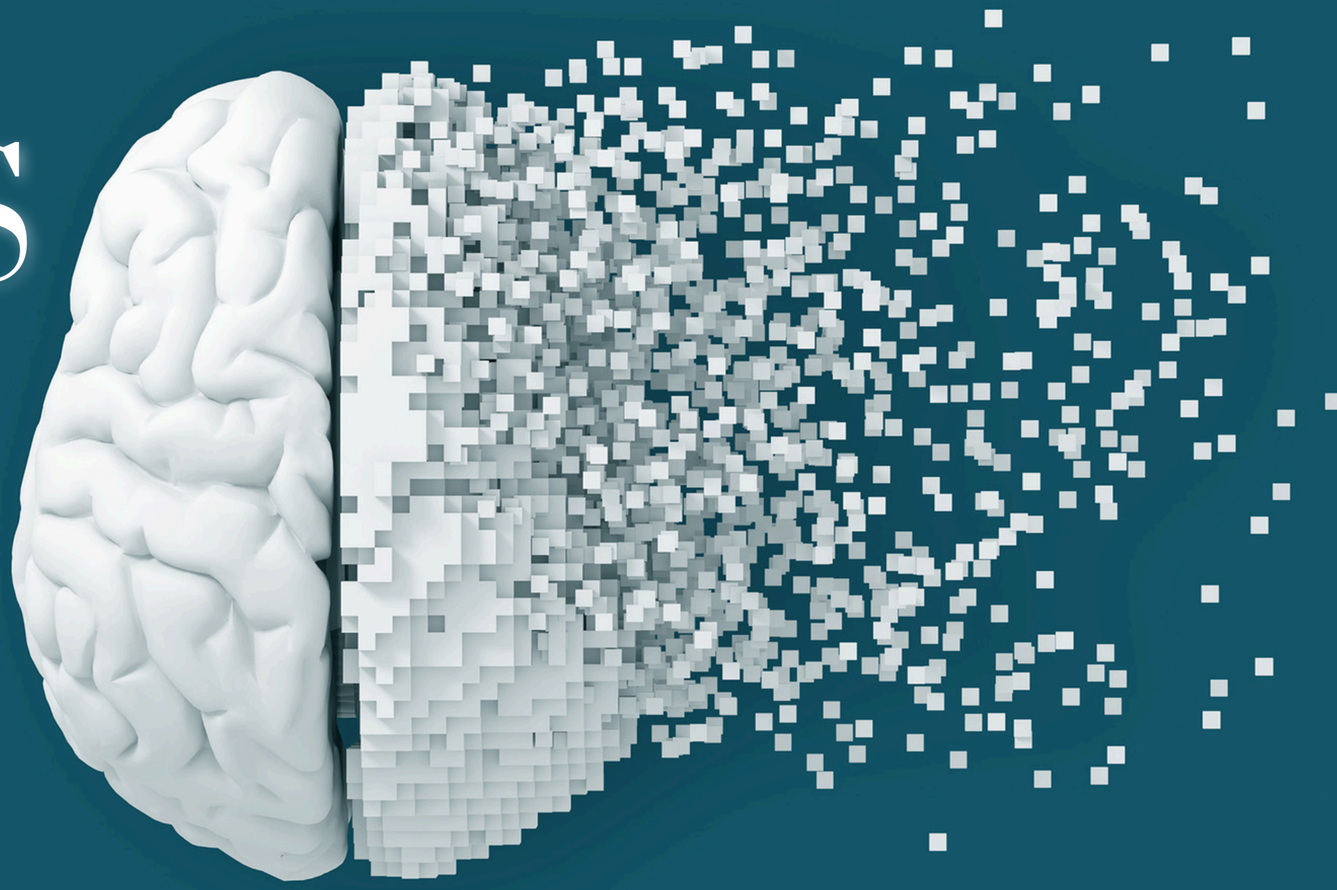
ALZHEIMER'S DISEASE

Providing Insights Through Data Analysis



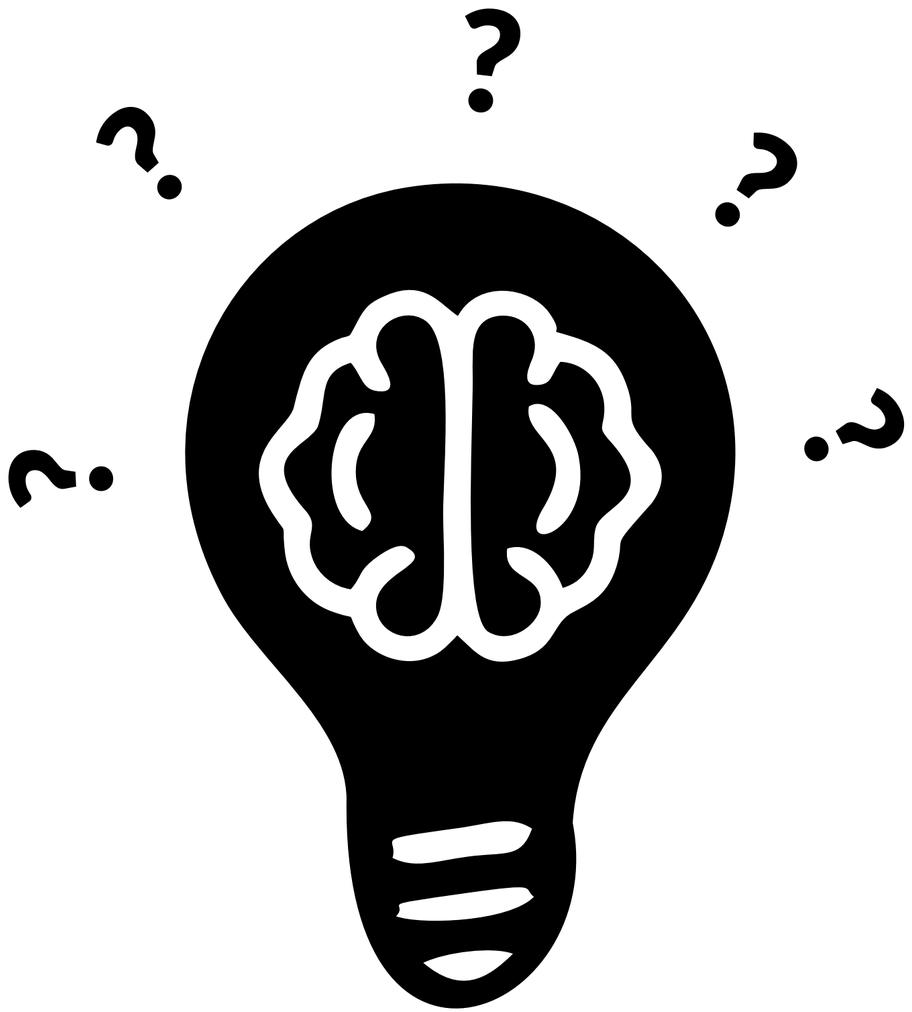


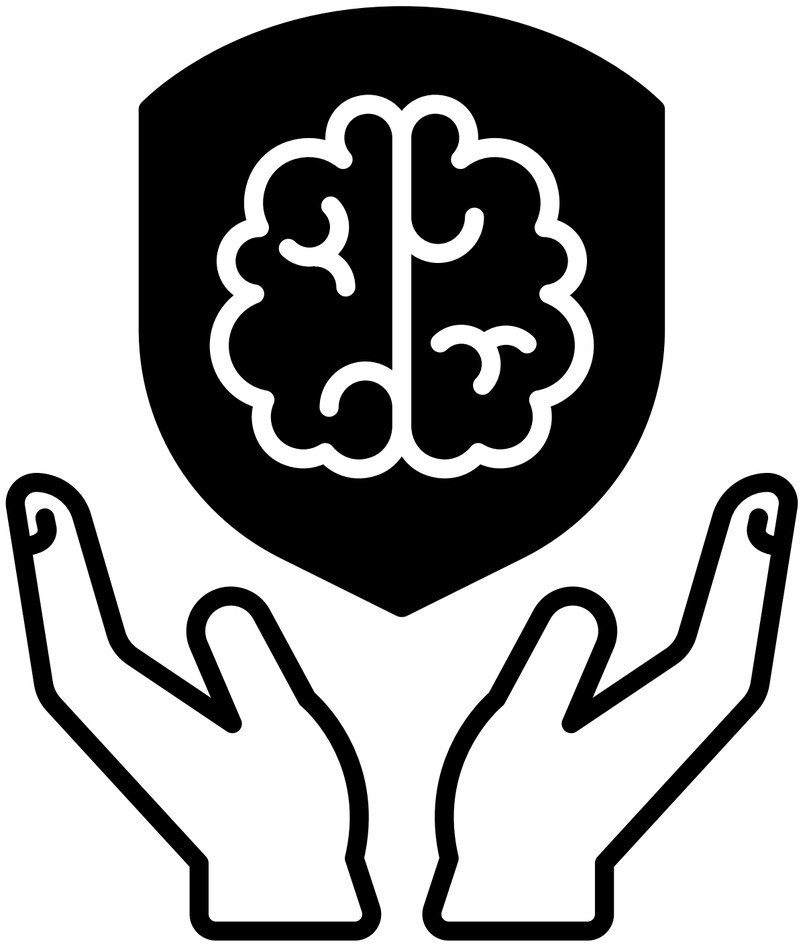
TABLE OF CONTENT

INTRODUCTION

Alzheimer's disease, now recognized as the most common cause of dementia, is a progressive neurodegenerative disorder primarily affecting the elderly, and it has a significant history marked by key discoveries and advancements.
It was first reported by Alois Alzheimer, a German psychiatrist in 1901 and named by Emil Kraepelin in 1910. Modern research, genetic discoveries and pharmaceutical research has been conducted since then.

ABOUT

This documentation provides a detailed analysis of the prevalence of Alzheimer's disease using a comprehensive dataset. The analysis includes geographic, temporal, data source, data type, confidence intervals, stratification, topic, class, and general data overview. The goal is to provide insights into the distribution and trends of Alzheimer's disease prevalence across various dimensions.



PROBLEM STATEMENT

Alzheimer's disease is a progressive neurological disorder that affects millions of people worldwide. Understanding its prevalence across different geographies, demographics, and time periods is crucial for public health planning and resource allocation. This documentation aims to analyze the available data to uncover patterns and trends in Alzheimer's disease prevalence, providing valuable insights for researchers, policymakers, and healthcare professionals.

DATA CLEANING

QLQuery2.sql - GL...r (GLORY\User (54))* | SQLQuery1.sql - GL...r (GLORY\User (53))*

```
--Identify NULL values--
Select *
FROM [dbo].[alzheimer]
WHERE [Low_Confidence_Limit] IS NULL

--Replace NULL values with mean--
UPDATE [dbo].[alzheimer]
SET High_Confidence_Limit = (
    SELECT ROUND(AVG(High_Confidence_Limit),1)
    FROM [dbo].[alzheimer]
    WHERE High_Confidence_Limit IS NOT NULL)
WHERE High_Confidence_Limit IS NULL;

UPDATE [dbo].[alzheimer]
SET[Low_Confidence_Limit] =(
SELECT ROUND(AVG([Low_Confidence_Limit]),1)
FROM [dbo].[alzheimer]
WHERE [Low_Confidence_Limit] IS NOT NULL)
WHERE [Low_Confidence_Limit] IS NULL

UPDATE [dbo].[alzheimer]
SET [Data_Value]=(
SELECT ROUND( AVG([Data_Value]),1)
FROM [dbo].[alzheimer]
WHERE [Data_Value]IS NOT NULL)
WHERE [Data_Value] IS NULL

-- Replace NULL values with ‘unknown’
UPDATE [dbo].[alzheimer]
SET [Stratification2] =  'unknown'
WHERE [Stratification2]IS NULL
```

