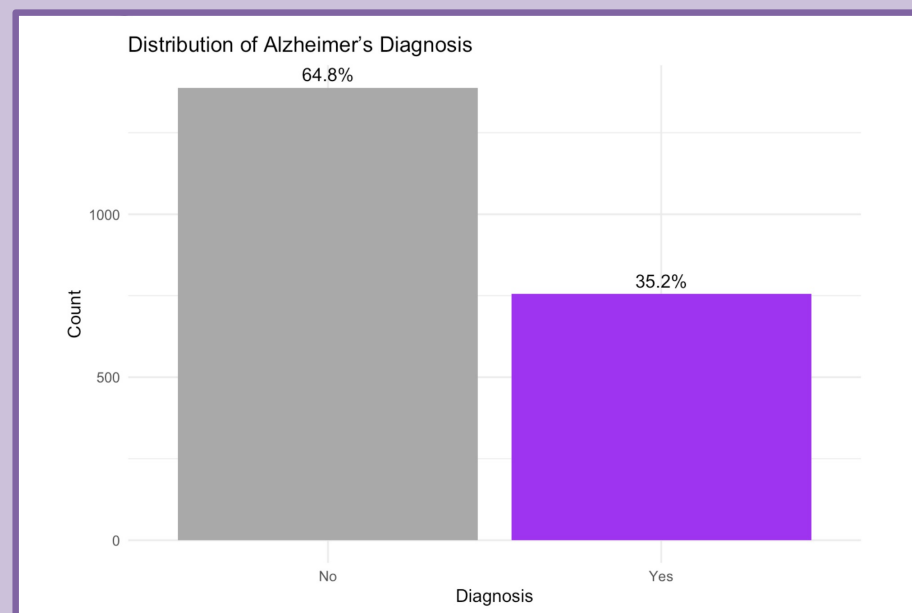
This dataset provides a variety of variables to explore predictors of Alzheimer's diagnosis and align with the research questions.

EXPLORATORY DATA ANALYSIS

✳ Data Wrangling, Key Cognitive Predictors, Additional Insights

Data Wrangling

Transforming raw data & preparing for analysis

Purpose:

To clean, organize, and refine the dataset for accurate and meaningful analysis.

Key Steps:

- **Filtered Variables**:
  - Reduced from 36 to 18 key variables by selecting those most relevant to demographics, lifestyle, health metrics, and cognitive measures.

- **Recoded Categories**:
  - Gender: Recoded as Male and Female.
  - Ethnicity: Converted numerical codes to descriptive categories (ex: Caucasian, African American, etc.).
  - Diagnosis: Changed binary values to "Yes" and "No" for clarity.

- **Rounded Continuous Variables**:
  - Variables like MMSE scores, BMI, physical activity, and diet quality were rounded to one decimal place for consistent visualizations and interpretation.

- **Checked for Missing Data**:
  - Minimal missingness was observed; no imputation was required.

Before:

Age	Gender	Ethnicity	EducationLevel	BMI	Smoking
73	0	0	2	22.92775	0

After:

Age	Gender	Ethnicity	Smoking	PhysicalActivity	FamilyHistoryAlzheimers
73	Male	Caucasian	No	6.3	No

Created a clean, organized dataset optimized for exploratory analysis and modeling.

Exploring Cognitive Predictors

**MMSE Scores and Memory Complaints in Alzheimer's Diagnosis**

1. MMSE Scores:

➢ MMSE (*Mini-Mental State Examination*) scores measure cognitive function and are a common tool for assessing Alzheimer's risk.

➢ Lower scores indicate more severe cognitive impairment, as reflected in diagnosed participants.

2. Memory Complaints:

➢ Memory complaints provide a subjective measure of early cognitive decline.

➢ This variable bridges the gap between clinical observation and patient-reported symptoms.

