

## General Project Description

### Action Justification Section

#### Adding

Length of Study field

The Confidence Interval Range Field:

#### Altering

LocationDesc and LocationAbbr

Name Alteration of Fields

Fields names were altered to create a better data analysis experience via more data relevant labels.

#### Deleting

Deleting of ID fields

#### Data Parsing

Negative and Positive Question Parsing

Separating by Class fields

# Data Subset Creation from Core Data Set

## General Project Description

The primary goal of the overall project is to separate data into meaningful subsets that allow for proper **cleaning**, **analysis** and **visualization**. This project is completed using **Python** via **Jupyter Notebooks** in combination with **Tableau Public**.

For this **Python Script** the goal was to clean and parse the data in order to separate the core data set into a date subsets that were manageable and made **analysis**, **addressing nulls and outliers** manageable while also allowing for appropriate data visualization

By **adding**, **deleting**, and **altering** existing fields as well as **parsing** the dataset by research question type and class data subsets were able to be created to continue to progress in the project via the exported CSVs.

# Action Justification Section

## Adding

### Length of Study field

The **length of study field** gives an idea of the amount of time that was put into the research and the potential robustness of the process and data used to get the research measures.

### The Confidence Interval Range Field:

This field was added in order to attach a level of potential **variance** in the result if the research was repeated and the level of accuracy of the measure relative to the actual population value.

## Altering

### LocationDesc and LocationAbbr

These fields were altered to exclude non 50 state territories to accomplish narrowing in on a more **standardized** population based on culture, regulation, and governance to allow for proper analysis and prevent large outliers in the data set.

### Name Alteration of Fields

Fields names were altered to create a better data analysis experience via more **data relevant labels**.

## Deleting

### Deleting of ID fields

Certain ID fields were full of **redundant data** and were not true Unique ID fields. They were removed to lean out the data set.

## Data Parsing

### Negative and Positive Question Parsing

With data values analyzed and visualized in aggregates it was important to separate them into these **connotation based questions**. This is because it allows the eventual viewer of the analysis and visualizations to understand that the values are coming from a **cohesive place** with a **straightforward insight** to understand.

## Separating by Class fields

With the eventual cleaning of the data with **null and outlier handling** being of the tasks it was important to separate these subsets by classes at the question connotation level. This would allow for the handling of nulls and removal of outliers to be handled at the class level rather than the overall level which could lead to **over generalized insights**.