Graph Ideas:

1.Difference in initial and final rating by number of people(7 columns on x and number of people on y axis).X axis represents the difference between initial and final rating.There can only be 7 such columns since max rating is 7 and minimum rating is 1.So,the absolute difference can only be 0,1,2,3,4,5,6.Y axis represents the number of debates for which we are measuring the absolute difference in initial and final rating.

2. bar graph for avg difference in rating. Y axis is the likert number change and x is the debate type(simple,argument,mixed,stats etc)

3. gender wise which persuasion strategy was most impactful. Bar graph. Each strategy has 2 bars in diff colors(for each gender). Each bar represents average change in rating for that particular prompt

Should I measure on avg how much was the persuasion for each gender by prompt or how many people were persuaded by xyz points for each prompt by gender?

4. same but by education. See some interesting results. Make 3 bars for starters(high school,bachelors,masters+)

5.time vs persuasion – slots of time taken to complete the debate(1hr – 1hr-15 mins,1:15-1:30 etc) vs the change in rating. Change in rating being y axis and time slots denoting x axis as bar graphs

6.

Reviews:

* Number 69 doesn’t really understand how likert ratings work ig. The person says that they had a change in opinion as per the final opinion,however,that’s not reflected in the final ratings(especially for the internet access debate).
* Number 83 was straight up copying from articles ig. “This article focuses on xyz” something was written in the rebuttal for the government and privacy debate. Was flagged for suspicious time tho.
* Number 54’s argument llm debate. His final opinion seems to be more nuanced and acknowledging of the other side,despite him reinforcing his opinion and having a negative liekrt change.

Normalization:

The choice between dividing by the average shift or subtracting 4 for weighting depends on the specific behavior you want to highlight in your analysis:

**1. Dividing by the Average Shift (Normalization Approach):**

* **Purpose**: This approach accounts for how much movement typically happens from any given initial score. If an initial score usually has large shifts (e.g., users who start at 4 tend to shift more), then a shift in that range should be given less weight than a shift in a range where shifts are less common (e.g., from 7 or 1).
* **Advantage**: It allows you to control for the variability of shifts across different initial ratings. This ensures that rare, more extreme shifts are weighted more, and common shifts (which may occur at more neutral points like 3, 4, 5) are weighted less.
* **Disadvantage**: Dividing by the average shift might mask how strongly opinions at the extremes are held. For instance, if many people with an initial score of 7 tend to shift, dividing by this larger average might reduce the weight of these shifts, even though they're still significant in terms of opinion change.

**2. Subtracting 4 (Distance from Neutral Approach):**

* **Purpose**: This method explicitly weights shifts based on how far they are from the neutral point (Likert score of 4). The idea is that the farther an initial opinion is from neutral, the stronger it is, and thus, the more meaningful any shift should be.
* **Advantage**: This method directly emphasizes strong opinions (e.g., 1 or 7), which are often more resistant to change. A shift from a 7 to a 6, or from a 1 to a 2, would carry more weight than a shift from 4 to 5 or vice versa. It highlights persuasion more effectively for extreme views.
* **Disadvantage**: It doesn't account for how common or uncommon shifts are in each category, so it treats all shifts equally based solely on their distance from neutrality. This can overlook natural variability in certain score ranges (e.g., scores around 4 may naturally have more shifts).

**Which is Better?**

It depends on what you want to prioritize:

* **If you care about the rarity of shifts** (i.e., you want to highlight unusual, unexpected shifts based on the initial score): **Dividing by the average shift** would be better. It normalizes for the common behavior, making unexpected changes stand out.
* **If you care about the strength of initial opinions** (i.e., you want to emphasize the strength of persuasion for users holding extreme positions): **Subtracting 4** is a simpler and more direct way to weight strong opinions. It clearly shows how much influence AI had on people with firmly held views.

**Recommendation:**

For a study on **persuasive capabilities of AI**, where you are likely trying to measure how effective AI is at changing strongly held opinions, **subtracting 4** (distance from neutrality) is likely a more appropriate choice. It emphasizes the shift in strong opinions, which are typically harder to change and thus more significant from a persuasion standpoint.