

* **Mohan Raj Thangavel Sekar**

## ABSTRACT:

In India, suicide and mental health crises claim thousands of lives every year, yet data on these tragedies often remains scattered, messy, and underutilized. This project wrangles a massive, 5,000+ row multi-table dataset—complete with inconsistencies, missing values, and nested JSON—to uncover hidden patterns behind the numbers. Using Python, SQL, and Power BI/Tableau, the analysis reveals where, why, and among whom the crisis is most severe, and pinpoints critical resource gaps. The findings aim to turn raw, chaotic data into actionable insights that can save lives.

**Multi-Table Data Cleaning, Transformation, and Visualization of Suicide & Mental Health Trends in India**

"A 5,000+ Row Multi-Table Relational Dataset in the Public Health Domain — Data Cleaning, Transformation, and Visualization Using Power BI to Derive Actionable Insights"

**Multi-Table Data Cleaning, Transformation, and Visualization of Suicide & Mental Health Trends in India**

|  |  |
| --- | --- |
| **Name:** | Mohan Raj T S |
| **Batch No:** | TN-DA-FNB03 |
| **Contact:** | 9095552548 |
| **Email ID:** | [mohanraj.sekar@gmail.com](mailto:mohanraj.sekar@gmail.com) |
| **Domain:** | Health Care |
| **Project Title:** | Multi-Table Data Cleaning, Transformation, and Visualization of Suicide & Mental Health Trends in India |
| **Submission Date:** | 20-8-2025 |
| **Mentor:** | Kumaran |
| **Dataset Link:** | <https://drive.google.com/drive/folders/1Nvnb3eEijvRap-mOG1CSL06mV9AMZ0rA?usp=drive_link> |

# **Table of Contents**

1. Executive Summary
2. Project Objectives
3. Data Sources & Methodology
4. Visual Representation (Power BI Exports)
5. Findings / Results
6. Insights & Discussion
7. Recommendations
8. Limitations
9. Conclusion
10. Appendices

# **Executive Summary**

This report analyses suicide data and helpline effectiveness between 2008 and 2022. The findings highlight declining trends in suicide growth, key state-level insights, and helpline program effectiveness. Recommendations are provided to strengthen prevention measures.

# **Project Objectives**

- Assess suicide trends from 2008 to 2022.  
- Evaluate Year-over-Year (YOY) Suicide Growth %.  
- Examine helpline call effectiveness and program coverage.  
- Provide actionable recommendations to reduce suicide rates.

# **Data Sources & Methodology**

**Data Sources:**  
- Government suicide statistics (2008–2022)  
- Helpline call logs  
- Program and facility records  
  
**Methodology:**

- Data Cleaning and Transformations  
- Data Visualization through charts  
- Calculative Measures and Columns for more accurate results  
- Forecasting suicides using Power BI’s forecast model  
- Comparative analysis to have more clearer understanding of problem

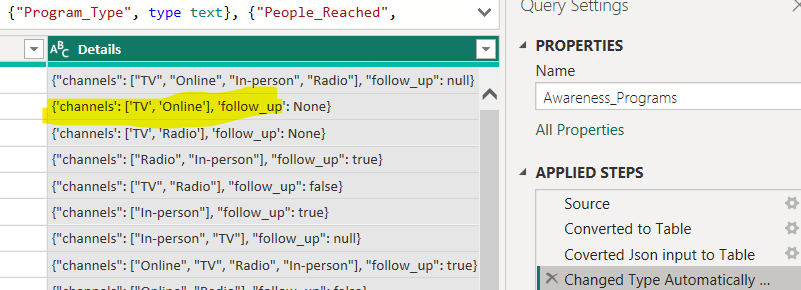
# **Visual Representation (Power BI Exports)**

# **4.1. Data Cleaning**

## **Awareness program Table:**

**Replaced all single quotes in json to double quotes for parsing the data into column values**

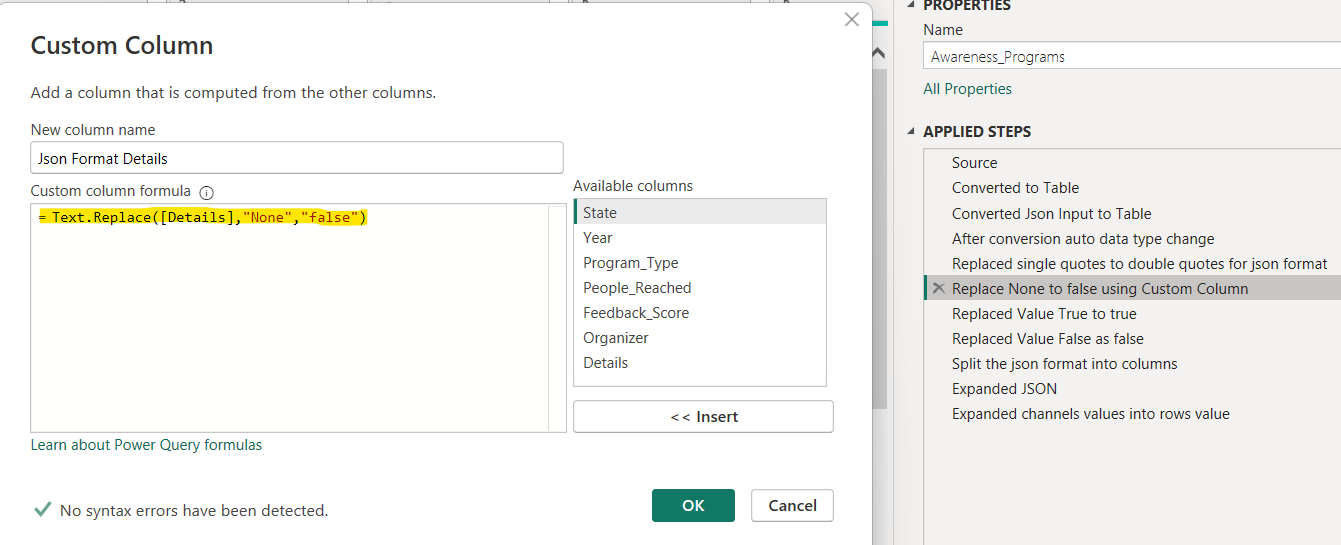
**Before:**



**After:**

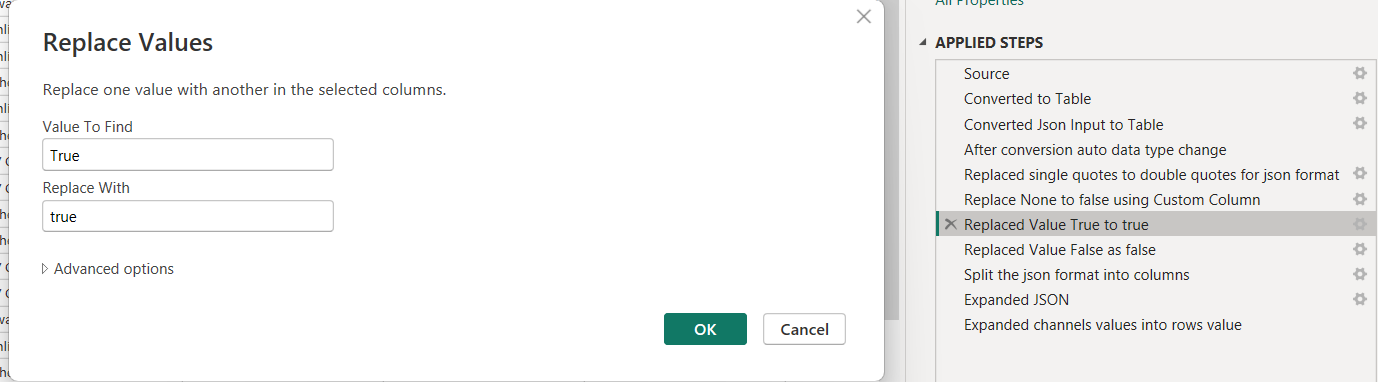


**Replacing None to false:**

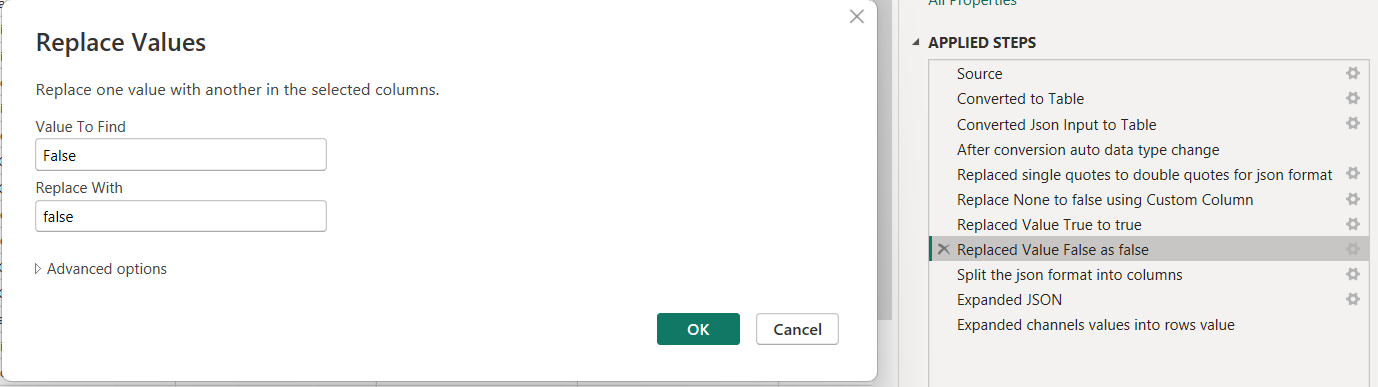


Reason for doing this while parsing json “None” does not fall within double quotes which implies to the structure of json. If a value isn’t inside a double quote, it implies a Boolean. So instead of adding double quotes changed it to Boolean value.

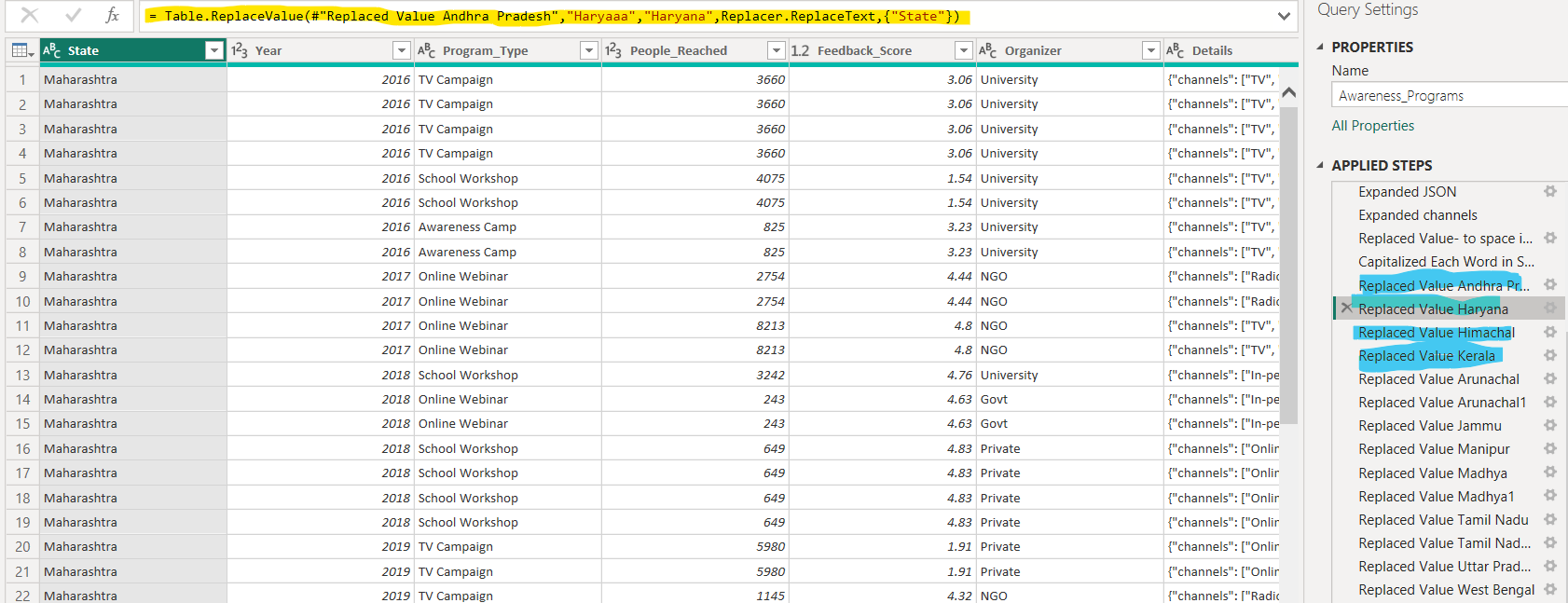
**Replaced Value True to true: in json all Boolean are lower case**



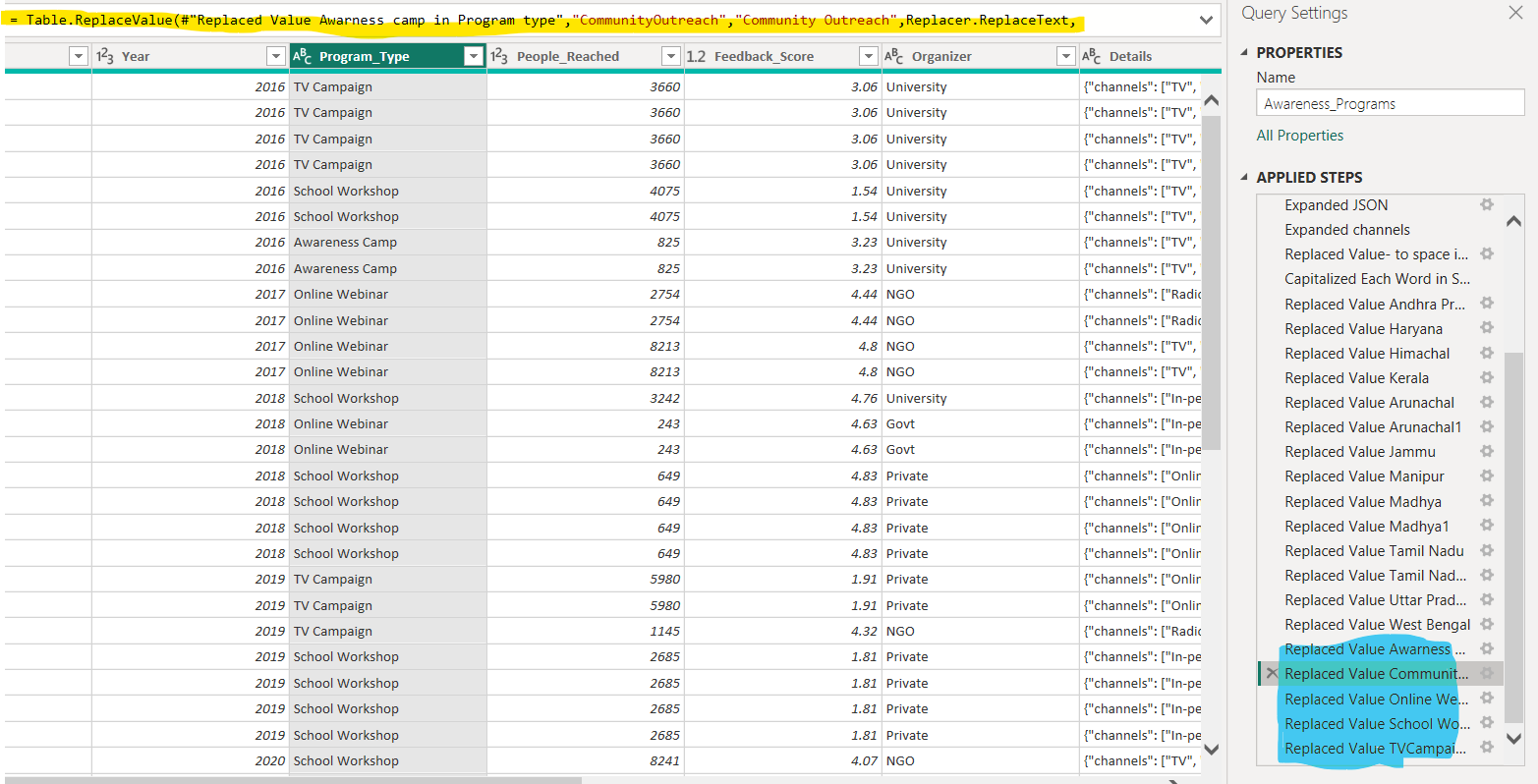
**Replaced Value False to false: in json all Boolean are lower case**