DATA EXPLORATION

SQLQuery2.sql - GL...r (GLORY\User (54))* | SQLQuery1.sql - GL...r (GLORY\User (53))*

```
WHERE [Stratification2]IS NULL

--General Analysis
--How many patients have an average of 20 or more hours of care per week
SELECT COUNT([Question])
FROM [dbo].[alzheimer]
WHERE [Question] ='Average of 20 or more hours of care per week provided to a friend or family member'

--How many patients are experiencing a subjective cognitive decline or memory loss in the past 1 year
SELECT COUNT([Question])
FROM [dbo].[alzheimer]
WHERE [Question] ='Percentage of older adults who reported subjective cognitive decline or memory loss that is happening more often or is getting wo


--Geographic Analysis
--How does the prevalence of Alzheimer's disease vary between different locations ?
SELECT [LocationDesc], COUNT([RowId]) AS location_count
FROM [dbo].[alzheimer]
GROUP BY [LocationDesc]
ORDER BY location_count DESC

--What are the distinct locations  and their corresponding abbreviations ?
SELECT DISTINCT [LocationDesc],[LocationAbbr]
FROM [dbo].[alzheimer]

--How does the data value distribution vary by geolocation ?
SELECT  [Geolocation],
ROUND(AVG([Data_Value]),1) AS avg_data_value,
ROUND(MIN([Data_Value]),1)AS min_data_value,
ROUND(MAX([Data_Value]),1) AS max_data_value,
ROUND(STDEV([Data_Value]),1) AS standard_data_value
FROM [dbo].[alzheimer]
GROUP BY [Geolocation]
```

DATA EXPLORATION

SQLQuery2.sql - GL...r (GLORY\User (54))* | SQLQuery1.sql - GL...r (GLORY\User (53))*

```sql
--How does the confidence interval vary by location ?
SELECT [LocationDesc],
ROUND(SUM(CAST([Low_Confidence_Limit] AS FLOAT)),1) AS low_limit,
ROUND(SUM(CAST([High_Confidence_Limit] AS FLOAT)),1) AS high_limit
FROM [dbo].[alzheimer]
GROUP BY [LocationDesc]

--How does the data value (Data_Value) compare between different geolocations (Geolocation)?
SELECT [Geolocation],
ROUND(SUM([Data_Value]),1) AS data_count
FROM [dbo].[alzheimer]
GROUP BY [Geolocation]

--What is the distribution of high confidence limits (High_Confidence_Limit) across different locations (LocationDesc)?
SELECT [LocationDesc],
ROUND(SUM([High_Confidence_Limit]), 1) AS high_limit
FROM [dbo].[alzheimer]
GROUP BY [LocationDesc]
ORDER BY high_limit DESC


--Temporal Analysis
--Temporal Analysis
--How does the data value (Data_Value) change over time between YearStart and YearEnd?
SELECT [YearStart],[YearEnd],
ROUND(AVG([Data_Value]),1) AS avg_value
FROM[dbo].[alzheimer]
GROUP BY [YearStart], [YearEnd]
ORDER BY [YearStart]
--How does the data value (Data_Value) change across different years for the same location (LocationDesc)?
SELECT [LocationDesc],[YearStart],[YearEnd],
ROUND(AVG([Data_Value]),1) AS avg_value
```

DATA EXPLORATION

SQLQuery2.sql - GL...r (GLORY\User (54))* | SQLQuery1.sql - GL...r (GLORY\User (53))*

```sql
--How does the confidence interval vary by location ?
SELECT [LocationDesc],
ROUND(SUM(CAST([Low_Confidence_Limit] AS FLOAT)),1) AS low_limit,
ROUND(SUM(CAST([High_Confidence_Limit] AS FLOAT)),1) AS high_limit
FROM [dbo].[alzheimer]
GROUP BY [LocationDesc]

--How does the data value (Data_Value) compare between different geolocations (Geolocation)?
SELECT [Geolocation],
ROUND(SUM([Data_Value]),1) AS data_count
FROM [dbo].[alzheimer]
GROUP BY [Geolocation]

--What is the distribution of high confidence limits (High_Confidence_Limit) across different locations (LocationDesc)?
SELECT [LocationDesc],
ROUND(SUM([High_Confidence_Limit]), 1) AS high_limit
FROM [dbo].[alzheimer]
GROUP BY [LocationDesc]
ORDER BY high_limit DESC


--Temporal Analysis
--How does the data value (Data_Value) change over time between YearStart and YearEnd?
SELECT [YearStart],[YearEnd],
ROUND(AVG([Data_Value]),1) AS avg_value
FROM[dbo].[alzheimer]
GROUP BY [YearStart], [YearEnd]
ORDER BY [YearStart]
--How does the data value (Data_Value) change across different years for the same location (LocationDesc)?
SELECT [LocationDesc],[YearStart],[YearEnd],
ROUND(AVG([Data_Value]),1) AS avg_value
FROM [dbo].[alzheimer]
```

