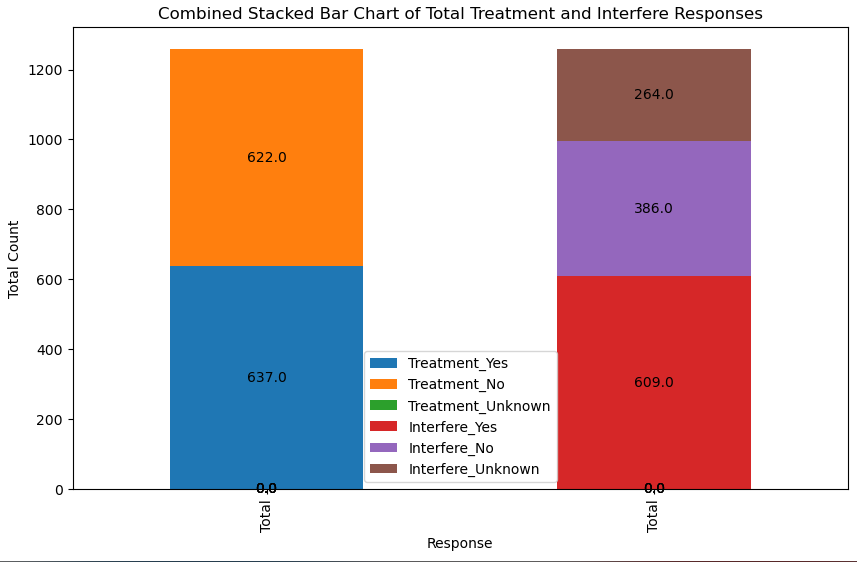
### **Project Overview**

**Our initial goal was to investigate mental health in the tech field—a topic that affects us all directly or indirectly. We were particularly interested in exploring data on this subject. We began with a survey titled "Mental Health in Tech."**

**We constructed a correlation matrix to analyze how the column "work\_interfere" (which asks if a mental health condition interferes with one's work) correlated with other variables. Aside from one notable correlation, the results were not particularly revealing. The exception was with the column "treatment" (which inquires if someone has sought treatment for a mental health condition). This showed a positive correlation of +0.40, suggesting that those who feel their mental health condition interferes with their work are more likely to seek treatment. This result is illustrated in the following graph:**

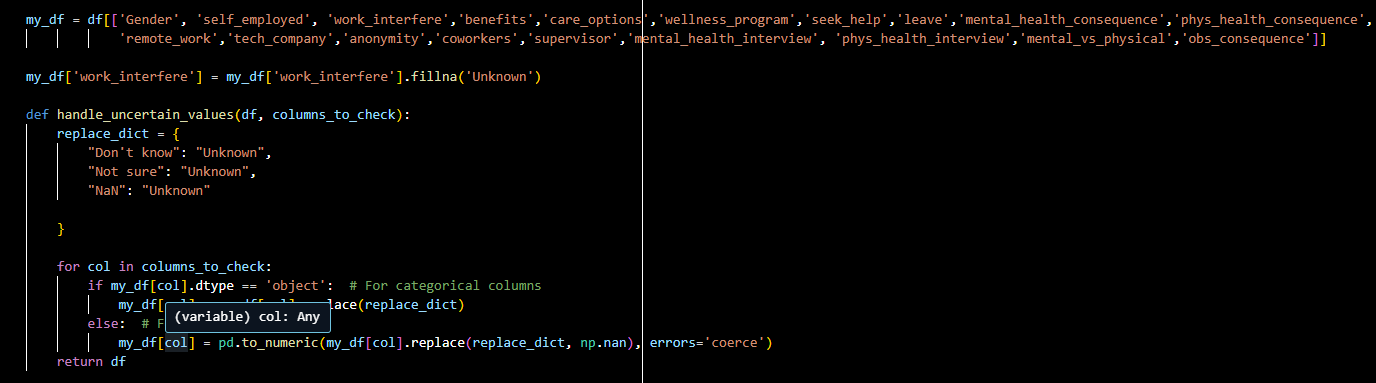
****

**While this was a promising start, we realized we needed a broader scope to obtain more conclusive data. We set a new goal: to fit every quantifiable column into a correlation matrix.**

