

Analysis of “Am I the Jerk” Survey Data

Daniel Yu

1 Introduction

On the popular social media platform Reddit, a subreddit (subforum) known by the abbreviation “AITA” allows users to post stories involving themselves and seek judgement on their actions from the wider Reddit community. This concept of crowd-sourced ethical rating was adopted by Max Morawski as a subject of analysis for his CMSC320 classes. In Max’s Fall 2024 CMSC320 class, students were surveyed on their opinion of over 14 questions provided by Max. They were then given the results of the survey along with two other surveys from Max and Fardina’s CMSC320 classes as separate uncleaned datasets. Each of the datasets were uncleaned and had slight differences. In Max’s 2024 dataset, the genders inside the questions were swapped, in Fardina’s 2023 dataset, a priming question was asked, and Max’s 2023 dataset had none of these additions. This report aims to observe whether or not students age and gender had an effect on their judgements of each Reddit post in the sample. It was confirmed via analysis using chi-square testing that for 3 questions age had a statistically significant effect on submitted feedback and that for 1 question gender had a significant effect.

2 Background

As mentioned previously, 3 uncleaned datasets were given containing student responses to 14 ethical scenarios inspired by the AITA subreddit format. These datasets are

- **Fall 2024 Dataset (Max Morawski):** Features gender-swapped scenarios to analyze potential gender biases in ethical judgments.
- **2023 Dataset (Fardina):** Includes a priming question before the main scenarios to examine the influence of pre-existing biases on ethical decisions.
- **2023 Dataset (Max Morawski):** Serves as a relative control group with original, unmodified scenarios.

Key variables include:

- Student age
- Responses to 14 AITA-style scenarios
- Gender (in gender-swapped dataset)
- Priming question responses (in Fardina’s 2023 dataset)

This combination of datasets enables analysis of factors influencing ethical decision-making among students. Comparisons across years and experimental conditions allow exploration of gender perception, priming effects, and temporal changes on moral judgments. The study focuses on the statistical significance of the relationship between students’ age and their survey responses, aiming to uncover generational trends in ethical reasoning and contribute to understanding how age and gender influences moral opinions.

3 Methodology

3.1 Data Cleaning

Several steps were taken to prepare the dataset for analysis:

1. Standardized the datatype of age columns to Integer
2. Dropped any duplicate responses
3. Dropped any responses that contained two or more empty columns
4. Used mode imputation to replace any other missing values
5. Removed any other data mishaps including age outliers, mislabelled genders, etc

3.2 Analysis of Age and Gender Correlation to Responses

To begin the analysis, exploratory data visualization was performed to understand the demographic composition of the sample. Two histograms were created to visualize the age and gender distribution of respondents across all three datasets, providing insight into the age ranges represented in the study (see figure 1 in appendix).

Following the initial visualization, statistical analysis was conducted through a series of chi-squared tests of independence. These tests were performed on each dataset examining the relationships between age groups and gender to responses for each AITA scenario, measured by a “Jerkiness rating” of “Not a Jerk”, “Mildly a Jerk”, and “Strongly a Jerk”. A significance level of $p < 0.05$ was established to identify statistically significant relationships.

For scenarios that demonstrated statistically significant correlations, visualizations were developed to illustrate these relationships. Heatmaps were generated for age-correlated scenarios, with age groups on the y-axis, response categories on the x-axis, and color intensity representing response frequency. This visualization method effectively highlighted patterns in how different age groups responded to specific scenarios. For gender analysis, the same heatmap visualization was used with different axis assignments, but a paired bar chart was created to offer an additional comparison between responses of each gender.

4 Findings

4.1 Age vs Jerkiness Correlations

The results of the analysis of chi-squared tests on the age column vs participant response to each question revealed age had a significant effect on three scenarios: the marriage scenario, the trust fund kid scenario, and the relatives/LGBTQ scenario. See appendix figure 2 for heatmaps that visualize the correlation for each situation.

4.2 Gender vs Jerkiness Correlations

The results of the analysis of chi-squared tests on the gender column vs participant response to each question revealed age had a significant effect on one scenario: the dentist/expensive school scenario. See appendix figure 3 for heatmap and paired bar chart that visualize the correlation for this situation.

5 Conclusion/Future Work

Overall, this study has provided valuable insights into the influence of age and gender on ethical judgments within the context of AITA-style scenarios. The analysis revealed statistically significant correlations between age and responses in three scenarios, while gender showed a significant effect in one scenario. These findings suggest that demographic factors do play a role in shaping moral opinions, albeit to varying degrees across different ethical dilemmas. The use of multiple datasets with different experimental conditions (gender-swapped scenarios and priming questions) allowed for a more comprehensive examination of these influences. Future research could expand on these findings by exploring additional demographic variables, conducting studies to track changes in ethical reasoning over time, or investigating the interplay between age, gender, and other sociocultural factors in moral decision-making. Such work would contribute to a deeper understanding of the complex dynamics underlying ethical judgments and could have implications for fields such as education, psychology, and social policy.