100 %

DATA EXPLORATION

SQLQuery2.sql - GL...r (GLORY\User (54))* | SQLQuery1.sql - GL...r (GLORY\User (53))*

```
GROUP BY [Stratification2]

--Topic and Class Analysis
--What are the distinct classes and topics covered in the dataset?
--Class
SELECT DISTINCT [Class]
FROM [dbo].[alzheimer]
--Topics
SELECT DISTINCT [Topic]
FROM [dbo].[alzheimer]

--What is the average data value for each class?
SELECT [Class],
ROUND(AVG([Data_Value]),1) AS avg_data_value
FROM [dbo].[alzheimer]
GROUP BY [Class]
ORDER BY avg_data_value

--What is the range of data values (Data_Value) for each topic (Topic)?
SELECT [Topic],
ROUND(AVG([Data_Value]),1) AS avg_data_value
FROM [dbo].[alzheimer]
GROUP BY [Topic]
ORDER BY avg_data_value

--What is the number of occurence between different questions (QuestionID)?
SELECT [Question], COUNT([Data_Value]) AS occurence_count
FROM [dbo].[alzheimer]
GROUP BY [Question]
ORDER BY occurence_count
```

DATA EXPLORATION

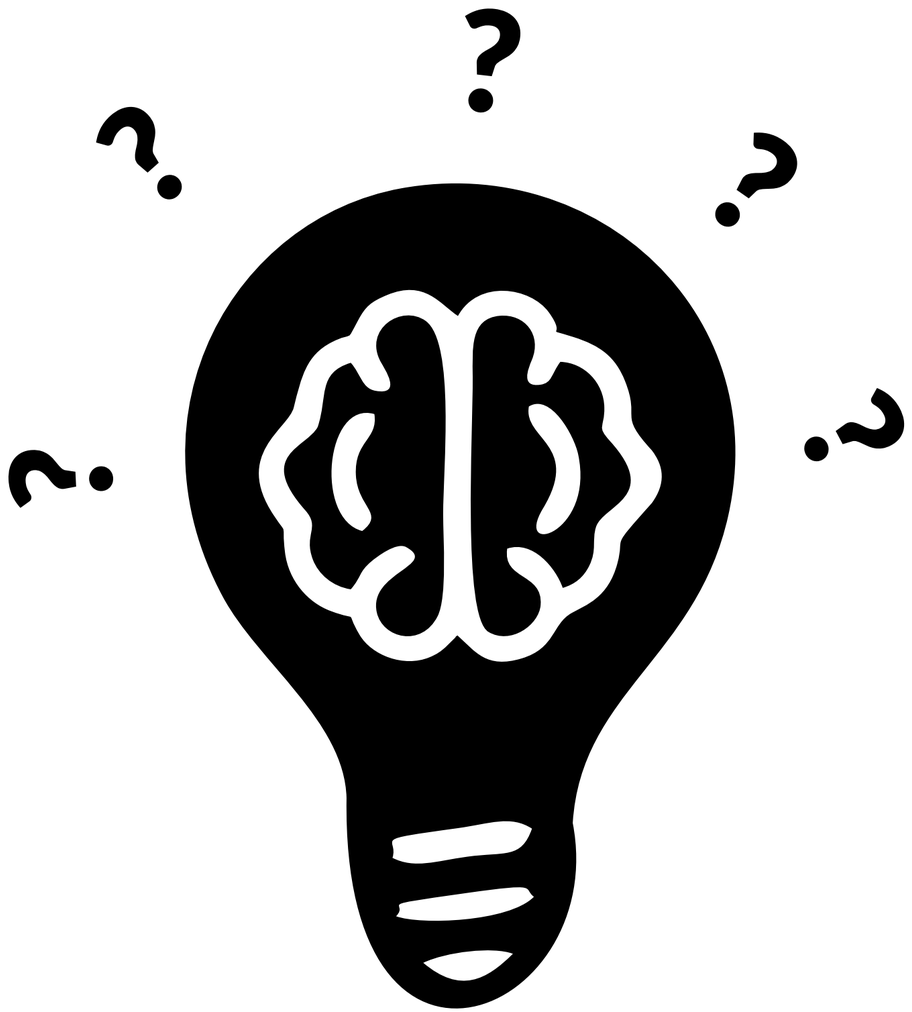
```sql
--Stratification Analysis
--How does the data value vary between different Age groups?
SELECT [Stratification1],
ROUND(AVG([Data_Value]),1) AS avg_data_value
FROM [dbo].[alzheimer]
GROUP BY [Stratification1]

--What is the frequency of different stratification categories (StratificationCategory1, StratificationCategory2)?
SELECT [Stratification1],[Stratification2],
ROUND(AVG([Data_Value]),1) AS avg_data_value
FROM [dbo].[alzheimer]
GROUP BY [Stratification1], [Stratification2]

--How does the data value vary between different races?
SELECT [Stratification2],
ROUND(AVG([Data_Value]),1) AS avg_data_value
FROM [dbo].[alzheimer]
GROUP BY [Stratification2]
ORDER BY avg_data_value DESC

--How does the data value vary between genders?
SELECT [Stratification2],
ROUND(SUM([Data_Value]),1) AS avg_data_value
FROM [dbo].[alzheimer]
WHERE [Stratification2] IN ('Female', 'Male')
GROUP BY [Stratification2]
```

INSIGHTS

1. Alzheimer is most common between the age group 65 years and older
2. Alzheimer is most common amongst the Hispanic race
3. Alzheimer is more prevalent in United States, DC & Territories
