CSCE 5320 - Scientific Data Visualization

Autism Spectrum Disorder for Toddlers

**Contents**

Team Members… 3

[Introduction… 4](#_TOC_250014)

[Domain… 4](#_TOC_250013)

[Approach… 4](#_TOC_250012)

[Data abstraction… 4](#_TOC_250011)

Dataset Attributes… 6

Data Info… 7

Data Transformations… 8

Task abstractions… 8

Target 8

[Actions… 8](#_TOC_250010)

[Exploratory Data Analysis… 10](#_TOC_250009)

Implementation Tools… 10

Preliminary results for Analysis… 12

[Python Library plots… 12](#_TOC_250008)

[Overview plots 14](#_TOC_250007)

[Area plots 17](#_TOC_250006)

[Heat Map Representation… 19](#_TOC_250005)

Tableau Visualization… 20

[Visualization 1… 20](#_TOC_250004)

[Visualization 2… 21](#_TOC_250003)

[Project Management… 21](#_TOC_250002)

[Conclusion… 23](#_TOC_250001)

[References… 24](#_TOC_250000)

**Final Report**

To

**Dr. Haili Wang**

By

**Group 27**

**Team Members:**

## Sri Ram Relangi - 11556267 Priyanka Jammigumpula – 11619024 Vishnu Priya Eupuri -11593758

## Introduction

Data visualization is increasingly recognized as a valuable tool for analyzing complex issues across various domains. Within the healthcare industry, data visualization plays a crucial role in understanding behavioral patterns and addressing significant challenges with precision.

## Domain:

Our analysis focuses primarily on Autism Spectrum Disorder (ASD), a well-documented neurological condition that affects individuals at different stages of life. ASD poses challenges to social interaction and behavior, impacting an individual's ability to communicate effectively and engage with others. Despite ongoing research, ASD lacks definitive diagnostic benchmarks, making it essential to delve deeper into its behavioral aspects. By gaining insights into these behaviors, we can better support families and communities affected by this disorder.

## Approach:

Here is the detailed approach how we are going to analyze this particular behavior using the data visualization scenarios. Following are the plots we are going analyze as a base for this ASD.

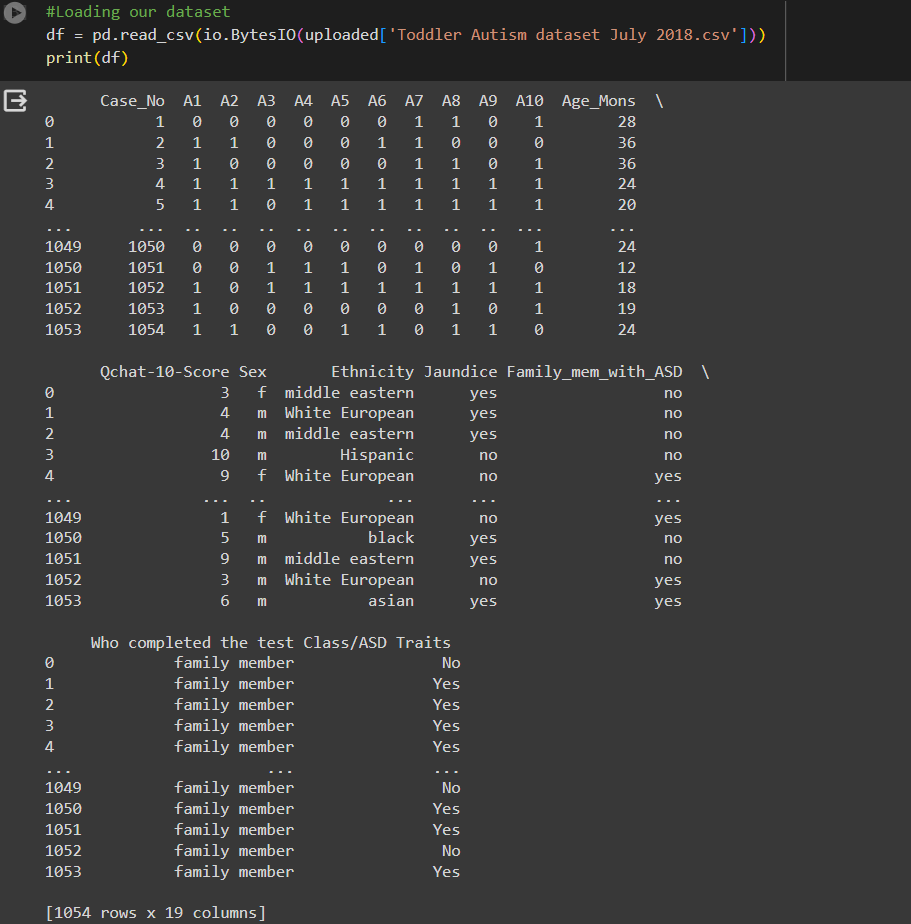
1. Bar graphs
2. Scatter plots
3. Heat maps
4. Geographic mapping
5. Pie charts

This data-driven approach to Autism Spectrum Disorder gives functionalities to enhance our collaborative understanding about the disorder and how we are going to treat and what are the insights that we are depicting and measures to treat them.

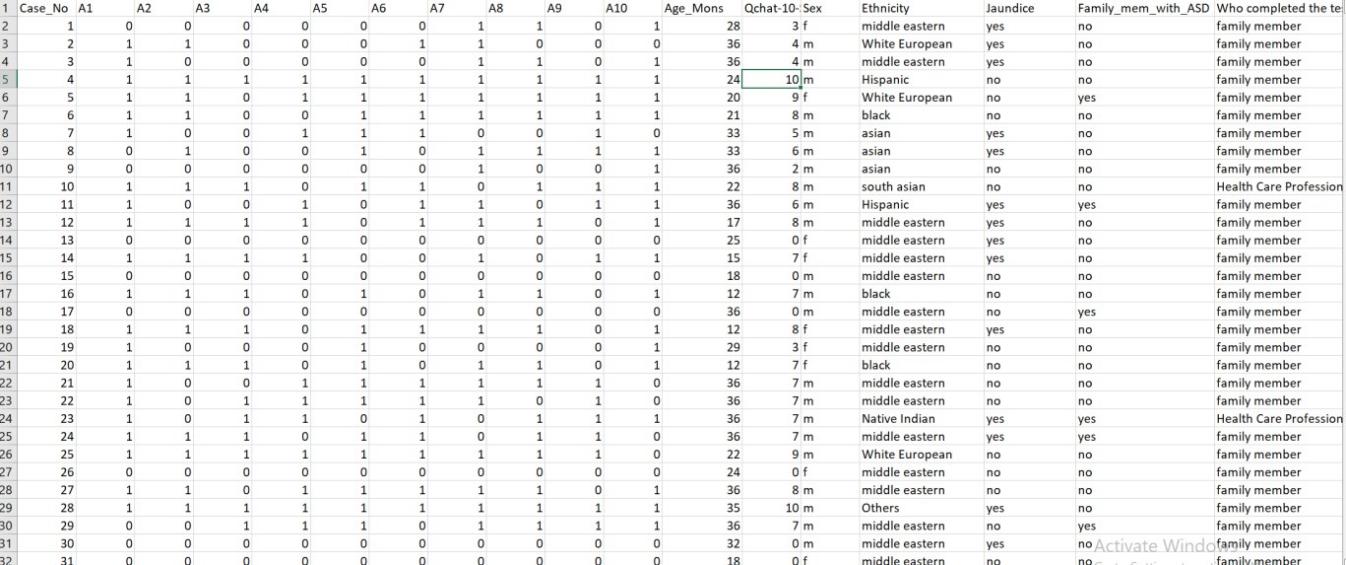
## Data Abstraction:

Dataset we obtained to do this autism analysis can be found in Kaggle <https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers>

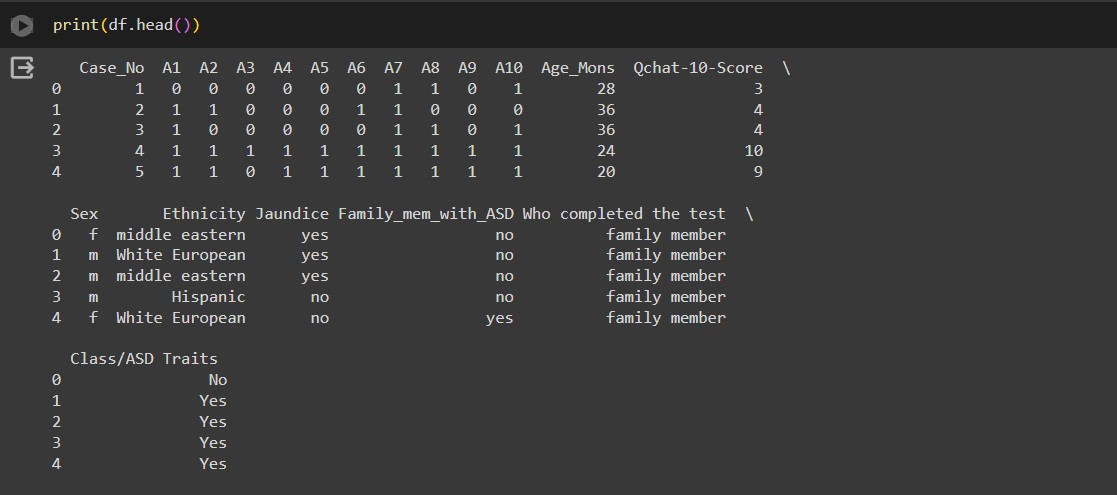
The data consists of 1054 rows which indicate they did the test 1054 toddlers with a series of questions and their responses have been recorded for the analysis. Including with these there are other attributes which are also considered like gender of the toddler, ethnicity of the toddler, whether the toddler is having jaundice or not, if it is inherited or not and who took the test for the analysis (family member/ medical



professional)



**Dataset Features**: Head of the dataset has been printed and we can see that there are multiple questions asked and the general information regarding the toddler also considered.



Based on the above responses they have been marked 0 or 1 in the respective columns how the toddler reacts to the specific questions.

## Data Attributes and Type:

Qualities or attributes which are unable to quantifiable are denoted by the nominal or categorical data type. A few examples are gender, and ethnicity. Nominal data exists in a couple of forms: binary and non-binary.