**Replaced inconsistent data in State Column:**



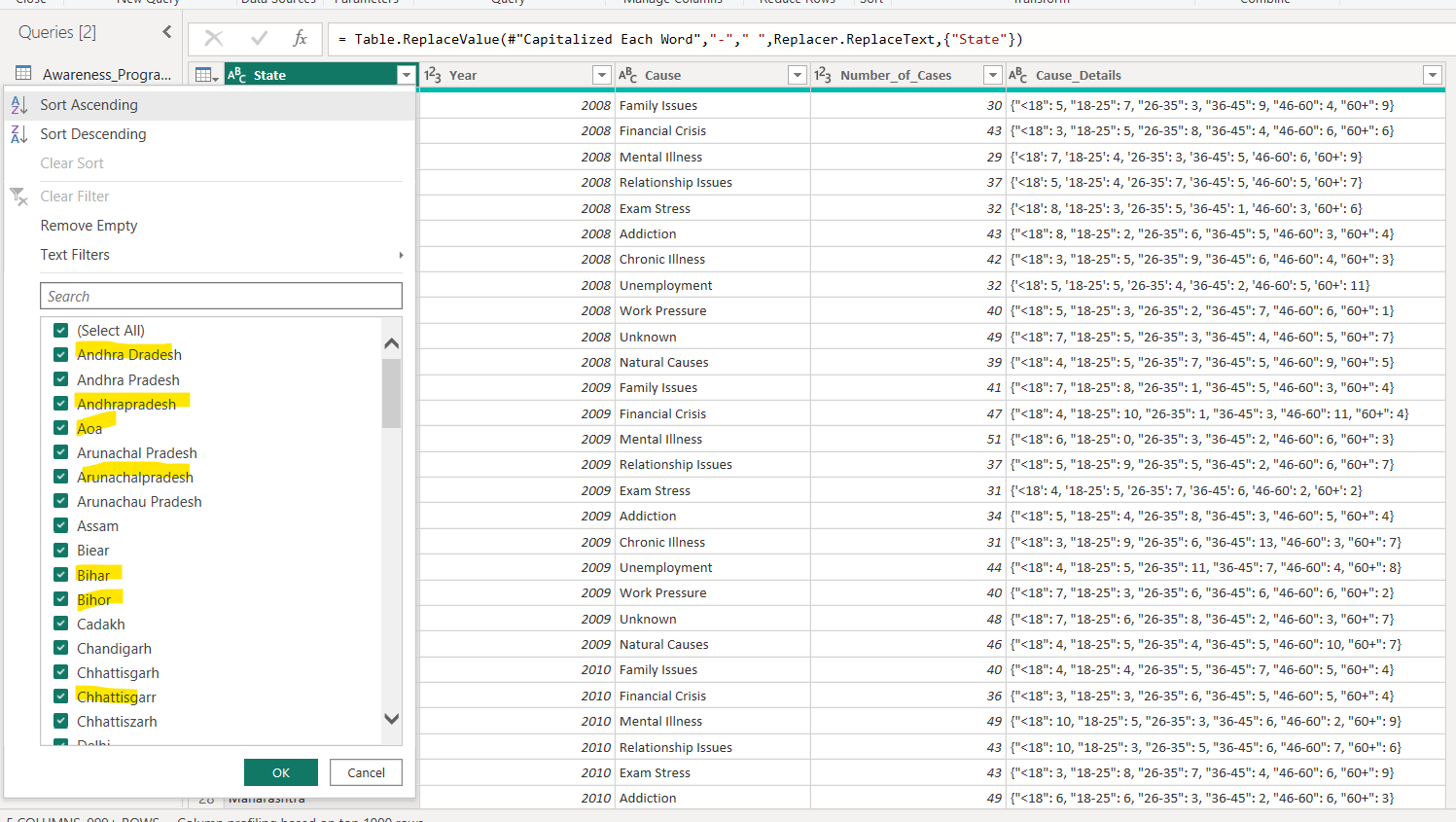
**Replaced inconsistent data in Program\_Type Column:**

****

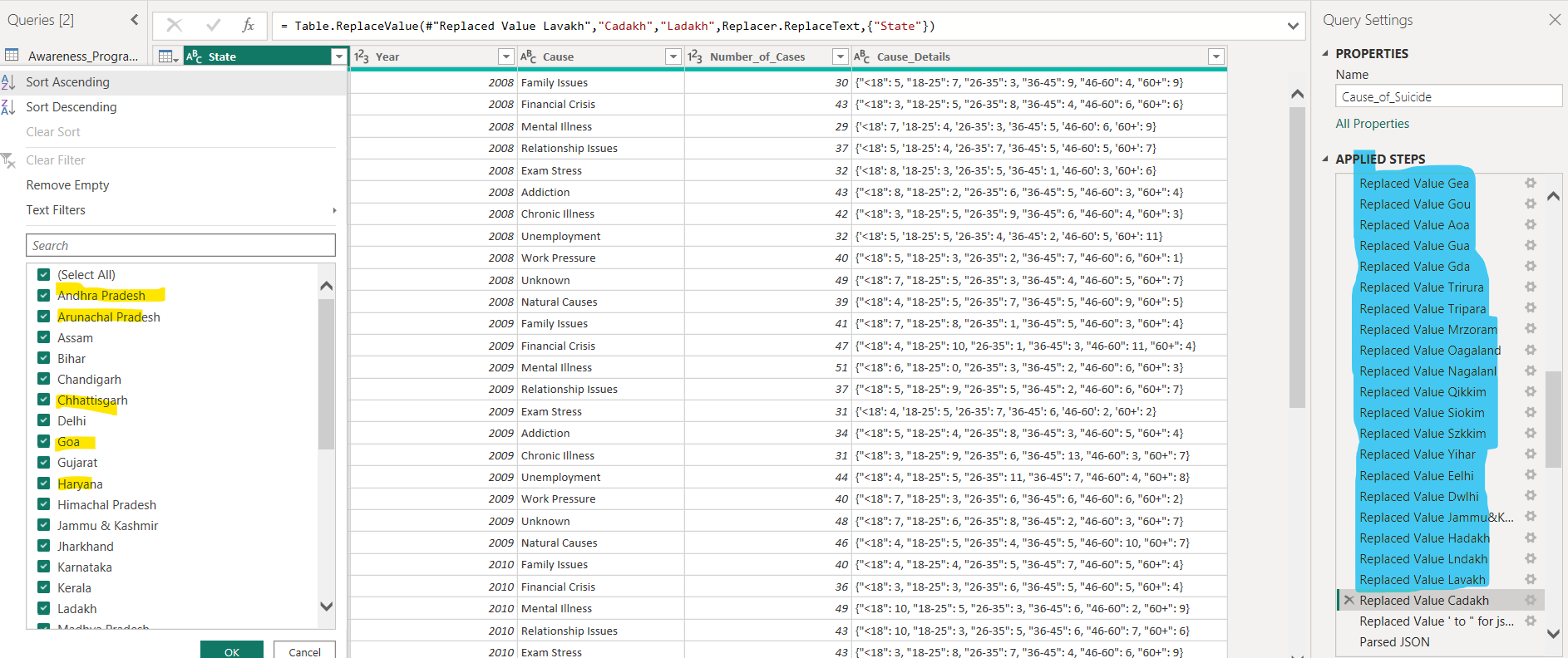
## **Cause\_of\_Suicide Table:**

**Data inconsistencies cleaning in State Column:**

**Before: Could see duplicates and inconsistencies**

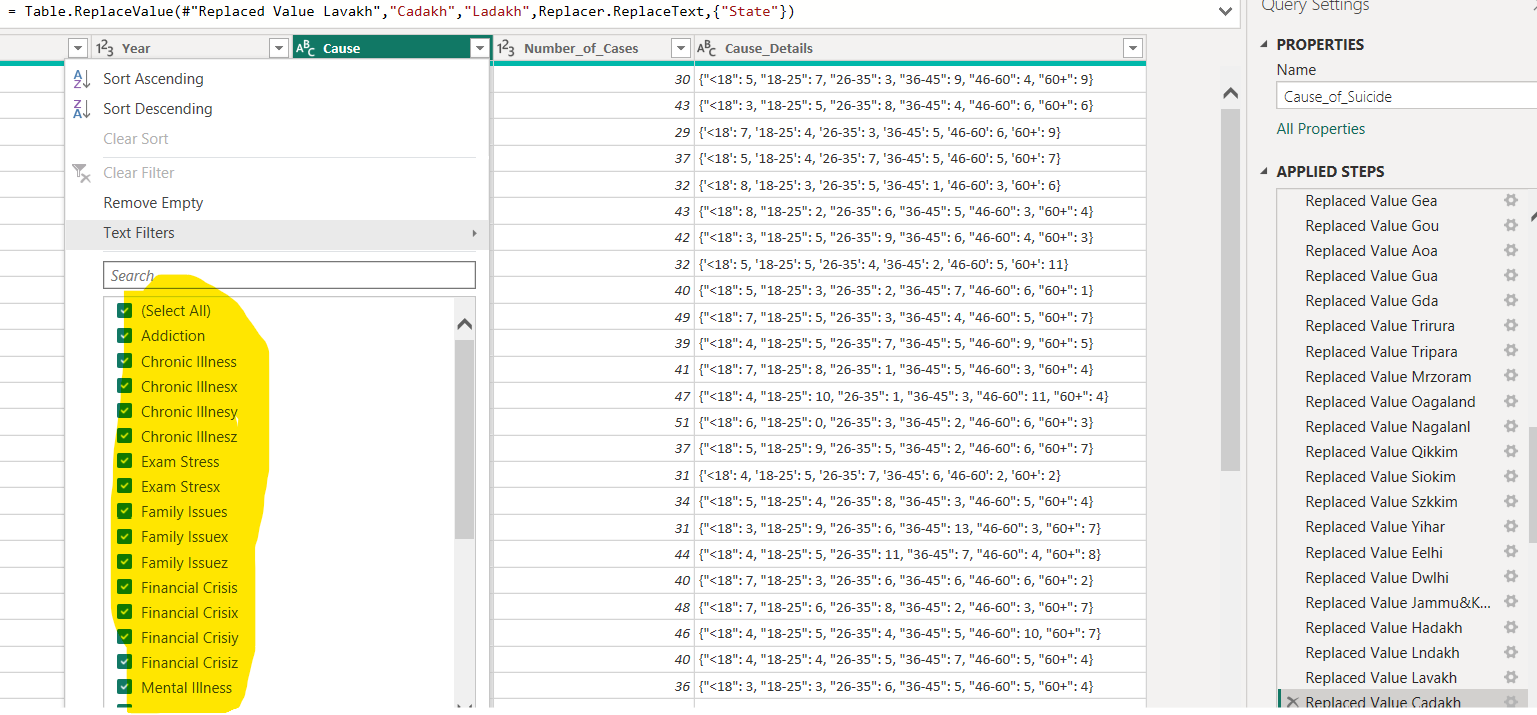
****

**After: Without duplicates and inconsistencies**

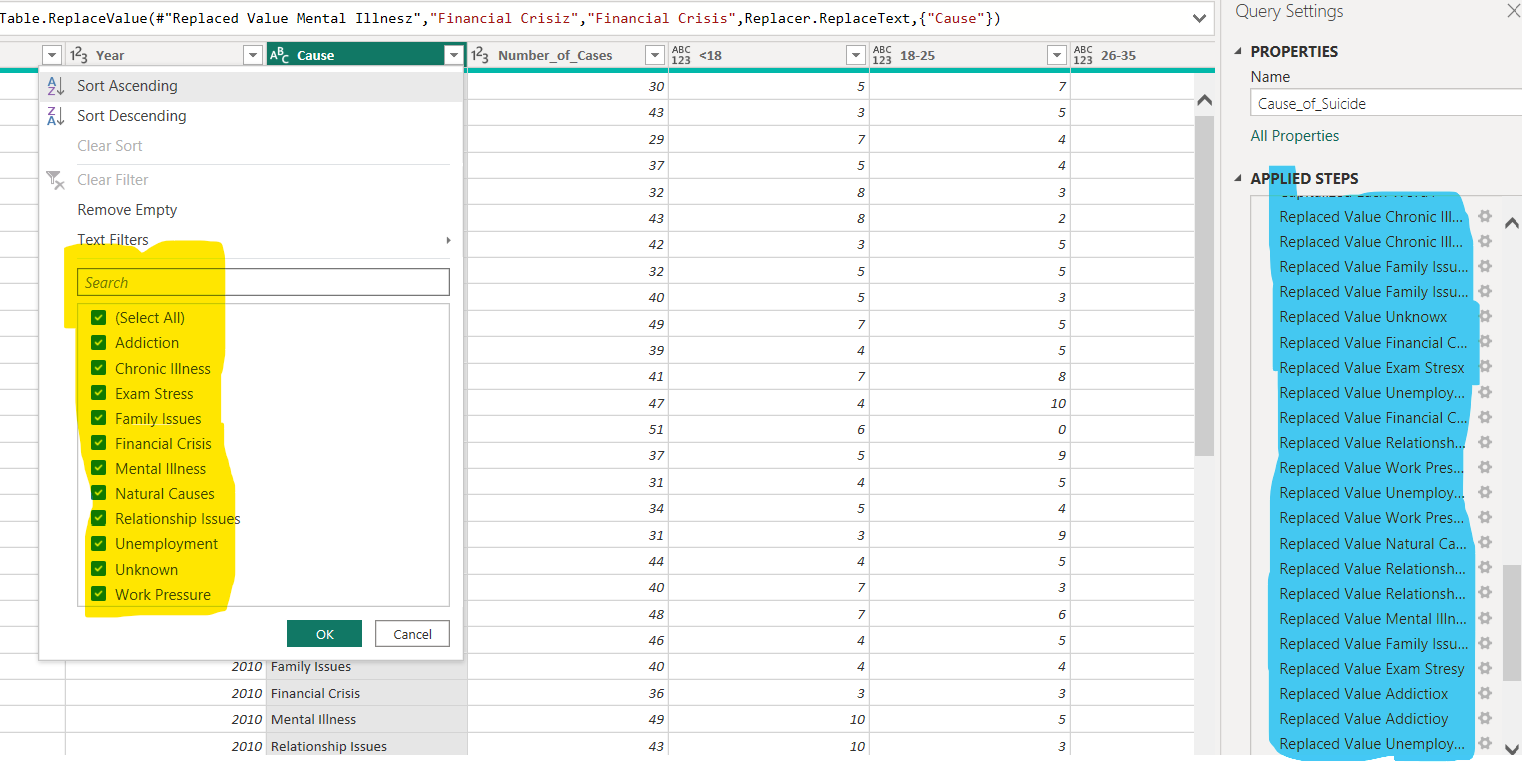
****

**Data inconsistencies cleaning in Cause Column:**

**Before: Could see duplicates and inconsistencies**

****

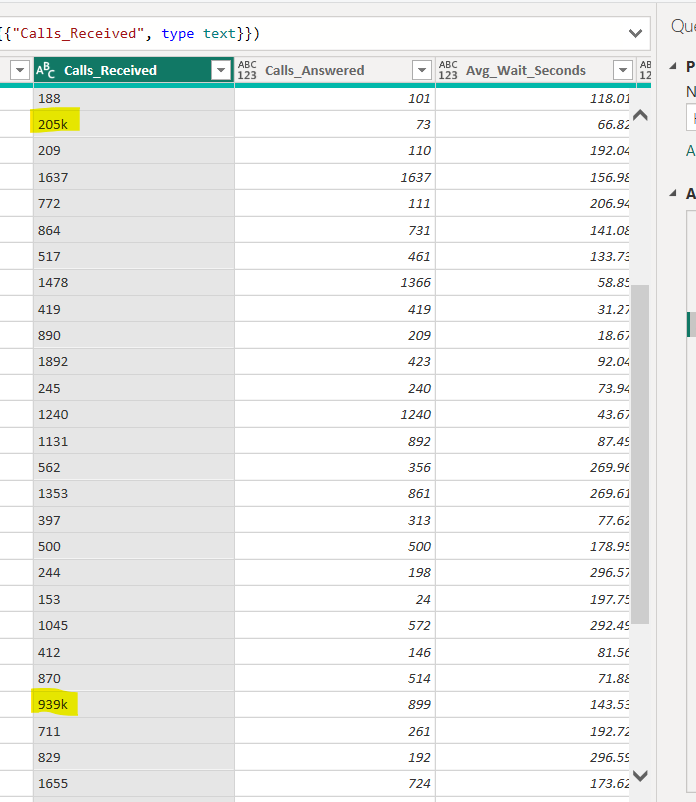
**After: Without duplicates and inconsistencies**

