CMSC 320, Homework 4

Peter Yan

October 16, 2023

1. Data Issues (Fardina)

While exploring the data, I found and corrected the following issues:

1.1 Missing Responses to Questions

I found eight students were missing responses to certain questions. I elected to fix it by using hot-deck imputation and mode imputation to find rows with data in the columns "You could describe yourself as...", "How would you rate your religiousness / spirituality?", "What bests represents your gender?" and "Would you describe yourself as compassionate?" most similar to the rows with missing data and substitute the most common values (Modes) in the columns from those rows into the rows with missing data. If no similar rows are found, I inserted the most common values across the entire dataset into the cells with missing values. I felt this was a valid way to fix it because students who have the same demographic will likely respond similarly to answers regarding morals.

1.2 Incorrect Data in Gender Column

I found 4 students listed their answers as "Famale" in the column for the question "What bests represents your gender?" I elected to fix it by changing all who students have the value "Famale" to "Female". I felt this was a valid way to fix it because it can be assumed that "Famale" was an error when typing and was intended to be "Female".

1.3 Empty Row

I found that row 138 was completely empty except for its Timestamp column. I elected to fix it by omitting this row from the dataset. I felt this was a valid way to fix it because one empty row out of 135 rows of data would not cause a significant difference in the data calculations.

2. My Questions

2.1 What is the impact of religiousness/spirituality on perceptions of non-traditional sexuality?
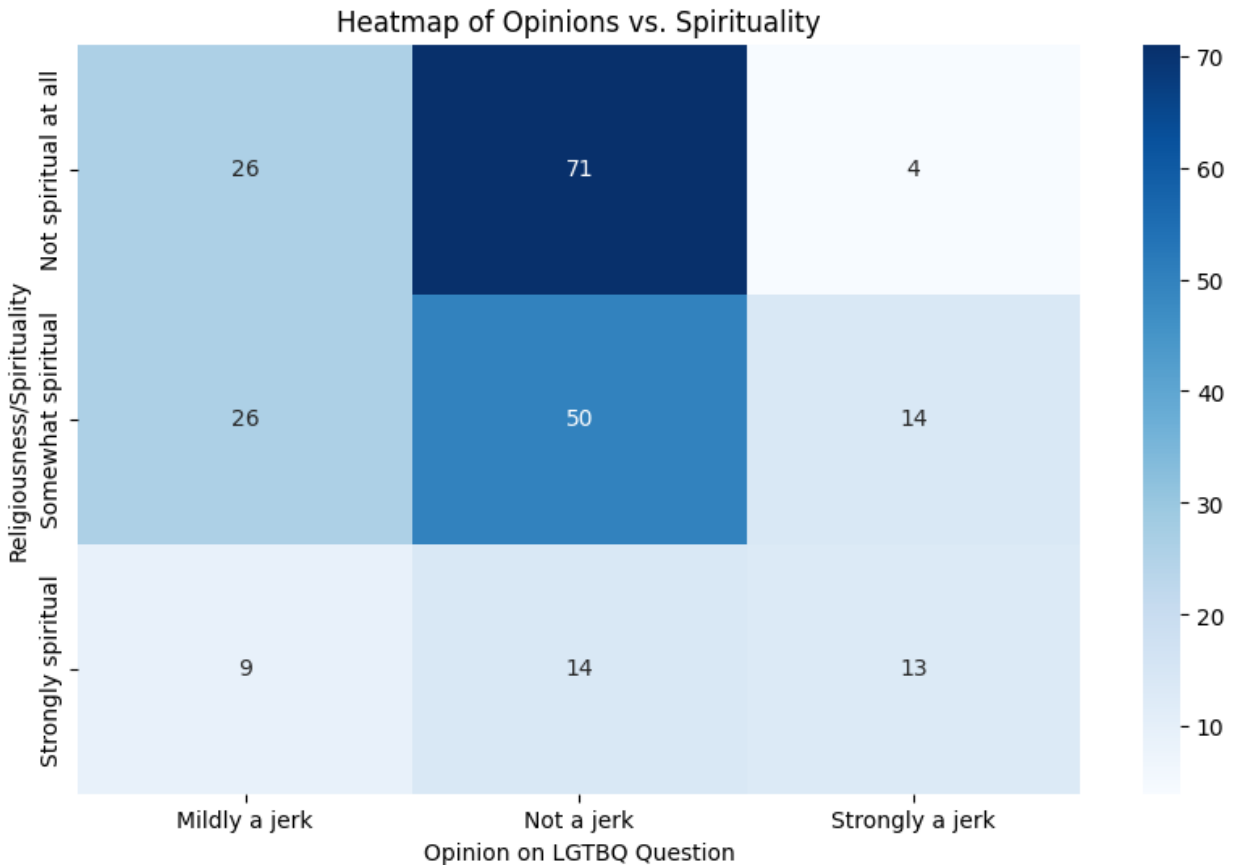
This question was of interest to me because I wanted to understand if an individual's level of religiousness/spirituality affects their perception of morality in a scenario that involves LGBTQ. It is generally believed that religious/spiritual people have a negative perception of non-traditional sexuality compared to non-religious people due to their beliefs in their holy books and traditions that may influence their sense of morality.

