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_70ct_EDATeam2.pdf

Codes:
Supported on Task 1:
https://github.com/arvindkrishna87/STAT390_LegalAid_Fall2025/blob/main/Internal%20work%20of%20Teams/EDA_simple_2/pres_1_task_1.R
My extra plot:
https://github.com/arvindkrishna87/STAT390_LegalAid_Fall2025/blob/main/Internal%20work%20of%20Teams/EDA_simple_2/extra_plot_nk.R

• What did you do?
I helped clean and organize the data, collaborating with my team to determine the most effective approach for structuring it. I also created an additional plot to further analyze and identify any discrepancies between the CAR and All Calls datasets.

• How does it help the project?
My work improves data consistency and visualization, allowing the team to better compare datasets and identify potential mismatches in call information.

• Issues faced (if any)
The main challenge was uploading large files to GitHub. I also supported the Python and R team in reviewing data categorizations, as the datasets initially appeared mismatched, particularly with their time variables.

• Attempts to resolve issues (if any)
I tried a GitHub command in the terminal to manage the large file uploads.

• Issues resolved (if any)
I standardized the time variables by converting them to the same time zone, which helped align the datasets.

• Next steps
To address the mismatch in inbound call numbers, I plan to further filter the All Calls dataset by the six phone lines. I will also apply my understanding of category names and use filtering by “User Type” to improve accuracy. Finally, I would like to take a closer look at timestamp discrepancies between “Activity Start Timestamp” and “Start Time/Answer.”

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