University of Science and Technology
School of computational Science and artificial intelligence
DSAI 202: Data governance

Name: Adham Ashraf Abdelhaq Mohamed

ID: 202200953

Sec: 1

Dataset name: Alzheimer's Disease and Healthy Aging Data

Dataset link: Click Here

Dataset description

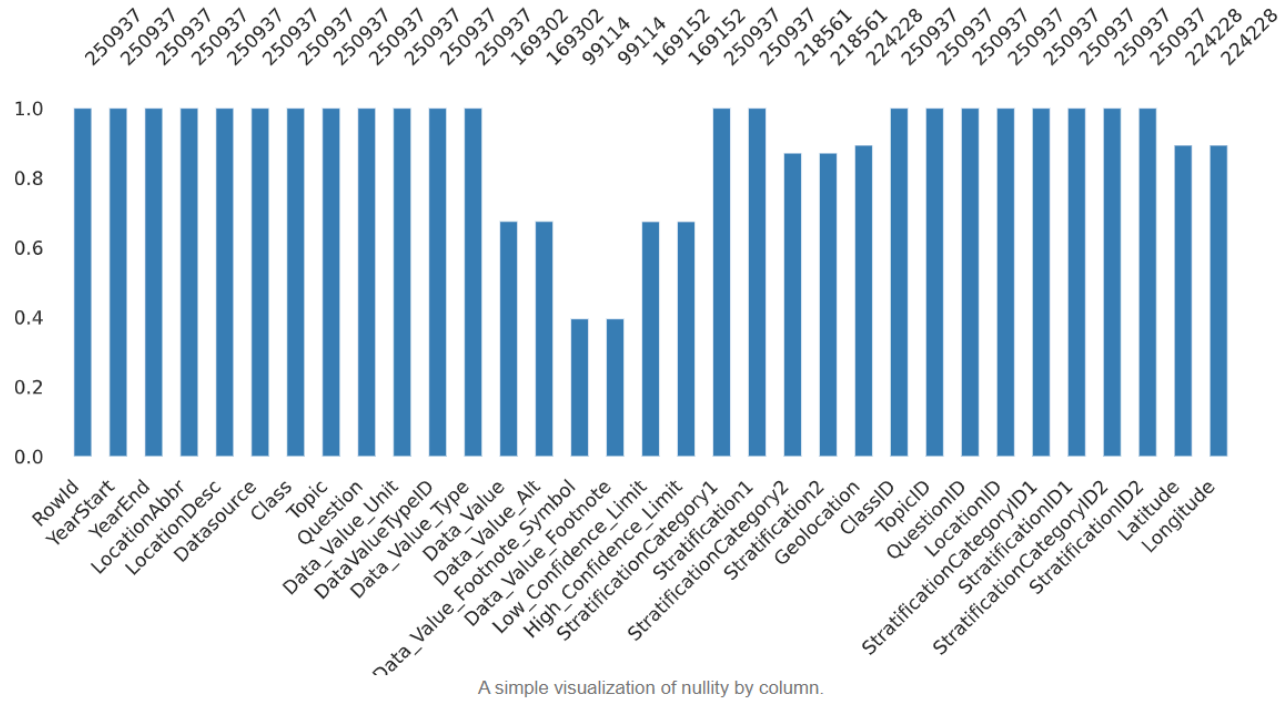
1- One or two paragraphs describing your dataset

This "Alzheimer's Disease and Healthy Ageing Data" collection is a goldmine for anyone who wants to understand Alzheimer's and what keeps people healthy as they age. It's like having a giant magnifying glass to see exactly how widespread Alzheimer's is and what things might raise your risk.

The data covers a bunch of years and different places, giving you a broad picture of the disease. It even digs into details like age, gender, and maybe even social background, so you can see how Alzheimer's affects different people in different ways. Plus, the numbers come with uncertainty ranges, which makes them even more reliable for analysis.

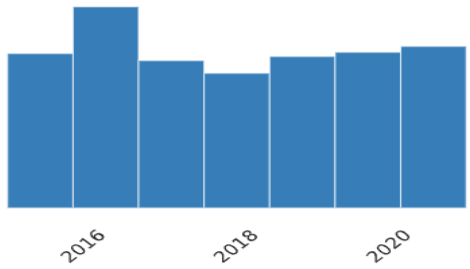
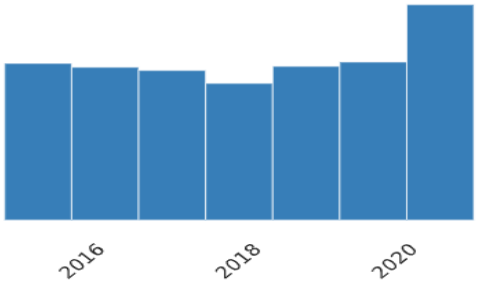
And that's not all! This data has location information too. So researchers can track where Alzheimer's seems to be more common and target those areas with better treatments. It's like having a map to fight back against the disease!