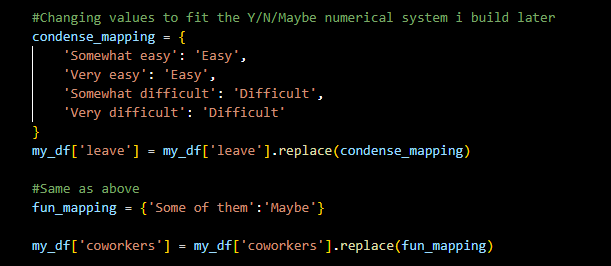
### 

### **Data Cleaning and Formatting**

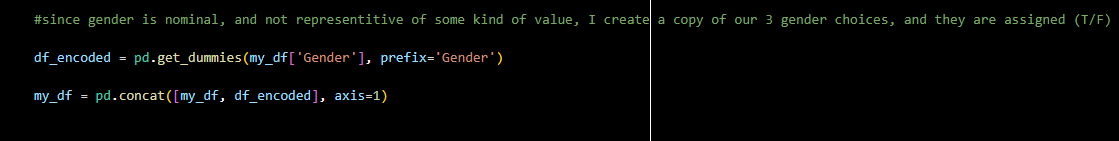
**To achieve this, we focused on data cleaning and formatting. For a correlation matrix, the data needs to be numerical. Our first task was to convert all non-numeric responses (such as "don't know" and "not sure") into a standardized format suitable for correlation analysis.**

****

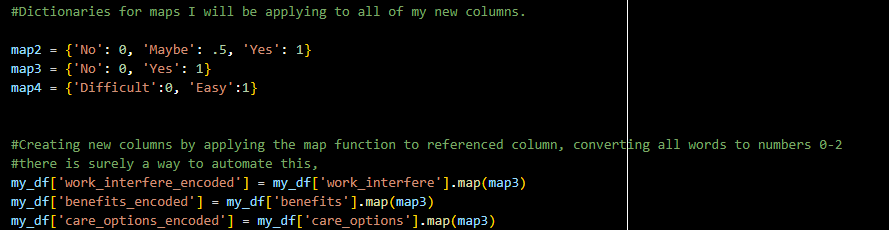
**We began by writing a function to standardize these non-answer columns. Next, we addressed specific issues, such as varying responses in the "leave" column, which were replaced with a uniform value.**

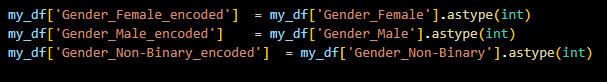
****

**We also handled the "gender" column, which contained three unique values: Male (M), Female (F), and Non-Binary. This did not fit the Y/N/Maybe format, so we used the pd.get\_dummies function to create three new binary columns representing these values. Each new column was assigned a True or False value.**