****

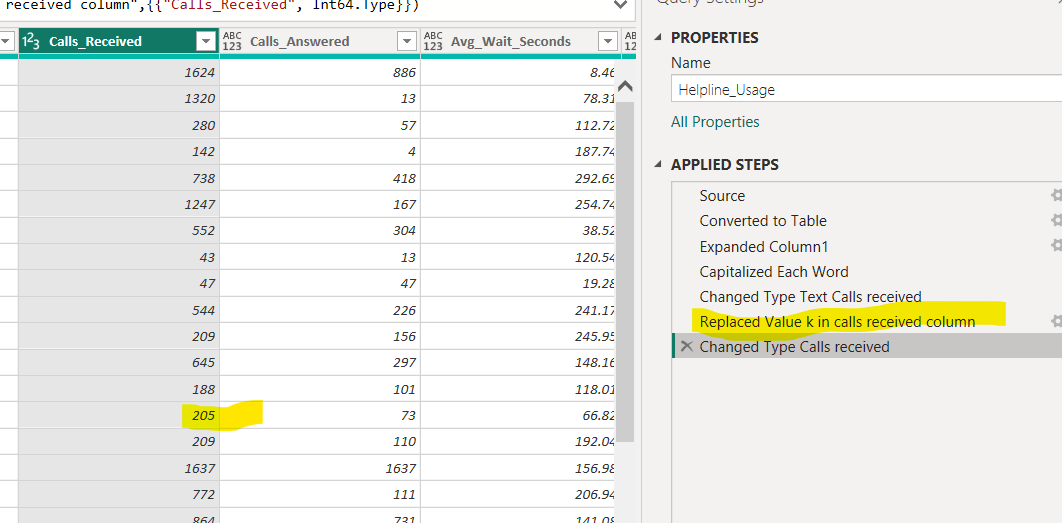
## **Helpline Usage Table:**

**Replaced letter k with empty data in Calls\_Received column.**

**Before:**

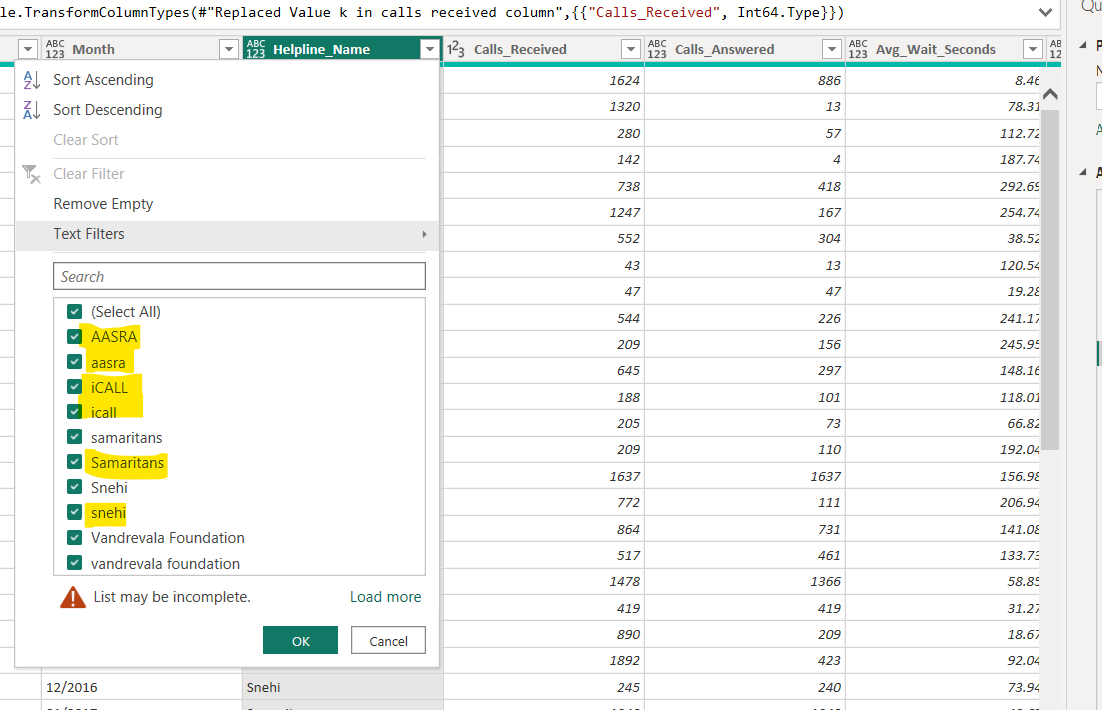
****

**After:**

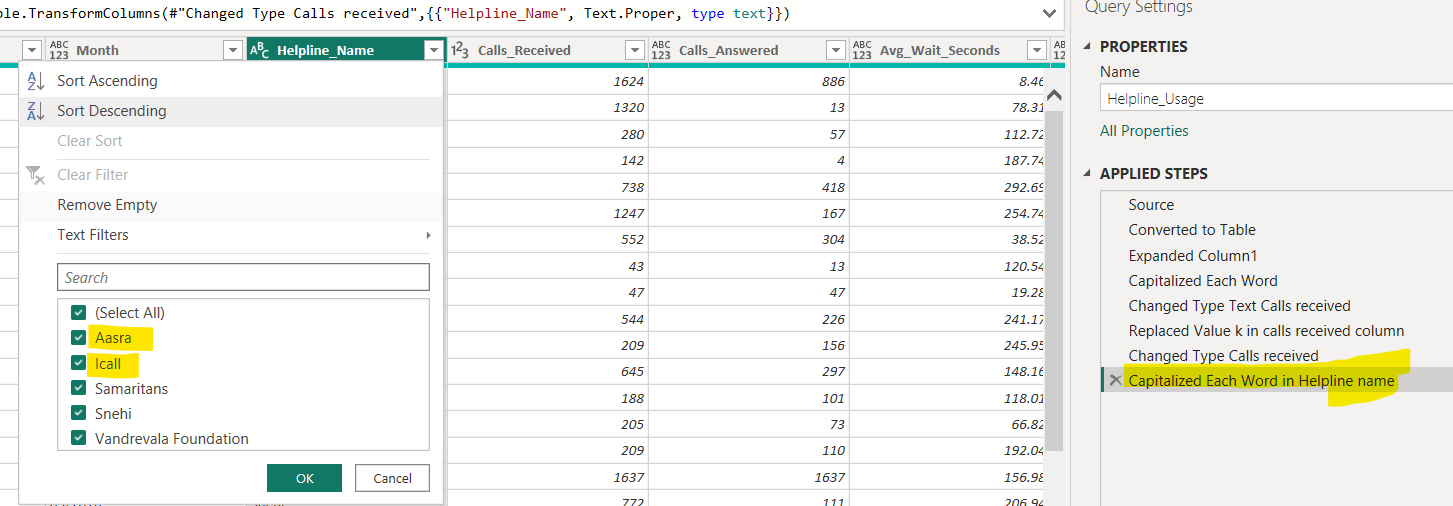
****

**Capitalized each word in Helpline name Column to remove duplicates**

**Before:**

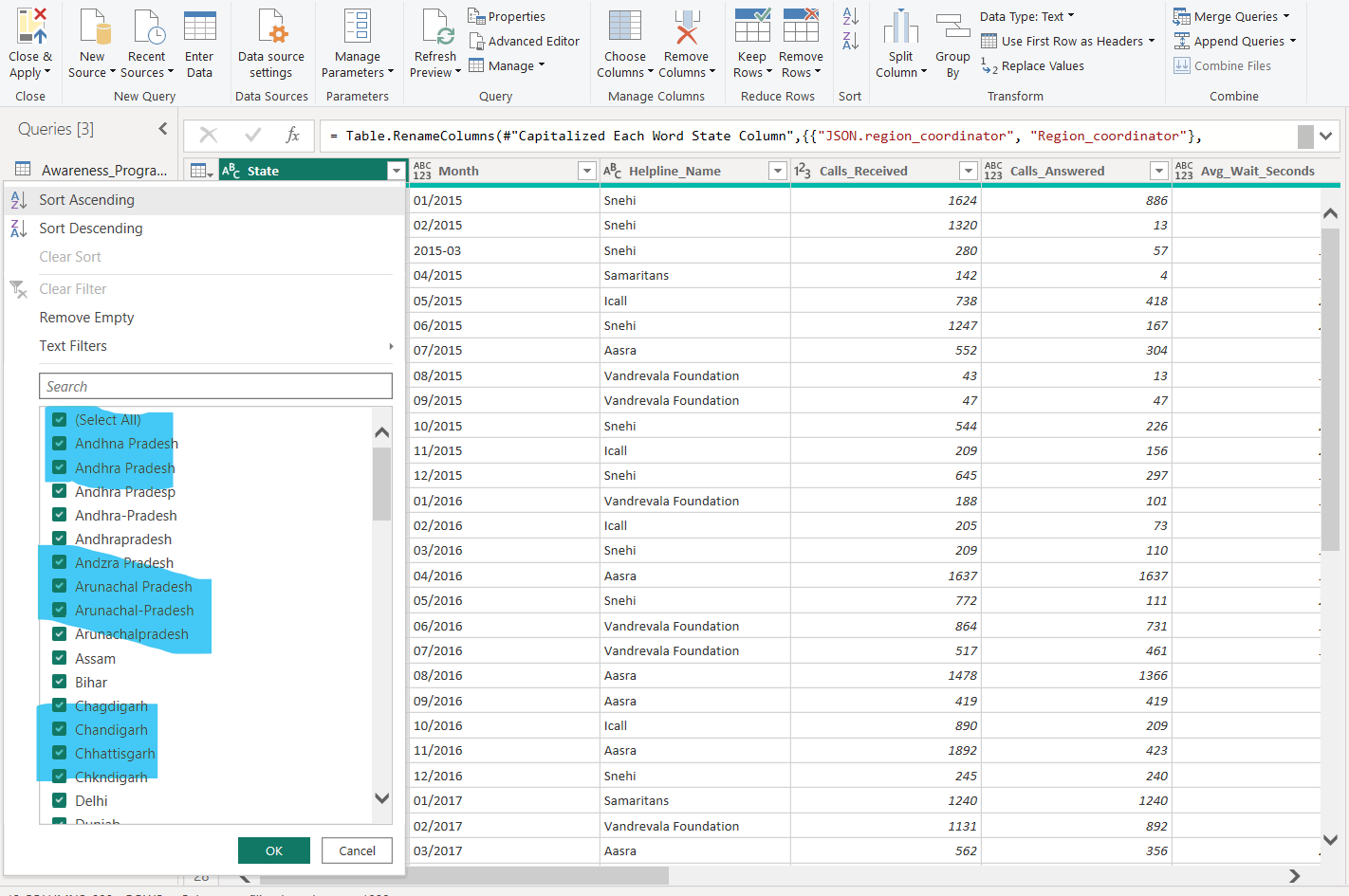
****

**After:**

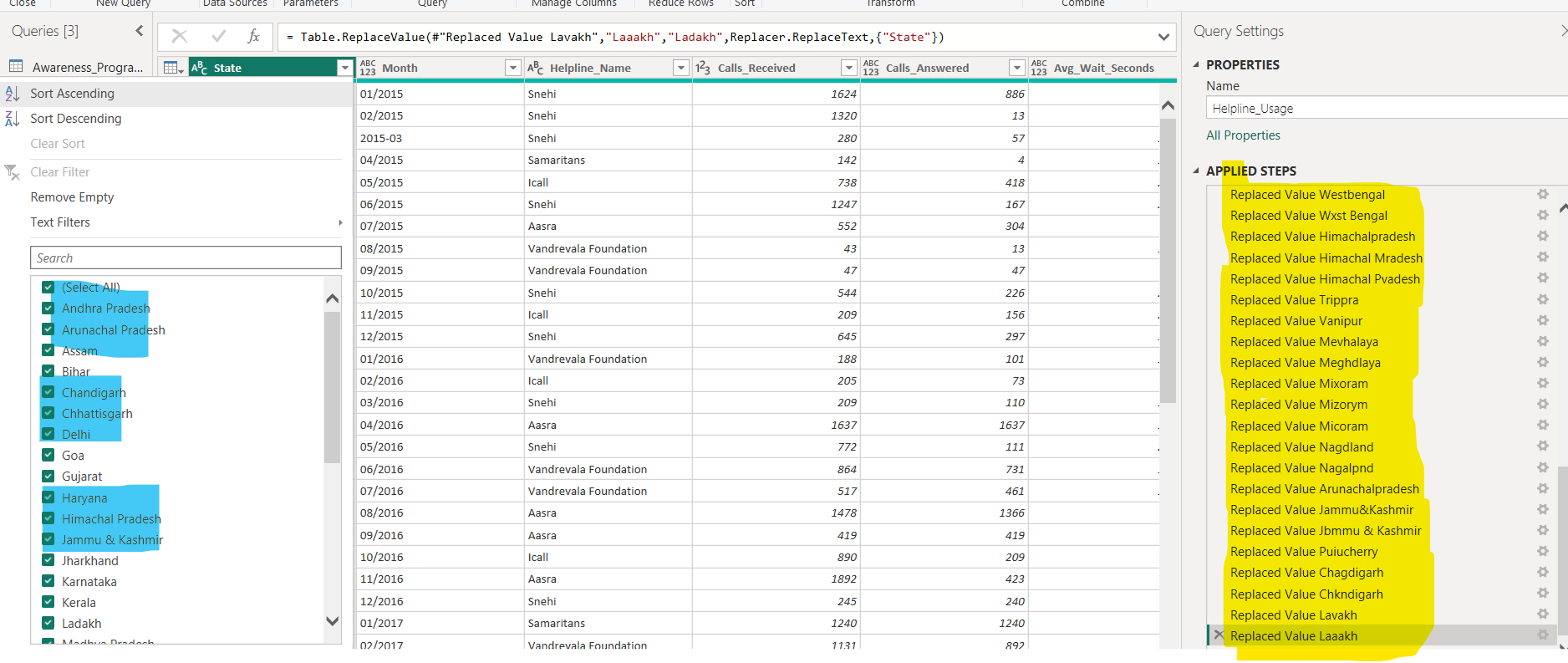
****

**Data inconsistencies cleaning in State Column:**

**Before: Could see duplicates and inconsistencies**

****

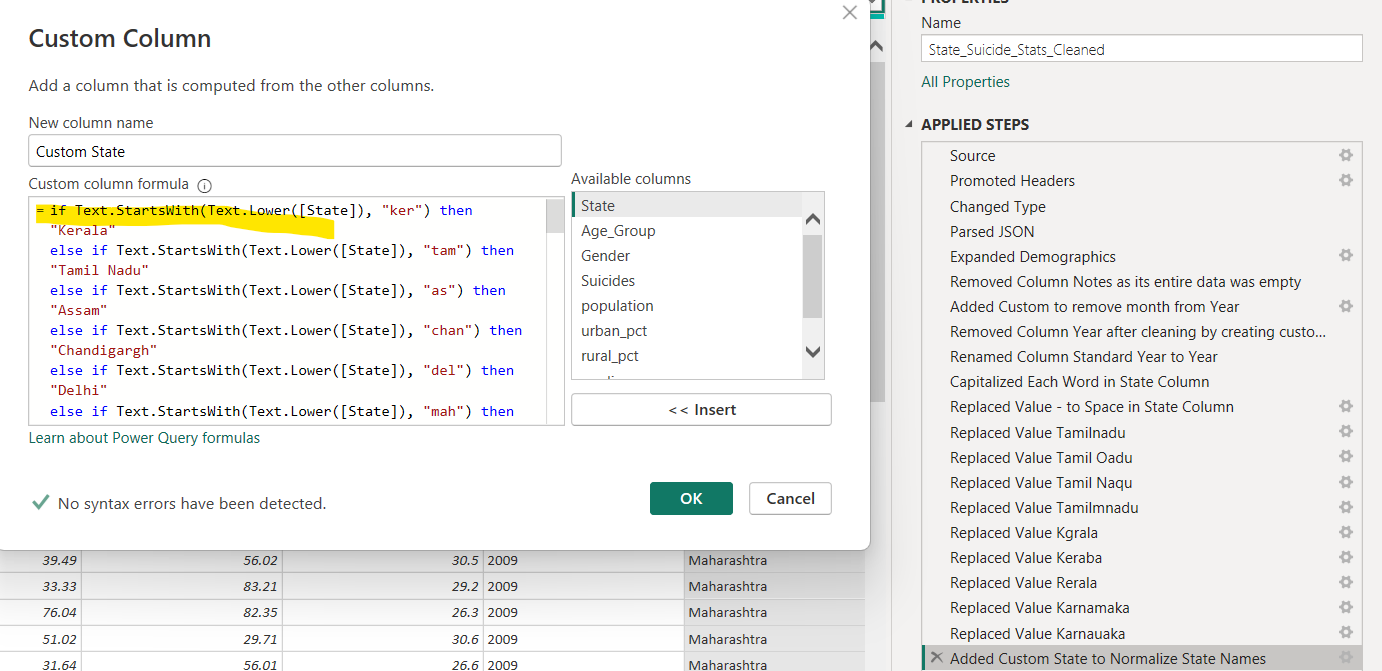
**After:**

****

## **State\_Suicide\_Stats Table:**

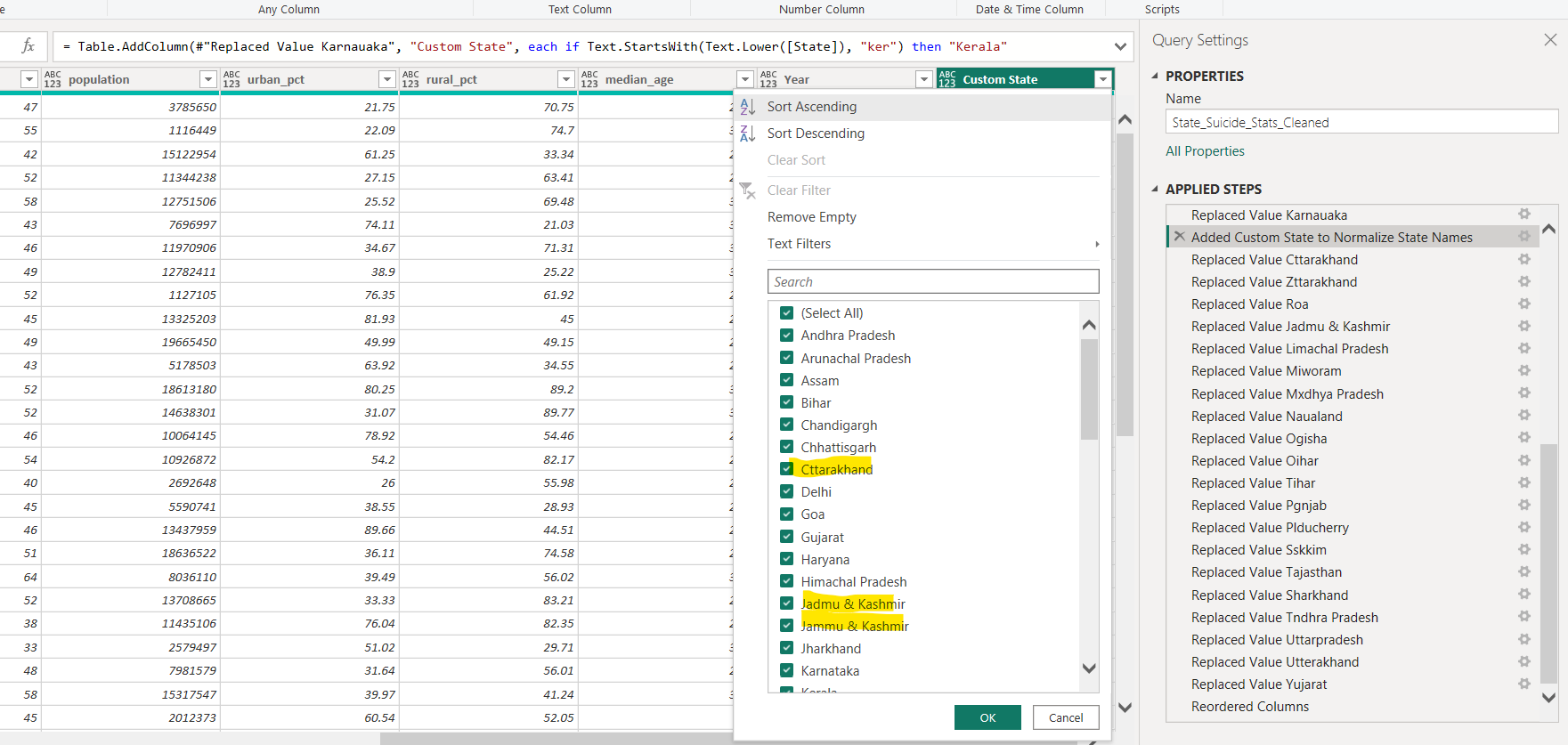
**Initially did excel cleaning single quote to double quote in Demographics Column to format JSON Structure and loaded CSV file into power BI.**