Missing Values Count

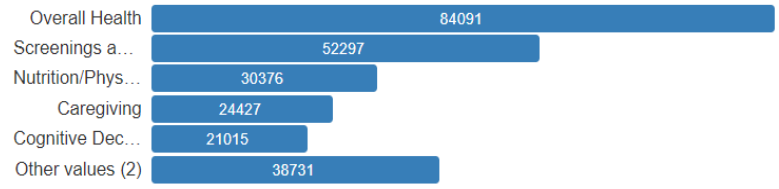


A simple visualization of nullity by column.

2- Create a table for ALL the columns in your dataset similar to the following table:

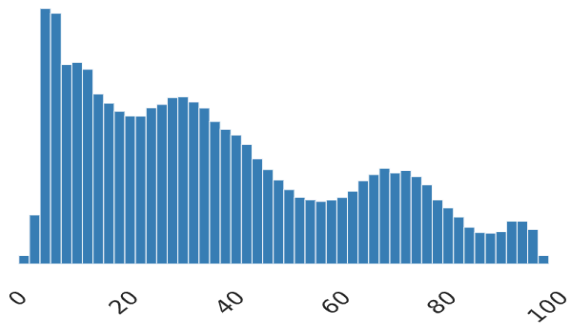
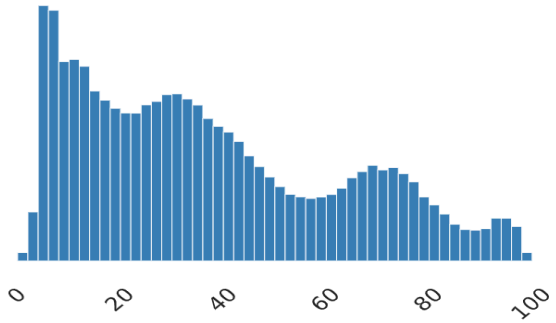
Column name	Column profiling	Notes
YearStart The starting year of the data collection period.	Distinct = 7 Distinct (%) < 0.1% Missing = 0 Missing (%) 0.0% Infinite= 0 Infinite (%) 0.0% Mean = 2017 (Months at last) Minimum = 2015 Maximum = 2021 Zeros = 0 Zeros (%) 0.0% Negative = 0 Negative (%) 0.0%	 **Column Dependencies:** • YearEnd
YearEnd The end year of the data collection period.	Distinct = 7 Distinct (%) < 0.1% Missing = 0 Missing (%) 0.0% Infinite= 0 Infinite (%) 0.0% Mean = 2018 (Months at the beginning) Minimum = 2015 Maximum = 2021 Zeros = 0 Zeros (%) 0.0% Negative = 0 Negative (%) 0.0%	 **Column Dependencies :** • YearStart

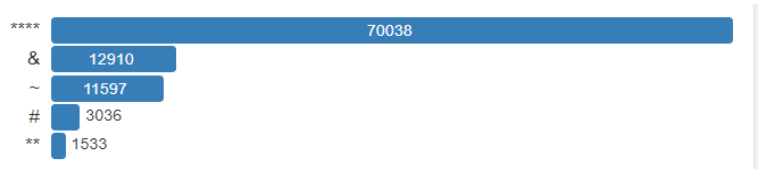
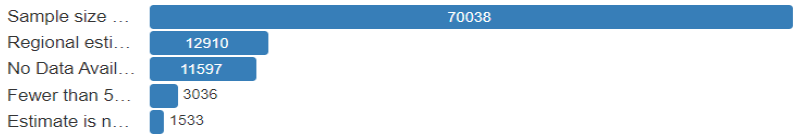
LocationAbbr The location where the data was collected.	Distinct= 59 Distinct (%) < 0.1% Missing= 0 Missing (%) 0.0% **Length** Max length =4 Median length=2 Mean length=2.106413 Min length=2 **Unique** Unique= 0 Unique (%) 0.0%	
LocationDesc Description of the location	Distinct= 59 Distinct (%) < 0.1% Missing= 0 Missing (%) 0.0% **Length** Max length =31 Median length=13 Mean length=8.885202 Min length=4 **Unique** Unique= 0 Unique (%) 0.0%	
Datasource The source of the data	Distinct= 1 Distinct (%) < 0.1% Missing= 0 Missing (%) 0.0% **Length** Max length =5 Median length=5 Mean length=5 Min length=5 **Unique** Unique= 0 Unique (%) 0.0%	