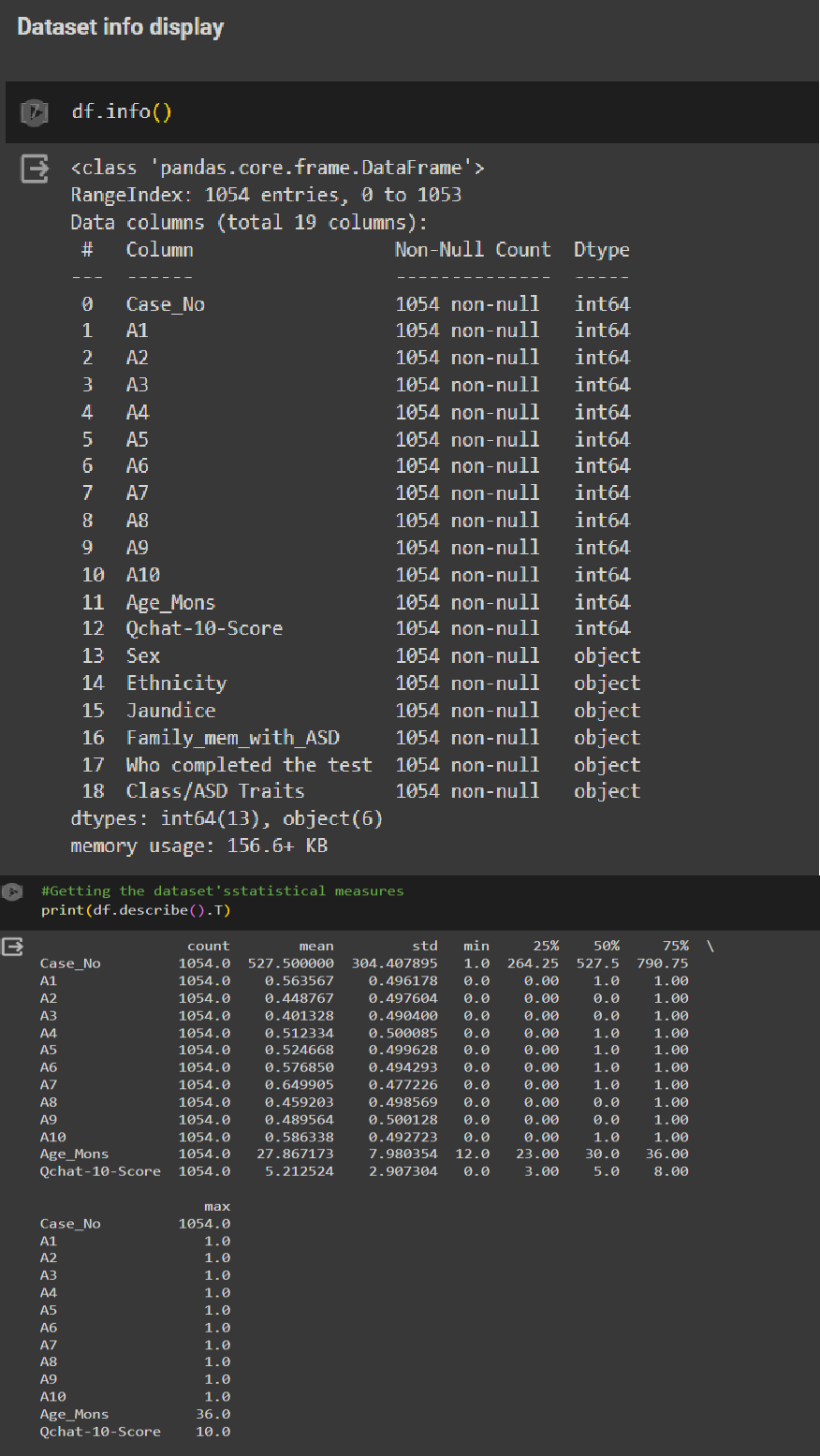
Merely just two variables in binary data, an example of nominal data: yes or no, true or false, or 1 or 0. Such as whether the consumer finished an order or whether a patient has an illness.

Quantitative in nature, continuous data comprises findings that can take on any amount within a particular range. Size, volume, and high levels of blood pressure are a few instances. Interval and ratio data are two additional groups into which continuous information can be classified.

Q-Chat-10 responds compose changes to the dataset; responses to questions A1 through A9 are transformed to a value of "1" depending on whether the response was "Sometimes," "Only occasionally," or "Never" in However irrespective of whether the answer was "Every time," "In general," or "At times," question A10 earns a "1."

The Q-Chat-10 score gets boosted by a point for each "1" that gets assigned to a question. Characteristics for ASD are present if the sum of the score is higher than three. On the other side, a total score of 3 or less indicates no evidence of any specific ASD characteristics. When a child exhibits characteristics that can suggest they have autism spectrum situation, this scoring method helps with early detection and caution.

## Dataset Info:



For the sake of visualization, a number of columns were replaced. There were several probably Yes or No fields in these columns. The No changed to a zero and the Yes became a 1. I have included a picture of the changes.

## Task Abstaction:

**Target:** Finding trends and patterns in the information that can guide the evaluation and therapy of autism behaviors among toddlers could be the ultimate objective of a data visualization study on the condition. This can be implemented by the following actions below.

## Actions:

**Data Cleaning**: Cleaning the data according to the requirements is the most important phase among all the preliminary steps that we are going to take in successive phases of the visualizing. Cleaning the data majorly involves identifying the missing values, extracting the outliers, standardizing the data into the respective order, handling inconsistent data and validating the data if it can be relied to perform the tasks.

**Statistics and visualization:** After cleaning the data we need to calculate the descriptive statistics to obtain the outline of the dataset. To understand the correlation and dependence about various columns in the dataset we are going to create some histograms, bar plots and pie charts. By plotting this kind of figures, we are going to understand the underline hidden trends between the variables that are conspired for the analysis.

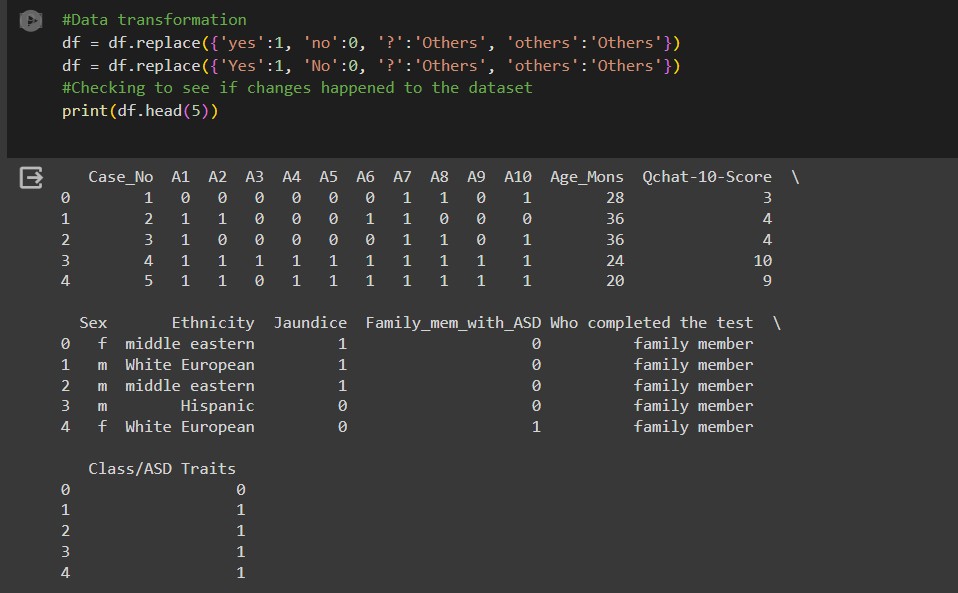
**Data Visualization**: By representing the data visually, it is easy to analyze the data like spotting the outliers, comparison of different features, and occurrence of one feature with another. Visualizing particularly helps in analyzing the data and to get meaningful observations from the intricate datasets.

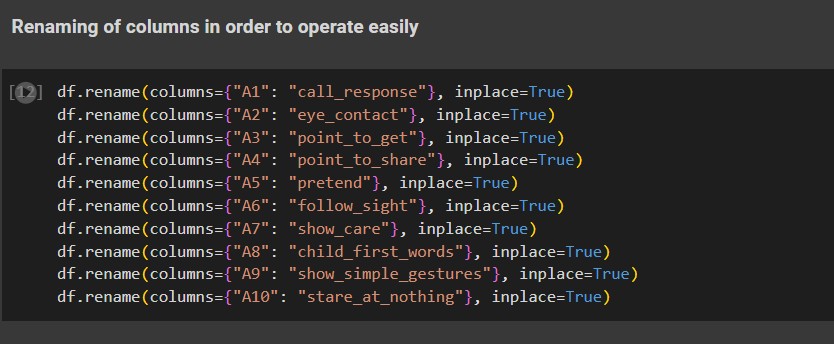
## Data Transformation:

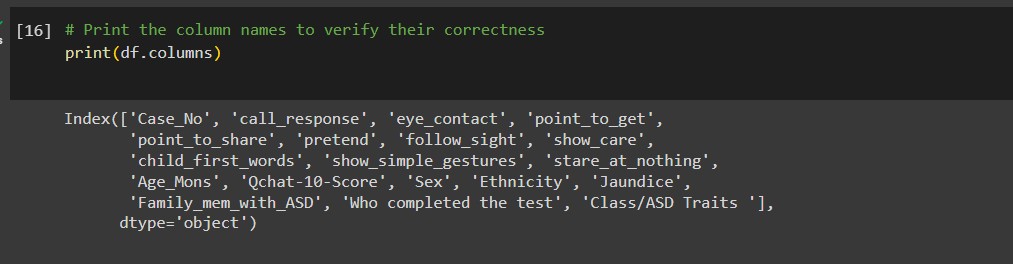
Some information transformation must be performed before any task abstraction on the EDA dataset.

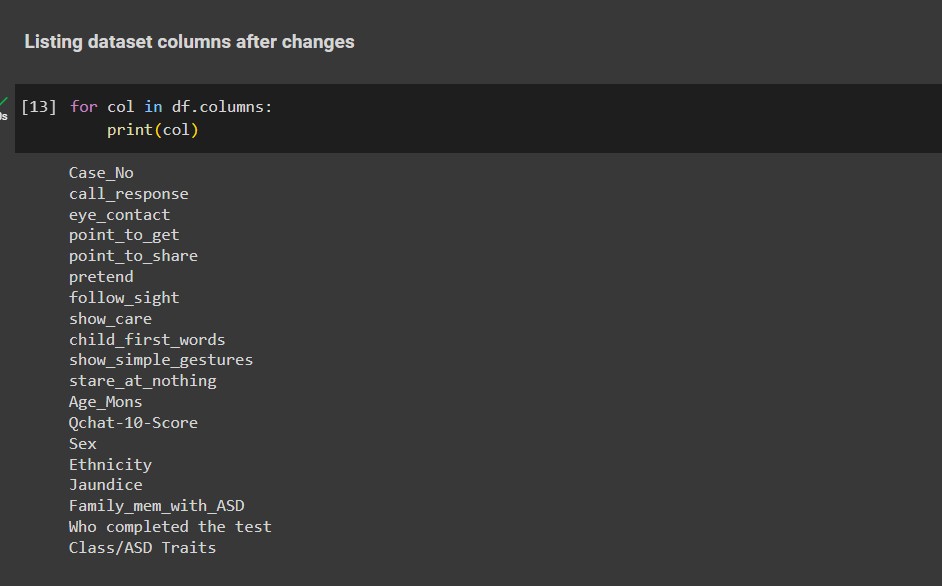
Finding the missing values, summing up the missing values, and eliminating the ones that are null from the dataset are the first elements in the procedure.

