**Carefully read the background and collection plan again. What types of potential bias exist in your team lead’s collection plan? Why was it biased? Please explain your answer. You may also think of biases that go beyond this reading (e.g., cultural bias).**

Collection and Sampling Bias is happening in this data collection process since it is targeting only one graphic region and demographic. Collection is happening only in the most southern region of the California and Mexico border within a 100 miles from the border, with all of the transactions happening exclusively in Mexico. Resulting in 75% of flagged transactions involving Mexican citizens, while only 11% of the transactions fit the suspicious activity logic, suggesting a bias related to nationality, socio-economic class. This might unintentionally run the risk of bias of teammates looking at the data with who might have preconceived notions or beliefs about certain regions and backgrounds making the project unusable because bias is being used within the project and doesn’t represent the true

Which disregards transactions happening across the U.S. and excluding different types of ATM transactions happening between different people.

**How might these biases distort the results? What could you do to avoid these biases?**

By targeting one area, and representing it as a larger portion of the population sku’s the results and doesn’t take into account different factors and other data from other areas of the population or country. This could falsify the information and give the wrong information to a wider range of people looking at the data and eventually the and the general public. If the wrong information is used we might not get the results we are looking for at the end of the project wasting resources.

To avoid biases, it's crucial to collect a broader range of transaction data from diverse geographic locations and demographics. This is important to not only target one demographic and one region but to work with broader data and work inward to be more specific regions and demographics to avoid these biases.

Also teaching this method during onboarding as one of the main themes and importances of data ethics will elevate this from happening. By having diverse training on this at the beginning of one’s duties will help all teams have the same code of ethics when collecting and analyzing data to come to conclusions. Having more than one analyst work on a project will help compare the ways they have collected their data to better understand and come up with processes to elevate these biases that might arise; ensures that decisions are based on relevant factors, promoting more accurate, fair, and reliable results.

**If you know that there is bias in the collection method, what could you do to communicate your concerns to your team lead? Please be as specific as possible.**

By communicating your concerns as early as possible during the collection process will help eliminate any concerns that might arise later. Vocalizing by asking questions on why certain processes are being used after gathering evidence that these methods are targeting only one particular section of the whole. Afterwards, express ways in which we can collect the data ethically without biases individually and within the team so everyone has a clear understanding of how we eliminate bias altogether.

**Analyze the bar chart showing the scores of individual analysts and see where their scores fall on the distribution curve. If the mean of the scores was 307 and the standard deviation is 166, which score or scores might you eliminate to control for bias? Why?**

Eliminating the analyst scores with 10 or 759 should be eliminated to obtain a more accurate representation of the data that doesn’t have a bias. This score is an outlier, and could skew the analysis if included without the adjustment in scale. Eliminating this score would help in maintaining a more balanced and accurate representation of the overall data distribution and more realistic conclusion.