Cognitive Predictors in Context - Combining objective tests like MMSE with self-reports of memory complaints improves early detection strategies.
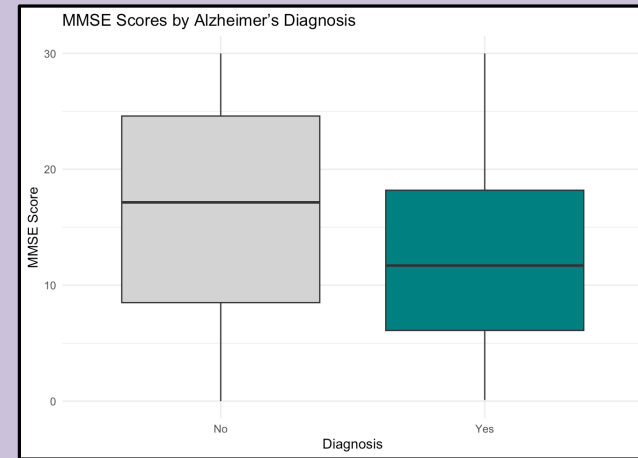
*Boxplot showing MMSE scores by Alzheimer's diagnosis. **Diagnosed** participants had significantly **lower** scores **(median = 14.5)** compared **to non-diagnosed** participants **(median = 26)**, reflecting cognitive decline.*
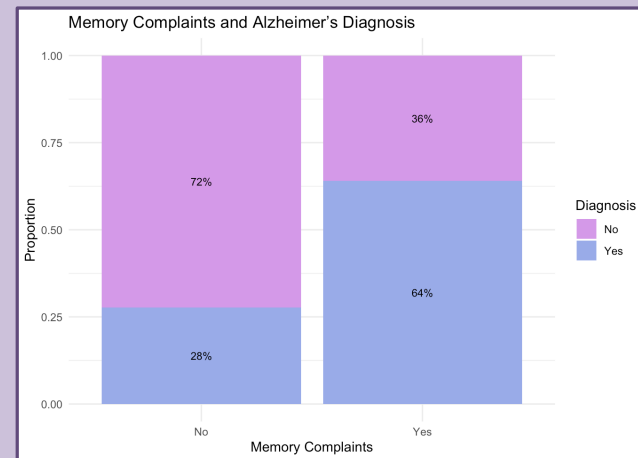
*Proportion of participants reporting memory complaints by diagnosis. **64% of diagnosed** participants reported memory complaints, compared to 28% of **non-diagnosed** participants.*

Additional Insights

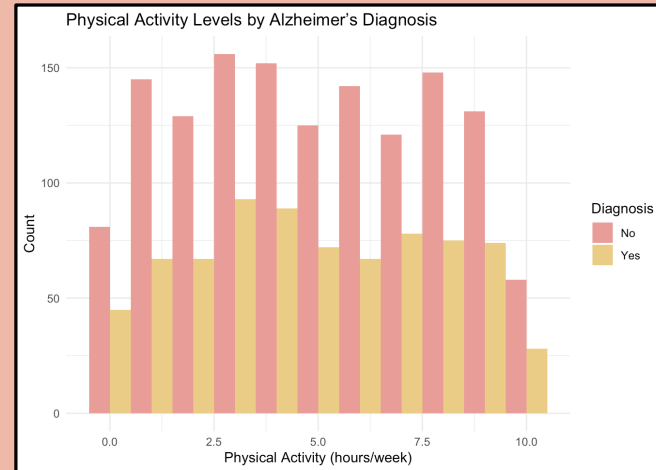**Physical activity and Education level in Alzheimer's Diagnosis**

1. Physical Activity:

➤ Diagnosed participants reported lower physical activity levels.