I investigated it by first proposing the following hypotheses:
- Null hypothesis (H0): Religiousness/Spirituality and Morality/Perception of LGBTQ Question are independent variables
- Alternate hypothesis (HA): There is a significant association between religiousness/spirituality and morality/perception of LGBTQ question

I created a contingency table to find the frequency of the answers to the question "Some of my relatives refuse to come to my wedding, since they don't approve of our 'lifestyle'. I would like to donate the money I will save to an LBGTQ organization in their name. I'm hoping that I or the organization will be able to send them a receipt/thank you for the donation. Would I be a jerk?" with correlation to the responses to the religiousness/spirituality question and determine whether to reject or fail to reject my null hypothesis. I then created a heatmap that shows how many students from each demographic group answered "Not a jerk", "Mildly a jerk", "Strongly a jerk".

By performing a chi-squared test on the contingency table, the resulting p-value was smaller than the 5% significance level, indicating to reject the null hypothesis. It seems that there is a statistically significant difference in opinion regarding the LGBTQ question among various religiousness/spirituality groups, with strongly spiritual people being more inclined to choose "Strongly a jerk" than people who are somewhat spiritual or not spiritual at all. However, the results are not definitive since the sample size is not large enough to prove a pattern.

Heatmap of Opinions vs. Spirituality

2.2 Does the political leaning of the adults one grows up with influence one's own political leaning?
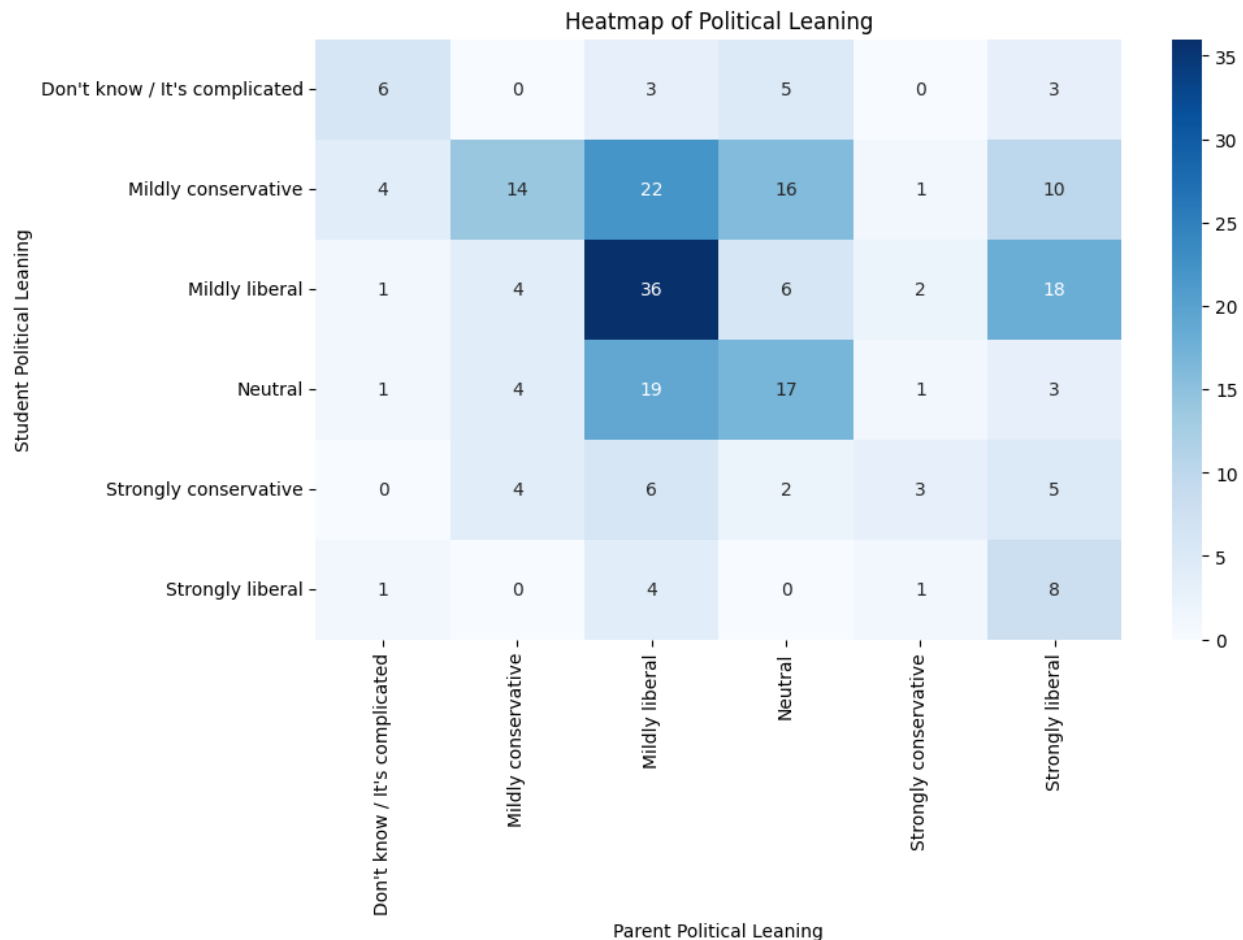
This question was of interest to me because I wanted to understand if there is a correlation between an individual's political leaning and the political leanings of the adults they grew up around. It is generally believed that a person's parents or caretakers when they were growing up has a heavy influence on the personality and ideology of an individual, which includes political leaning.