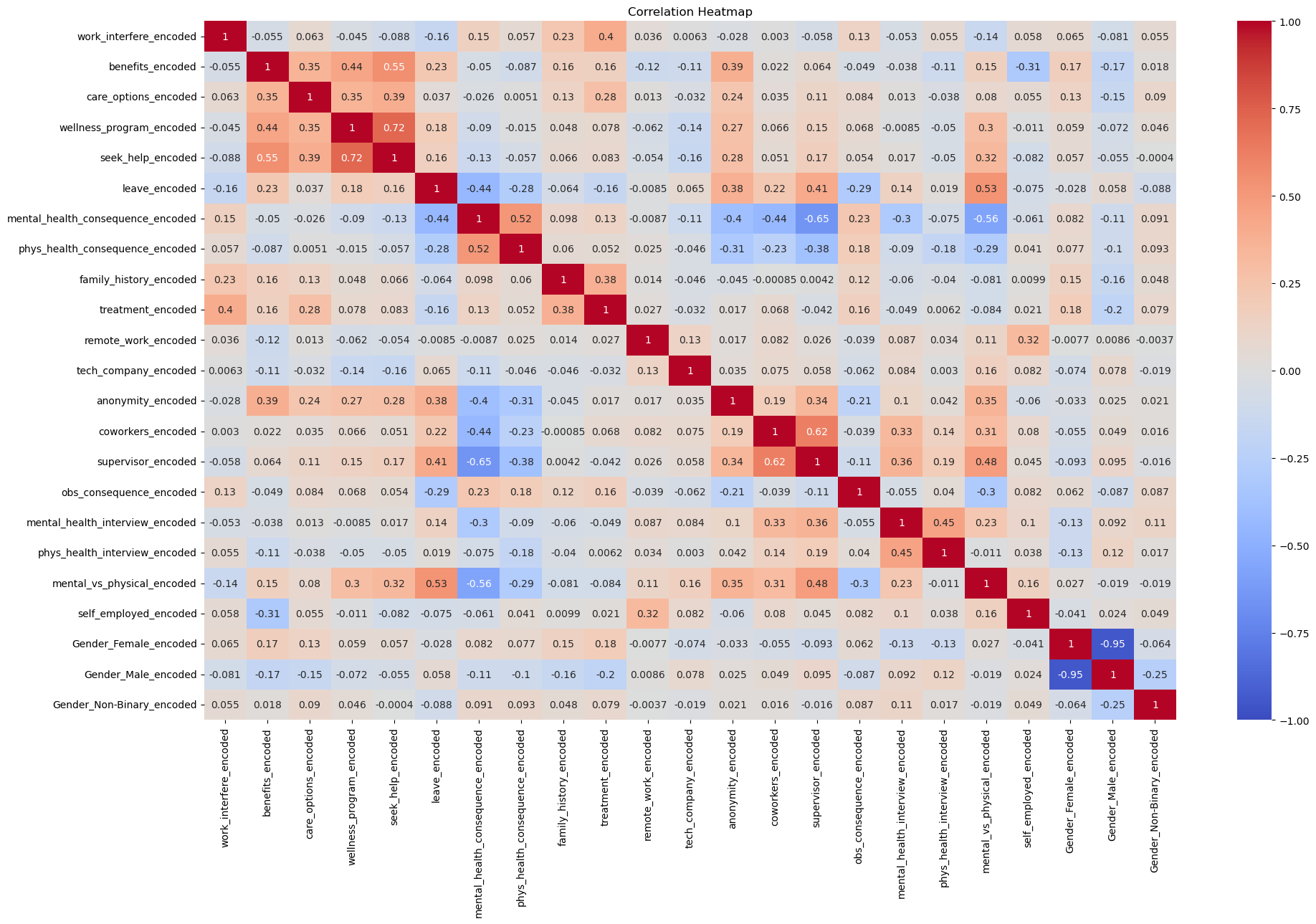
****

**Our aim was to convert all data into a Y/N/Maybe format for the correlation function. We created two mapping functions: one for Y/N columns and another for Y/N/Maybe columns. We encoded the columns, converting binary True/False values into 0s and 1s for Gender using .astype(int).**