**Added a custom column to reduce multiple replace steps into one by giving text,startswith function**

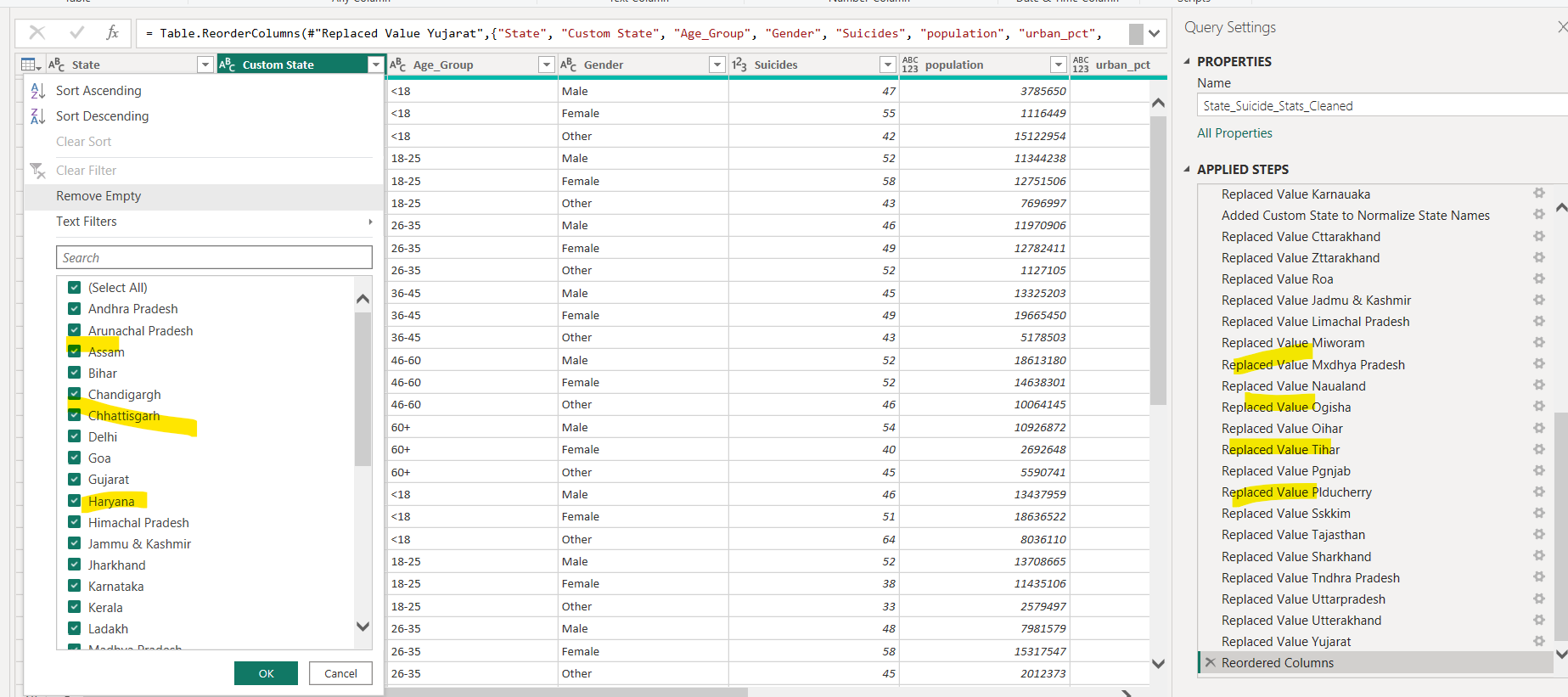
****

**Data inconsistencies cleaning in Custom State Column:**

**Before: Could see duplicates and inconsistencies**

****

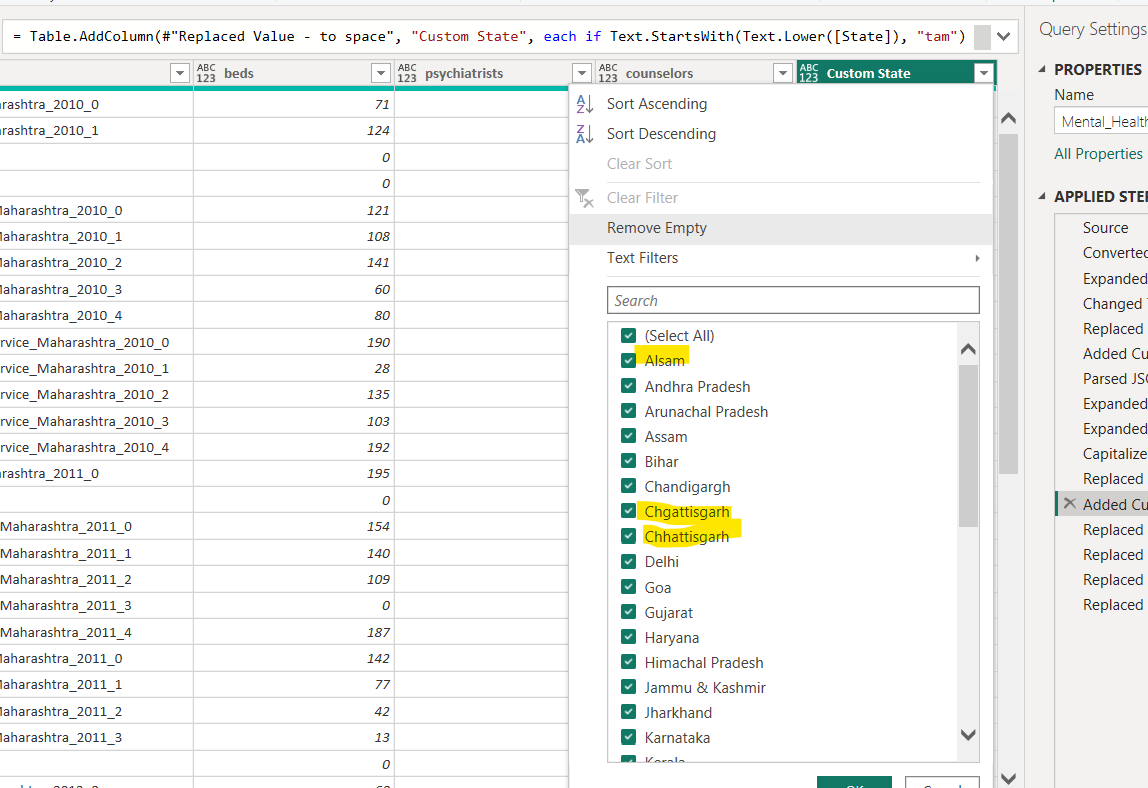
**After:**

****

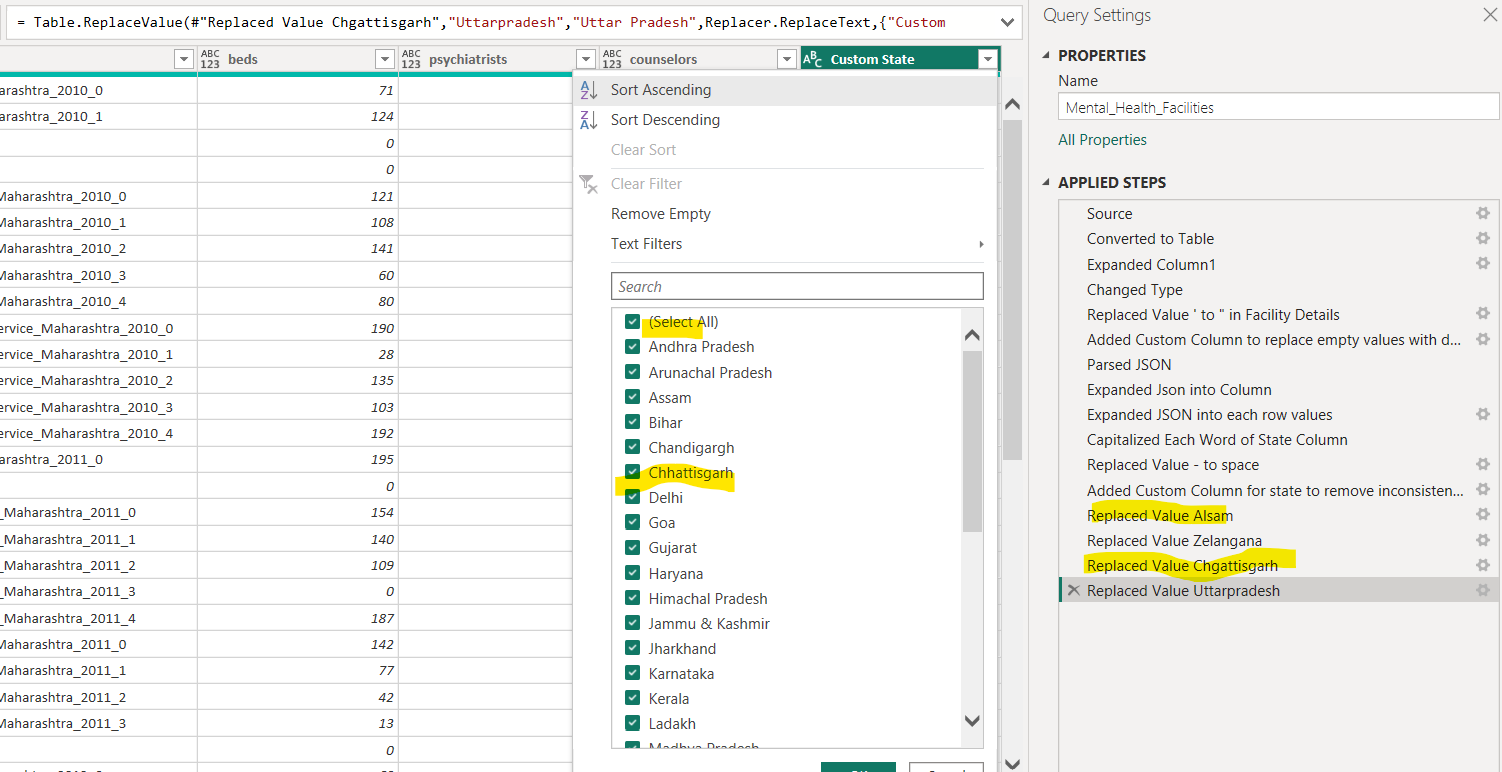
## **Mental\_Health\_Facilities Table:**

**Data inconsistencies cleaning in State Column:**

**Before: Could see duplicates and inconsistencies**

****

**After:**

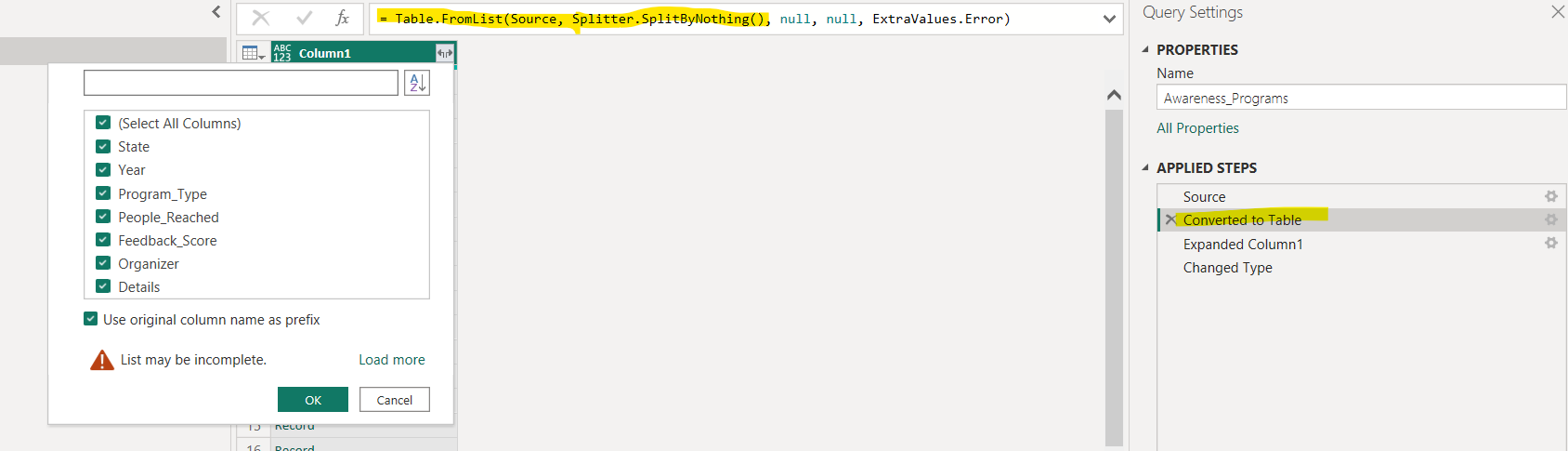
****

# **Data Transformation:**

## **Awareness program Table:**

**Converting Json data into table in power query editor in Power BI**

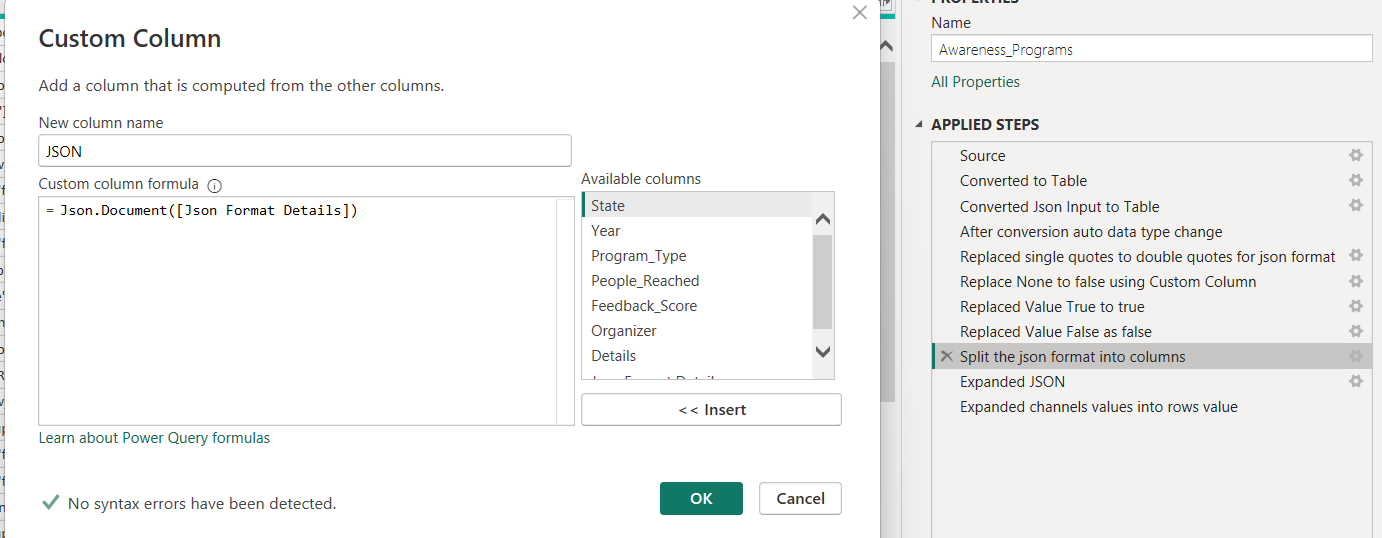
**Before:**



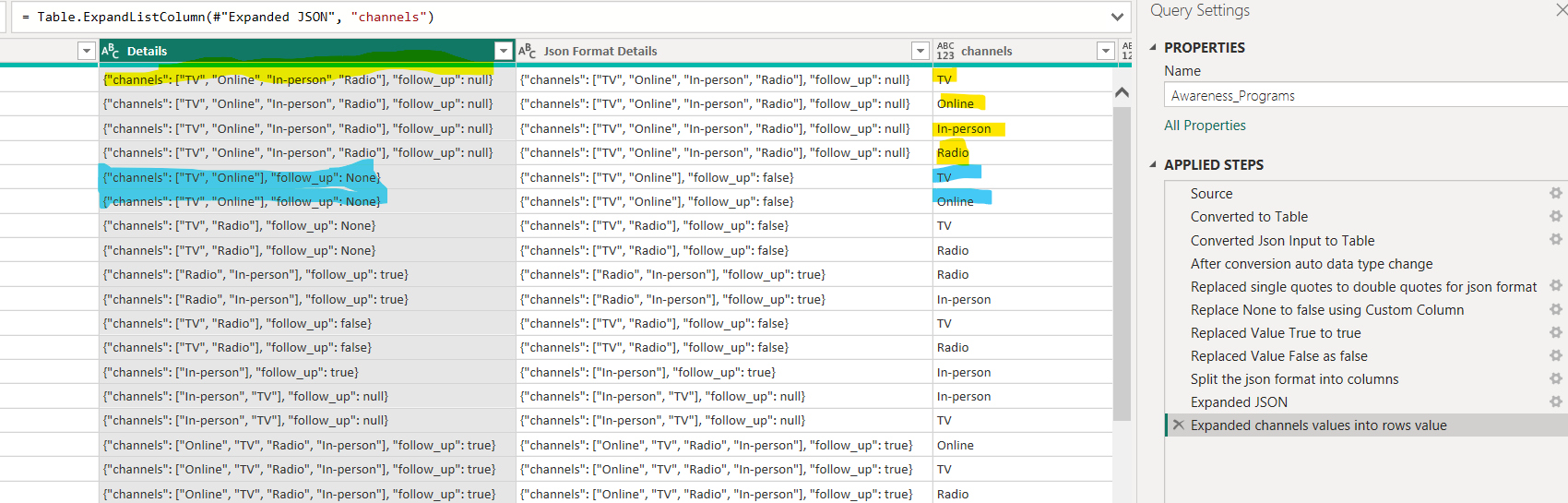
**After:**



**Split JSON data into column: Used Parse feature in Transform tab**



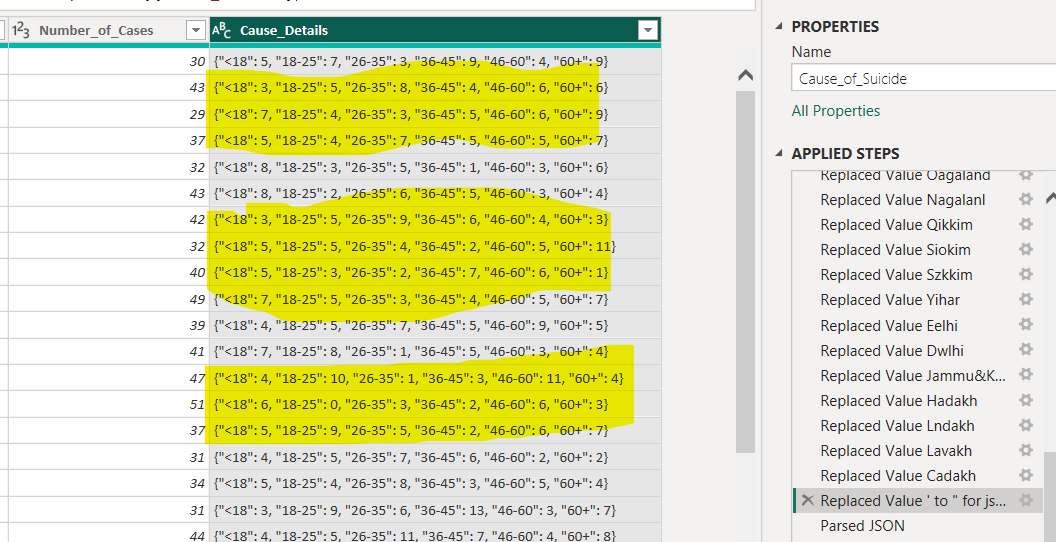
**Expanded channels multiple values into each row value**