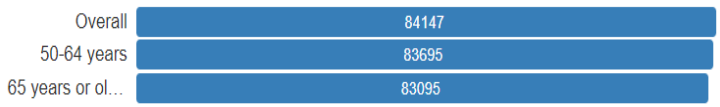
Class Classification of the Data	Distinct= 7 Distinct (%) < 0.1% Missing= 0 Missing (%) 0.0% **Length** Max length =35 Median length=23 Mean length=18.8969 Min length=10 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • ClassID • Question • QuestionID • Topic • TopicID
Topic The topic or subject area of the data.	Distinct= 7 Distinct (%) < 0.1% Missing= 0 Missing (%) 0.0% **Length** Max length =35 Median length=23 Mean length=18.8969 Min length=10 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • Class • ClassID • DataValueTypeID • Data_Value • Data_Value_Alt • Data_Value_Type • Data_Value_Unit • QuestionID • Topic • TopicID

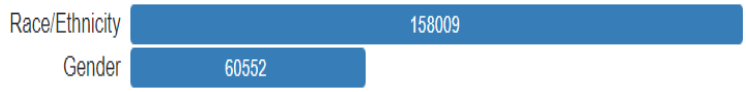
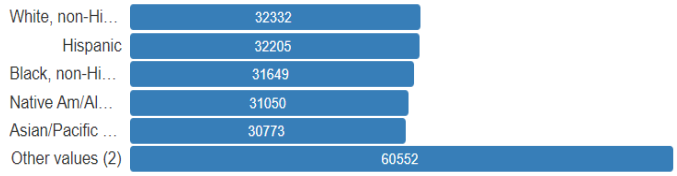
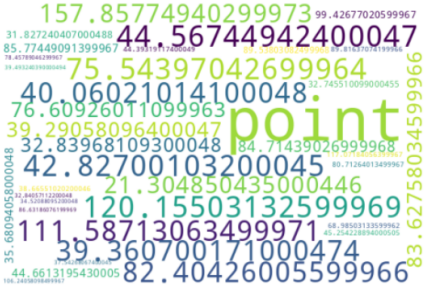
Question The specific question being investigated	Distinct= 7 Distinct (%) < 0.1% Missing= 0 Missing (%) 0.0% **Length** Max length =35 Median length=23 Mean length=18.8969 Min length=10 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • Class • ClassID • DataValueTypeID • Data_Value • Data_Value_Alt • Data_Value_Type • Data_Value_Unit • QuestionID • Topic • TopicID
Data_Value_ Unit Unit of measurement for the data value	Distinct= 2 Distinct (%) < 0.1% Missing= 0 Missing (%) 0.0% **Length** Max length =6 Median length=1 Mean length=1.384239 Min length=1 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • DataValueTypeID • Data_Value • Data_Value_Alt • Data_Value_Type • Question • QuestionID • Topic • TopicID

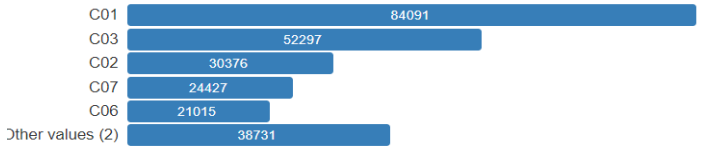
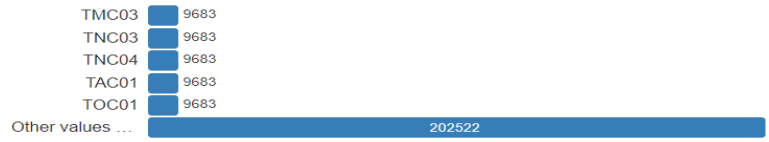
DataValueTypeID Identifier for the type of data value	Distinct= 2 Distinct (%) < 0.1% Missing= 0 Missing (%) 0.0% **Length** Max length =5 Median length=5 Mean length=4.923152 Min length=4 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • Data_Value_Type • Data_Value_Unit • Question • QuestionID • Topic • TopicID
Data_Value_Type Type of data value.	Distinct= 2 Distinct (%) < 0.1% Missing= 0 Missing (%) 0.0% **Length** Max length =10 Median length=10 Mean length=9.538912 Min length=4 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • DataValueTypeID • Data_Value • Data_Value_Alt • Data_Value_Unit • Question • QuestionID • Topic • TopicID