To make them suitable for visualization, multiple rows needed modifications. A 'Yes' or 'No' field might be found in these columns. 'Yes' converted into a one, and 'No' transformed into a zero. After the change, this picture was taken.









## Exploratory Data Analysis:

A crucial step in the data analysis procedure is exploratory data analysis (EDA), and Python provides a wide range of modules and tools that help with its execution. During this time frame, the following objectives were completed:

**Data Visualization:** To recognize trends, patterns, and anomalies that would not be obvious from the figures alone, we rendered the data using Python tools like Matplotlib, Seaborn.

**Data Transformation:** We converted the data into a format appropriate for evaluation using methods involving scales, normalization, and feature selection.

**Correlation Analysis:** We were able to gauge the amount of connection among variables using correlation analysis and illustrate the results using Python tools such as Pandas and Seaborn.

**Outlier Detection:** Any outliers that could have had an influence on analysis have been identified and treated with using Python modules like Pandas and Scikit-learn.

## Implementation using tools:

**Tools Description:**

## Google Colab:

1. Web- based platform which is extensively used for the data analysis and visualizations using the python programming language
2. Interactive environment to execute python code and supports vast amount of python libraries
3. It has free access to CPU, GPU’s which are used to run and visualize larger datasets to get meaningful observations.
4. Data can be directly handled from the google drive for the analysis purpose and the plots can be able to store in the cloud platforms for future purposes.

**Python Libraries:** Python is very popular programming language in the sea of data science for the data analysis and data visualizations. Python have some rich libraries like seaborn, matplotlib and plotly for data visualizations which will be discussed below:

## Matplotlib:

It is a basic and low level library used for the plotting purpose and to create static, interactive and animated visualizations. With this library we can create line charts, scatter plots and histograms.

## Seaborn:

It is a high and advance level data visualization technique when compared to the matplotlib and it is also a python library. It is build with the matplotlib ground but it can do some tasks which are difficult and cannot be done using the basic library matplotlib.

## Plotly:

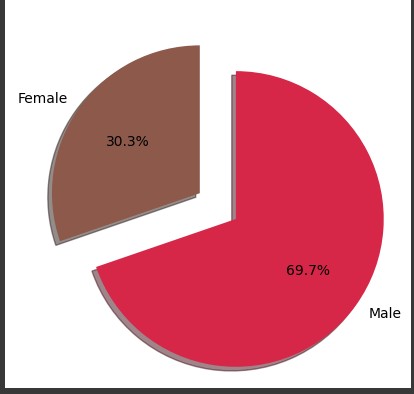
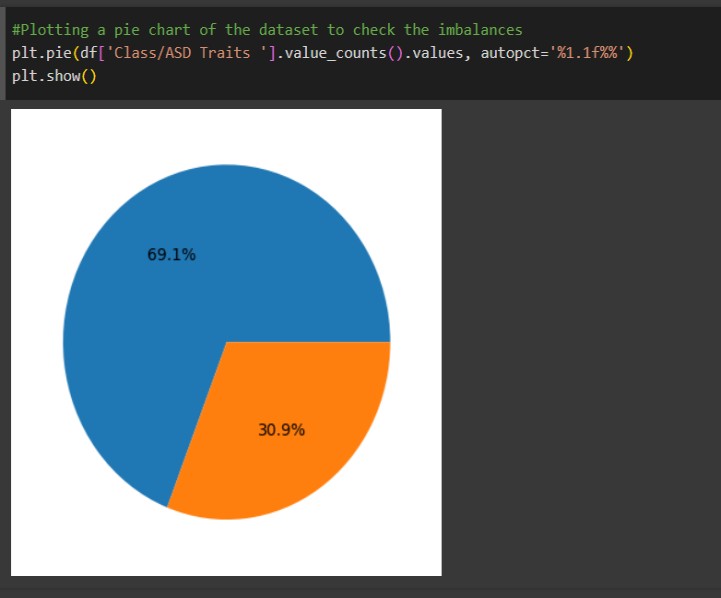
It is a open source plotting library which us specifically used for the creation of interactive and web based data visualizations. It has real time collaboration with the peers and wide range of sharing capabilities which is very much useful for the academia purpose

**Tableau:** It is powerful visualization tool and majorly useful for the industry specific purposes and the visualization we can create here can be dynamic in nature. It also supports real-time analysis and sharing capabilities. Below are few efficient aspects of tableau:

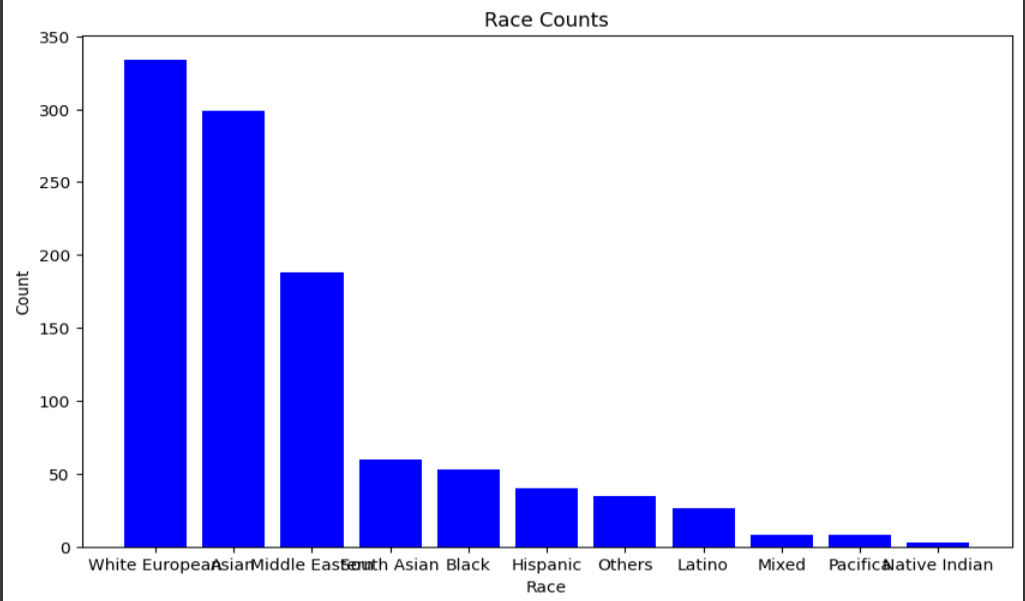
1. **Data Connectivity:** It supports wide range of data sources including spreadsheets, data bases, cloud platforms, and big data platforms. It also provides data preparation tools and helps with the cleaning of the data.
2. **Visualization types:** Wide range of plots can be done in tableau not limiting to charts, graphs and maps. The plots can be customized using the labels and the highlights and many more for the user interaction and convenience
3. **Real Time Data analysis:** It allows user to visualize and analyze the data as it was generated.

# Preliminary Results of the Visualization:

## Python Library Plots:



**Analysis:** In the above pie chart the class ASD traits has been plotted and by observing the chart we can conclude that 69.1% of the toddlers are having this problem of autism and 30.9% of the people are not encountering this phenomenon even though there are some behaviors observed in the toddler. According to the data that has been considered it is imbalance and might lead to some incorrect assumptions while training the models.



**Analysis:** Here in this bar plot we can observe that among the samples that we are having one third of the samples have been taken from White Europe and it is followed by Asian and the remaining regions are having a smaller number of samples compared to other. According to our analysis we have a doubt that major samples are coming from two geographic locations and completely relying on this data might not be appropriate. Situations which are specific to these locations may also lead to the toddler behavior and we need to look for extra information to make some concrete points while analyzing.

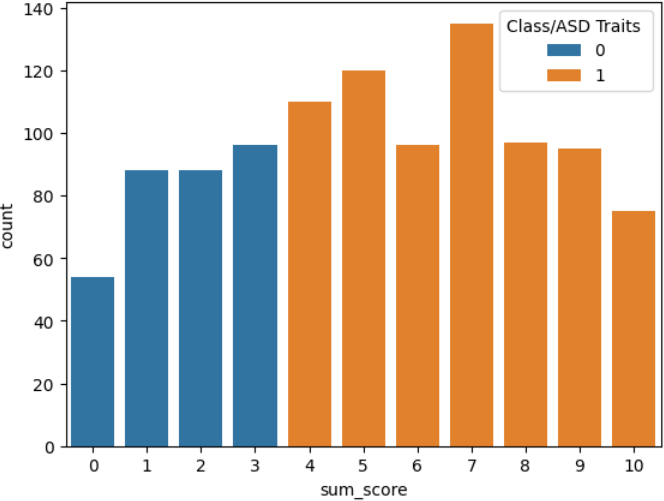


Figure : Positive and negative comparisons between Q1 and Q10.

The above graph compares with both positive and negative ASD measurements from Q1 to Q10. We analyze the total outcome which each examination applicants receives. The study values with quality members have been determined that they were higher. A 7 on these characteristics could score someone anywhere between 120 and 140. Besides five, there will be approximately 120 in all. 10 was the smallest amount, according to a 80–60 count. The highest-ranking class had an overall score of nearly 100. 90 was the value at which the total amount of 1,2 reached itself.0 is the total score of the smallest total, 57.