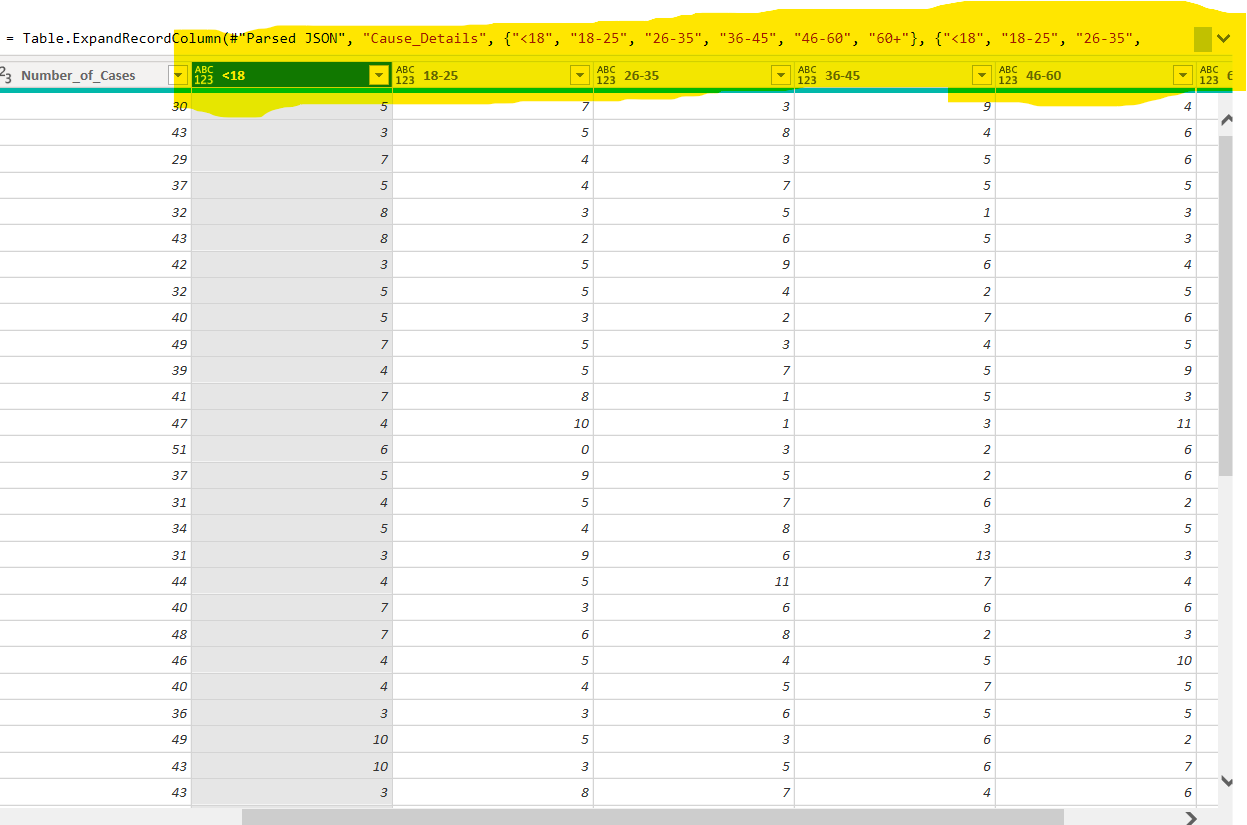
**Cause\_of\_Suicide Table:**

**Split JSON data into column: Used Parse feature in Transform tab**

**Before:**

****

**After:**

****

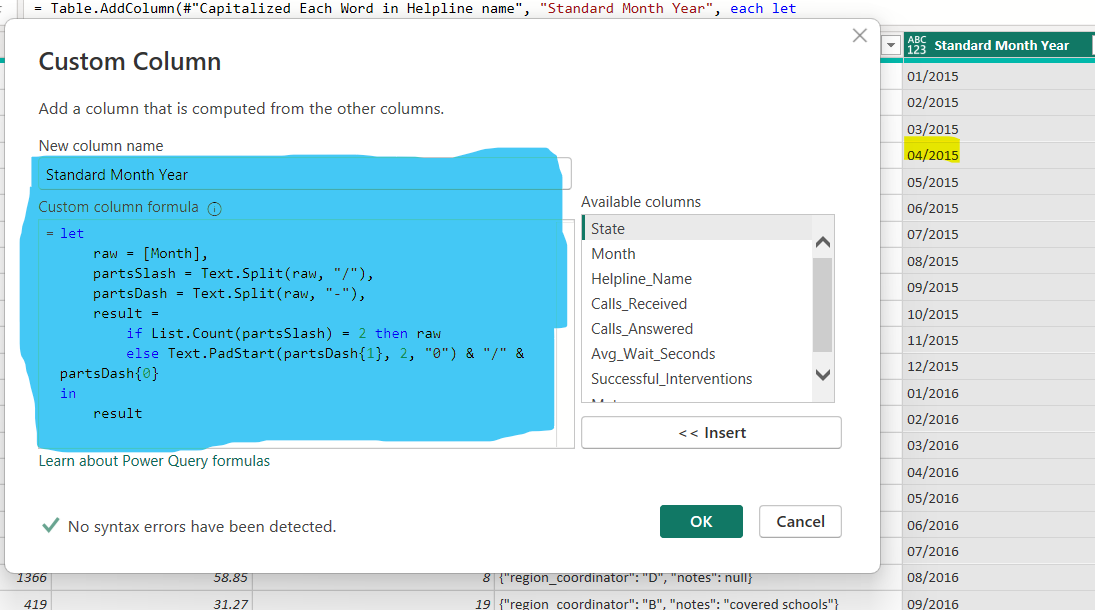
## **Helpline Usage Table:**

**Month** **column has data with two different format 01/2023 and 2022-03. Transformed it into one format 01/2023.**

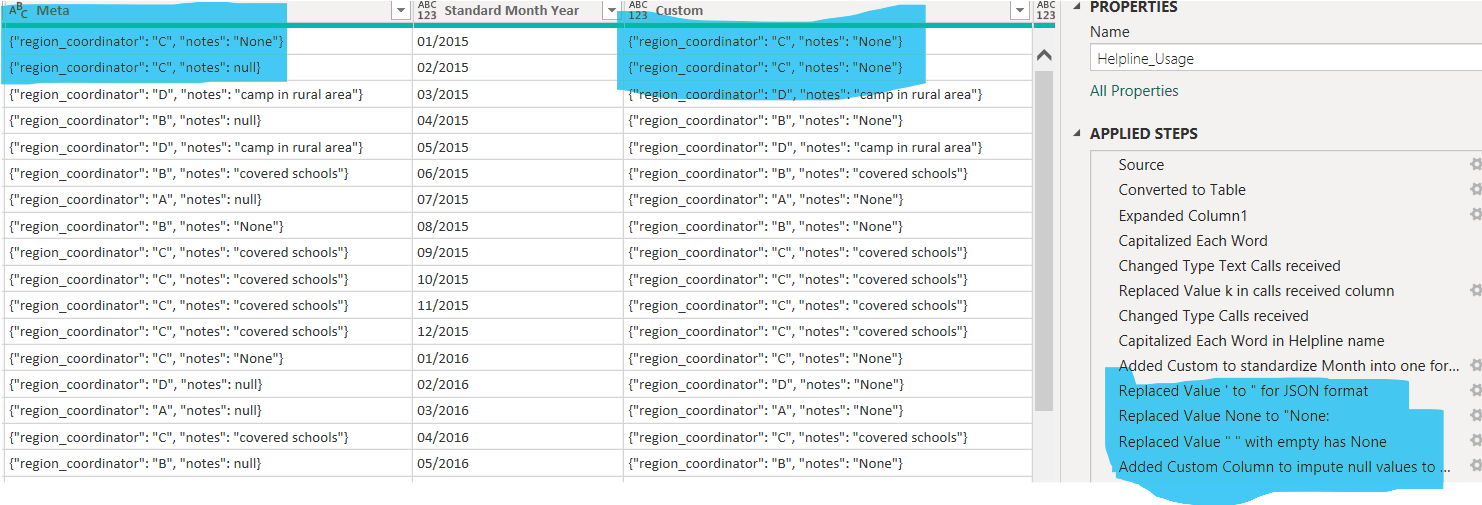
**Before:**



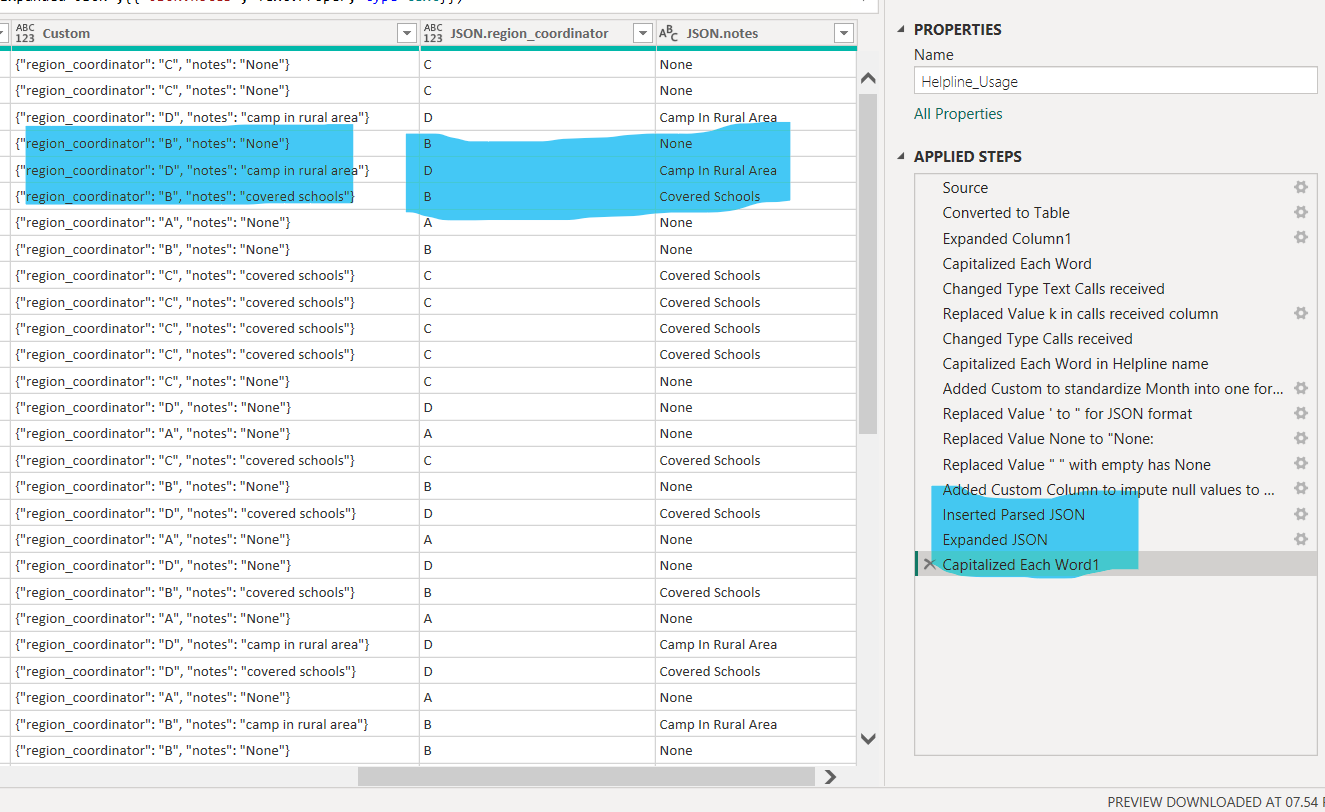
**After:**



**Corrected Inconsistencies to form a proper Json Structure and consistent data:**

****

**Split JSON data into column: Used Parse feature in Transform tab**

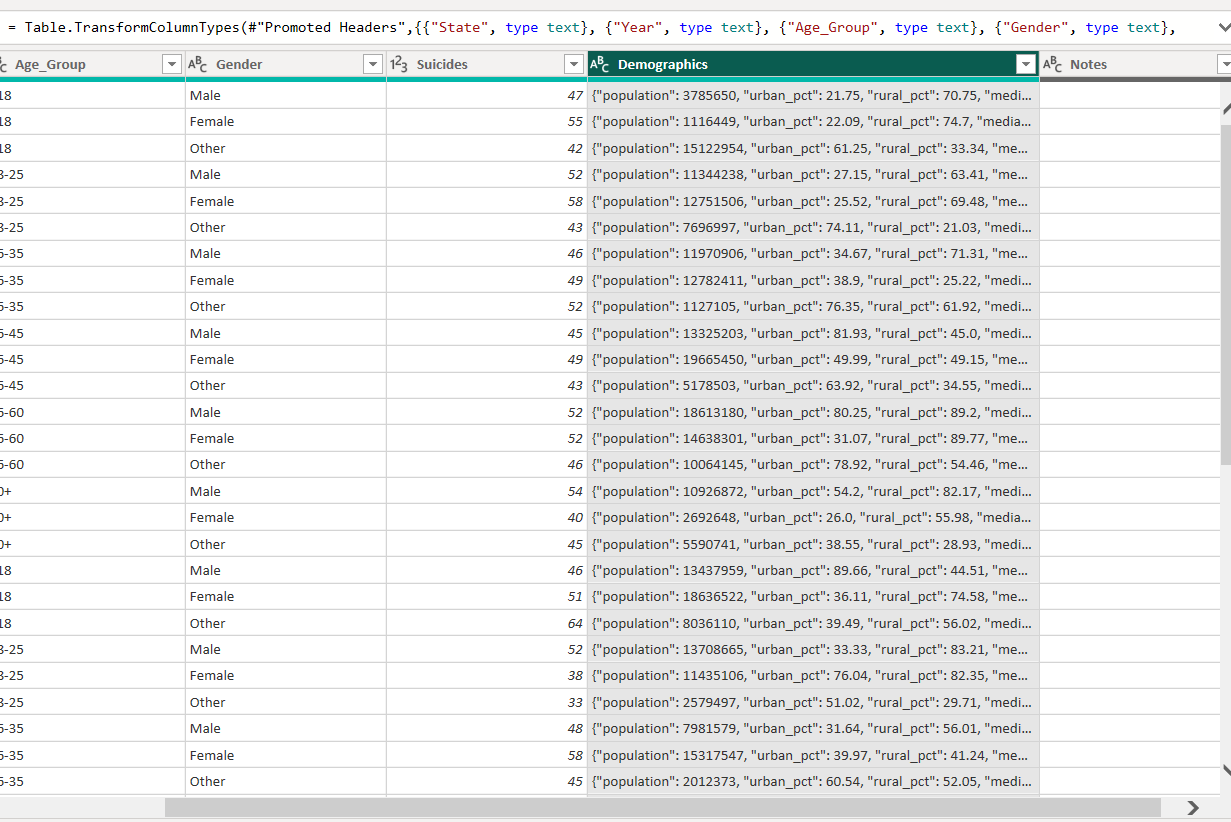
****

## **State\_Suicide\_Stats Table:**

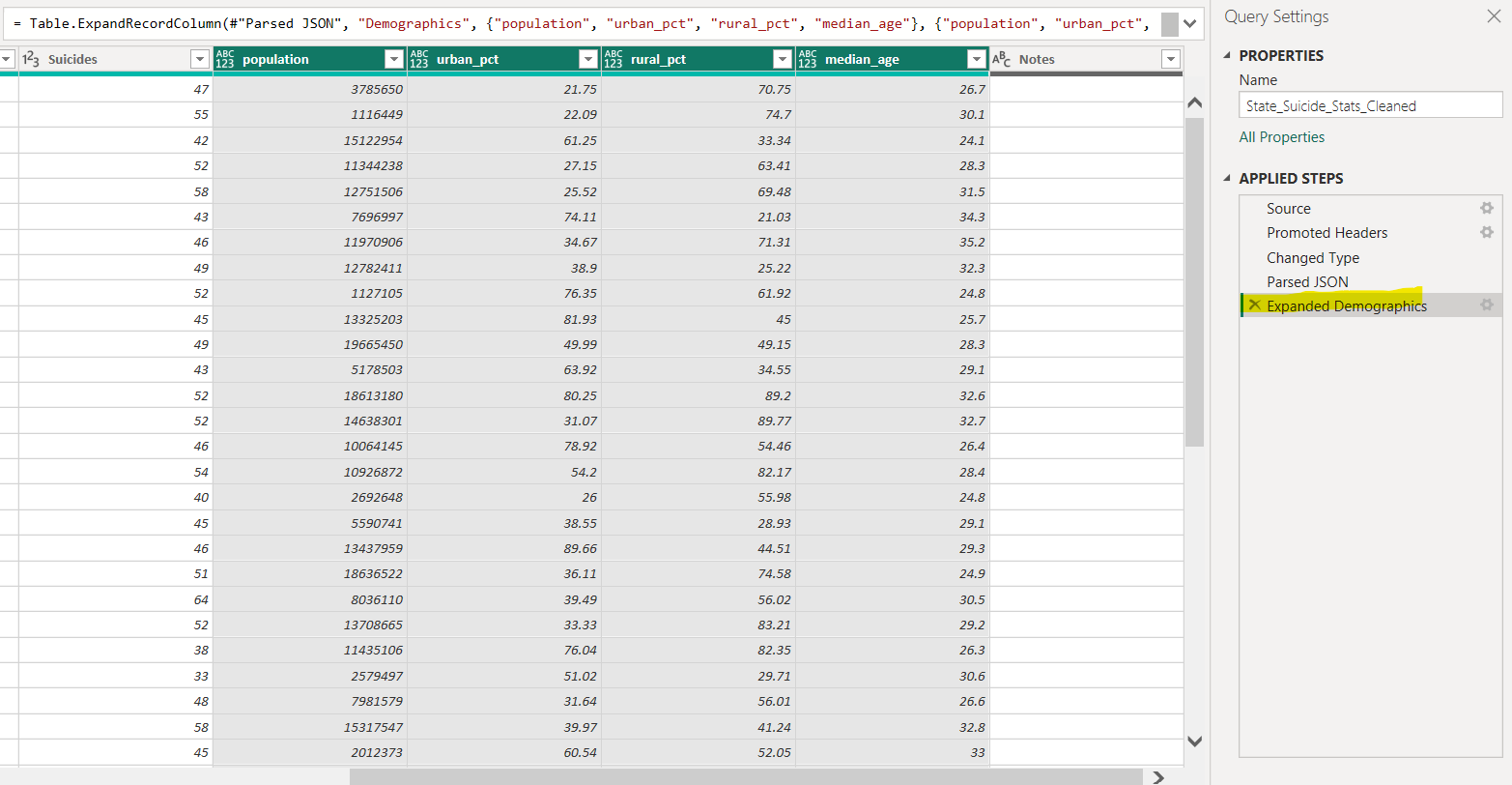
**Initially did excel cleaning single quote to double quote in Demographics Column to format JSON Structure and loaded CSV file into power BI.**

