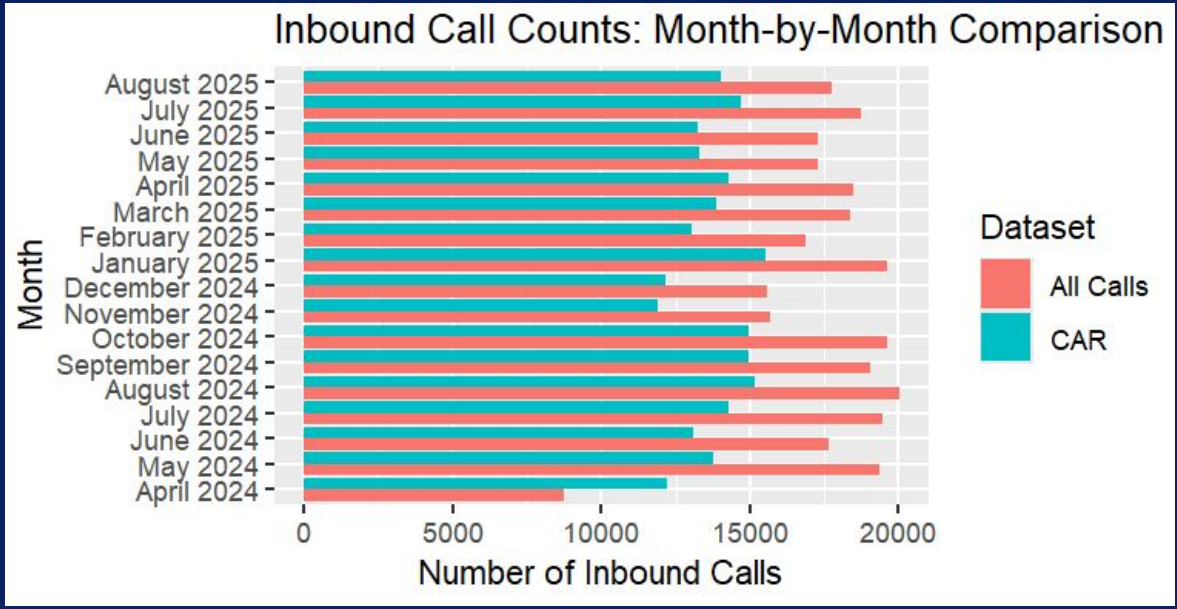
LEGAL AID CHICAGO

Stat 390 Simple EDA Team 2: Sabrina, Vivienne, Caroline, Hailee, Nikole

Objective 1

- Filtered all calls by keeping calls > 0 seconds, terminating direction, and “Call Tower” PSTN vendor name
- Only unique Correlation IDs/Contact Session IDs kept from both datasets
- All Calls count exceeds CAR count for every month from May ‘24 to August ‘25 – suggesting we need to tweak our classification of an inbound call in All Calls

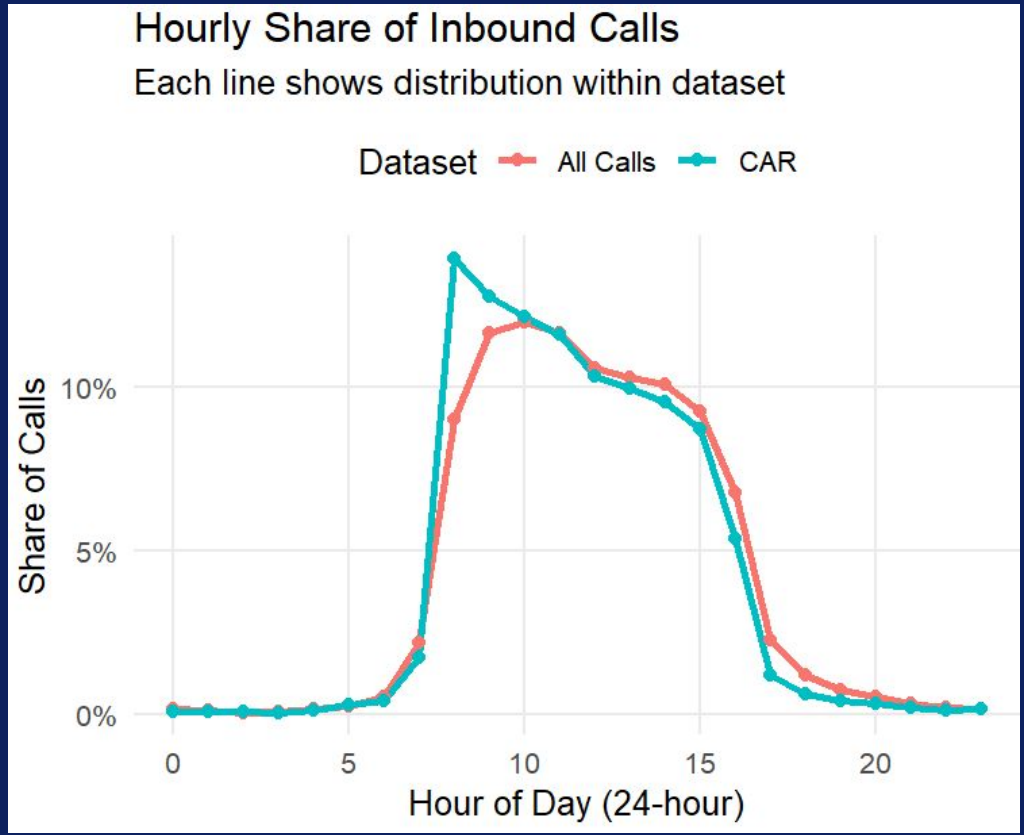
Compare and visualize the number of calls between the CAR dataset and corresponding inbound calls in All Calls dataset