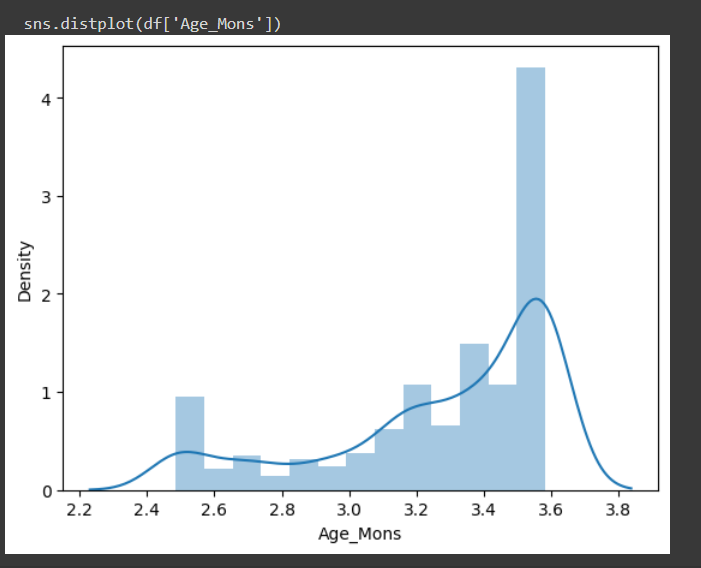
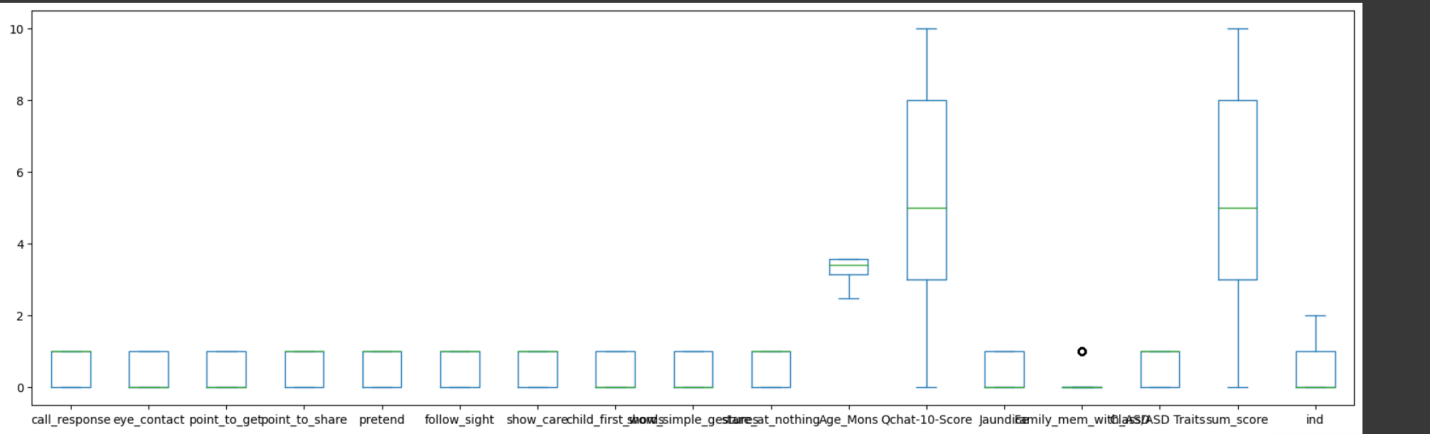


Figure : Plot of Skewness on the Age\_Mons column.

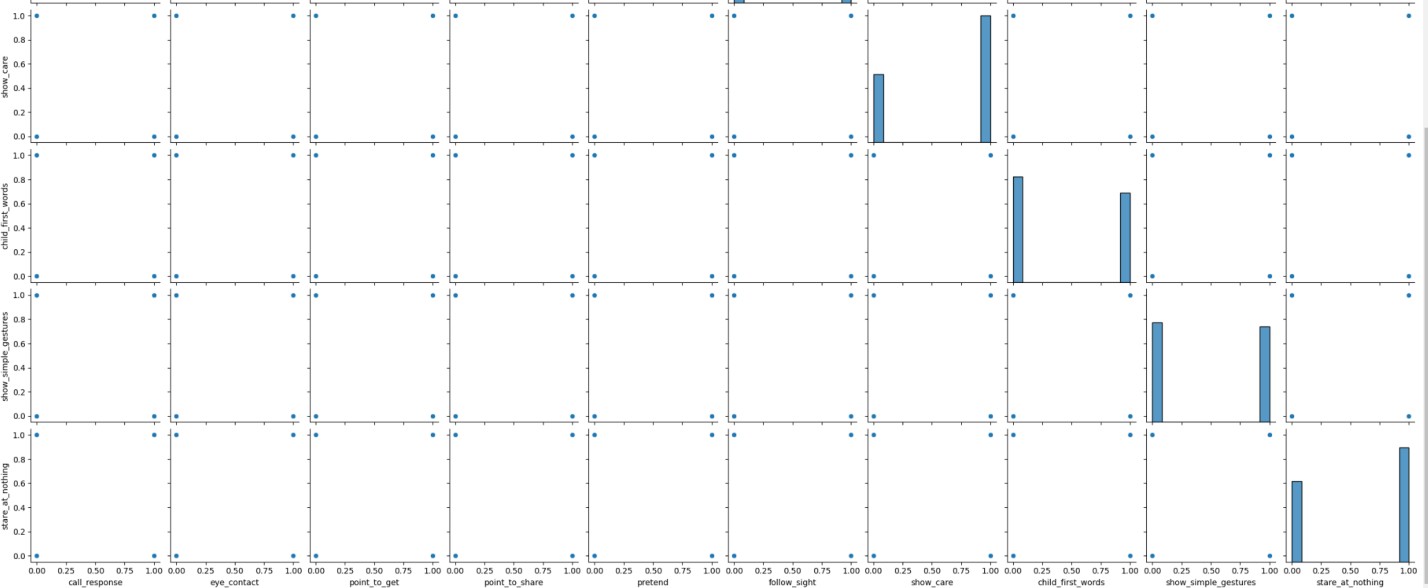
The Age in months skewness graph is displayed in the graph above. We can observe an important density of the problem around 3.6 years old. The age group with the lowest density of people with the disease is 2.8 years old. It was approximately 3.2 years old and its population density were a little more than 1. Because the density line reached 3.6 is obvious. Between 2.2 to 3.6 next, the density of the line reduced. Many ages, like 3.3, 2.6, and 3.5, have still not achieved the density limit.

# Overview Plots:

## Some Plots for understand the data effectively:

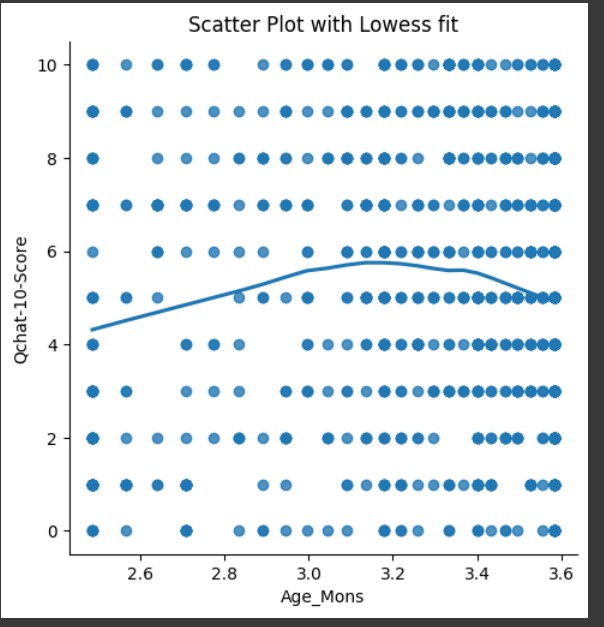


**Displaying the scatter Plot using Seabron Library:**

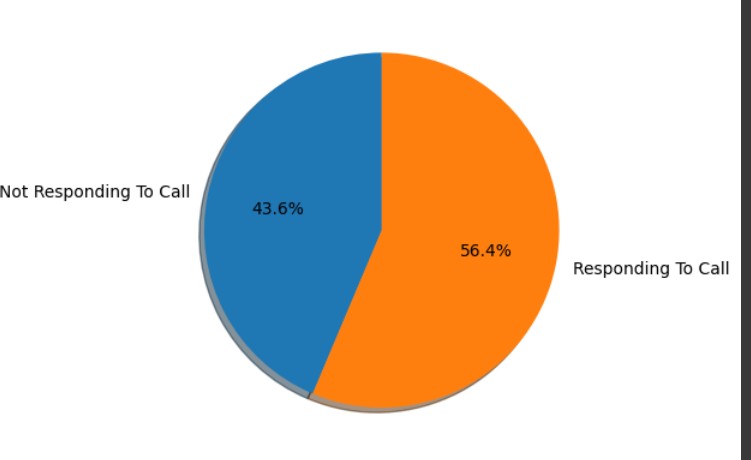


**Analysis:** Scatter plot for different features related to each other

Note: All the plots are not attached because of the space constraint and they can be found in ipynb

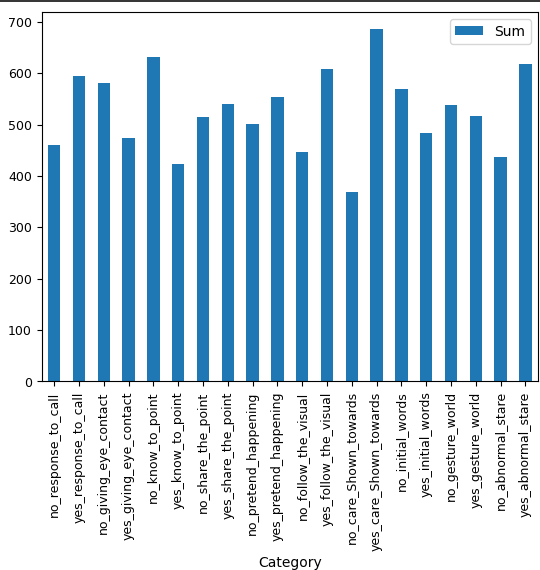


## Displaying response of child to a call:

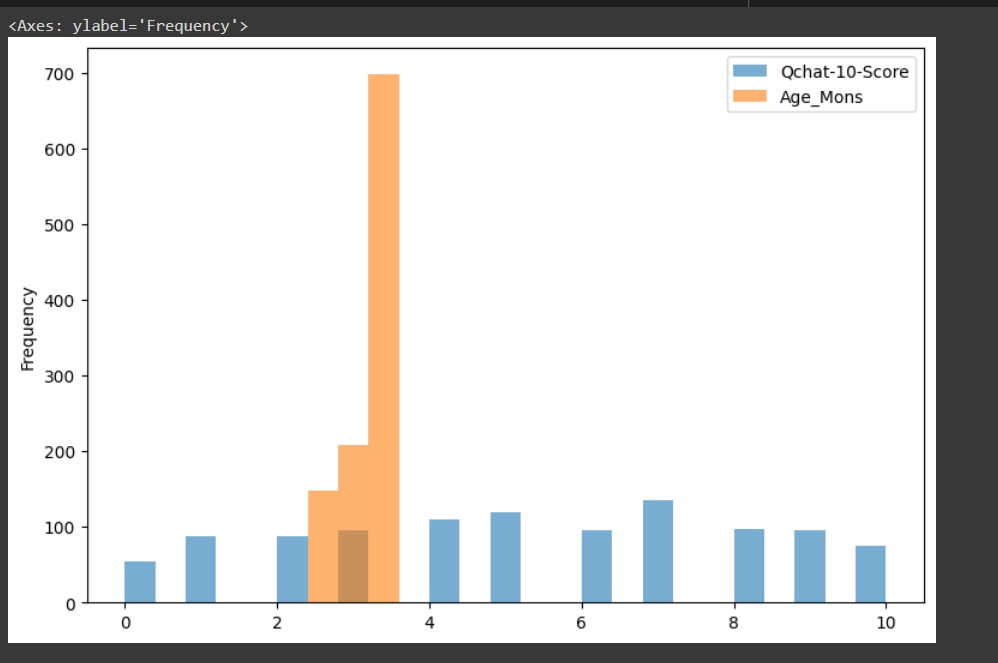


**Analysis:** As the graph clearly explains the record of response by the toddler when we call them. It is one of the features when defining the test statistics and one of the important observation that we are considering while doing the test. According the chart there is no big difference in the two segments which are only differ by 12%. Considering only this feature while doing the test may not be that much effective and we need to consider other options also.