**Expanded the Json value Demographics column into Column values.**

**Before:**

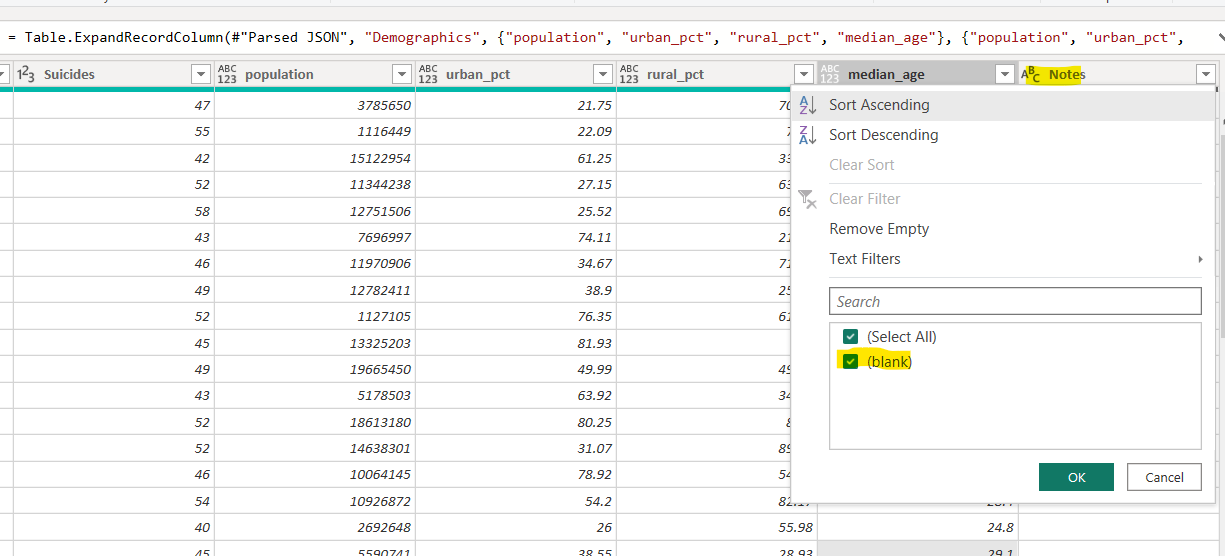
****

**After:**

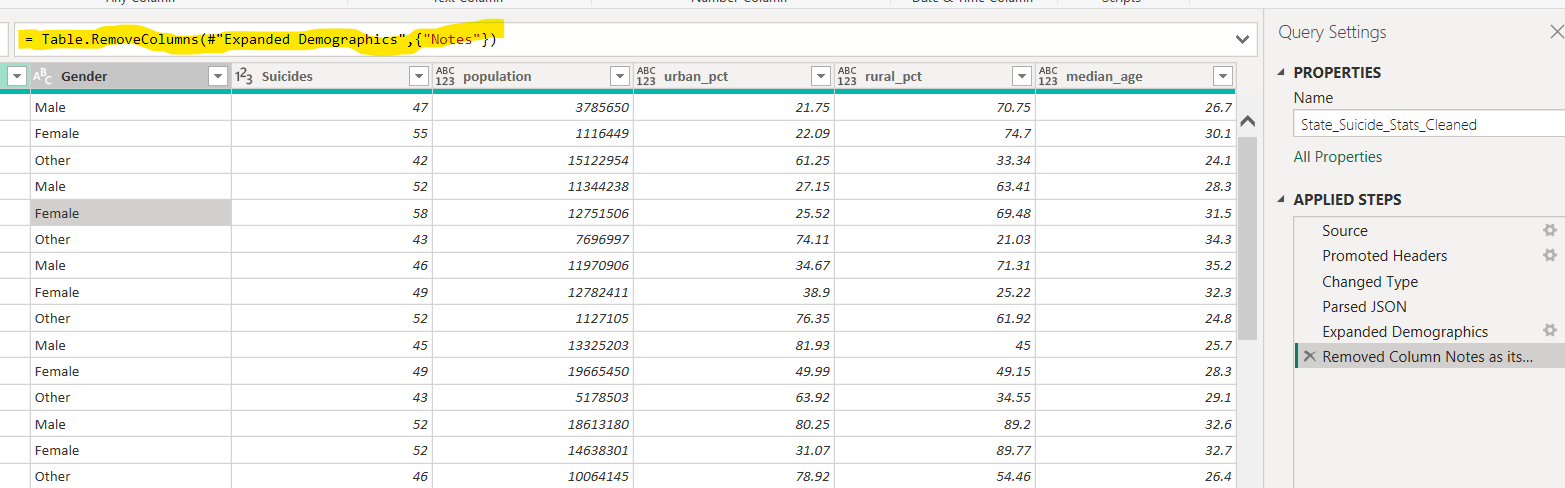
****

**Removed column Notes from Table as Notes columns entire data was empty.**

**Before:**

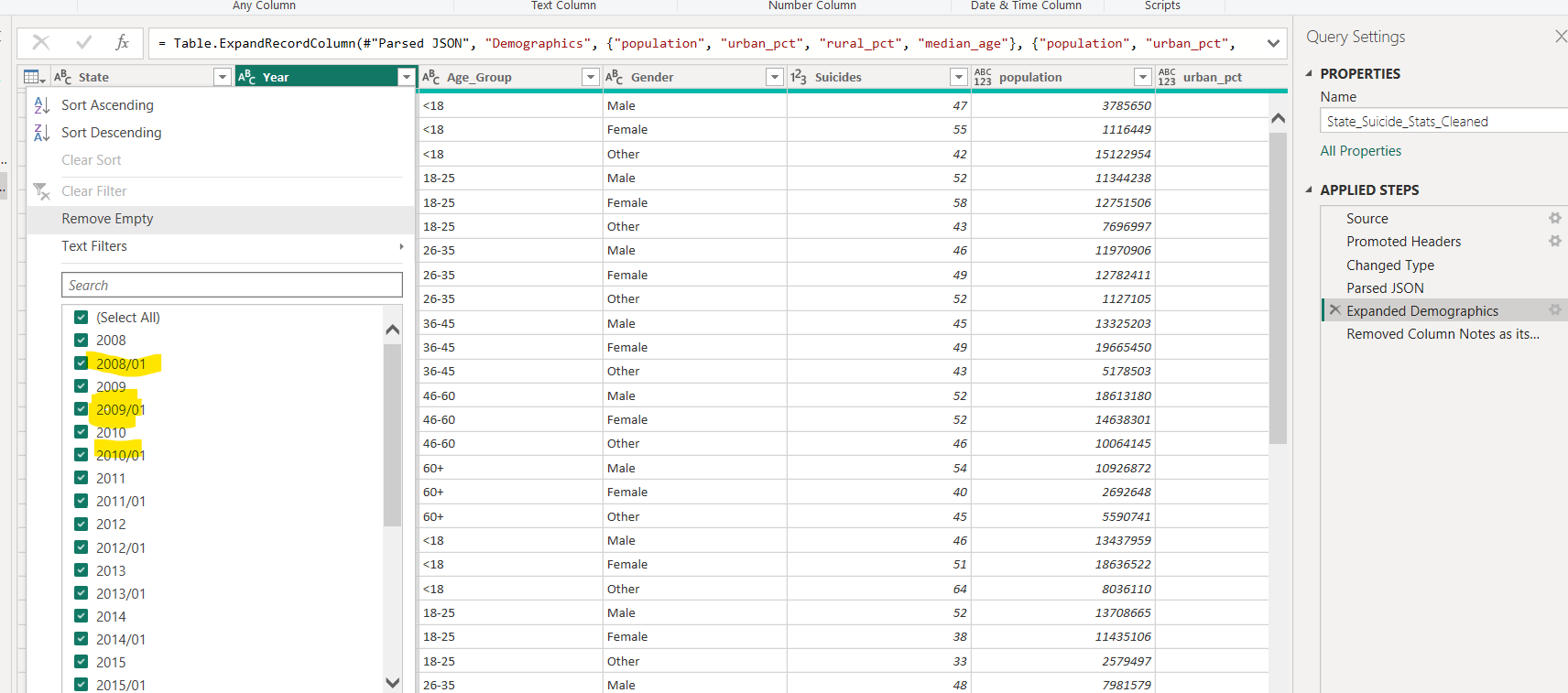
****

**After:**

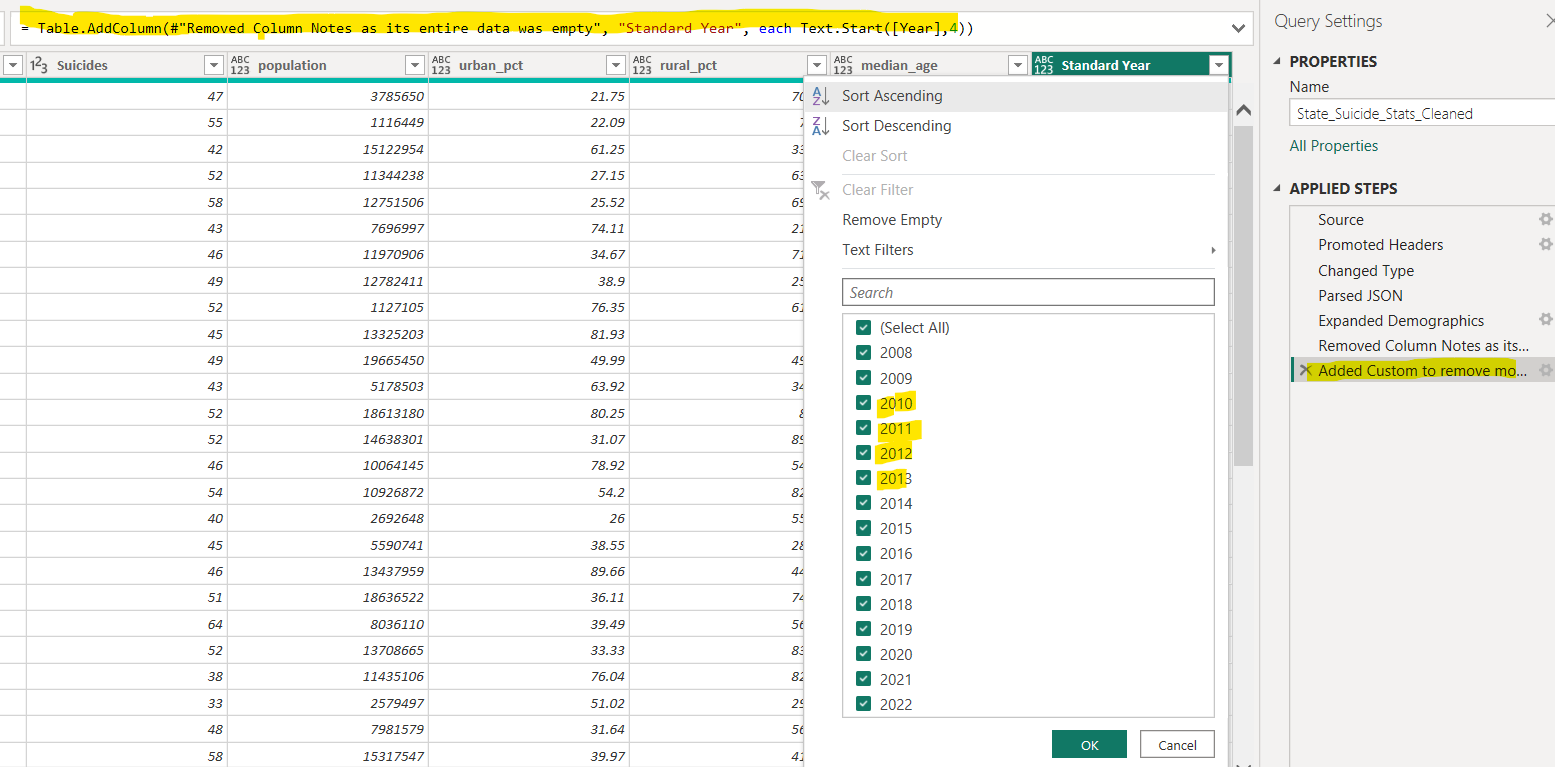
****

**Standardized year Column format as its few data had month appended so added custom column to have only year using text.start function**

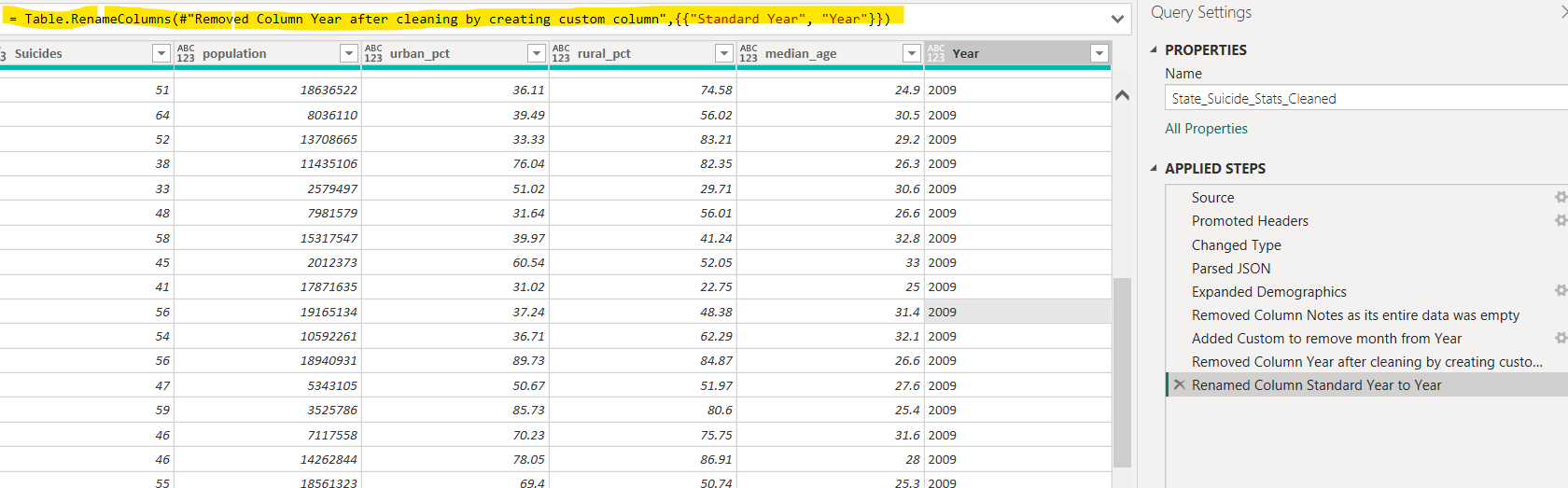
**Before:**

****

**After:**

****

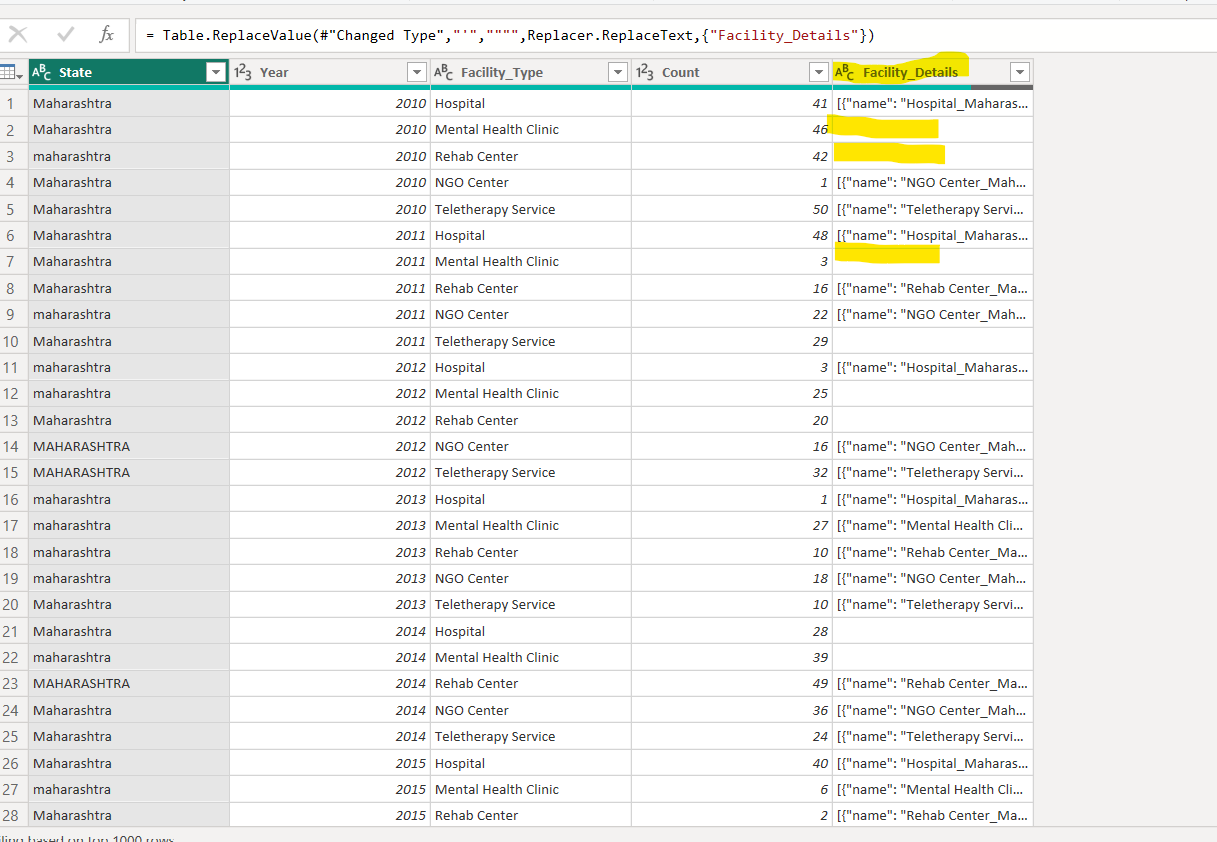
**Removed Column year after cleaning and Renamed Standard Year to Year**