➤ Regular exercise may reduce Alzheimer's risk by promoting brain health.
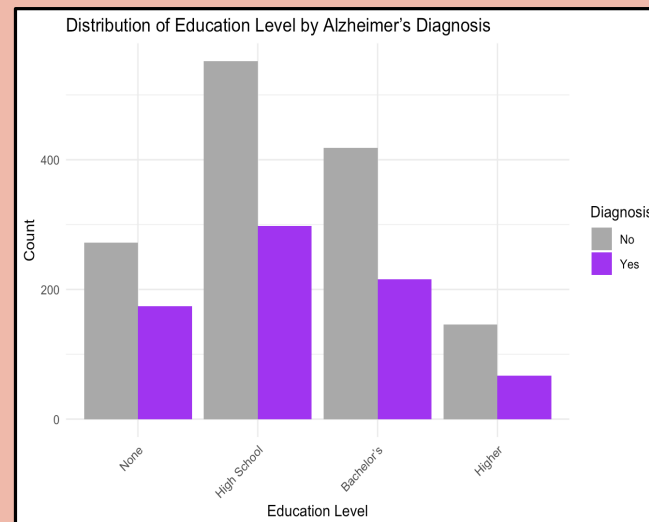
2. Education Level:

➤ High school education was the most common among diagnosed participants.

➤ Higher education may provide cognitive reserve, potentially delaying Alzheimer's symptoms.

Key Takeaway: Lifestyle and cognitive factors are crucial in understanding Alzheimer's risk.

Physical Activity Levels by Alzheimer's Diagnosis

Histogram showing physical activity levels by Alzheimer's diagnosis. **Diagnosed** participants reported **lower** levels (**median = 3** hours/week) compared to **non-diagnosed** participants (**median = 6** hours/week).

Distribution of Education Level by Alzheimer's Diagnosis

Bar chart showing education levels by Alzheimer's diagnosis. **High school** education was the most common (**39.5%**), followed by Bachelor's degree (**28.6%**), with only 8.87% pursuing higher education.

Statistical Methods

*Statistical Approach, Model Evaluation, Key Predictors

Statistical Approach

Why Logistic Regression?

➢ **Evaluates** associations between **predictors** and the likelihood of diagnosis.

➢ **Evaluates** associations between **predictors** and the likelihood of diagnosis.

➢ **Handles** both **continuous** and **categorical** predictors.

Model Fitting:

➢ **Response Variable:** Diagnosis (**0 = No, 1 = Yes**).

➢ Predictors selected based on EDA and theoretical relevance.

➢ **Training data (80%)** used to fit the model; **testing data (20%)** used for evaluation.

Model Fit:

➢ **Reduced null deviance** (2223.3 to 1738.5).

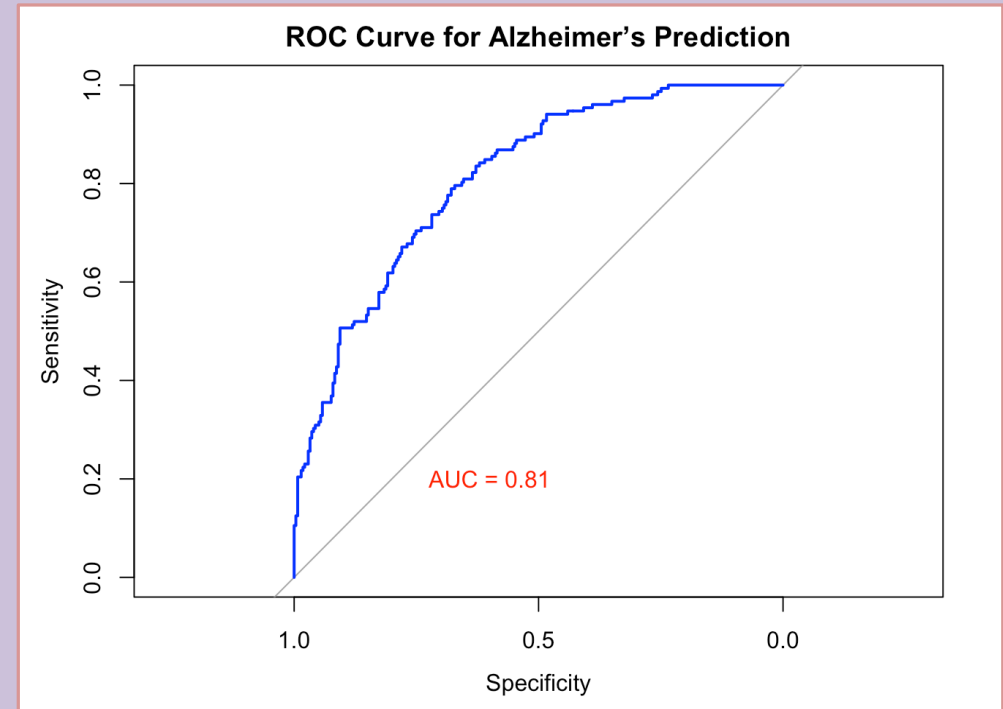➢ **AIC** (Akaike Information Criterion): **1782.5** indicates model quality

