Exploring Alzheimer's Disease Risk Factors: A Data-Driven Approach

By Jesus Medina

Alzheimer's disease is a progressive brain disorder that affects memory, thinking skills, and the ability to perform simple tasks. It is a leading cause of dementia, significantly impacting patients, families, and healthcare systems. I chose to analyze a dataset on Alzheimer's patients to explore potential risk factors and patterns associated with the disease. This report details my analysis, key questions, methodology, and findings.

Dataset Information

The dataset used for this analysis was sourced from Kaggle.com and consists of 2,149 patient records across 35 categories. It includes demographic information, lifestyle factors, medical history, clinical measurements, cognitive and functional assessments, symptoms, and diagnosis data.

Key Data Categories:

- Demographics: Age, gender, ethnicity
- Lifestyle Factors: BMI, smoking, alcohol consumption
- Medical History: Family history, cardiovascular disease, diabetes, depression
- Clinical Measurements: Blood pressure, cholesterol, triglycerides
- Cognitive and Functional Assessments: Functional assessment, memory complaints
- Symptoms: Confusion, disorientation, personality changes
- Diagnosis: Alzheimer's diagnosis (Yes/No)

Purpose and Goals

My primary goal was to identify patterns and correlations among various risk factors associated with Alzheimer's disease. By leveraging this dataset, I aimed to gain insights into the demographic, lifestyle, and medical conditions that contribute to the disease, which could help improve early detection and prevention strategies.

Key Research Questions

Understanding the key factors that contribute to Alzheimer's disease is essential for improving early detection and intervention strategies. By analyzing this dataset, I aim to uncover relationships between various risk factors and Alzheimer's diagnosis, providing insights that could aid healthcare professionals and researchers. It is important to note that the results of this study are based solely on the analyzed dataset and should not be generalized as definitive conclusions about Alzheimer's disease.

The main questions I aim to answer while analyzing and studying this dataset were:

1. What is the relationship between age and the likelihood of receiving an Alzheimer's diagnosis?
2. Is there a correlation between ethnicity and the likelihood of developing Alzheimer's disease?
3. Does a family history of Alzheimer's significantly increase an individual's risk of being diagnosed with the disease?
4. How do lifestyle factors, such as BMI, smoking, and physical activity, influence the risk of Alzheimer's?

Metrics and KPIs

To track my analysis effectively, I defined the following key performance indicators (KPIs):

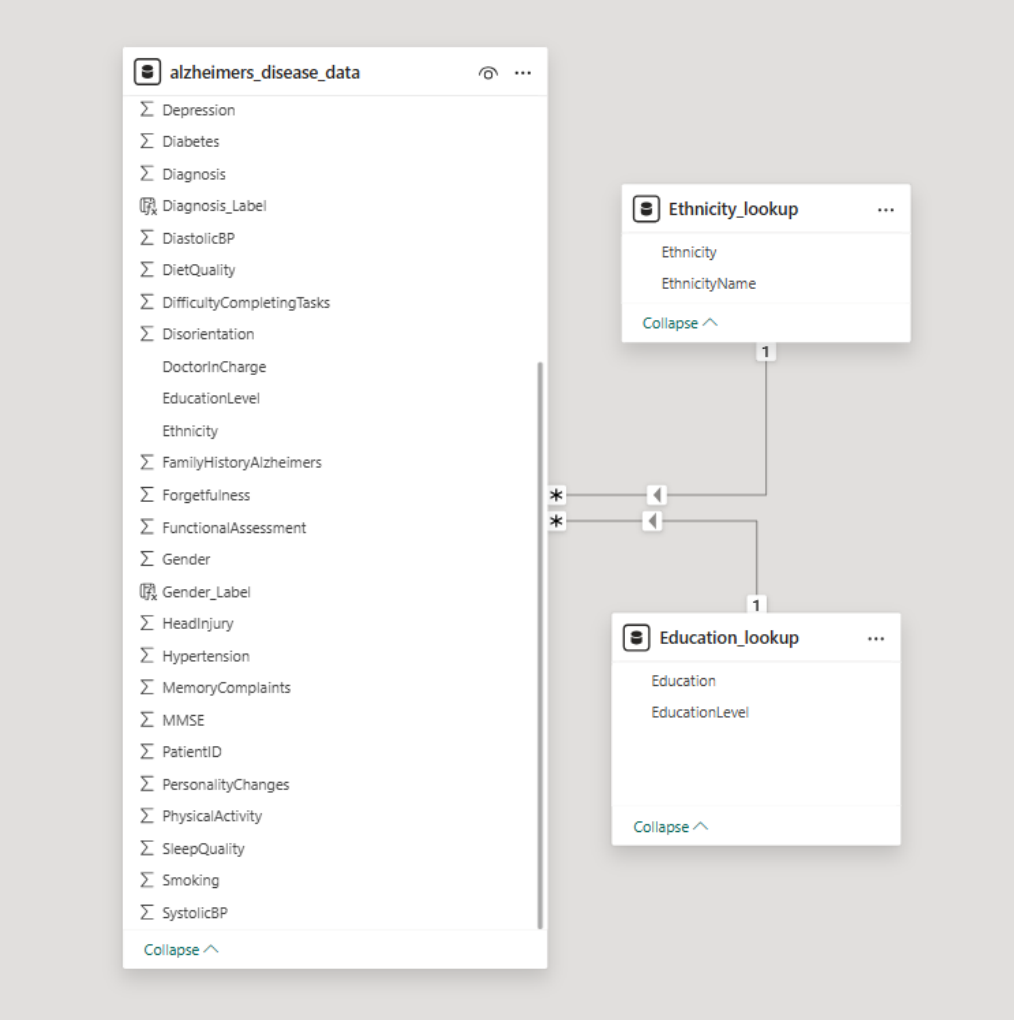
- Prevalence Rate: Percentage of diagnosed cases in the dataset.
- Age Distribution: Comparison of diagnosis rates across different age groups.
- Ethnicity-Based Diagnosis Rate: Percentage of Alzheimer’s cases within each ethnic group.
- Impact of Family History: Percentage of diagnosed individuals with a family history of the disease.
- Lifestyle Influence Metrics: Relationship between BMI, smoking, and alcohol consumption with Alzheimer’s diagnosis.

