Garv Gaur

Differential Privacy Report

1. The TA must have a musical taste of classical, house, country, hip hop, pop, rock, or metal. To narrow this down, we know that Kinan is a grad student, so his age is likely <22. Thus, this rules out house as a possible answer.
2. On Kinan's Facebook page, we find a post about Metallica. This likely means that he is a metal fan. Additionally, his website at babman.io indicates that he is a metal fan. Thus, he must be either 29 or 30, as these are the only two metal observations in the data. Upon further investigation, his LinkedIn indicates that his birthday is on January 13, and his Facebook indicates that his birth year is 1994. Thus, he must be 29 years old.
3. Since there is only one observation of a 29 year old metal fan, we can assume this is Kinan. After running a count query on age, music, and color, we find that his favorite color is black. This is a fairly easy attack to do since we know that he has a unique set of observations in the dataset.
4. This year, in the 25 or more age group, there exist observations with baseball, basketball, E-Sports, and soccer, so one of these must be Kinan's favorite sport.
5. Since Kinan was 28 years old last year, his favorite sport is either American Football, Soccer, or E-Sports. Thus, his favorite sport must be either E-Sports or soccer. If we increase the granularity of the query on this year's data, (using count age sport), we find that the favorite sports of the 29 year olds in the data are soccer and basketball. Thus, Kinan's favorite sport is soccer.

6. As the privacy parameter grows larger, the noised data looks increasingly like the non-noised data, decreasing privacy. As the privacy parameter grows smaller, we have increasingly noisy distributions, which increase privacy but decrease accuracy by a lot.
7. The most likely value is 1, and the average value is 1.21. These are the same as the actual value. For different values of the privacy parameter, the graphs have flatter distributions with higher variance around 1.
8. We can deduce that the TA probably has 10+ years of experience, since the average age for the 10+ group is approximately 28.2, where the other groups are around 21-22. We are not completely confident, as there are scenarios where Kinan has less experience but the average age of his experience group is brought down by younger members of the group in higher quantities.
9. The exposed counts show that the count for the 10+ years programming group is approximately 4. Taking 4 individuals with an average age of 28.2, we have that the total age of all the people is approximately 112.8. The combination of individuals that would lead to this average likely includes both 29 year olds. In addition, the other plausible group is the 8-10 years of experience group, which has an average age of approximately 22.5. If Kinan were included in this 4-person group, the other individuals would have to be quite young (19-20ish) and would have to have started programming at 11-12 years of age. This is less likely, and considering that Kinan is a grad student and his LinkedIn profile showing his experience, he probably has 10+ years of experience.

   In summation, we know from the data that Kinan is 29 years old, his favorite music is metal music, his favorite sport is soccer, and his favorite color is black. He has 10+ years of programming experience.

10. This class doesn’t suffice to truly enforce that a dataset is never used beyond a certain privacy budget. Developers can overuse the dataset by simply changing the budget for their dataset or by using multiple hosts to use the budget multiple times. A good way to solve these problems is to centralize the privacy budget in some trusted party that can bookkeep all the budget subtractions. Also, data with proper DP calibration should be accurate enough that multiple draws on the same data are not required, so perhaps either noised data could be exposed to everyone and/or multiple draws are not permitted.