6 Appendix

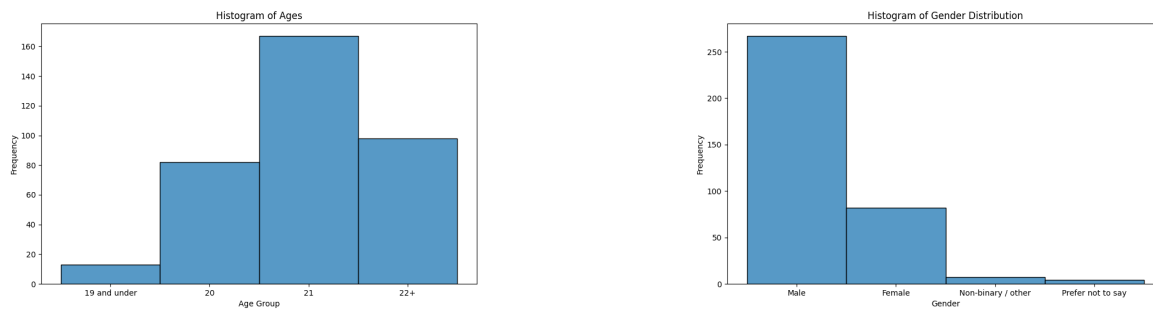


Figure 1: Histograms Displaying Gender and Age Distributions

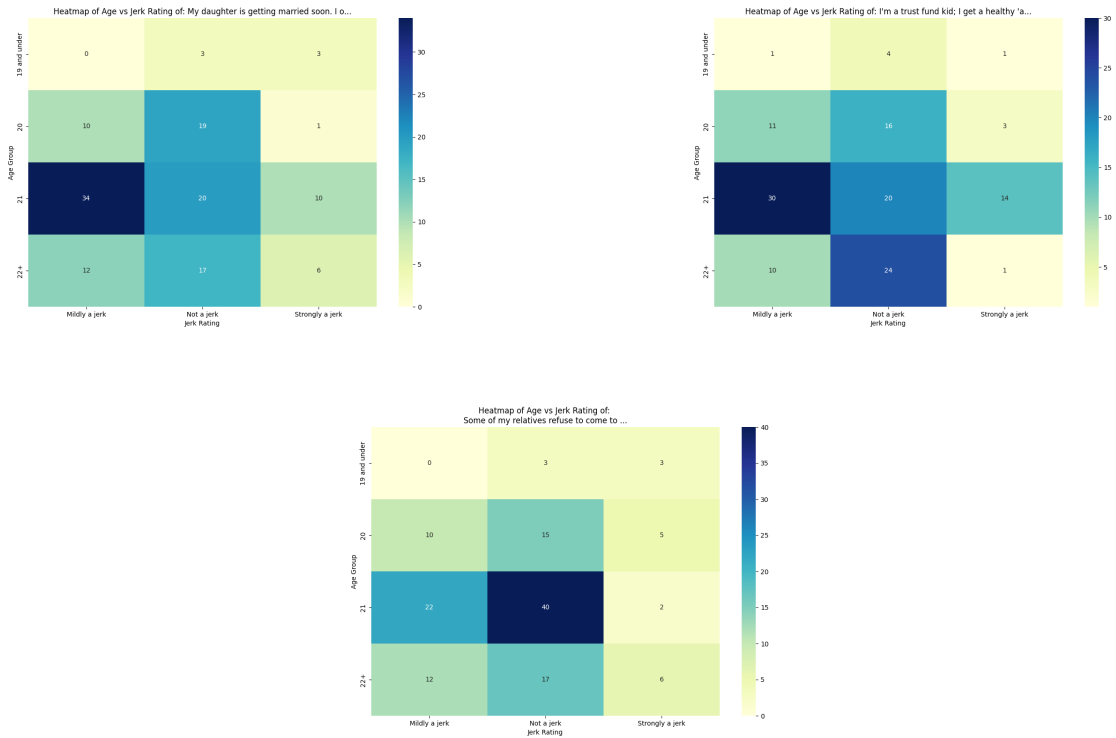


Figure 2: Heatmaps for Age vs Questions

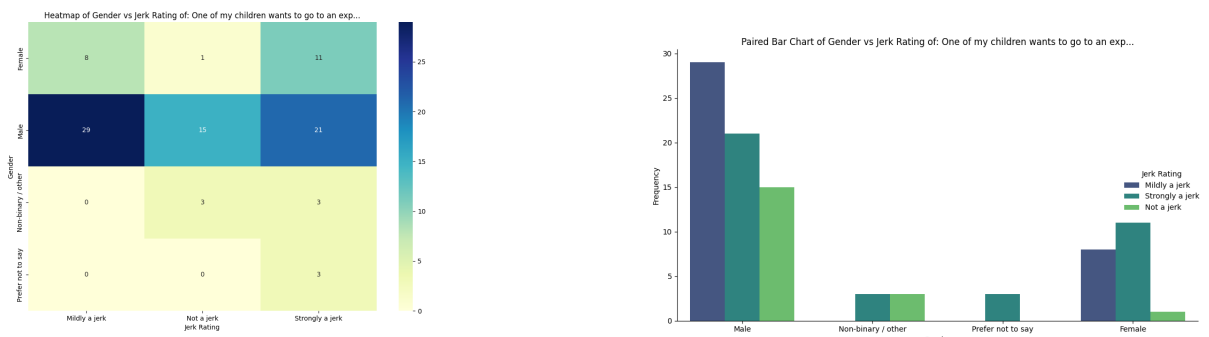


Figure 3: Heatmap and Paired Bar Chart for Gender vs Questions