****

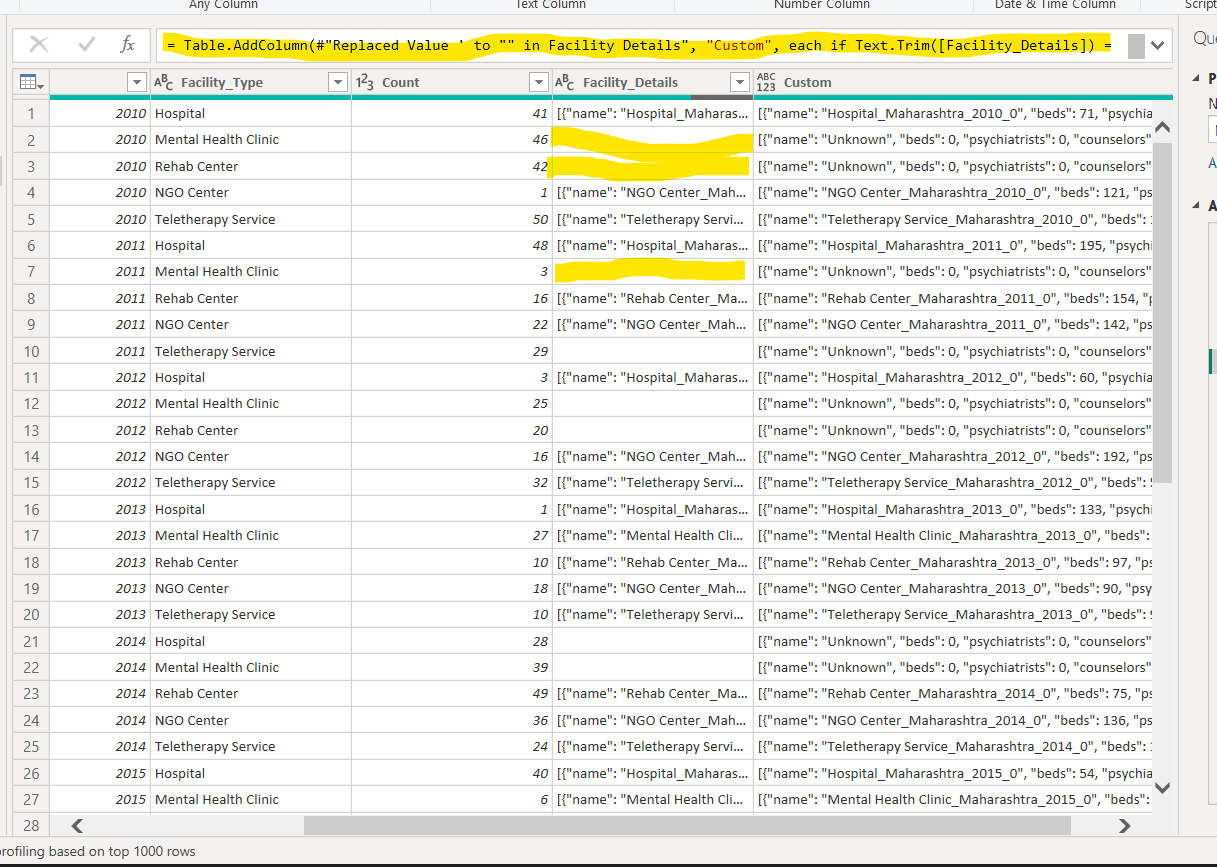
## **Mental\_Health\_Facilities Table:**

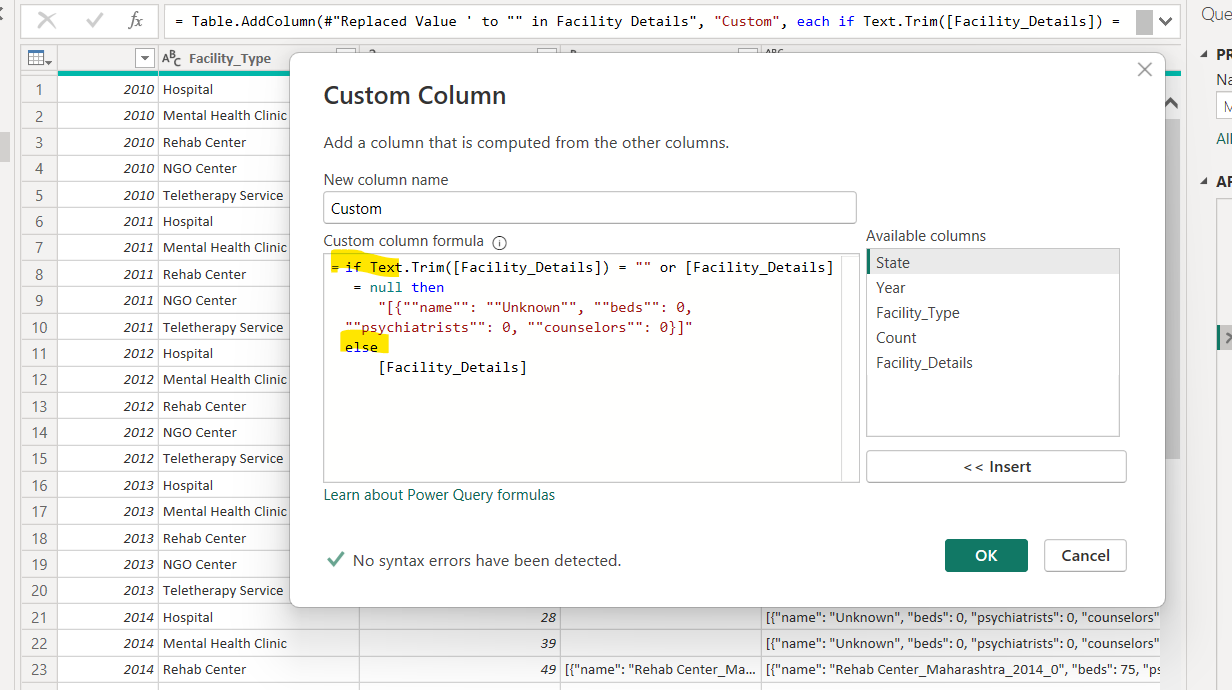
**Added Custom Column to replace empty values with default json string.**