Objective 1

Visualize number of inbound calls versus calling hour in both datasets

- Converted timestamps to Central Daylight Time for both datasets to ensure consistency.
- Both datasets follow the same daily pattern: low overnight, rising after 8 AM, peaking mid-morning, then tapering off in the afternoon.
- The CAR dataset mirrors the All Calls dataset closely, confirming consistent call-time behavior.
- Slightly higher CAR share during morning hours may reflect differences in call routing or reporting logic.



Next steps

Questions for Professor/Cynthia

No call duration information in the CAR data? Are there any other resources we can use to obtain this data?

Does the 'Activity Start Timestamp' correspond to the 'Start Time', the 'Answer Time' in the All Call data, or neither?

In the “User Type” column of the All Call data, what do the categories “Place”, “User”, “Virtual Line”, “Voice Mail Retrieval”, and “WCCAdaptor” represent?

- We’re assuming that filtering out some of these categories might help

Are we understanding the unique identifiers (Correlation ID and Contact Session ID) properly? Should one ID correlate to a single inbound call, or are there exceptions to this?

Objective 2

Recreate the call outcomes aggregated in the Legal Menu Summary utilizing the CAR dataset

We specifically focused on “Queue Timeout” and “Abandoned Calls ”

Queue Timeout:

- Which queue the caller was in when their call was terminated

Abandoned Calls:

- Filtered data for calls with ‘Customer Left’ as ‘Termination Reason’ and aggregated by ‘EP Name’ and ‘Queue
- Began to apply to ‘Total Calls’ and ‘Open Queue Calls’

Issues we faced in these analyses:

- Tried to look at which menu the caller was in when reaching a closed queue, but dataset is not clear
- Duplicate entries meant last step in their call journey did not necessarily reflect the path

Objective 2

Issues We Faced + Questions for Cynthia

Issues

- Ambiguous naming within EP Name, Flow Name, and Activity Name
- Duplicate entries with varying information
- Uncertainty in how to systematically track journey and categorize calls

Questions

- What does “EP Name” stand for in CAR dataset?
  - “Endpoint”, “Entry point”, “Ephone”
- What is the significance between “Flow Name”, “EP Name” and “Activity Name”?

Objective 3

Analyze the number of repeat callers and identify the issues for which they call

- Repeat callers can only be identified for Courtesy Callbacks
- We feel that it makes sense to integrate the analysis of Courtesy Callbacks into Objective 2 since it is already aggregating the CAR data
- Once our approach to tracking the call journeys has been refined, we will apply it to identify and analyze any patterns in terms of legal areas or specific queues where callers tend to receive callbacks

Thank you!

Appendix

Objective 1

Call Data Integration and Preparation Workflow

Combined the data into one large dataset

```
combined_data <- map_dfr(file_list, read_and_tag)

combined_data <- combined_data %>%
  mutate(
    activity_datetime = ymd_hms(`Activity Start Timestamp`),
    weekday_number = wday(activity_datetime),
    Weekend_Weekday = ifelse(weekday_number %in% c(1, 7), "Weekend", "Weekday")
  )
```

Adjust timestamps to Chicago timezone

```
all_calls |> janitor::clean_names() |>
  mutate(report_time = with_tz(report_time, tzone = 'America/Chicago'),
         month = month(report_time, label = TRUE, abbr = FALSE),
         year = year(report_time),
         month_year = paste0(month, ' ', year),
         )
```

Extract only valid inbound calls from All Calls — calls lasting more than 0 minutes, with direction “TERMINATING”, and PSTN vendor “CallTower”

```
all_calls |>
  filter(duration > 0,
         direction == "TERMINATING",
         pstn_vendor_name == 'CallTower') |>
  distinct(correlation_id, .keep_all = TRUE) |>
```

Count distinct “contact_session_id” in All Calls and “correlation_id” in CAR

```
car |>
  distinct(contact_session_id, .keep_all = TRUE) |>
```

Objective 2

Result of “Queue Timeout” Analysis

```
start_date = "2025-03-16"
end_date = "2025-09-23"
df_short = df_main[(df_main['date'] >= start_date) & (df_main["date"] <= end_date)]
```

