

Presentation 1

- Links to presentation(s) and code(s) on GitHub

Presentation:

https://github.com/arvindkrishna87/STAT390_LegalAid_Fall2025/blob/main/Presentations/Data%20integrity%20checks/CARAllCallsIntegrity_7Oct_EDATeam2.pdf

Code:

https://github.com/arvindkrishna87/STAT390_LegalAid_Fall2025/blob/main/Code/Data%20integrity%20checks/AllInboundCalls_6Oct_CarolineCorr.R

- What did you do?

I pulled relevant code from the previous quarter and developed strategies to tackle Objective 1 (comparing and visualizing the number of calls between the CAR dataset and the corresponding inbound calls in the All Calls dataset). I also designed the team presentation and compiled the emerging questions that highlighted the issues we faced and informed potential next steps.

- How does it help the project?

It led to a solid visualization of Objective 1 and helped identify potential reasons for the remaining mismatches between the All Calls and CAR datasets.

- Issues faced (if any)

We had trouble importing the data into the R environment and understanding the different structures of the All Calls and CAR datasets, including mismatches in timestamps, time zones, and file formats (.csv vs. .xlsx). There is still a discrepancy in the inbound call numbers.

- Attempts to resolve issues (if any)

I pulled relevant code from the previous quarter and collaborated with team members who worked in Python. I reviewed the data row by row to closely examine the mismatches and consulted ChatGPT for help with syntax-related issues.

- Issues resolved (if any)

We successfully imported the data by leveraging code from the spring quarter and collaborating with team members who worked on Objective 2 using Python.

- Next steps

To address the mismatch in inbound call numbers, we plan to further filter the All Calls data by the six phone lines and explore whether understanding the category names and filtering by “User Type” could improve accuracy. We can also take a closer look at the timestamp discrepancies between “Activity Start Timestamp” and “Start Time/Answer Time.”

- References (Mention if you built up on someone else’s work)

We used the R data import code from spring quarter to get started loading in the CAR data.

Link:

https://github.com/arvindkrishna87/STAT390_SP25_LegalAid/blob/main/Presentations/explore.

R