## Displaying graph for different kind of activities:

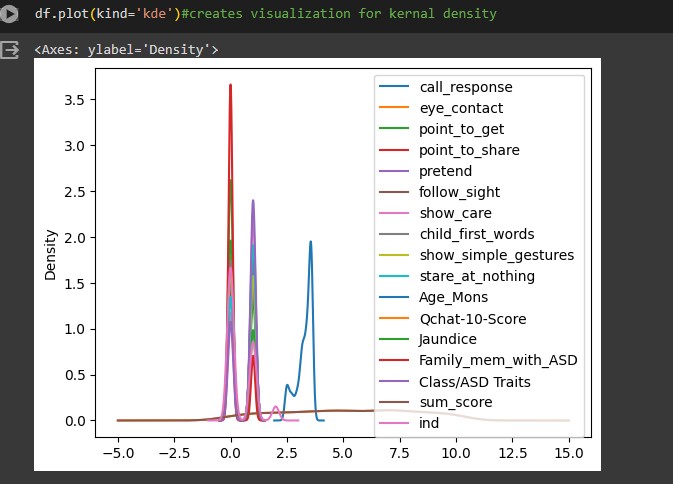


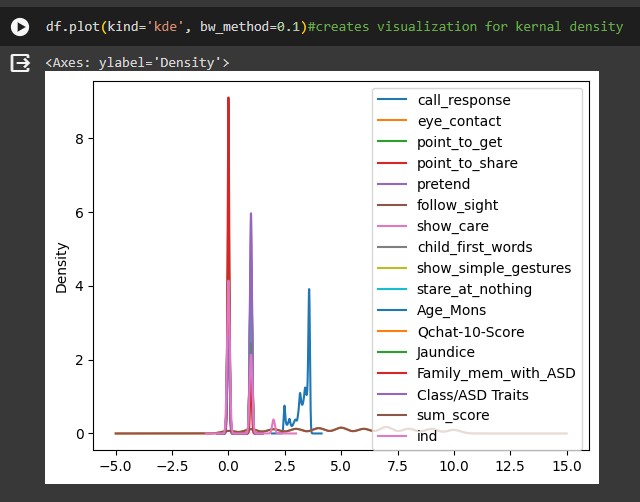
**Analysis:** So, there are total 10 questions in the test and every questions has been marked separately with two categories implies the yes or no for the response. We are having a total of 1000+ samples and response for each toddler has been recorded. The responses have been segregated and they are count as the yes/no values and the bar plot has been plotted for that scenario. As we can see in the above graph we got the pictorial representation of the responses count.



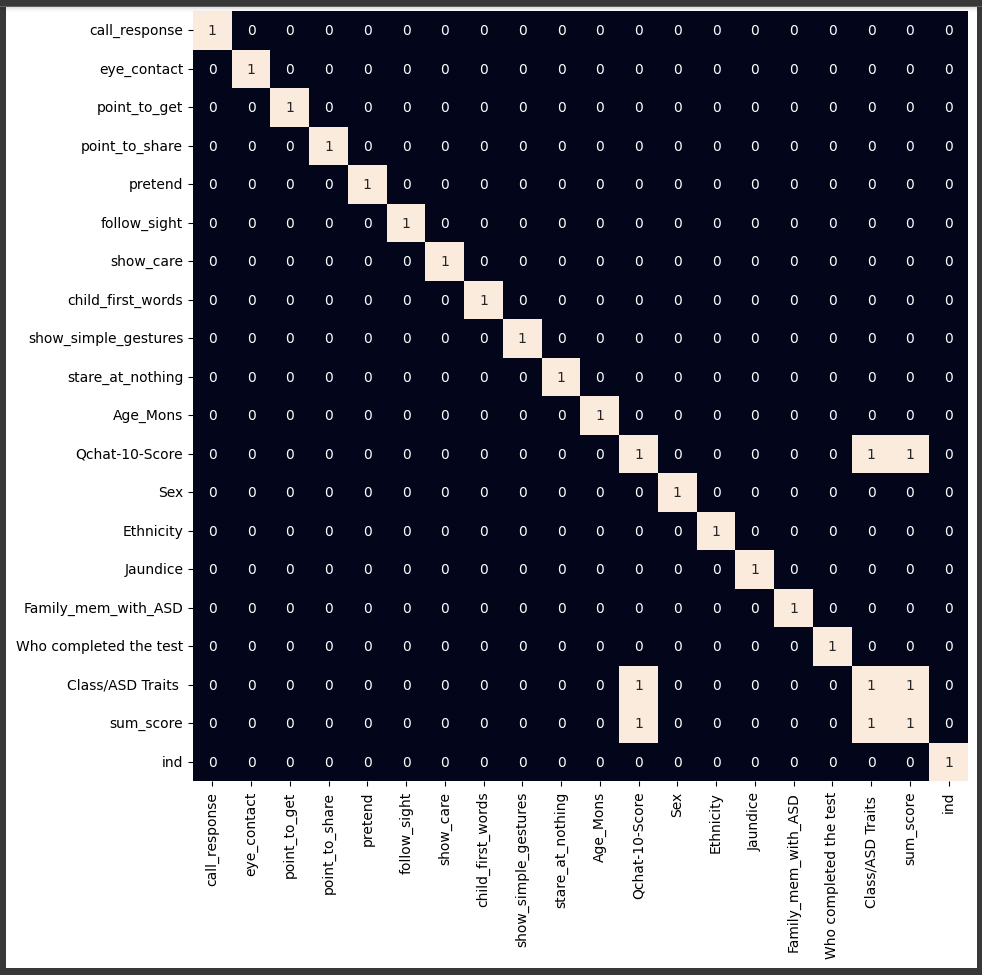
# Area Plots:







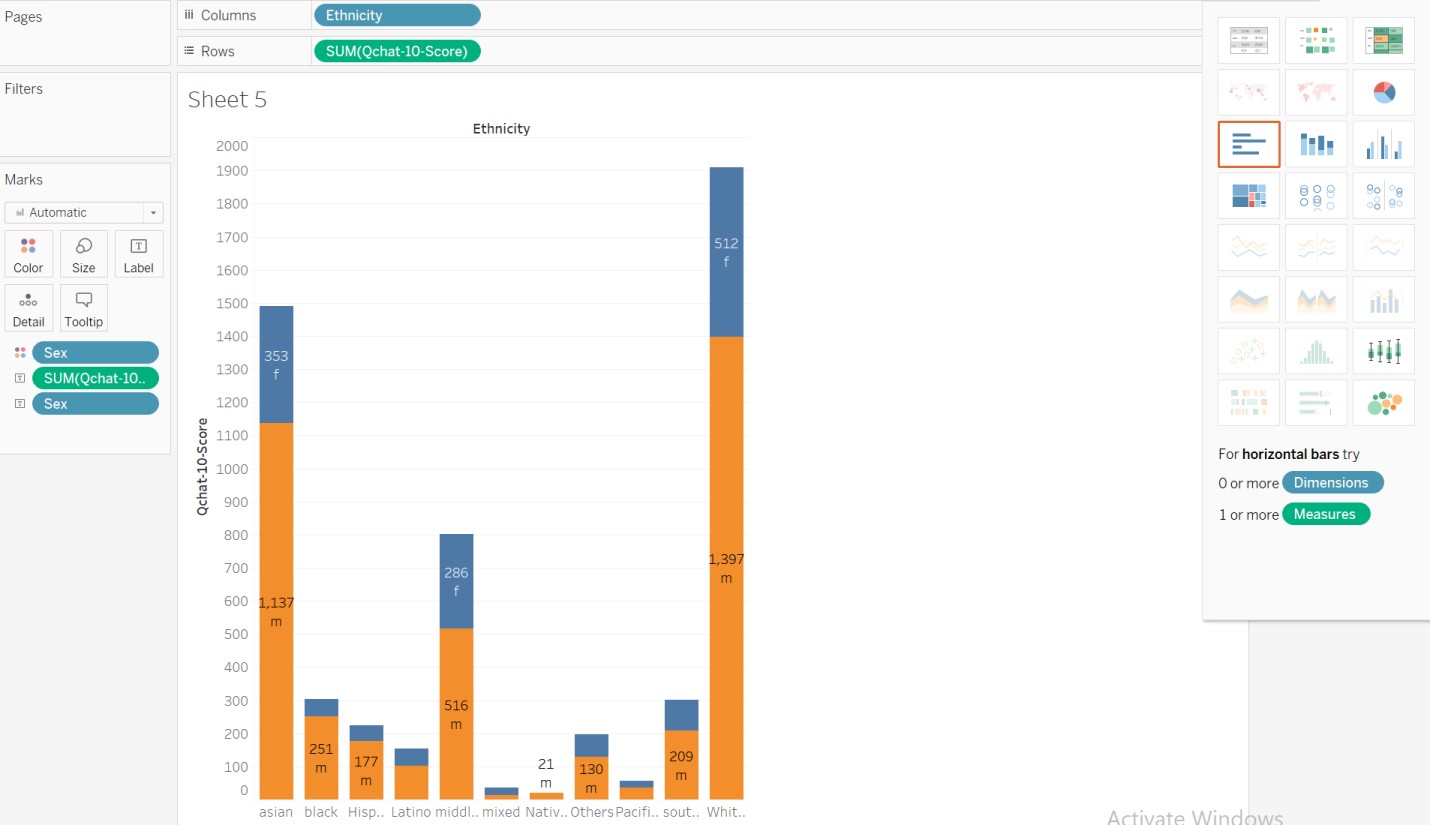
**Heat map representation:**



Data values are shown as color-coded cells in a heat map, that helps the detection of movements and trends in the information being analyzed. It can be seen that both the coordinates of the x- and y-axes, which correspond to the current value of the heat map, are same. The horizontal axis of the heat map displays the various colors. Once the parameters and scores have been combined, each class/Asd has a value of 1. The Qchat 10 score, also the same as 1, corresponded with both the Asd characteristics and overall score.