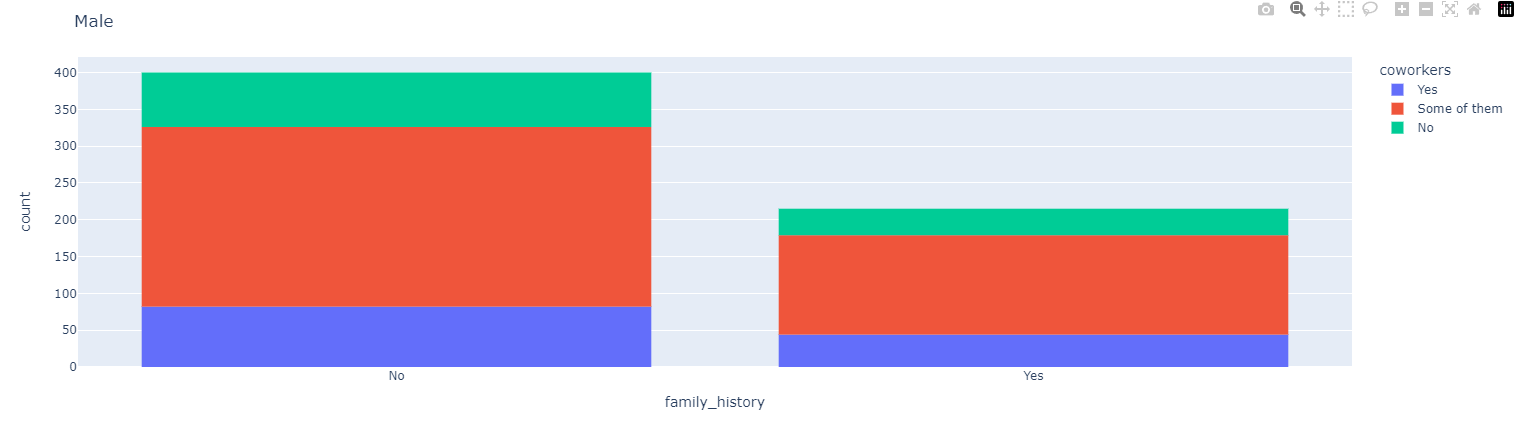
****

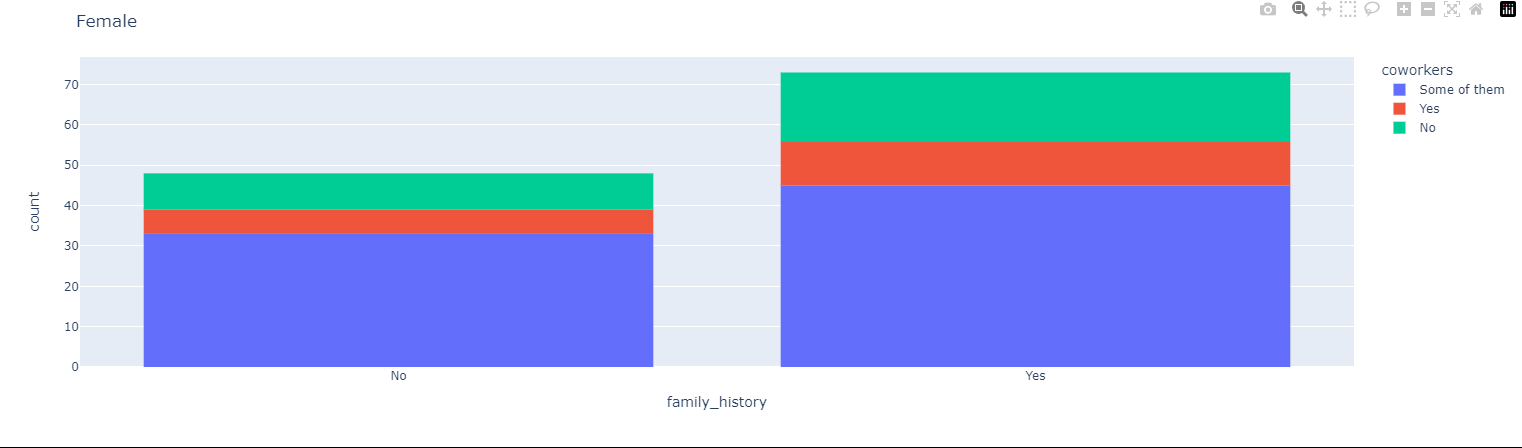
****

**These processed columns were extracted from the original DataFrame and saved into a new DataFrame called "my\_encoded\_df". We then ran the correlation function on this new DataFrame and visualized the results with a heatmap:**

****

**Noticing that "family\_history" had a high correlation with the previously mentioned "work\_interfere," we decided to delve deeper into the demographics. We visualized the relationship between family history of mental health issues and the willingness to talk with coworkers about potential conditions. We created separate graphs for males and females and found that there was no significant correlation between family history and willingness to speak about it, regardless of gender.**

****

****

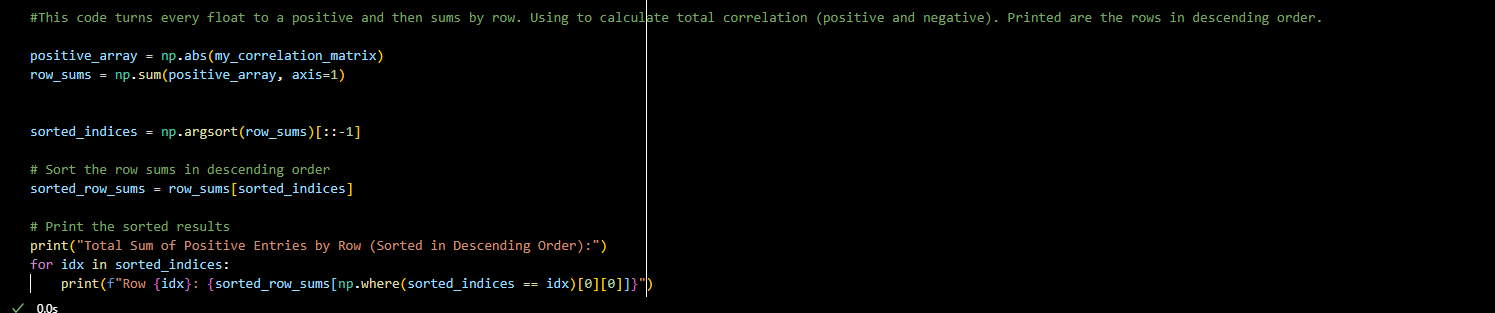
**The heatmap revealed several clusters of high correlation. Notably, the top left section contained a group of columns related to company support, such as wellness programs and benefits. It makes sense that companies offering wellness programs are also likely to provide additional support.**