Model Evaluation

Key Metrics:

✓ **Accuracy:** 76.2% overall

✓ **AUC (Area Under the Curve):** 0.81

Insights:

✓ The **ROC Curv**e demonstrates strong performance, showcasing a good balance between sensitivity (True Positive Rate) and specificity.

✓ An **AUC value of 0.81** indicates the model correctly ranks diagnosed cases higher than non-diagnosed cases 81% of the time.

✓ The model is effective at identifying Alzheimer's cases but has room for improvement in reducing misclassifications.

ROC Curve for Alzheimer's Prediction

AUC = 0.81

- Sensitivity (y-axis): 0.0, 0.2, 0.4, 0.6, 0.8, 1.0
- Specificity (x-axis): 1.0, 0.5, 0.0

○ AUC closer to 1 indicates excellent predictive performance.
○ AUC of 0.81 signifies the model correctly ranks Alzheimer's and non-Alzheimer's cases 81% of the time.

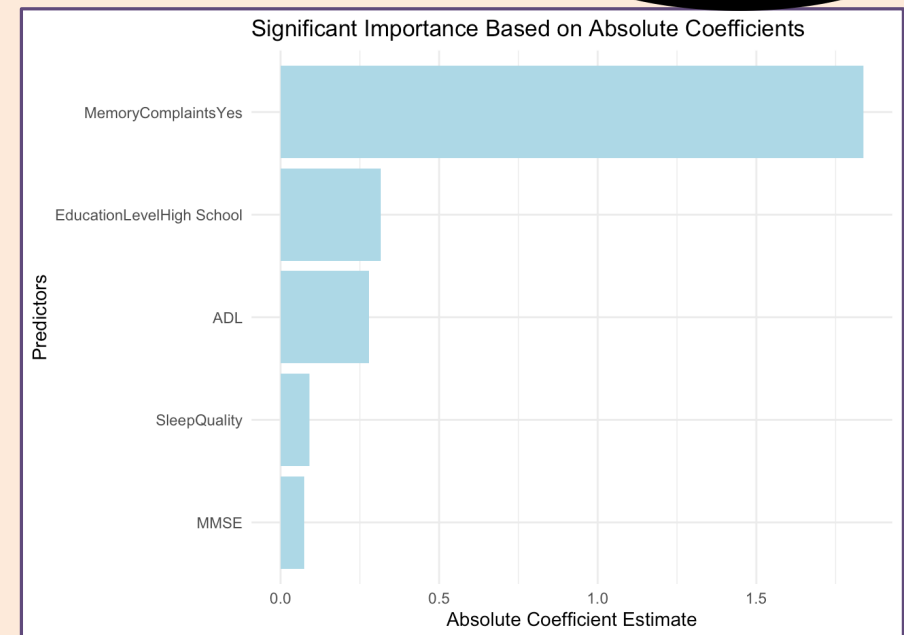# Key Predictors in Alzheimer's Diagnosis

Memory Complaints: Strongest predictor, significantly increases the likelihood of diagnosis.