# Tableau Visualizations:

## Visualization 1:

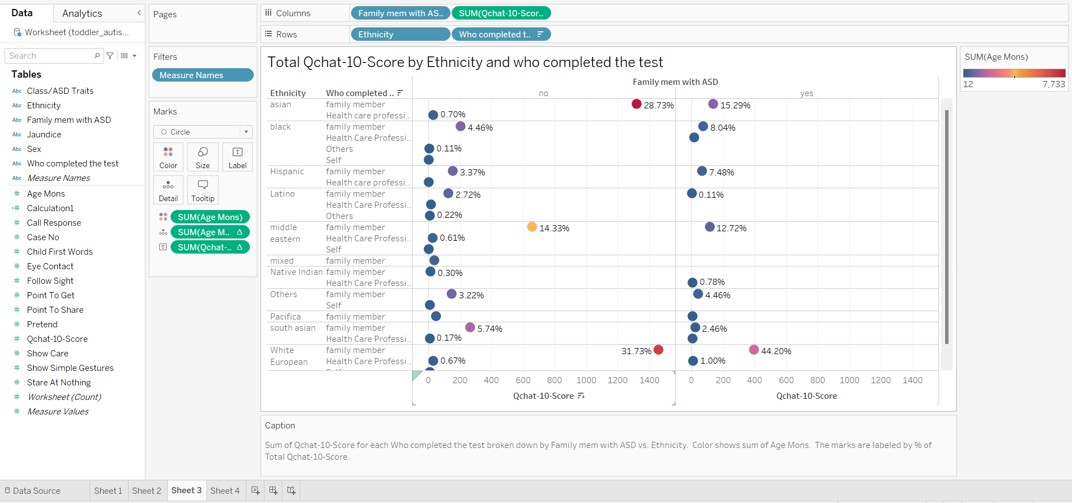


In line with the graph above, Asians and people of various ethnic backgrounds were the next groups with the highest Q-10 scores, being followed by White Europeans for both men and women. This suggests there might be a correlation between gender, ethnicity, and Q-10 scores. Justification: The data in Qchat-10 scores provide credibility to the assumption that there could have a connection involving ethnicity and autistic traits. Therefore, the theory is correct.

1) Can an exam taker's ethnicity or if they're related to a family member with autism spectrum disorder have an effect on the total Qchat-10 results?

Using color-coded circles to symbolize the entire age in months, this picture opposes the overall Qchat-10-Score for each ethnicity and the amount of test takers. The ease with which variations and patterns could be identified with this method promotes awareness of distinctions in Qchat- 10-Score throughout age groups and racial groups. The figure's utilization of circles for displaying the data efficiently reveals variations due to age and race.

## Visualization 2:



**Analysis:**

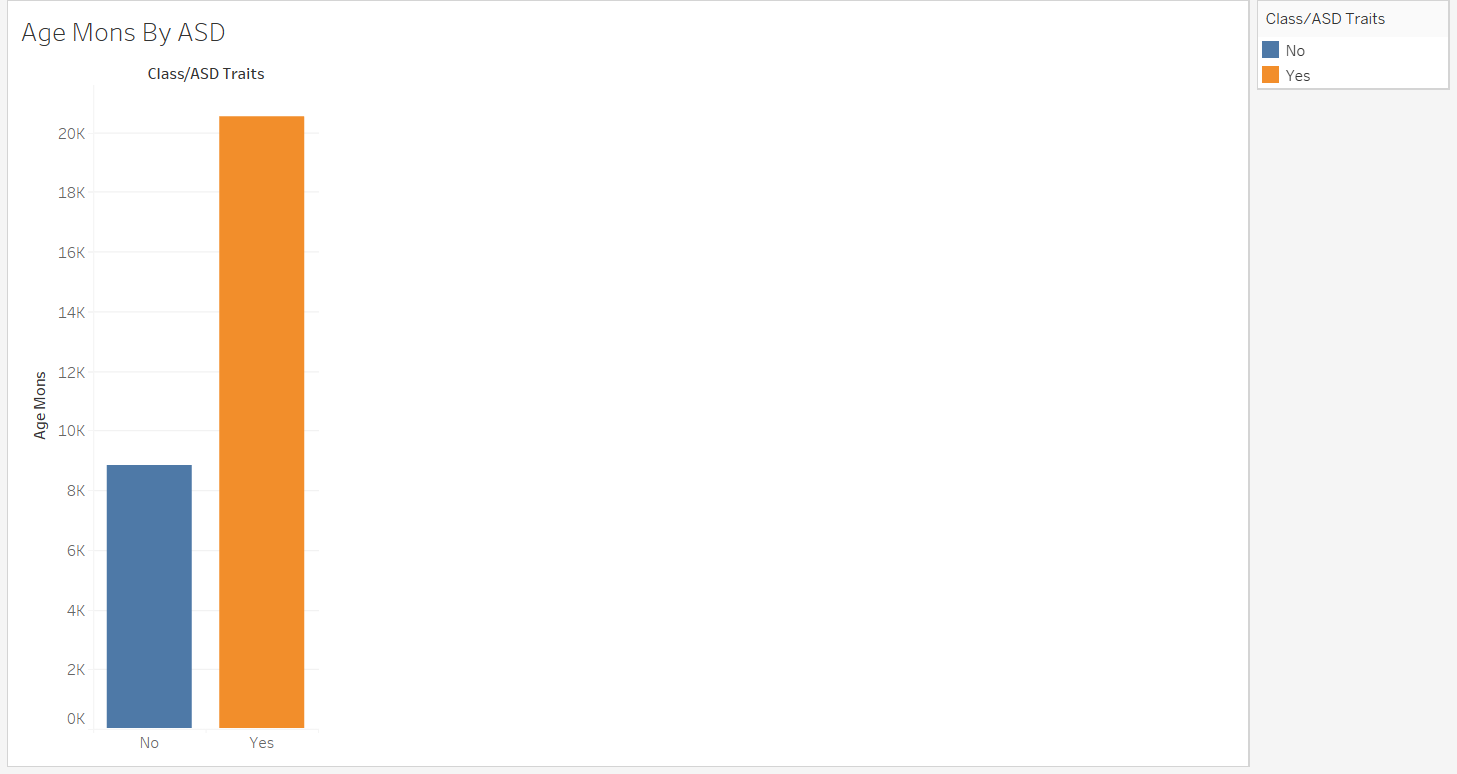
We can infer from the above image shows the family members with autism experienced the greatest percentage of test participation and also the highest Q-10 score.

Reasoning The Qchat-10 score variation based on ethnicity and family history of ASD is a valid theory. Higher scores are observed among people who have a family member with ASD, suggesting a possible genetic connection. Moreover, various ethnic populations rate differently, indicating that genetic and cultural factors may be involved. Gaining insight into those connections can aid in our comprehension of the subtle elements and hallmarks of autism.

## Visualization 3:

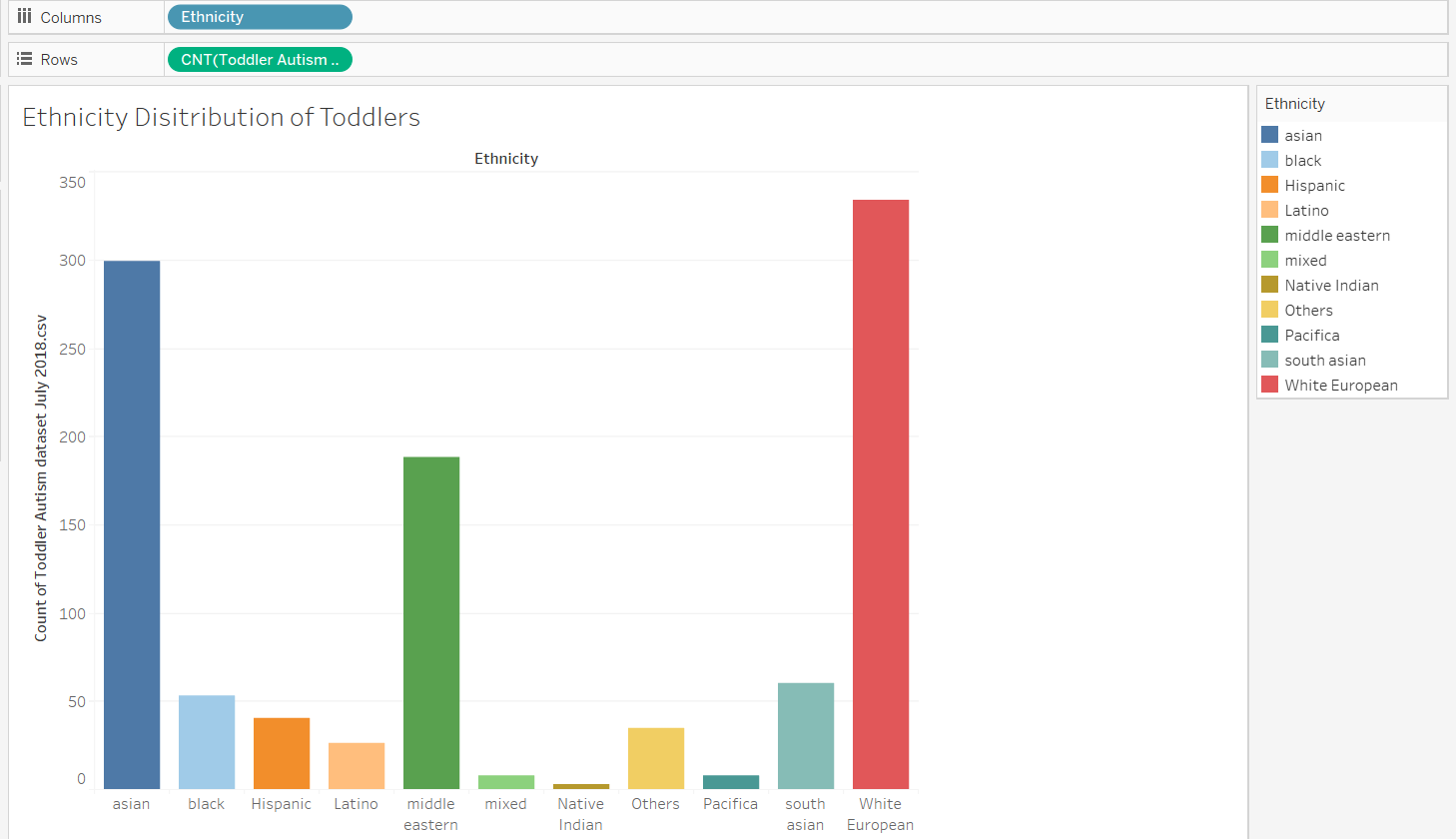
### Age Distribution by ASD Traits

This bar chart visualizes the distribution of age (in months) among toddlers based on their ASD traits. The x-axis represents the ASD traits ('Yes' for ASD traits present and 'No' for no ASD traits), while the y-axis displays the count of toddlers. The bar chart shows two bars corresponding to 'Yes' and 'No' ASD traits, with their respective counts (20,547 for 'Yes' and 8,825 for 'No'). This visualization provides a clear comparison of age distribution between toddlers with ASD traits and those without.