**Similarly, another cluster at the other end of the heatmap related to whether employees believed they would be punished for discussing their health conditions and their willingness to talk about their condition with coworkers or supervisors. This aligns with the expectation that employees who fear punishment are less likely to discuss their mental health issues.**

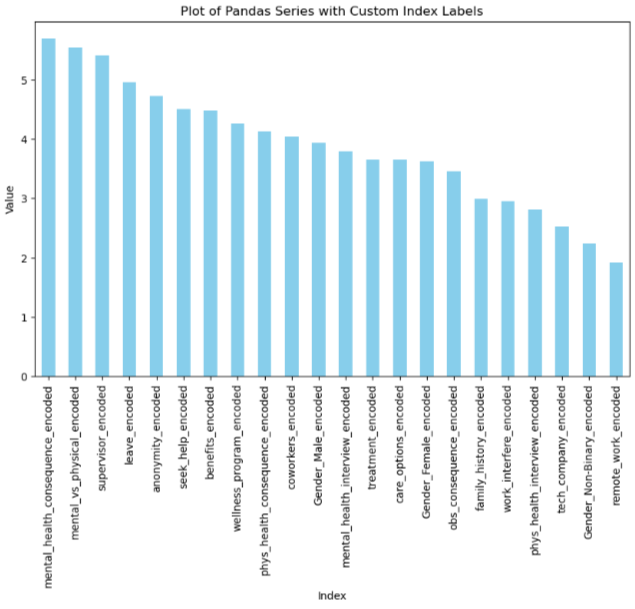
### **Analyzing Correlation Patterns**

**We observed that our heatmap was effective in highlighting these clusters. To gain deeper insights, we investigated which columns had the highest total correlation (both positive and negative) with the rest of the data.**

**We used code to turn all correlation values into positive numbers and summed them to determine the total correlation of each column.**

****

**A resulting bar graph displayed the highest totals:**

****

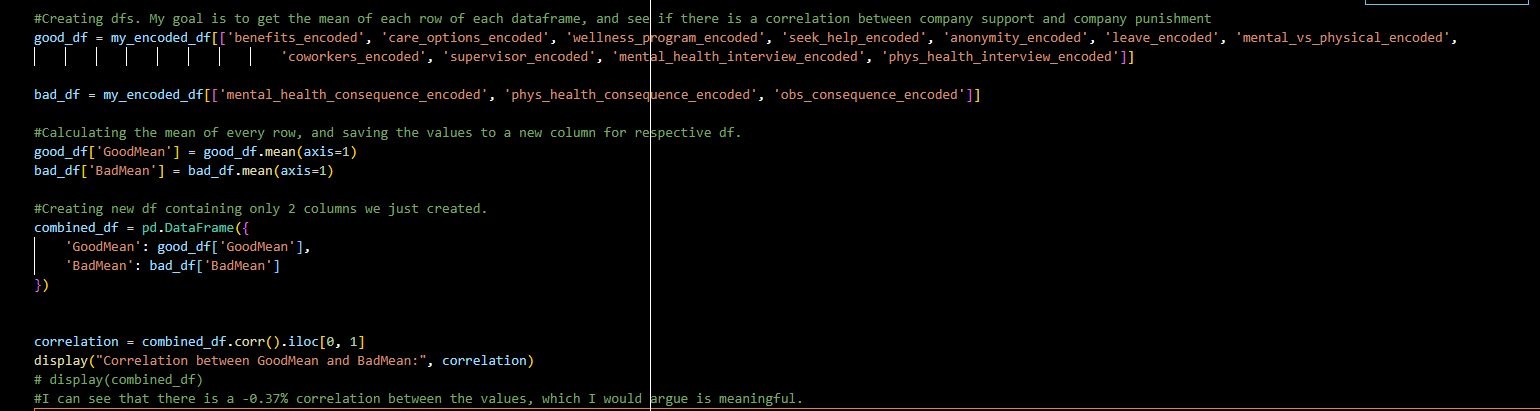
**From this analysis, we noticed that the "mental\_health\_consequence" column, among others, showed strong relationships with factors of trust in the employer. This suggests that while trust in the employer might seem a minor aspect of mental health at work, it significantly impacts various related factors. The "leave" column, in particular, demonstrated positive correlations with trust factors and negative correlations with distrust factors..**