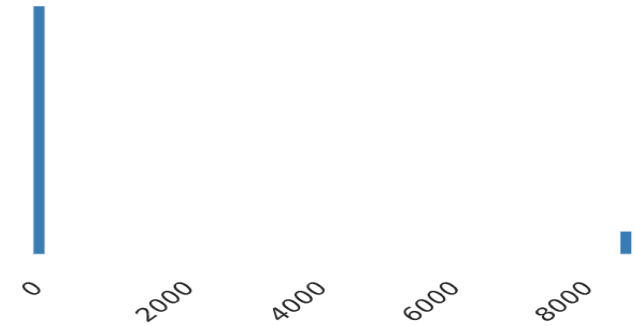
Data_Value The actual value of the data	Distinct = 1000 Distinct (%) < 0.6% Missing = 81635 Missing (%) 32.5% Infinite= 0 Infinite (%) 0.0% Mean = 37.328349 Minimum = 0 Maximum = 100 Zeros = 156 Zeros (%) 0.1% Negative = 0 Negative (%) 0.0%	 **Column Dependencies :** • Data_Value_Unit • Data_Value_Alt • Data_Value_Type • Question • QuestionID • Topic • TopicID
Data_Value_Alt Alternate representation of the data value	Distinct = 1000 Distinct (%) < 0.6% Missing = 81635 Missing (%) 32.5% Infinite= 0 Infinite (%) 0.0% Mean = 37.328349 Minimum = 0 Maximum = 100 Zeros = 156 Zeros (%) 0.1% Negative = 0 Negative (%) 0.0%	 **Column Dependencies :** • Data_Value_Unit • Data_Value • Data_Value_Type • Question • QuestionID • Topic • TopicID

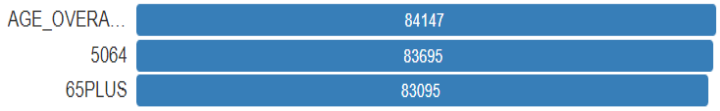
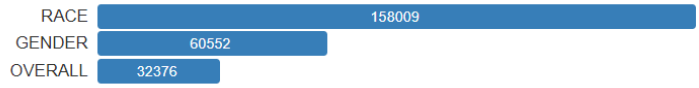
Data_Value_Footnote_Symbol Symbol indicating a footnote related to the data value.	Distinct=5 Distinct (%) < 0.1% Missing=151823 Missing (%) 60.5% **Length** Max length =4 Median length=4 Mean length=3.135389 Min length=1 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies:** ● Data_Value_Footnote
Data_Value_Footnote Additional Value for the data footnote	Distinct=5 Distinct (%) < 0.1% Missing=151823 Missing (%) 60.5% **Length** Max length =127 Median length=127 Mean length=102.3292 Min length=17 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies:** ● Data_Value_Footnote_Symbol
Low_Confidence_Limit The lower bound of the confidence interval for the data value	Missing=81785 Missing (%) 32.6%	**This Column Contains:** ● Missing Data ● Data types that are not supported or recognized by the data profiling tool ● Data quality issues, inconsistency

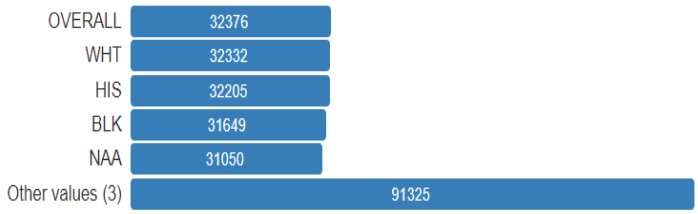
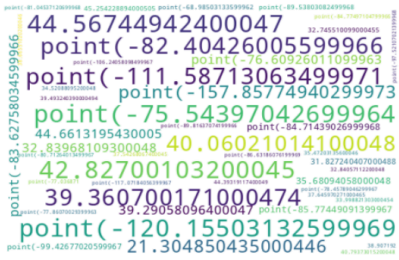
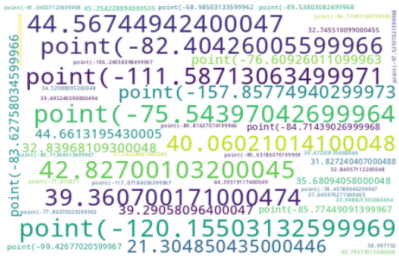
High_Confidence_Limit The upper bound of the confidence interval for the data value.	Missing=81785 Missing (%) 32.6%	**This Column Contains:** • Missing Data • Data types that are not supported or recognized by the data profiling tool • Data quality issues, inconsistency
Stratification Category1 First level of categorization for the data stratification.	Distinct=1 Distinct (%) < 0.1% Missing=0 Missing (%) 0% **Length** Max length =9 Median length=9 Mean length=9 Min length=9 **Unique** Unique= 0 Unique (%) 0.0%	
Stratification 1 The first level of stratification.	Distinct=3 Distinct (%) < 0.1% Missing=0 Missing (%) 0% **Length** Max length =17 Median length=11 Mean length=11.64550 Min length=7 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies:** • StratificationID

