Identifying Trends in (Theoretical) Alternative Sentencing Programming

Prepared for the Center for Justice Innovation by Caroline Cullinan

August 2, 2023

The following document outlines answers to the posed questions in the provided Data Task document for the Research and Data Associate role. Section 1.0 provides direct key findings to the posed questions. Section 2.0 puts forth recommendations.

Please also see the Appendix for access to the code behind this analysis and for a complete explanation of assumptions made, data concerns, and analysis methods.

Click **here** to access the code for the analysis.

1.0 Answers to Posed Data Task Questions

Based on the data, please tell us the answers to the posed questions, grouped by people younger than 25 and those 25 and older (at the time their case opened):

1.1 Question 1:

How many individuals attended at least one individual counseling session in 2019?

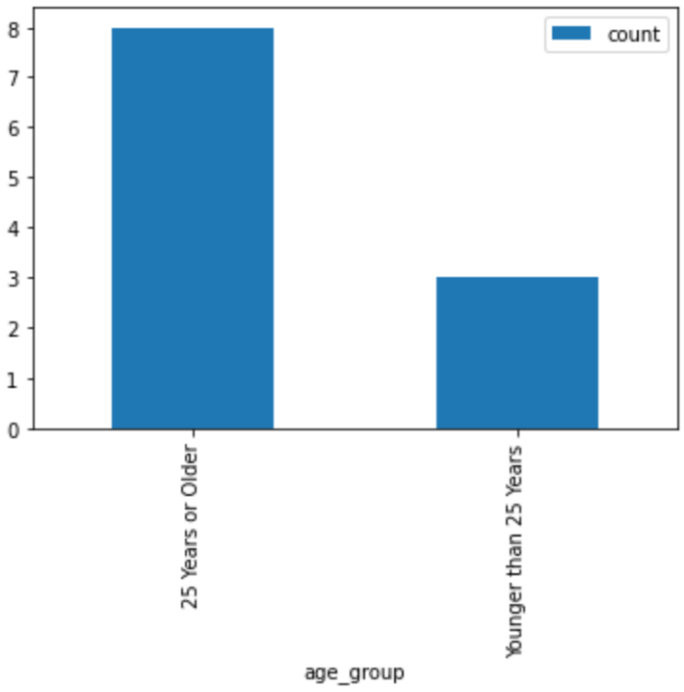
	age_group	count
0	25 Years or Older	3
1	Unidentified Age	1
2	Younger than 25 Years	9

- **3 clients aged 25 years or older** at the time in which their cases were opened attended at least one Individual Counseling session in 2019.
- Similarly, **9 clients** who were **younger than 25 years old** when their cases opened attended at least one Individual Counseling session in 2019.
- Finally, at least one Individual Counseling Session was also attended in 2019 by **1 client without an identified age**.

1.2 Question 2:

How many individuals attended at least one group counseling session in 2019?

	age_group	count
0	25 Years or Older	8
1	Younger than 25 Years	3

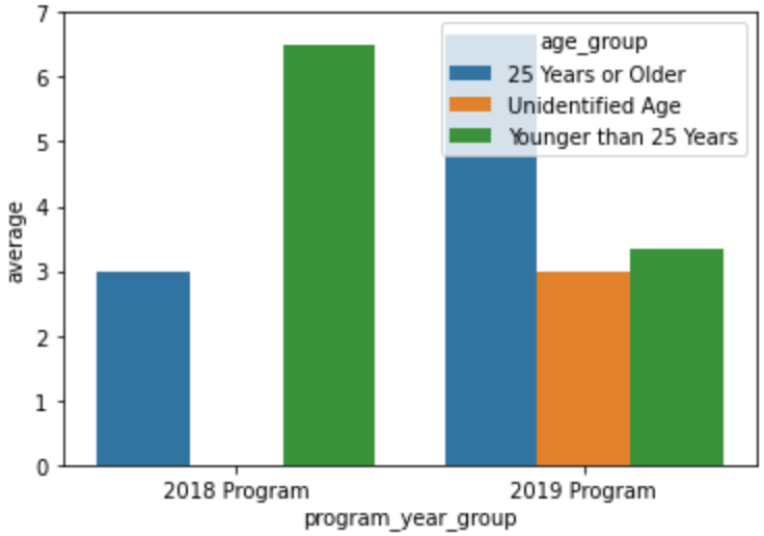


- **8 clients aged 25 years or older** at the time in which their cases were opened attended at least one Group Counseling session in 2019.
- Similarly, **3 clients** who were **younger than 25 years old** when their cases opened attended at least one Group Counseling session in 2019.
- Finally, **0 clients** with an **unidentified age** attended any Group Counseling sessions in 2019.

1.3 Question 3:

Of the individuals who attended at least one individual counseling session, what is the average number of individual counseling sessions attended by year?

	age_group	program_year_group	average
0	25 Years or Older	2018 Program	3.000000
1	25 Years or Older	2019 Program	6.666667
2	Unidentified Age	2019 Program	3.000000
3	Younger than 25 Years	2018 Program	6.500000
4	Younger than 25 Years	2019 Program	3.333333



2018

- In 2018, for clients who were **25 years or older** when their cases were opened, the average number of Individual Counseling sessions attended was **3 sessions**.
- In 2018, for clients who were **younger than 25 years old** when their cases were opened, the average number of Individual Counseling sessions attended was **6.5 sessions.**

2019

- In 2019, for clients who were **25 years or older** when their cases were opened, the average number of Individual Counseling sessions attended was roughly **6.67 sessions**.
- Similarly, in 2019, for clients who were **younger than 25 years old** when their cases were opened, the average number of Individual Counseling sessions attended was roughly **3.33 sessions**.

- Finally, in 2019, clients with an **unidentified age** at the time in which their cases were opened attended, on average, **3** Individual Counseling **sessions**.

1.4 Question 4:

How has the average individual counseling sessions attended changed between 2018 and 2019?

Clients 25 Years or Older

- The percent change in Individual Counseling session attendance from 2018 to 2019 for folks **aged 25 years or older** was roughly **122%** (i.e., [(6.67 - 3) / 3] * 100 = 122.3).

Clients Younger than 25 years

- The percent change in Individual Counseling session attendance from 2018 to 2019 for folks aged **younger than 25 years old** was roughly **-49%** (i.e., [(3.33 - 6.5) / 6.5] * 100 = -48.77).

Clients with an Unidentified Age

- Seeing as clients without an **unidentified age** did not attend any Individual Counseling sessions in 2018, a percent change for this age group **cannot be calculated**.

1.5 Question 5:

Among closed cases, which individual had the highest group counseling session attendance rate, excluding canceled sessions?

Among closed cases, clients **"Road Runner"** (Person ID: 1000040_00 (first case); 1000040_01 (second case)) and **"Wiley Coyote"** (Person ID 1000050) had the highest attendance rates for Group Counseling sessions (i.e., both "Road Runner" and "Wiley Coyote" had **attendance rates of 100% for Group Counseling sessions**).

Notably, however, client **"Road Runner"** differs from other clients, because they are associated with more than one case in the dataset. It is important to note that their first case (concerned with Person ID: 1000040_00) was closed and is associated with a 100% Group Counseling session attendance rate. However, their second case (concerned with Person ID: 1000040_01), while also closed, has an associated Group Counseling attendance rate of 0%. While there was only a single Group Counseling session that **"Road Runner"** could have attended from when their second case was open and closed, they still did end up missing this session, hence contributing to a 0% attendance rate for this case.

1.6 Question 6:

Among closed cases, identify the individual(s) that had an individual counseling session attendance rate (excluding canceled sessions) in the bottom quartile.

Among closed cases, clients that had an Individual Counseling Session Attendance Rate that was in the bottom quartile include:

- **"Wiley Coyote"** (Person ID: 1000050) had an Individual Counseling Session Attendance Rate of **25%**.
- **"Roger Rabbit"** (Person ID: 1000090) had an Individual Counseling Session Attendance Rate of **40%**.
- **"Charlie Dog"** (Person ID: 1000180) had an Individual Counseling Session Attendance Rate of roughly **57%**.
- **"Ralph Phillips"** (Person ID: 1000220) had an Individual Counseling Session Attendance Rate of roughly **33%**.

1.7 Question 7

Explain how you would approach summarizing or visually presenting the above findings to program staff.

In order to summarize and visually present the above findings to program staff, I would upload the exported data from this analysis to an Observable Notebook. Then, with JavaScript, I would prepare data visualization elements (e.g., tables and graphs) which would help communicate the outlined findings per posed question. In order to help program staff understand and interpret these visualization elements, I would also provide clear and concise text describing the data visualization elements for each posed question.

Once answers to all questions were clearly outlined, I would provide an overall text summary which would summarize all of the findings in all of the posed questions in a simple paragraph.

Given this summary of findings, I would also provide some interpretation as to what these findings may mean for programming. I would also put forth recommendations for current programming activities and potential future programming activities given the research findings. In addition, the findings to this research would also allow me to put forth recommendations for future program-supporting research.

Ultimately, this document is a lengthy rough draft of what I would provide to program staff.

2.0 Recommendations

The following recommendations for continued analysis should be considered in order to develop a deeper understanding of the data that has the potential to inform current and future programming.

1. Visualize results for deeper exploratory analysis that could ultimately provide greater insight into current programming and guide future programming and research.
2. Dive deeper into analysis that considers the cancelation of programming over time (i.e., which types of programs are canceled and when (e.g., which months))
3. Similarly, conduct analysis that looks to understand attendance and non-attendance over time in different months for both younger and older client populations.
4. Try to develop insight into the duration of cases in relation to attendance / non-attendance for various programming for various age groups.
5. Consider smaller age groups in future analysis (i.e., instead of being concerned with just folks who are aged younger than 25 years old or 25 years and older, consider age groups that concern folks aged younger than 18 years old, 18-24 years old, 25-34 years old, 35-44 years old, etc.)
6. Finally, client “Road Runner” was the only client in the provided dataset that had more than one case. Try to develop insight into what factors may have led to “Road Runner” having more than one case.

Appendix

1. 0 Assumptions

In conducting this preliminary analysis, the following assumptions have been made:

1.1 It is assumed that a “Cancelled” value for “Attended” means that the nonprofit agency canceled the programming for that day. Thus, it is assumed that this does NOT mean that the client canceled their session.

1.2 It is assumed that if a case does not have a “Case Close Date” listed, than it cannot be confirmed to be closed (i.e., the case is not confirmed closed). Alternatively, it is assumed that if a case has a “Case Close Date” listed, than it can be confirmed to be closed (i.e., the case is confirmed closed).

1.3 Seeing as client “Road Runner” has two different cases in our dataset, this individual has been assigned two different “Person ID”s for each case (i.e., Person ID: 1000040_00 (case 1); Person ID: 1000040_01 (case 2). Thus, “Road Runner” is associated with two different cases and different Person IDs, regardless of the fact that they are a single client. This has implications for the proposed analysis, as Person ID is used to quantify the responses pertaining to each posed question.

1.4 Two clients (“Marvin Martian” and “Tasmanian Devil”) had a “Date of Birth” with a 2023 year listed, meaning that these two clients would be aged “- 4 years old” at the time in which their cases were opened. This is obviously an error in data entry, because it is impossible to be “negative years old”. Thus, an additional age group, “Unidentified Age” was used in this analysis. While this serves as a work-around for the sake of our analysis, it is also a limitation in the dataset.

1.5 In the raw data “Bugs Bunny” had a “Case Open Date” of May 25, 2109 and a “Case Close Date” of September 12, 2019. It was assumed obvious that this “Case Open Date” year was a typo. As such, this date was manually changed to May 25, 2019, instead.

2.0 Data

All data used for this analysis was provided by the Center for Justice Innovation and outlines client case information and alternative programming attendance for the years 2018 and 2019.

3.0 Analysis

In order to analyze this data, the Python programming language was used. With Python, data was wrangled, analyzed, and summarized. Analysis was performed on the basis of the posed questions in the ‘Data Task’ assignment provided. Exploratory data analysis was also performed in order to develop a deeper understanding of the dataset.

4.0 Analysis Code

Click **here** to access the code for the analysis.