### 

### 

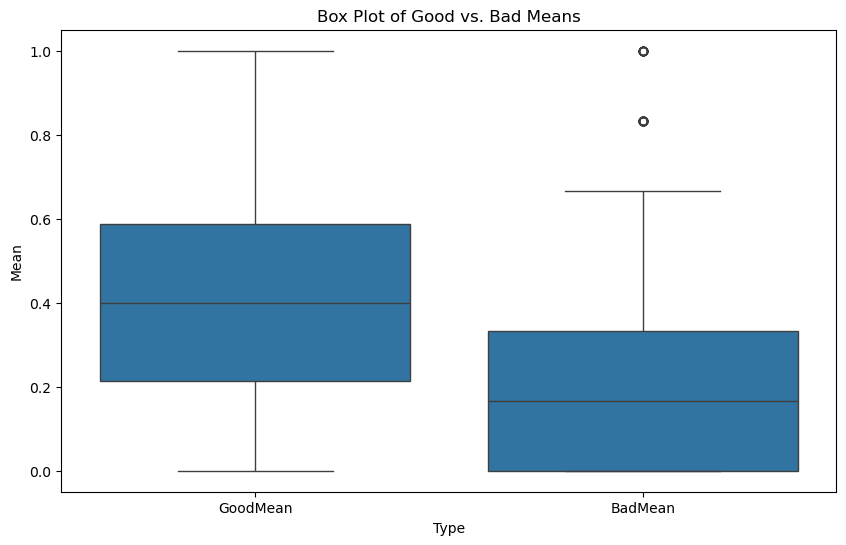
### **Further Examination**

**Noting correlations between perceived positive and negative entries led us to a deeper examination. We created two new DataFrames: one for "good" values (e.g., company support) and one for "bad" values (e.g., fear of punishment). We calculated the mean of each row for these DataFrames, yielding an average "good score" and "bad score" for each entry.**

****

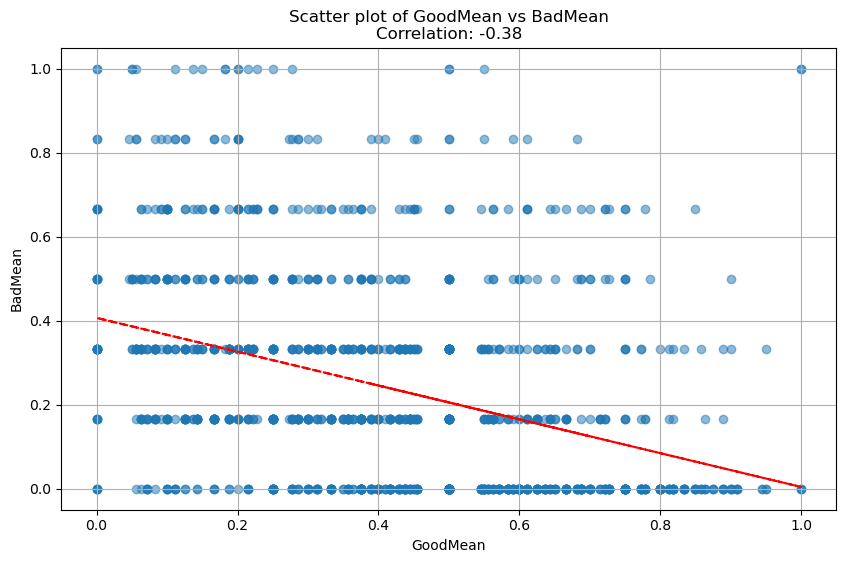
**Combining these results into a new DataFrame, we ran another correlation function, which showed a negative correlation of -0.37. While not exceptionally strong, this suggests that companies more likely to punish employees for discussing mental health are also less likely to offer support.**

**We visualized this relationship with a quick box plot:**

****

**This box plot shows the relationship between the average of GoodMean and BadMean. We observe that the average BadMean score is (thankfully) about 0.2 lower than that of GoodMean. Additionally, GoodMean tends to be more centrally distributed, with wide whiskers reaching the extremes of the graph.**

**We also made a scatter plot:**

****

**The scatter plot shows the relationship between "GoodMean" and "BadMean." The top left quadrant, although small in sample size, indicates that employees expecting punishment are often at companies that do not actively support mental health. The trendline confirms that employees anticipating punishment are more likely to be at companies that do not provide adequate support.**