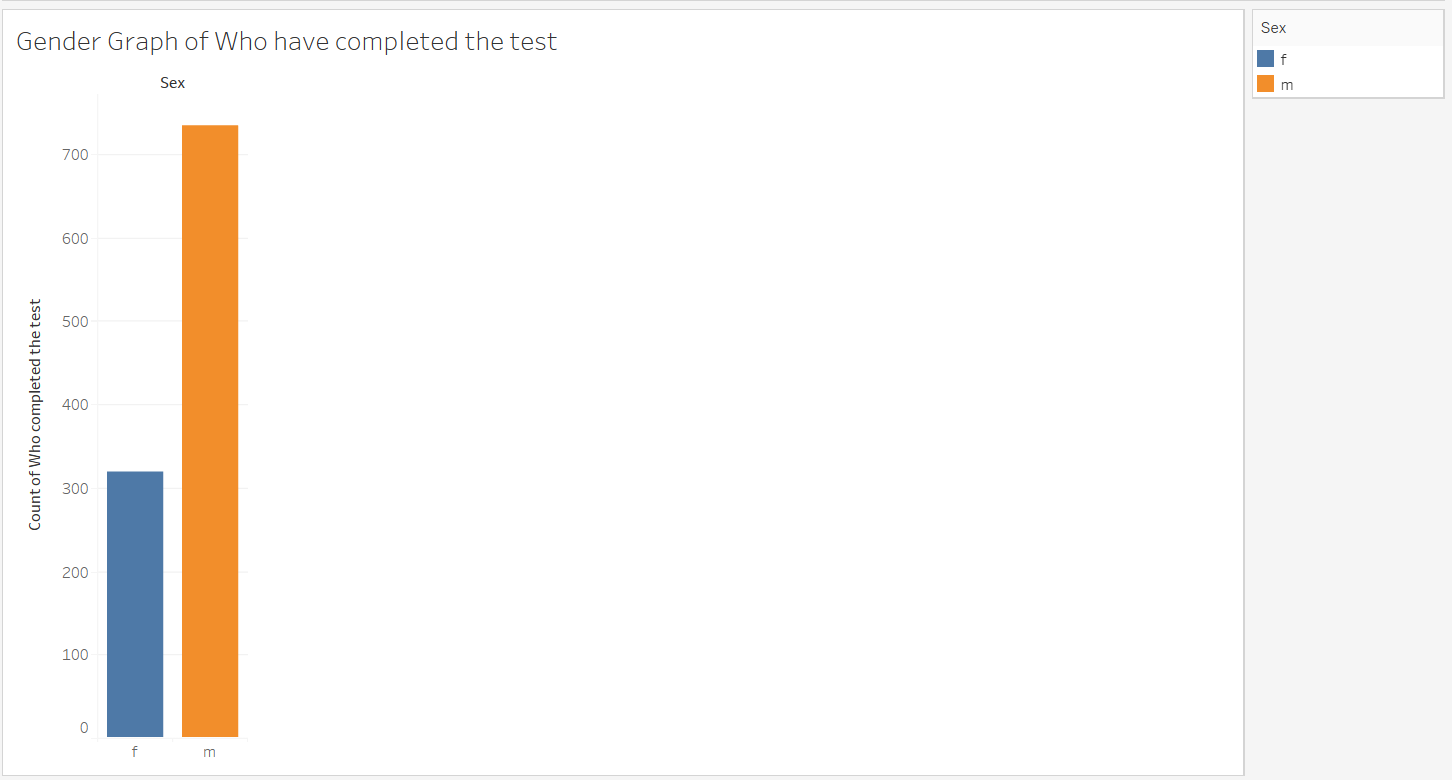
### Ethnicity Distribution of Toddlers

This bar chart illustrates the distribution of toddlers across different ethnicities within the dataset. Each bar represents a specific ethnicity, and the height of the bar corresponds to the count of toddlers belonging to that ethnicity. The ethnicities are listed on the x-axis, and the y-axis shows the count of toddlers. For example, 'White European' has the highest count (334 toddlers), followed by 'Asian' (299 toddlers) and 'Middle Eastern' (188 toddlers). This visualization offers insights into the diversity of ethnic backgrounds represented in the toddler dataset.



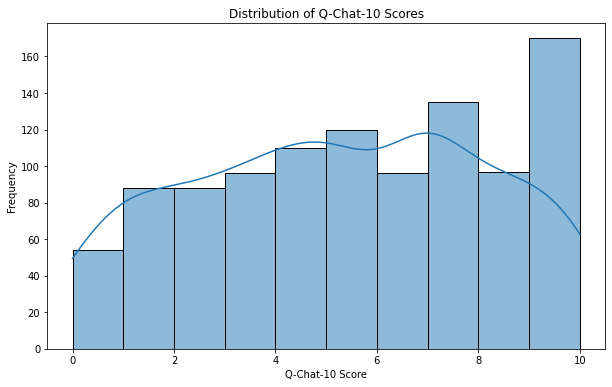
### Gender Distribution of Test Completers

This bar chart displays the gender distribution of individuals who completed the ASD screening test. The x-axis represents the genders ('m' for male and 'f' for female), while the y-axis shows the count of test completers. The bar chart has two bars corresponding to male and female genders, with their respective counts (735 for male and 319 for female). This visualization provides a clear overview of the gender composition among individuals who underwent the ASD screening test.



### **1. Distribution of Q-Chat-10 Scores**

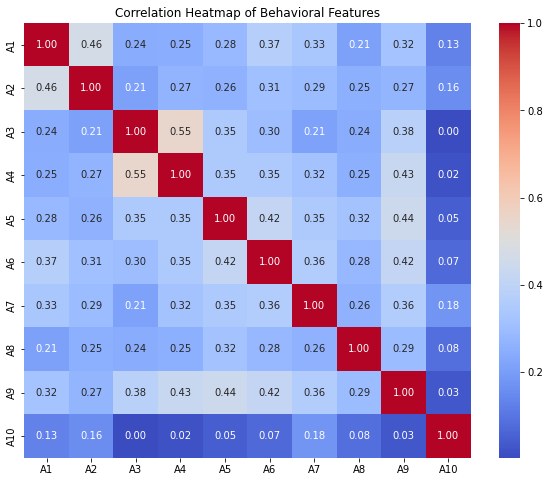
* **Visualization:** Histogram showing the distribution of Q-Chat-10 scores.



* **Interpretation:**
  + Most individuals scored between 4 to 8 on the Q-Chat-10 scale.
  + A significant number of people scored very high (10) on the scale.

### **2. Correlation Heatmap of Behavioral Features**

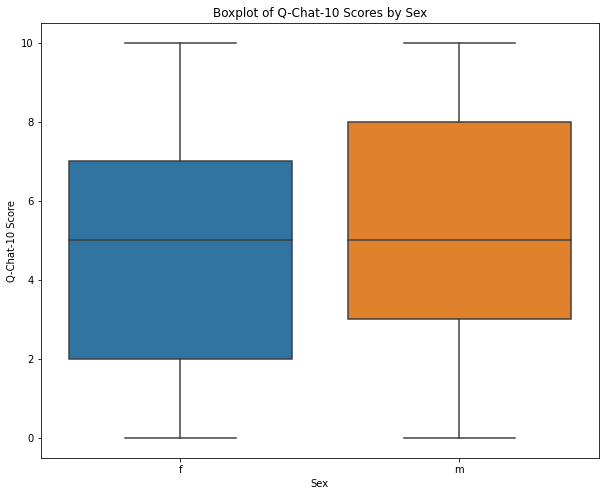
* **Visualization:** Heatmap displaying the correlation matrix between behavioral features (A1 to A10).



* **Interpretation:**
  + Features along the diagonal show perfect positive correlations with themselves.
  + Moderate positive correlations observed between some pairs of features (e.g., A4 and A3).
  + Overall, low or negative correlations between different behavioral features suggest relative independence.

### **3. Boxplot of Q-Chat-10 Scores by Sex**

* **Visualization:** Boxplot comparing Q-Chat-10 scores between sexes (male vs. female).



* **Interpretation:**
  + Both sexes have similar median Q-Chat-10 scores.
  + Females show a narrower interquartile range (IQR), indicating less variability around the median.
  + Males exhibit a wider IQR, suggesting more diverse scores with some higher and some lower than females.

### **Potential Implications:**

* **Q-Chat-10 Score Distribution:** The concentration of scores around certain values and the spike at the highest score (10) could indicate distinct groups within the dataset, potentially useful for identifying individuals with strong autistic traits.
* **Behavioral Feature Correlations:** Understanding the independence of behavioral features helps in designing targeted interventions or assessments that consider multiple traits without redundancy.
* **Sex-Based Score Variability:** Variability in Q-Chat-10 scores between sexes could prompt further investigation into the underlying factors contributing to these differences.

# Chapter 1: Identifying Real-World Problems

#### **"Life" Story Overview:**

The dataset on Autistic Spectrum Disorder (ASD) screening for toddlers sheds light on a pressing issue affecting healthcare systems globally. ASD, a neurodevelopmental condition, carries significant healthcare costs and challenges due to lengthy diagnosis procedures. The rising prevalence of ASD worldwide underscores the urgent need for efficient and accessible screening methods to aid early diagnosis and intervention.

**Who:** Toddlers at risk of ASD, their families, healthcare providers, and communities are affected by the challenges of timely diagnosis and intervention. Families face emotional and financial burdens, while healthcare professionals struggle with limited resources for effective ASD screening.

**What:** The problem lies in lengthy waiting times for ASD diagnosis, exacerbating stress for families and burdening healthcare systems with high costs. The delay in diagnosis delays early intervention, impacting the long-term outcomes for children with ASD.

**When:** The urgency of this issue is evident in the rising global prevalence of ASD cases, demanding immediate attention to develop streamlined and effective screening methods.

**Where:** The challenge is both societal and clinical. Families reside in communities seeking improved healthcare access, while diagnosis procedures occur within healthcare settings, requiring improvements in screening efficiency.

**Why:** The economic impact of ASD, coupled with its increasing prevalence, necessitates innovative screening solutions to alleviate healthcare burdens and enable timely interventions. Limited availability of resources and ASD-specific datasets further complicates the situation.