Stratification Category2 Second level of categorization for the data stratification.	Distinct=2 Distinct (%) < 0.1% Missing=32376 Missing (%) 12.9% **Length** Max length =14 Median length=14 Mean length=11.78361 Min length=6 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • Stratification2 • StratificationID2 • StratificationCategoryID2
Stratification 2 The second level of stratification.	Distinct=7 Distinct (%) < 0.1% Missing=32376 Missing (%) 12.9% **Length** Max length =24 Median length=19 Mean length=14.63319 Min length=4 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • StratificationCategory2 • StratificationID2 • StratificationCategoryID2
Geolocation Geographical coordinates of the location.	Distinct=54 Distinct (%) < 0.1% Missing=26709 Missing (%) 10.6% **Length** Max length =46 Median length=45 Mean length=43.69910 Min length=28 **Unique** Unique= 0 Unique (%) 0.0%	

ClassID Identifier for the data class.	Distinct= 7 Distinct (%) < 0.1% Missing= 0 Missing (%) 0% **Length** Max length =3 Median length=3 Mean length=3 Min length=3 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependecies :** • Class • Question • QuestionID • Topic • TopicID
TopicID Identifier for the data topic.	Distinct= 39 Distinct (%) < 0.1% Missing= 0 Missing (%) 0% **Length** Max length =5 Median length=5 Mean length=3 Min length=5 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • Class • ClassID • DataValueTypeID • Data_Value • Data_Value_Alt • Data_Value_Type • Data_Value_Unit • QuestionID • Topic • Question

QuestionID Identifier for the data question	Distinct= 39 Distinct (%) < 0.1% Missing= 0 Missing (%) 0% **Length** Max length =3 Median length=3 Mean length=3 Min length=3 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • Class • ClassID • DataValueTypeID • Data_Value • Data_Value_Alt • Data_Value_Type • Data_Value_Unit • Question • Topic • TopicID
LocationID Identifier for the location.	Distinct = 59 Distinct (%) < 0.1% Missing = 0 Missing (%) 0.0% Infinite= 0 Infinite (%) 0.0% Mean = 793.86644 Minimum = 1 Maximum = 9004 Zeros = 0 Zeros (%) 0.0% Negative = 0 Negative (%) 0.0%	

Stratification CategoryID1 Identifier for the first level of stratification category.	Distinct= 1 Distinct (%) < 0.1% Missing= 0 Missing (%) 0% **Length** Max length =3 Median length=3 Mean length=3 Min length=3 **Unique** Unique= 0 Unique (%) 0.0%	
Stratification ID1 Identifier for the first level of stratification.	Distinct= 3 Distinct (%) < 0.1% Missing= 0 Missing (%) 0% **Length** Max length =11 Median length=6 Mean length=7.009596 Min length=4 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • Stratification1
Stratification CategoryID2 Identifier for the second level of stratification category.	Distinct= 3 Distinct (%) < 0.1% Missing= 0 Missing (%) 0% **Length** Max length =7 Median length=3 Mean length=4.869668 Min length=3 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • StratificationCategory2 • Stratification2 • StratificationCategoryID2

Stratification ID2 Identifier for the second level of stratification.	Distinct= 8 Distinct (%) < 0% Missing= 0 Missing (%) 0% **Length** Max length =7 Median length=3 Mean length=3.998688 Min length=3 **Unique** Unique= 0 Unique (%) 0.0%	 **Column Dependencies :** • StratificationCategory2 • Stratification2 • StratificationCategoryID2
Latitude Latitude coordinate of the location.	Distinct= 54 Distinct (%) < 0.1% Missing= 26709 Missing (%) 10.6% **Length** Max length =45 Median length=44 Mean length=42.69910 Min length=27 **Unique** Unique= 0 Unique (%) 0.0%	
Longitude Longitude coordinate of the location	Distinct= 54 Distinct (%) < 0.1% Missing= 26709 Missing (%) 10.6% **Length** Max length =45 Median length=44 Mean length=42.69910 Min length=27 **Unique** Unique= 0 Unique (%) 0.0%