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/Call%20outcome%20analysis/CallOutcomeAggregation_7Oct_Vivienne.ipynb

- What did you do?

My focus for this presentation was beginning to see if the CAR dataset could be utilized to replicate the Legal Menu Summary's information about call outcomes so that future analysis could rely only on the CAR dataset. I began by using the 'CAR_import_data' file to read in the data, and I also specifically created a folder that only contained the months where the CAR dataset included 'Termination Reason' as a column. I began by looking at the unique values for 'EP Name,' 'Queue Name' and 'Flow Name' to try to gauge what each step represented. My initial approach to tracking the call journey (in order to approximate which menu they had selected and which queue they entered; shown in Legal Summary Menu through the 'Legal Menu Option,' 'Menu Selection,' 'Queue Selection,' and 'Group suboption' columns) was to use the combinations of 'EP Name' and 'Queue Name' but I recognize this as a rough method until I can develop a more nuanced understanding of how the CAR data represents the call pathways. My first step was to create a dataframe with the total calls for each menu and queue combination by grouping by 'EP Name' and 'Queue Name' and then aggregating by unique 'Contact Session ID' values. I also filled the NaNs in the 'EP Name' and 'Queue Name' columns with 'Unknown' so that if there was a significant number of calls missing one of these columns, it could hopefully be used to help identify in which step or queue was causing issues. This was intended to begin recreating the 'Total Calls' column in Legal Menu Summary and also provide a

baseline for comparison as I aggregated other call outcomes. For abandoned calls, I began by creating a subgrouping of the CAR dataset that only includes the periods where ‘Termination Reason’ was included as a column in the data as this creates the basis for determining the call outcome. I filtered the dataframe to only include the last instance of each ‘Contact Session ID’ as it seemed that this would be the best way to identify unique calls and the final step of the caller’s journey. However, I do recognize that there are potentially issues with this method depending on how call outcomes are being defined but I will discuss this further in the ‘Issues faced’ section. For open queue calls, I filtered the dataset for rows that had a ‘Queue Name’ value and did not have a ‘Termination Reason’ as I reasoned this would reflect ongoing calls where the caller was in a queue. The same method for aggregating total calls was applied to abandoned and open queue calls.

- How does it help the project?

This work helps the project because it helps further the goal of being able to conduct analysis on call outcomes while working to streamline the process by only utilizing one dataset. Additionally, it seems that the aggregation methods behind the Legal Menu Summary dataset are unclear so having an aggregation where the methods are known could help facilitate further analysis about trends in call outcomes.

- Issues faced (if any)

The main issue faced at this stage in analysis is understanding how these journeys through the different menus and queues are reflected in each dataset, and then how to effectively use that to replicate the summarized information from the Legal Menu Summary dataset with the CAR dataset. This also raised issues of how to define different outcomes based on the values available in the dataset, for example, ‘Customer Left’ could be taken as the customer abandoning their call, but it could also represent that they spoke to an agent and were successful with getting help with their legal issue. At this time, I felt I did not have enough familiarity with the dataset and the call flow to effectively define these outcomes through a

combination of values. Additionally, there seems to be certain issues with the data itself and how it is being compiled or how the phone system itself is interpreting different actions. In the image below, it can be seen that the customer navigates through the menu and eventually selects receiving a courtesy callback about their issue. After several attempts to reach the customer, it seems that the agent and customer are both able to get on the call. However, after the customer hangs up, there are two more entries for the call (one at the same time as the customer is shown leaving, and the other two minutes later) but neither has an ‘EP Name’ or a ‘Termination Reason’ so if these were looked at as the last step in the call journey, all information about the call back and why the call ended would be lost. I recognize this issue is partially as result of my approach of only looking at the last instance of each ‘Contact Session ID,’ however, I’m unsure of how to systematically address this issue if it is a larger issue with nonconsecutive entries for a single call or if I just need a better understanding of the procedure used by the phone system to track and compile the steps in the journey.

