Autistic Spectrum Disorder (ASD) is a neurodevelopmental illness associated with large healthcare expenses, which can be significantly reduced with early detection. Unfortunately, ASD diagnosis wait times are extensive, and procedures are not cost effective.

We offer a new dataset for autism screening of toddlers that contains influential aspects that can be used for further analysis, particularly in defining autistic traits and enhancing ASD case classification. In this dataset, we record ten behavioral traits (Q-Chat-10) as well as other individual characteristics that have been shown in behavior science to be beneficial in distinguishing ASD cases from controls.

Data Type: Predictive and Descriptive: Nominal / categorical, binary and continuous

Task: Classification but can be used for clustering and association or feature assessment

Attribute Type: Categorical, continuous and binary

Area: Medical, health and social science

Missing values? No

Number of Instances (records in your data set): 1054

Number of Attributes (fields within each record): 18 including the class variable

Attribute Information: For Further information about the attributes/feature see doc file attached

The dataset contains the following variables:

Case_No: Case number or identifier for each individual.

A1-A10: Answers to 10 questions from the ASDTests mobile app. These questions are used to screen for autism in toddlers.

Age_Mons: Age of the toddler in months.

Qchat.10.Score: Score obtained on the Q-CHAT-10 questionnaire, which assesses the risk of autism spectrum disorder.

Sex: Gender of the toddler.

Ethnicity: Ethnicity or racial background of the toddler.

Jaundice: Indicates whether the toddler had jaundice at birth.

Family_mem_with_ASD: Indicates whether there are other family members with autism spectrum disorder.

Who.completed.the.test: Information about the person who completed the screening test.

Class.ASD.Traits.: The class label indicating whether the toddler exhibits traits of autism spectrum disorder (ASD).

Attributes:

A1-A10: Items within Q-Chat-10 in which questions possible answers : "Always, Usually, Sometimes, Rarely & Never" items' values are mapped to "1" or "0" in the dataset. For questions 1-9 (A1-A9) in Q-chat-10, if the response was Sometimes / Rarely / Never "1" is assigned to the question (A1-A9). However, for question 10 (A10), if the response was Always / Usually / Sometimes then "1" is assigned to that question. If the user obtained More than 3 Add points together for all ten questions. If your child scores more than 3 (Q-chat-10- score) then there is a potential ASD traits otherwise no ASD traits are observed.

These variables can be used for various descriptive and predictive analyses, such as classification, clustering, and regression, to estimate the predictive power of machine learning techniques in detecting autistic traits.

Certainly! To begin, consider a categorical variable, "Sex," and a numerical variable, "Age_Mons." We can design appropriate charts to display the dataset's distribution and linkages.

To begin, we'll assume you've previously imported the dataset and given it the name "df."

Examining the Categorical Variable "Sex": To depict the distribution of the "Sex" variable, we can make a bar plot.

2nd : Consider the variables "Age_Mons" (numerical) and "Qchat.10.Score" (numerical) from the dataset to run an analysis on multiple variables. A scatter plot and a correlation analysis can be used to investigate the relationship between these two variables. Here is a R example: