

Breakout Exercise & Discussion Prompts

For today's breakout activity you will be reviewing sample intake and exit data from our sample program, Project Safe Streets (PSS). We'll be conducting a quality review of this PSS intake data and identifying potential quality issues with the intake and exit data. For each data quality issue you identify in the dataset, please also identify the corresponding data quality metric that best describes the issue. For this exercise, we will be focusing on the four intrinsic data quality metrics of accuracy, lineage, semantic, and structure.

Data Quality Metrics:

Accuracy	How well do data values depict reality / correctness?
Lineage	How trustworthy is the originating source of data?
Semantic	Are data values true to their meaning?
Structure	Do data values exist in the correct pattern and / or format?

As you perform data quality checks on the PSS Q4 2023 intake and exit data, reflect on how your agency would resolve the data quality issues. Refer to the PSS description below and note which of the data quality issues we discussed that you have discovered in the dataset and what steps you might take to resolve them.

Take notes on your data quality findings and potential solutions and be prepared to share your thoughts when we rejoin the main Zoom meeting room.

PSS Description: *This spreadsheet should include Intake and Exit data for participants served through PSS during Q4 (October through December) 2023. PSS aims to serve young people ages 14-19. PSS provides both prevention services to youth from any referral source and intervention services to youth with prior juvenile justice system involvement who are referred to the program through the courts, probation, or police.*

In Q4 2023, PSS served 100 individuals. Historically, an equal number of males and females have been served; however, qualitative feedback from previous years suggests a greater need for services among females as opposed to males. As a result of this feedback, in 2023 PSS promoted gender-based services at community events and during educational presentations with partner agencies with the goal of reaching more female and female-identifying youth .

At the time of enrollment, staff complete referral source information and the youth are asked to complete a self-assessment asking about their age, gender, race / ethnicity, school attendance, and arrest history. These youth self assessments are captured on a paper survey and entered into the system by a caseworker. When a youth is discharged, staff enter program exit, graduate, and indicate whether the youth achieved their goal(s). When a youth has achieved their goals, they successfully complete the program and graduate.

Discussion Prompts:

1. What data points possibly have accuracy issues (meaning they may not represent reality)? What can be done to improve the accuracy of the data?
2. Are there any metrics that may not be trustworthy (i.e., you may question how reliable they are based on their lineage or source)? If so, which one(s) and explain why the data are not trustworthy. Should these data be collected from another source? If so, what makes one data source better than another?
3. How would you define Service Date? How could "Service date" be clearer to where everyone has a shared understanding of what "Service date" measures?
4. Did you identify any data structure issues? In other words, are the text fields collecting text information, are the date fields collecting date information, and are the numeric fields collecting numeric information? What data validation techniques could help with this at the point of data entry?
5. Is there a correlation between "Reason for program exit" and "Graduated program"? What would you expect to see across these two data points? Based on the results, do you want to investigate the data further?
6. According to the program model, youth who complete their goals should graduate from the program. Looking at the cross tabulation between Completed All Goals and Graduation Status, do you have any data quality concerns? How would you increase the quality of the data?
7. According to the program model, in addition to prevention services, PSS provides intervention services to youth with a history of arrest who are referred to the program through the Courts/Probation/Police. Looking at the cross tabulation between Referral Source and Arrest History, do you have any data quality concerns or questions? How should your program process the data (keep the data as is, complete an audit, follow-up with court records, etc.)?
8. What specific changes would you make in your program to increase the quality of the data collected? Consider the following prompts:
 - a. Source of the data
 - b. Use of data dictionaries
 - c. Training and CQI processes