4. There is a correlation between the confidence interval and prevalence rate i.e the higher the rate of Alzheimer disease in a certain location, the higher the prevalence rate
5. The female gender have a higher rate of Alzheimer disease compared to the male gender
6. There are a total of 5014 patients who have an average of 20 or more hours of care per week provided to a friend or family member
7. There are a total of 5854 patients experiencing a subjective cognitive decline or memory loss in the past 1 year
8. There are a total of 6148 who has reported high blood sugar but don't have diabetes within the past 3 years
9. There are a total of 11092 percentage of Alzheimer patients with a lifetime diagnosis of depression
10. There are a total of 11092 percentage of Alzheimer patients who have smoked at least 100 cigarettes in their entire life and still smoke every day or some days

RECOMMENDATIONS

1. Targeted Awareness and Screening: Focus on older adults, Hispanics, and high-prevalence areas for early detection and intervention.
2. Support for Caregivers: Enhance resources and support systems for informal caregivers.
3. Gender-Specific Strategies: Develop healthcare strategies and research initiatives tailored to the needs of women.
4. Mental Health Integration: Incorporate mental health services into Alzheimer's care to address the high incidence of depression.
5. Lifestyle Interventions: Promote healthy lifestyle changes to manage risk factors such as high blood sugar and smoking.
6. Improved Data Collection: Enhance precision in data reporting to better understand regional variability and target public health strategies effectively



CONCLUSION

Alzheimer's disease predominantly affects individuals aged 65 and older, with a higher prevalence among the Hispanic population and in the United States, DC, and Territories. Females are more affected than males. Significant caregiving burdens exist, with many patients receiving extensive care. There is a notable presence of cognitive decline, high blood sugar (without diabetes), smoking history, and depression among patients.

THANK YOU!