Q1A

Biased training data may cause the machine learning system to unfairly target certain racial groups, creating disparities in predictions.

Discriminatory policing practices reflected in historical data can perpetuate racial disparities in the system's predictions.

Correlated variables like income or education level may indirectly influence the system to target specific neighborhoods based on race if used improperly or with bias.

Q1B

Counterfactual fairness creates a hypothetical world without race to evaluate predictions by removing any race correlations in the input data. This ensures the system doesn't use race as a factor in its predictions.

Interpretable models like decision trees or linear models can help identify driving variables, allowing closer scrutiny of potential biases or discriminatory practices. This prevents the system from targeting certain neighborhoods based on race.

Q1C

The data used in the study may be incomplete or biased: The study relies on data from the Oakland Police Department and the National Survey on Drug Use and Health, both of which may be incomplete or biased in some way. This could potentially impact the validity of the study's conclusions.

The study may not fully account for all confounding factors: While the study attempts to control for confounding factors, there may be other factors that are not accounted for that could impact the conclusions drawn. This could potentially impact the generalizability of the study's findings.

Q2A

Female & non-binary other group may be disadvantaged. Their mean experience level is higher than other groups, and imputing their missing values with overall mean could lower their ranking even if they have high skills test scores.

Q2B

An alternative imputation method is to use a group-specific mean.

Q2C

Technical bias introduced by imputation method can amplify pre-existing bias in hiring practices, reinforcing bias and making it difficult to detect and correct emergent bias in the process.

Q3A

There are 24 possible age value assignments for respondents in group 2D. Given the mean and median are both 24, we can get the sum of three persons are 72. 72-24 is 48. Thus the sum of the other two people should be 48. So we can list the ages from 1,47 all the way to 24,24 and thus makes 24 possible age values.

Q3B

One possible combination of ages for respondents in group 2D is (1,24,47).

Q3C

Dropping statistics about groups 2A and 2B increases the number of feasible satisfying assignments because it reduces the amount of information available to an attacker. If one of

the groups is dropped, the attacker still has enough information to eliminate some possible age assignments.

Q4

This has a epsilon value of Ln(2).

Pr[$q=y$]=Pr[$q=y$|coin shows "tails"]·Pr[coin shows "tails"]+Pr[$q=y$|coin shows "heads "]·Pr[coin shows "heads"]

Pr[$q=p$|coin shows "tails"]=1

Pr[$q=p$|coin shows "heads"]=1/2

Pr[$q=$"*yes*"|coin shows "tails"]=0

Pr[$q=$"*yes*"|coin shows "heads"]=1/2

Pr[$q=p$]/ Pr[$q=$"*yes*"]= Pr[$q=p$|coin shows "tails"]/ Pr[$q=$"*yes*"|coin shows "heads"] =1/(1/2)=2.