**How:** The issues arise due to the complex nature of ASD diagnosis, often reliant on labor-intensive procedures such as extensive behavioral assessments and clinical evaluations. The lack of standardized, accessible screening tools contributes to delays and inefficiencies in healthcare delivery.

The pressing need for accessible and efficient ASD screening methods is underscored by the challenges faced by toddlers at risk of ASD, their families, and healthcare providers. Timely diagnosis and intervention are crucial for improving outcomes and reducing the economic and emotional burdens associated with ASD. As ASD cases continue to rise globally, addressing these challenges becomes imperative to ensure better healthcare access and outcomes for affected individuals and communities.

# Chapter 2: Linking Data with Real-World Problems

In this chapter, the objective is to connect large-scale datasets with real-world problems identified in Assignment 1, focusing on data cleaning, preprocessing, and management using established tools. This process aims to identify relevant datasets that address the challenges outlined in Assignment 1 and analyze their characteristics and applicability.

#### **Who is the dataset about?**

The dataset primarily targets toddlers at risk of Autism Spectrum Disorder (ASD) and includes relevant individuals such as their families and caregivers. The sampling process likely involves individuals undergoing ASD screening or assessments, potentially leading to specific sampling biases. Understanding sample representation is critical to assess the dataset's applicability and potential limitations.

#### **What events, activities, behaviors, and observations are recorded?**

The dataset records various behavioral features and demographic attributes relevant to ASD screening. These include responses to specific questions related to communication, social interaction, and repetitive behaviors, aligning with the diagnostic criteria for ASD. Other recorded variables encompass personal and familial characteristics that may influence ASD traits.

#### **When did the events, activities, behaviors, and observations take place?**

The temporal structure of the dataset is crucial for understanding the developmental context of ASD traits. Data collection timestamps indicate when assessments were conducted, shedding light on whether the data is longitudinal or cross-sectional. This information informs the dataset's generalizability over time and its relevance to ongoing diagnostic practices.

#### **Where did the events, activities, behaviors, and observations take place?**

Data collection locations are significant, reflecting the geographical coverage and context of ASD screening efforts. Understanding the dataset's geographic scope—whether local, regional, national, or global—facilitates assessing its external validity across different settings. Geographic information systems (GIS) may be included to map spatial distributions within the dataset.

#### **Why were the events, activities, behaviors, or observations recorded?**

Data collection aims to facilitate early identification of ASD traits, contributing to timely interventions and healthcare planning. The dataset addresses the pressing need for efficient ASD screening methods to mitigate healthcare costs and support individuals at risk of ASD.

#### **How did the events, activities, behaviors, or observations occur?**

The dataset captures responses through standardized screening methods, reflecting what, when, and where specific behaviors and interactions occurred. The data's integrity and reliability depend on rigorous assessment protocols and documentation practices, ensuring consistency and accuracy in recording ASD-related traits.

By examining these aspects of the dataset, we gain insights into its relevance, representativeness, and potential implications for addressing real-world challenges in ASD screening and intervention. Leveraging advanced data analysis techniques and tools, we aim to extract meaningful patterns and insights to inform effective healthcare strategies and support systems for toddlers at risk of ASD. This hands-on exploration integrates theoretical knowledge with practical applications, bridging the gap between data science and healthcare innovation.

# Chapter 3: Users

#### **Who:**

The primary users of the ASD screening application and visualization include healthcare professionals, caregivers, and parents of toddlers at risk of Autism Spectrum Disorder (ASD). These individuals are pivotal in early detection and intervention efforts for ASD.

#### **What can the application do? What does the visualization show?**

The application allows users to input behavioral data related to toddlers' responses to specific questions (A1 - A10) from the Q-Chat-10 screening method. This data is used to generate visualizations that highlight ASD risk factors and traits. The visualizations show patterns, trends, and correlations in behavioral responses, aiding in the early identification of potential ASD traits.

#### **When can the user use the application/visualization?**

Users can access the application/visualization at any time, enabling continuous monitoring and screening of toddlers' behaviors. Regular usage allows for ongoing assessment and intervention, especially during critical developmental stages.

#### **Where will the visualization and applications be deployed?**

The visualization and application can be deployed on various platforms, including mobile devices (iOS, Android) and web browsers. This ensures accessibility for healthcare professionals and caregivers in clinical settings, homes, and community healthcare centers.

#### **Why is the visualization or application useful to the user?**

The visualization and application provide several key benefits:

* Early Detection: Allows for the early identification of ASD risk factors based on behavioral responses.
* Proactive Intervention: Supports timely intervention and support strategies tailored to individual developmental patterns.
* Data-Driven Decision-Making: Enables healthcare professionals and caregivers to make informed decisions based on visualized behavioral data.
* Accessibility: Provides a user-friendly interface accessible via mobile and web platforms, enhancing usability and adoption.

#### **How will the people/the community use this application or visualization to make changes?**

The application and visualizations empower users to:

* Monitor Developmental Milestones: Track toddlers' developmental progress and identify deviations that may indicate ASD traits.
* Customize Interventions: Tailor interventions and support strategies based on individual behavioral patterns and identified ASD risk factors.
* Promote Awareness: Raise awareness about ASD risk factors and the importance of early screening within communities and healthcare settings.
* Facilitate Collaboration: Foster collaboration between healthcare professionals, caregivers, and parents to implement proactive ASD screening and intervention measures.

In summary, the ASD screening application and visualizations serve as powerful tools for early detection, intervention, and community engagement in addressing Autism Spectrum Disorder. They leverage data-driven insights to empower users to make informed decisions and positively impact the lives of toddlers and families affected by ASD. The accessibility and usability of these tools contribute to enhanced healthcare delivery and support for individuals at risk of ASD.

# Project Management:

## Implementation Status Report:

**Work Completed:**

## Description:

* Identify the data sources and collect the data that is required to analyze and visualize
* Cleaned the data and the preprocessing techniques have been applied to ensure that the data is complete for the further tasks.
* Exploratory Data analysis has been done on the data to gain the insights how the data has been distributed and how the features are linked with each other.
* Concluded the tools that are needed to visualize the results.
* Created the initial data visualization to gain in-depth understanding of the data and how it is segmented

**Responsibilities of Team Members:**

**Priyanka Jammigumpula (33%)**

* Improve the preliminary visualizations created from the cleaned and preprocessed data.
* Create Tableau visualizations and enhance plots for better communication and understanding.
* Design the layout and features of the dashboard, ensuring usability and clarity.

**Vishnu Priya Eupuri (33%)**

* Explore the data across different variables to gain comprehensive insights into their interactions.
* Determine appropriate data visualizations beyond the initial ones to deepen understanding.
* Further clean the data, selecting key features for specific analysis purposes and addressing edge cases.

**Sri ram Relangi (34%)**

* Document and present the results in a compelling manner, integrating the findings from visualizations.
* Validate visualizations against ground truth and external sources to ensure accuracy.
* Conduct statistical analysis to assess the reliability and applicability of the data in real-world scenarios.

Concerns/Issues Addressed:

1. **Data Diversity:** Acknowledged the limitation of data being primarily focused on two geographic locations, emphasizing the need to consider this in concrete statements.
2. **Data Privacy and Security:** Recognized challenges in collecting more diverse data due to the sensitivity of the topic and ongoing nature of behavioral research.
3. **Visualization Tools:** Discussed challenges in aligning visualizations with analytical insights due to the data's nature, emphasizing the need for tools suitable for mathematical and logical analysis.
4. **Alignment of Visualizations:** Highlighted difficulties in translating analytical insights into meaningful visual representations that satisfy the team's understanding.

**Conclusion:**

By means of our examination of the Autism Spectrum Disorder (ASD) screening dataset and the ensuing visual aids, we have acquired significant understanding of toddler conduct and the elements impacting ASD characteristics. One important statistic that showed how infants react to certain interactions and situations is the Q-Chat-10 score.

A noteworthy discovery was the influence of geographical location on behavioral characteristics. Our data mostly concentrated on two regions, highlighting the necessity of taking regional differences into account when extrapolating conclusions. This realization emphasizes how crucial context is to comprehending and interpreting behavioral patterns linked to ASD.

Our visuals, which we created with Tableau and Python, have been really helpful in turning complicated data into insights that can be put into practice. We have been able to recognize patterns, trends, and possible topics for more research thanks to these visual representations, which range from correlation heatmaps of behavioral variables to histograms of Q-Chat-10 scores.

Even with the advancements, a number of problems and issues arose. Data diversity is still a concern because the dataset might not adequately represent the range of behaviors associated with ASD worldwide. Data collecting is also constrained by ethical and privacy concerns, particularly in delicate study domains.

In order to investigate edge cases and improve our comprehension of toddler behavior, our team realizes that more advanced visualizations and analysis are required in the future. With the use of interactive dashboards and sophisticated statistical techniques, we want to shed light on the subtle differences between ASD features and offer research and clinically useful insights.

To sum up, this effort is a first step toward a deeper comprehension of ASD and toddler behavior. Even if the data we used to make our conclusions is limited, this sector is dynamic, therefore more research and development are necessary. Our goal is to provide significant insights that assist families, researchers, and medical professionals in managing the intricacies of autism spectrum diseases by fusing meticulous analysis with careful visualization.

# References:

* <https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers>
* Wetherby, A. M., Brosnan, S. M., & Lord, C. (2018). Early indicators of autism spectrum disorder in the second year of life. Journal of Developmental & Behavioral Pediatrics, 39(3), 233-241.
* Zwaigenbaum, L., Bauman, M. L., Stone, W. L., King, B., Fein, D., Pierce, K., ... & Buie, T. (2015). Early identification of autism spectrum disorder: recommendations for practice and research. Pediatrics, 136(5), e1263-e1283.
* Ozonoff, S., Iosif, A. M., & Rose, T. (2010). Reduced coordinated eye-head movements during social referencing in infants later diagnosed with autism spectrum disorders. Journal of Autism and Developmental Disorders, 40(10), 1241-1250
* 1) Tabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. Proceedings of the 1st International Conference on Medical and Health Informatics 2017, pp.1-6. Taichung City, Taiwan, ACM.
* 2) Thabtah, F. (2017). ASDTests. A mobile app for ASD screening. [www.asdtests.com](http://www.asdtests.com/) [accessed December 20th, 2017].
* 3) Thabtah, F. (2017). Machine Learning in Autistic Spectrum Disorder Behavioural Research: A Review. Informatics for Health and Social Care Journal.
* 4) Thabtah F, Kamalov F., Rajab K (2018) A new computational intelligence approach to detect autistic features for autism screening. International Journal of Medical Infromatics, Volume 117, pp. 112-124.