**Before:**

****

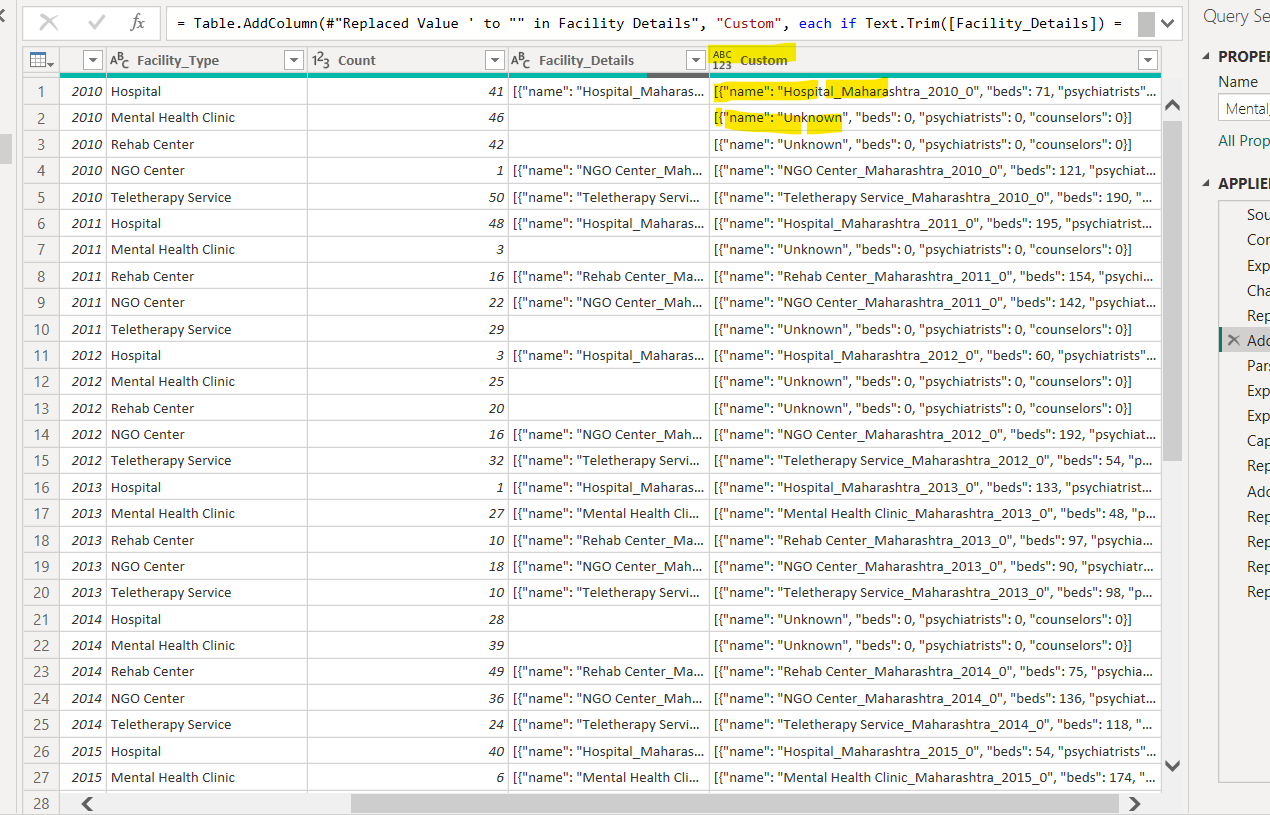
**After:**

****

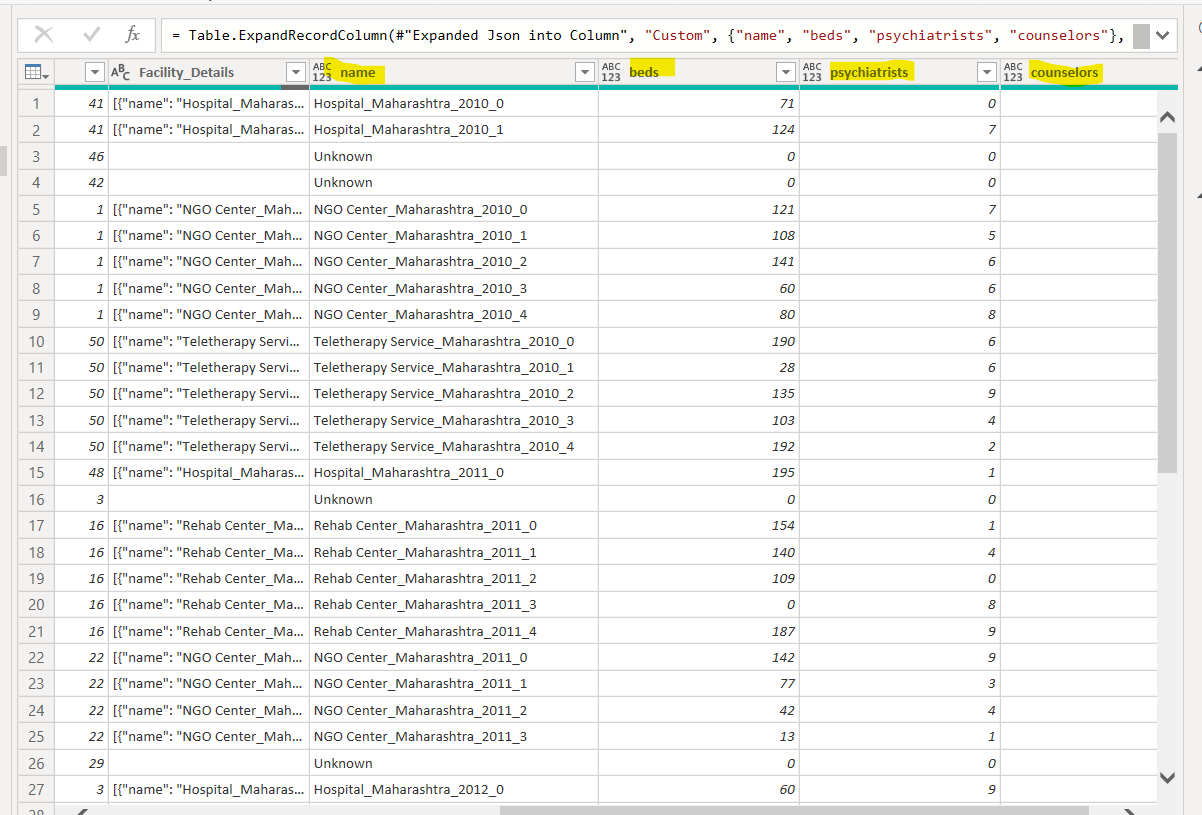
****

**Expanded the Json value Custom column into Column values and then into row values.**

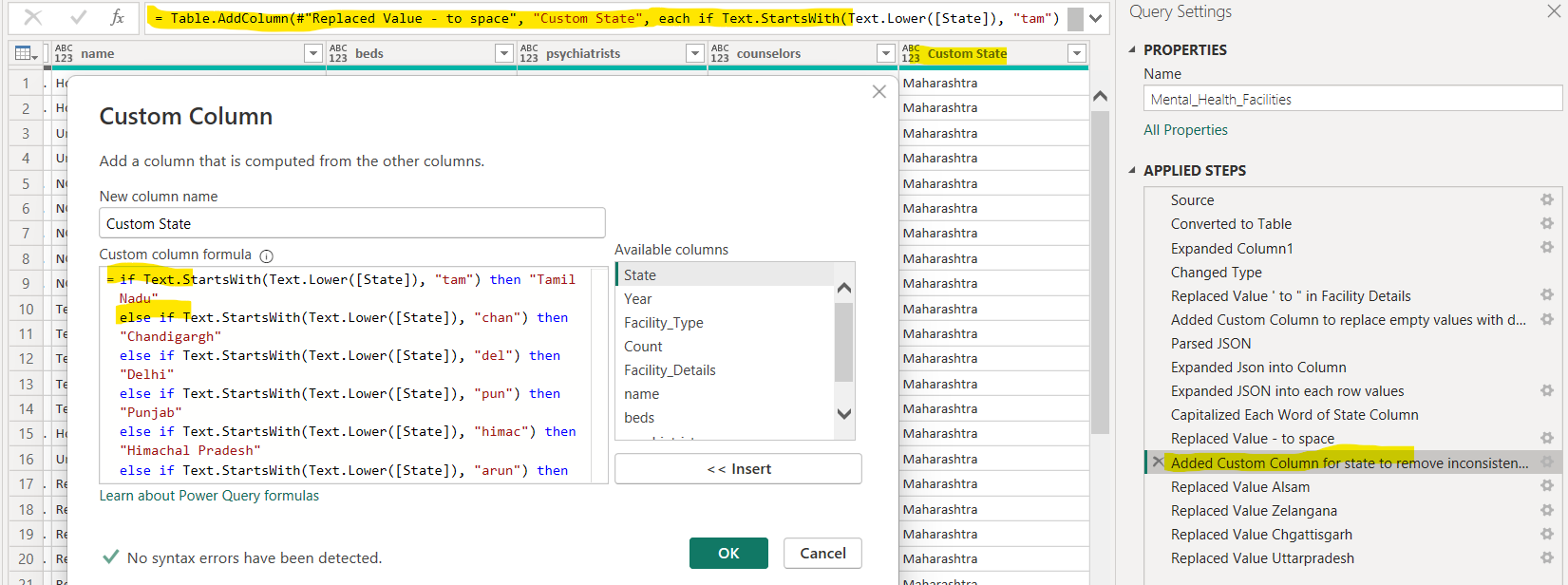
**Before:**



**After:**

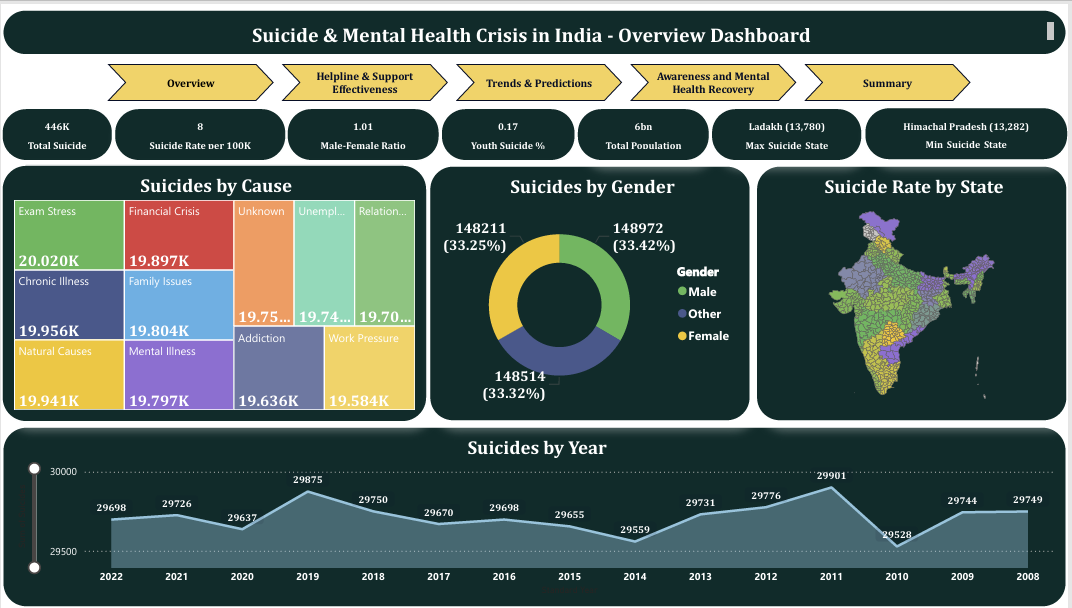
****

**Added Custom Column for state to remove inconsistencies**

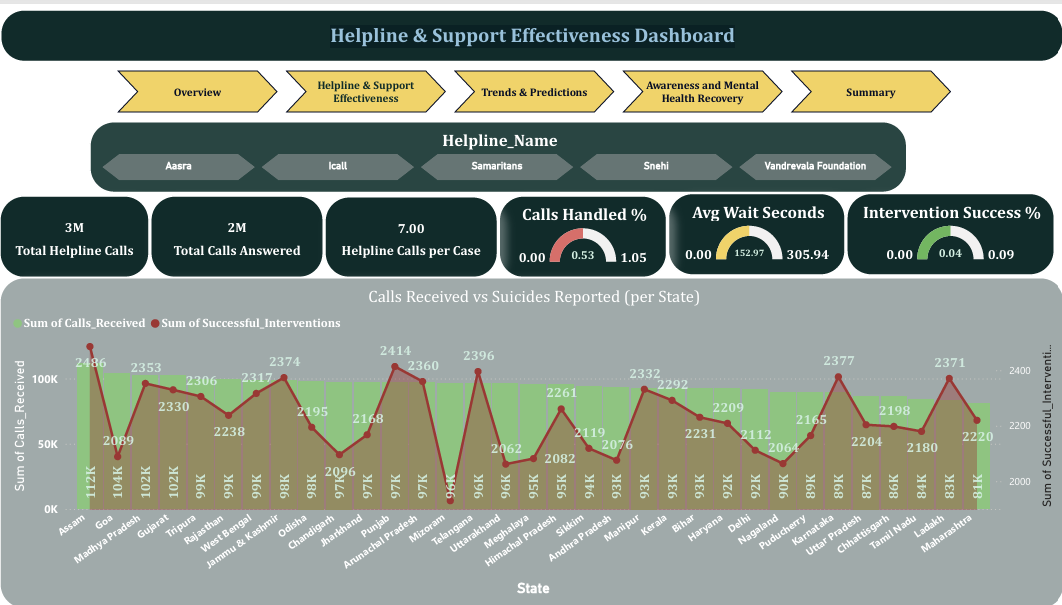
****

# **Dashboards Visualization**

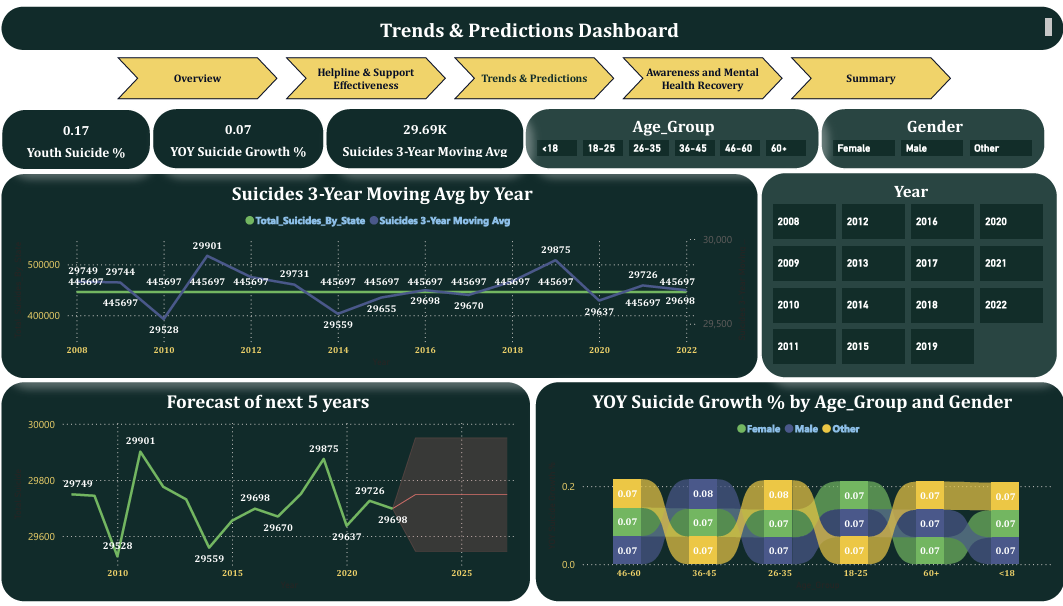
**Overview Report:**



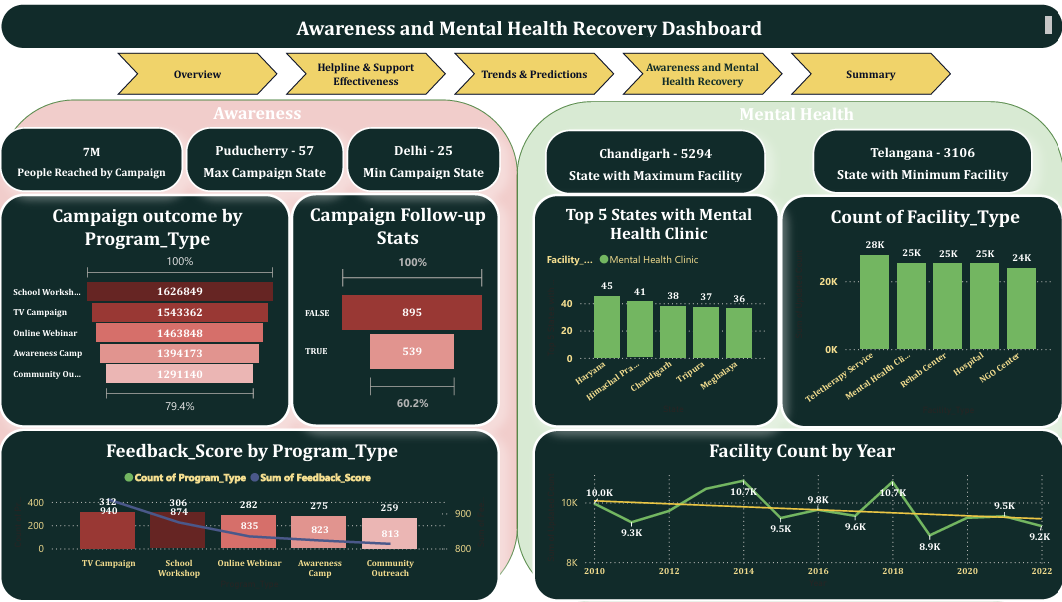
**Helpline & Support Effectiveness Dashboard:**

****

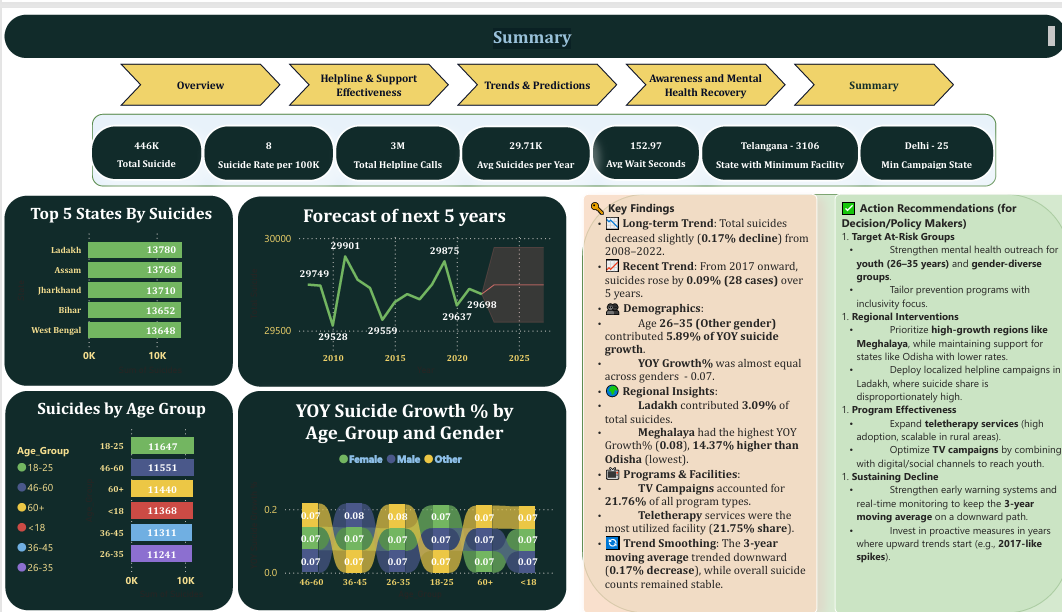
**Trends & Predictions Dashboard:**



**Awareness and Mental Health Recovery Dashboard:**



**Summary:**



# **Findings / Results**

- Total suicides trended down, resulting in a 0.17% decrease between 2008 and 2022.  
- From 2017, suicides started trending up, rising by 0.09% in 5 years.  
- Age group 26–35 in gender 'Other' accounted for 5.89% of YOY suicide growth %.  
- Ladakh accounted for 3.09% of total suicides.  
- Meghalaya had the highest YOY Suicide Growth % (0.08), 14.37% higher than Odisha (lowest).  
- TV Campaign accounted for 21.76% of all program types.  
- Teletherapy service accounted for the highest facility usage at 21.75%.

# **Insights & Discussion**

- Youth suicide due to academic pressure is rising in Tamil Nadu.  
- Financial distress-related suicides are high in Maharashtra and Kerala.  
- Rural access to helplines remains low despite improvement in call response times.  
- Gender 'Other' shows highest proportional YOY growth, requiring targeted support.

# **Recommendations**

- Strengthen rural helplines with improved outreach.  
- Expand financial counseling and support programs.  
- Implement stress management and academic support initiatives for students.  
- Increase awareness campaigns through media and community networks.  
- Enhance training for helpline staff to improve interventions.

# **Limitations**

- Data gaps exist in reporting for certain categories (e.g., 'Unknown').  
- Helpline records may underrepresent rural distress cases.  
- Forecast accuracy depends on historical data reliability.

# **Conclusion**

The analysis shows a mixed trend in suicides with improvements in some areas but challenges in others. By enhancing helpline coverage, addressing financial and academic pressures, and increasing awareness, suicide prevention efforts can be strengthened in the coming years.

# **Appendices**

- List of DAX Formula Used in measures and calculated and custom columns in project

**Awareness\_Programs**

**Campaign outcome =** CALCULATE(SUM(Awareness\_Programs[People\_Reached]),ALLEXCEPT(Awareness\_Programs,Awareness\_Programs[Program\_Type]))

**State with Maximum Campaign =**

VAR SummaryTable = SUMMARIZE ('Awareness\_Programs','Awareness\_Programs'[State],

        "ProgramTotal", COUNTROWS ( 'Awareness\_Programs' ))

VAR MaxPrograms =    MAXX ( SummaryTable, [ProgramTotal] )

VAR TopStates =    FILTER ( SummaryTable, [ProgramTotal] = MaxPrograms )

VAR StateList =    CONCATENATEX ( TopStates, 'Awareness\_Programs'[State], ", " )

RETURN

    StateList & " - " & MaxPrograms

**State with Minimum Campaign =**

VAR SummaryTable = SUMMARIZE ('Awareness\_Programs', 'Awareness\_Programs'[State],

        "ProgramTotal", COUNTROWS ( 'Awareness\_Programs' ))

VAR MaxPrograms = MINX ( SummaryTable, [ProgramTotal] )

VAR TopStates = FILTER ( SummaryTable, [ProgramTotal] = MaxPrograms )

VAR StateList =  CONCATENATEX ( TopStates, 'Awareness\_Programs'[State], ", " )

RETURN

    StateList & " - " & MaxPrograms

**Cause\_of\_Suicide**

**Avg Suicides per Year =** AVERAGEX(VALUES(Cause\_of\_Suicide[Year]),[Total Suicides])

**Helpline\_Usage**

**Avg Response Time =** AVERAGE(Helpline\_Usage[Avg\_Wait\_Seconds])

**Calls Handled % =** DIVIDE([Total Calls Answered],[Total Helpline Calls])

**Helpline Calls per Case =** DIVIDE(SUM(Helpline\_Usage[Calls\_Received]), [Total Suicide])

**Helpline Effectiveness Index =**

(([Calls Handled %] \* 0.4) + ([Intervention Success %] \* 0.4) +(1 - [Avg Response Time] / MAX(Helpline\_Usage[Avg\_Wait\_Seconds])) \* 0.2) \* 100

**Intervention Success % =**

DIVIDE(SUM(Helpline\_Usage[Successful\_Interventions]),

SUM(Helpline\_Usage[Calls\_Answered]), 0)

**Total Calls Answered =** SUM(Helpline\_Usage[Calls\_Answered])

**Total Helpline Calls =** SUM(Helpline\_Usage[Calls\_Received])

**Mental\_Health\_Facilities**

**State with Maximum Facility =**

VAR SummaryTable =    SUMMARIZE ( 'Mental\_Health\_Facilities','Mental\_Health\_Facilities'[State],

        "ProgramTotal", SUM( 'Mental\_Health\_Facilities'[Updated Count] ))

VAR MaxPrograms =  MAXX ( SummaryTable, [ProgramTotal] )

VAR TopStates =  FILTER ( SummaryTable, [ProgramTotal] = MaxPrograms )

VAR StateList =    CONCATENATEX ( TopStates, 'Mental\_Health\_Facilities'[State], ", " )

RETURN

    StateList & " - " & MaxPrograms

**State with Minimum Facility =**

VAR SummaryTable =    SUMMARIZE ( 'Mental\_Health\_Facilities','Mental\_Health\_Facilities'[State],

        "ProgramTotal", SUM( 'Mental\_Health\_Facilities'[Updated Count] ))

VAR MaxPrograms =  MINX ( SummaryTable, [ProgramTotal] )

VAR TopStates =  FILTER ( SummaryTable, [ProgramTotal] = MaxPrograms )

VAR StateList =    CONCATENATEX ( TopStates, 'Mental\_Health\_Facilities'[State], ", " )

RETURN

    StateList & " - " & MaxPrograms

**State\_Suicide\_Stats\_Cleaned:**

**Avg Suicides per Year1** = AVERAGEX(VALUES(State\_Suicide\_Stats\_Cleaned[Standard Year]),[Total Suicide])

**Male-Female Ratio =**

DIVIDE(CALCULATE(SUM(State\_Suicide\_Stats\_Cleaned[Suicides]), State\_Suicide\_Stats\_Cleaned[Gender] = "Male"),

CALCULATE(SUM(State\_Suicide\_Stats\_Cleaned[Suicides]), State\_Suicide\_Stats\_Cleaned[Gender] = "Female"))

**Max\_Suicide\_Count =**

CALCULATE(

    MAXX(VALUES(State\_Suicide\_Stats\_Cleaned[State]), [Total\_Suicides\_By\_State])

)

**Max\_Suicide\_State =**

VAR MaxCount = CALCULATE(

    MAXX(VALUES(State\_Suicide\_Stats\_Cleaned[State]), [Total\_Suicides\_By\_State])

)

VAR States = CONCATENATEX(

    FILTER(

        VALUES(State\_Suicide\_Stats\_Cleaned[State]),

        [Total\_Suicides\_By\_State] = MaxCount

    ),

    State\_Suicide\_Stats\_Cleaned[State],

    ", "

)

RETURN

States & " (" & FORMAT(MaxCount, "#,##0") & ")"

**Min\_Suicide\_State =**

VAR MinCount = CALCULATE(    MINX(VALUES(State\_Suicide\_Stats\_Cleaned[State]), [Total\_Suicides\_By\_State]))

VAR States = CONCATENATEX( FILTER(VALUES(State\_Suicide\_Stats\_Cleaned[State]),

        [Total\_Suicides\_By\_State] = MinCount ), State\_Suicide\_Stats\_Cleaned[State], ", ")

RETURN

States & " (" & FORMAT(MinCount, "#,##0") & ")"

**National Avg Suicide Rate =** AVERAGEX(ALL(State\_Suicide\_Stats\_Cleaned[State]), [Suicide Rate])

**Suicide Rate =** DIVIDE([Total Suicide],State\_Suicide\_Stats\_Cleaned[sum of population])\*100000

**sum of population =** CALCULATE(SUM(State\_Suicide\_Stats\_Cleaned[population]),VALUE(State\_Suicide\_Stats\_Cleaned[Standard Year])=2022)

**Total\_Suicides\_By\_State =**

CALCULATE( SUM(State\_Suicide\_Stats\_Cleaned[Suicides]),

    ALLEXCEPT(State\_Suicide\_Stats\_Cleaned, State\_Suicide\_Stats\_Cleaned[State]))

**Youth Suicide % =**

DIVIDE(CALCULATE(SUM(State\_Suicide\_Stats\_Cleaned[Suicides]), State\_Suicide\_Stats\_Cleaned[Age\_Group] IN {"18-25"}),[Total Suicide])

**YOY Suicide Growth % =**

DIVIDE( [Total Suicide] - CALCULATE([Total Suicide], DATEADD(State\_Suicide\_Stats\_Cleaned[Year], -1, YEAR)),

    CALCULATE([Total Suicide], DATEADD(State\_Suicide\_Stats\_Cleaned[Year], -1, YEAR))