Data Model

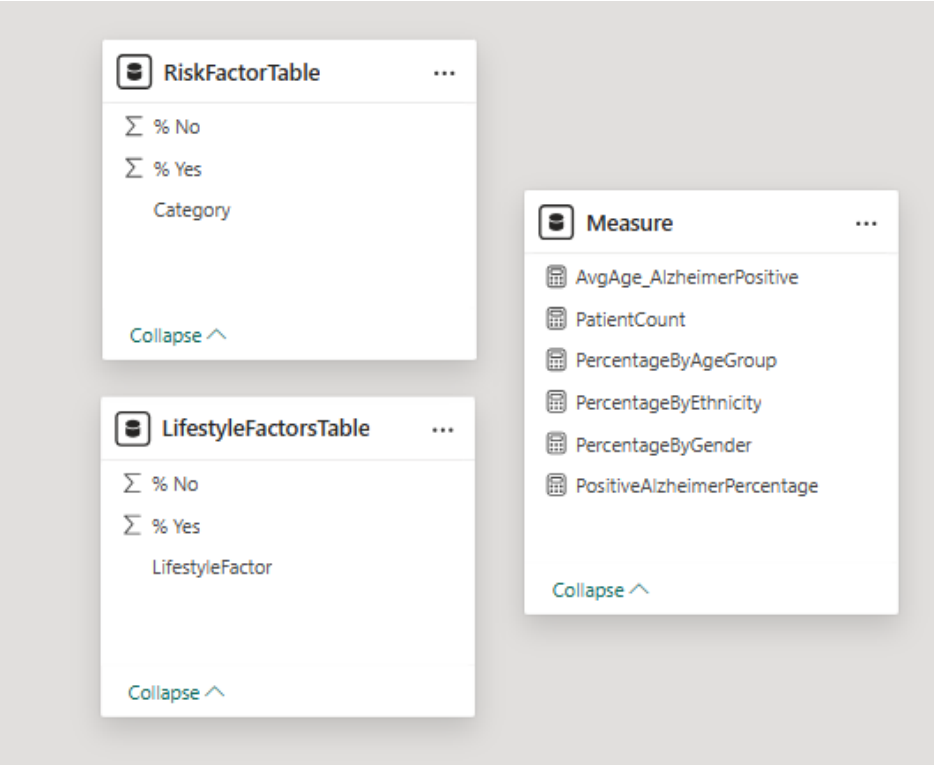
I structured my analysis using a CSV dataset containing patient records. The dataset primarily consisted of:

- One main table containing patient details and medical data.

- Two lookup tables for reference data.



- Measures and calculated columns to derive insights and comparisons.



Data Analysis in Power BI

To analyze the data and generate visual insights, I used Power BI for data modeling and visualization. The analytical techniques applied can be used to analyze any population that contains the same types of information as the columns in this dataset, making it a versatile approach for studying Alzheimer's disease risk factors across different datasets.

To analyze the data and generate visual insights, I used Power BI for data modeling and visualization. The key steps in my analysis included:

1. **Data Cleaning and Preparation:**

- Removed missing or inconsistent data entries.

- Standardized column names and data formats.

2. **Data Transformation:**

- Created calculated columns for better insights, such as age group classifications and risk factor categorizations.

- Established relationships between tables for a structured analysis.

