Cognitive Decline Analysis Readme Document:

Associated Tableau Public : [Cognitive Decline Research Insights. | Tableau Public]

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 the data **by addressing nulls and outliers** using statistics and abbreviating question field values to improve **data visualization**.

[By **adding, deleting,** and **altering** existing fields to continue to progress in the project via the exported CSVs.]

Action Justification Section

Adding

Affect/Stage Field

By creating a function to alter all questions values in the cognitive decline subset to a shorter summarized values for better data visualization. The function was used to create a new field called ‘Affect/Stage’ to house these abbreviated values for question.

Altering

Stratification Type and Stratification Value Fields

These fields contained nulls and these nulls were filled with their

Stratification Type and Stratification Value Fields

The values of these fields were crucial for data visualization. And the fact that these fields did not include values made these records seem very unreliable for analysis that was intended for.

Statistics

Null Values were filled in the confidence_range and the Data_Value with their median values. This was done before identifying the z-scores for each.

Once the median values were filled in for the nulls the z-scores were calculated using spicy stats. The z-scores were calculated after the filling of the null values because a z-score test cannot be calculated with null values within the fields.

Two new fields were created: confidence_range_z, Data_Value_z, representing the z scores of the confidence_range, Data_Value outliers of the field values of Data_Value and confidence_range.

The outliers were removed on the basis of their z-score of 3 or greater and -3 or less. This is basic z-score related outlier removal.

Deleting

Confidence_range_z and Data_Value_z

After deleting the Outliers the Z-score fields were useless and were then removed.