Education Level: Lower education levels (ex: high school) associated with higher diagnosis likelihood.

ADL Scores: Poorer functionality linked to greater risk.

Sleep Quality: Lower scores predict higher risk of Alzheimer's.

MMSE Scores: Lower cognitive scores correlate with increased likelihood.

Significant predictors from the logistic regression model based on absolute coefficient values. Memory complaints and education level exhibit the strongest influence.

Key Findings and Insights

* **Key Insights from the Analysis**

Key Insights from the Analysis

Demographic Observations:

➤ **Gender -** A higher proportion of Alzheimer's diagnoses were observed in women.
➤ **Ethnicity -** Caucasians were the most represented group among diagnosed participants.

Significant Predictors:

➤ **Memory Complaints -** Strongest predictor, highlighting the importance of self-reported cognitive symptoms.
➤ **Education Level -** Lower education levels increased diagnosis risk, supporting the "cognitive reserve" hypothesis.
➤ **ADL -**Better functional independence reduced Alzheimer's risk.
➤ **Sleep Quality -** Poor sleep quality linked to higher diagnosis risk.

Lifestyle & Health Observations:

➤ **Physical Activity -** Lower in diagnosed participants but not statistically significant.
➤ **Cholesterol -** Elevated median levels, weak link to diagnosis.

Model Performance:

➤ **ROC-AUC -** Score of 0.81 indicates strong diagnostic accuracy.

Gender Distribution

Ethnicity Distribution (Sorted by Count)
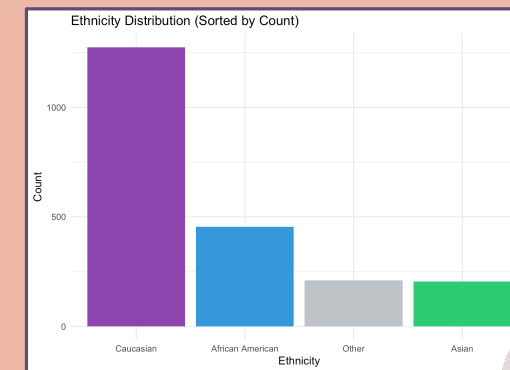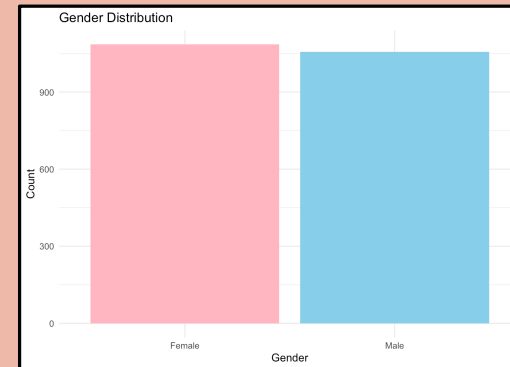
Conclusion

* **Conclusion and Future Directions**

Conclusion and Future Directions

Answering Research Questions:

1. *Which demographic factors are most strongly associated with Alzheimer's diagnosis?* Demographics like age and gender had minimal impact, but education level was significant.

2. *How do lifestyle and health metrics relate to diagnosis?* Lifestyle factors such as sleep quality emerged as predictors, while physical activity and cholesterol levels showed limited influence.

3. *Are behavioral and cognitive assessments significant predictors of Alzheimer's diagnosis?* Cognitive assessments (MMSE) and behavioral symptoms (memory complaints) were the strongest predictors.

Future Directions:

- ❑ **Enhanced Data Collection:** Incorporate additional variables such as genetic markers, socioeconomic status, and environmental factors for a more comprehensive analysis.
- ❑ **Early Detection Models:** Develop targeted models using significant predictors to aid in early diagnosis and intervention.

Thank you!