492	3e0c8f9b-f617-4444-8b74-63d829f18c98	NaN	NaN	NaN	2025-03-17 08:03:32	Family	NaN	NaN	8
493	3e0c8f9b-f617-4444-8b74-63d829f18c98	All LAC Queues Telephony EP	NaN	NaN	2025-03-17 08:03:32	NaN	NaN	NaN	8
494	3e0c8f9b-f617-4444-8b74-63d829f18c98	All LAC Queues Telephony EP	NaN	NaN	2025-03-17 08:03:32	Family	NaN	NaN	8
495	3e0c8f9b-f617-4444-8b74-63d829f18c98	All LAC Queues Telephony EP	NaN	PreQueueMessage2	2025-03-17 08:03:32	NaN	NaN	NaN	8
496	3e0c8f9b-f617-4444-8b74-63d829f18c98	NaN	CourtesyCallback	NaN	2025-03-17 08:03:42	NaN	NaN	NaN	8
497	3e0c8f9b-f617-4444-8b74-63d829f18c98	Courtesy Callback Telephony EP	NaN	ReadANI	2025-03-17 08:03:42	NaN	NaN	NaN	8
506	3e0c8f9b-f617-4444-8b74-63d829f18c98	Courtesy Callback Telephony EP	NaN	CCB	2025-03-17 08:03:58	NaN	NaN	NaN	8
507	3e0c8f9b-f617-4444-8b74-63d829f18c98	NaN	NaN	NaN	2025-03-17 08:03:58	Family	NaN	NaN	8
508	3e0c8f9b-f617-4444-8b74-63d829f18c98	Courtesy Callback Telephony EP	NaN	PlayCCBConfirmation	2025-03-17 08:03:58	NaN	NaN	NaN	8
512	3e0c8f9b-f617-4444-8b74-63d829f18c98	Courtesy Callback Telephony EP	NaN	DisconnectContact1	2025-03-17 08:04:08	NaN	NaN	NaN	8
838	3e0c8f9b-f617-4444-8b74-63d829f18c98	Courtesy Callback Telephony EP	NaN	NaN	2025-03-17 08:14:20	Family	Marisol Guadarrama	Customer Left	8
840	3e0c8f9b-f617-4444-8b74-63d829f18c98	Courtesy Callback Telephony EP	NaN	LegalServerScreenPop	2025-03-17 08:14:21	NaN	NaN	NaN	8
849	3e0c8f9b-f617-4444-8b74-63d829f18c98	Courtesy Callback Telephony EP	NaN	NaN	2025-03-17 08:14:35	Family	Marisol Guadarrama	Customer Left	8
937	3e0c8f9b-f617-4444-8b74-63d829f18c98	Courtesy Callback Telephony EP	NaN	NaN	2025-03-17 08:21:06	Family	Marisol Guadarrama	Customer Left	8
938	3e0c8f9b-f617-4444-8b74-63d829f18c98	NaN	NaN	NaN	2025-03-17 08:21:06	Family	Marisol Guadarrama	NaN	8
970	3e0c8f9b-f617-4444-8b74-63d829f18c98	NaN	NaN	NaN	2025-03-17 08:23:06	Family	Marisol Guadarrama	NaN	8

I also apologize for modifying Objective 3 after our group’s mentor meeting - it had not been my intention but as I was working, I just felt that it may be ineffective analysis as I was still unsure in my understanding of how to aggregate the CAR dataset so I felt it may be better approached in

later work. I will be sure in the future to ensure that any potential changes to objectives are raised in the mentor meeting.

- Attempts to resolve issues (if any)

At this point, I did not have any attempts to resolve the issues because I felt that my understanding of the dataset was not robust enough to effectively recognize and address the issues in the representation of the calls.

- Issues resolved (if any)

I did not resolve any issues at this stage, but the confirmation that ‘EP Name’ refers to the entry point will help inform future work.

- Next steps

For next steps, I would like to refine the approach used to define the menu-queue combinations to ensure that they are accurate to the pathways actually available to customers and the respective outcomes. I would also like to see if it is possible to create a systematic approach to identifying the last entry that is relevant to the outcome of interest, and then work to verify that the method is effective across all entries; this may involve further exploration of the data to identify if there are trends in how different types of entries appear or if it varies on a case-by-case basis.