I investigated it by first proposing the following hypotheses:
- Null hypothesis (H0): The parents' political leaning and the individual's political leaning are independent variables
- Alternate hypothesis (HA): There is a significant association between the parents' political leaning and the individual's political leaning

I created a contingency table to find the frequency of the answers to the question "You could describe yourself as..." with correlation to the responses to the question "You could describe the adults you grew up with as..." and determine whether to reject or fail to reject my null hypothesis.

By performing a chi-squared test on the contingency table, the resulting p-value was smaller than the 5% significance level, indicating to reject the null hypothesis. It seems that there is a statistically significant association between a person's political identity and their parents' political identity. Furthermore, the heatmap shows that a high percentage of respondents answered the same or similar political leaning as their parents', regardless of what ideology.

Heatmap of Political Leaning

2.3 Is there a difference in opinion from people of different genders on gender-specific questions?

This question was of interest to me because I wanted to understand if an individual's gender affects their perception of morality in a scenario that hypothetically determines the user's gender. It is generally believed that people have a difficult time imagining themselves in a given scenario where they must assume a different identity.

I investigated it by first proposing the following hypotheses:
- Null hypothesis (H0): Gender and Morality/Perception of Question With Assumed Gender are independent variables

- Alternate hypothesis (HA): There is a significant association between gender and morality/perception of question with assumed gender

I created a contingency table to find the frequency of the answers to the question "I'm a single mom with four kids, one of whom has a different father from the other three. I get a lot of child support for one of them, and none for the other 3. 2/3rds of the child support for the eldest go towards his expenses and college fund. The other third goes towards my other children. My friend recently found out and told me it was inappropriate to use the child support intended for the one on the other three. Am I a jerk?" with correlation to the responses to the gender question and determine whether to reject or fail to reject my null hypothesis. I then created a heatmap that shows how many students from each demographic group answered "Not a jerk", "Mildly a jerk", "Strongly a jerk".

By performing a chi-squared test on the contingency table, the resulting p-value was larger than the 5% significance level, indicating a failure to reject the null hypothesis. It seems that there is no statistically significant difference in opinion regarding the assumed gender question among various gender groups. The heatmap created from the contingency table also shows a similar distribution for all genders alike, indicating that gender had no factor in determining the response to the quesiton.

Heatmap of Response to Gender-Specific Question