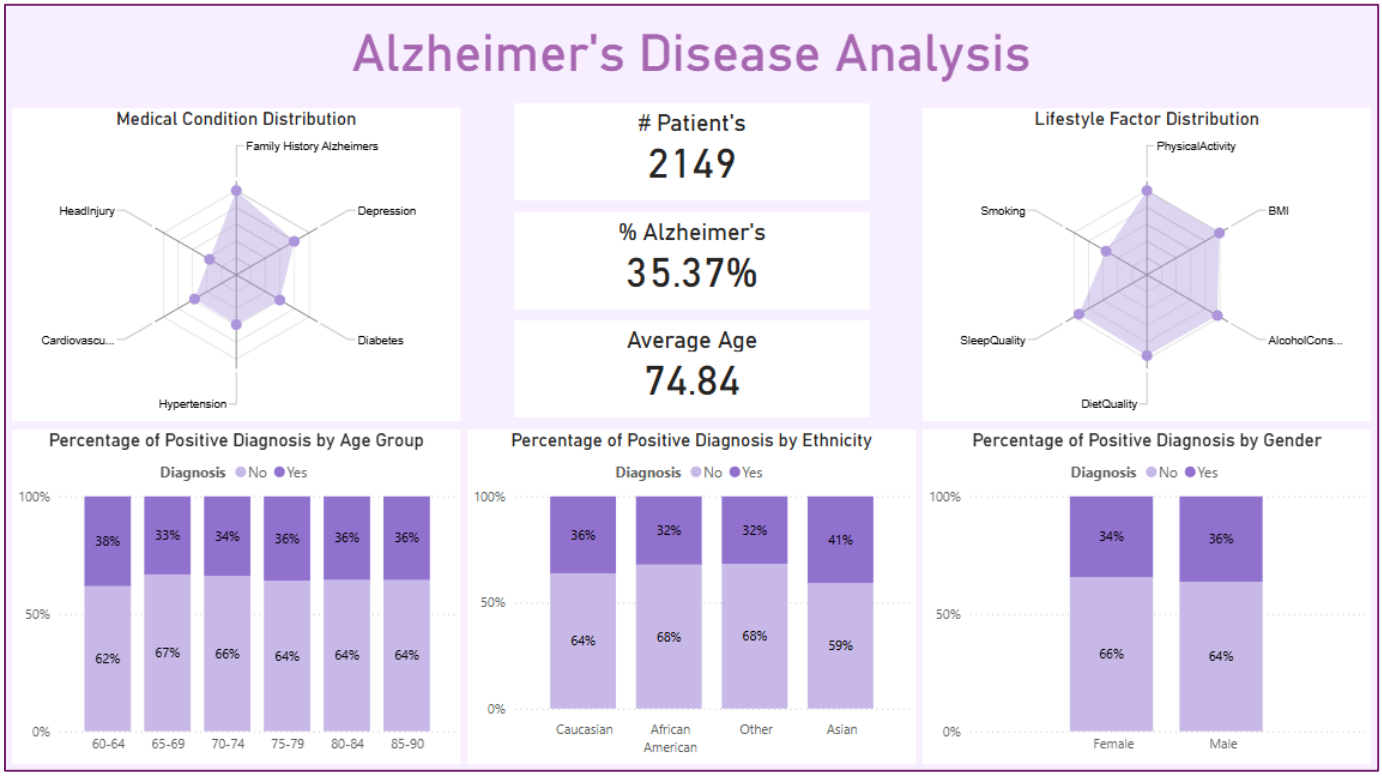
3. Visualizations and Dashboards:

- Created bar charts to display diagnosis rates by age and ethnicity.

- Used radar charts to show the impact of multiple risk factors on Alzheimer's diagnosis.

- Designed a summary dashboard to highlight key findings in an interactive format.

PowerBI Dashboard



Results

My analysis yielded several key findings:

1. Prevalence Rate: 35.37% of the dataset population was diagnosed with Alzheimer’s, highlighting the significant impact of the disease.

2. Ethnicity and Diagnosis: Asian patients had the highest diagnosis rate (41%), whereas African Americans and "Other" ethnic groups had the lowest (32%).

3. Gender Differences: Males (36%) had a slightly higher diagnosis rate than females (34%), suggesting potential biological or lifestyle influences.

4. Family History Impact: Patients with a family history of Alzheimer’s showed a strong correlation with higher diagnosis rates, indicating genetic predisposition as a major factor.

5. Lifestyle and Medical Factors: High BMI, smoking, and cardiovascular diseases were found to increase the likelihood of an Alzheimer’s diagnosis.

Conclusion

This study highlights several critical factors influencing Alzheimer's disease. However, it is crucial to emphasize that these findings are based on the dataset of studied patients and do not represent a general conclusion about Alzheimer's disease. The results should be interpreted as insights into trends within this specific population rather than definitive medical conclusions.