[10] ✓ 0.2s Python

```
closed_q = df_short[df_short["Termination Reason"] == "Queue Timeout"]
closed_q = closed_q.drop_duplicates(subset=["Contact Session ID"], keep="last")
print((closed_q["Contact Session ID"].nunique()), closed_q.shape[0])
```

[ ] Python

```
376 376
```

```
num_calls_q = pd.DataFrame(closed_q.groupby(["Queue Name"]).nunique())
num_calls_q.iloc[:,:1]
```

Python

Queue Name	Contact Session ID
ADAPT	21
ADAPT SP	2
ADAPT SubSeniors SP	1
Benefits	38
Benefits SP	5
Consumer	67
Consumer SP	6
Education	3
Education SP	1
Employment	25
Employment SP	6
Family	76
Family SP	19
Family SubSeniors	1
Family SubSeniors SP	2
Homeowner SubSeniors	1
Housing	31
Housing SP	1
Immigration	2
Immigration SP	9
Other SubSeniors	4
Other SubSeniors SP	1
SubSenior ADAPT SP	1
SubSenior Benefits	3
SubSenior Benefits SP	5
SubSenior Consumer SP	10
SubSenior Employment SP	4
SubSenior Family	1
SubSenior Family SP	3
SubSenior Homeowner	2
SubSenior Homeowner SP	3
SubSenior Other	4
SubSenior Other SP	11
SubSenior Tenant	3
SubSenior Tenant SP	4

Objective 2

Issues with Analyzing Queue Menu Selection

When looking at when a customer is at the Closed Queue Menu, there data can only indicate that it is in the Closed Queue Menu EP. There is no information on menu selection name (Flow Name) or what Queue they were placed in (Queue Name).

This is different from “Queue Timeout” because those clients we can identify which queue they were placed in, but when the queue is closed, we can not identify which queue they were trying to reach.

```
closed_q = df_short[df_short["Activity Name"] == "ClosedQueueMenu"]
closed_q = closed_q.drop_duplicates(subset=["Contact Session ID"], keep="last")
print((closed_q["Contact Session ID"].nunique()), closed_q.shape[0])
```

✓ 0.0s

```
16575 16575
```

```
num_calls_q = pd.DataFrame(closed_q.groupby(["EP Name"]).nunique())
num_calls_q
```

✓ 0.0s

EP Name	Contact Session ID	Flow Name	Activity Name	Activity Start Timestamp	Queue Name	Agent Name
Closed Queue Menu Telephony EP	16575	0	1	16533	0	0

Generate + Code + Markdown

Objective 2

Initial Aggregation for Total Calls, Abandoned Calls, and Closed Queue Calls

Total Calls

```
# Grouping by 'EP Name' and 'Queue Name' - only including unique Contact Session IDs
calls_by_queue = pd.DataFrame(df_main.groupby(['EP Name','Queue Name'])['Contact Session ID'].nunique().reset_index(name='Total Calls'))

calls_by_queue.head(15)
```

[60] ✓ 0.1s Python

	EP Name	Queue Name	Total Calls
0	All LAC Queues Telephony EP	ADAPT	154
1	All LAC Queues Telephony EP	ADAPT SP	16
2	All LAC Queues Telephony EP	ADAPT SubSeniors	7
3	All LAC Queues Telephony EP	ADAPT SubSeniors SP	1
4	All LAC Queues Telephony EP	Benefits	314
5	All LAC Queues Telephony EP	Benefits SP	32
6	All LAC Queues Telephony EP	Benefits SubSeniors	7
7	All LAC Queues Telephony EP	Benefits SubSeniors SP	4
8	All LAC Queues Telephony EP	Consumer	701
9	All LAC Queues Telephony EP	Consumer SP	50
10	All LAC Queues Telephony EP	Consumer SubSeniors	12

Abandoned Calls

```
# Removing duplicate entries so only include the final entry for each Contact Session ID
last_instance = pd.DataFrame(df_main.drop_duplicates(subset=['Contact Session ID'], keep='last'))

    # Filtering to only include calls where it ended because the customer left
abandoned_calls = last_instance[last_instance['Termination Reason'] == 'Customer Left']

    # Grouping by 'EP Name' and 'Queue Name'
aban_calls_df = pd.DataFrame(abandoned_calls.groupby(['EP Name', 'Queue Name'])['Contact Session ID'].nunique().reset_index(name='Num Calls'))

aban_calls_df.head(15)
```

[47] ✓ 0.0s Python

	EP Name	Queue Name	Num Calls
0	All LAC Queues Telephony EP	ADAPT	20
1	All LAC Queues Telephony EP	ADAPT SP	5
2	All LAC Queues Telephony EP	ADAPT SubSeniors	2
3	All LAC Queues Telephony EP	Benefits	36
4	All LAC Queues Telephony EP	Benefits SP	2
5	All LAC Queues Telephony EP	Benefits SubSeniors	2
6	All LAC Queues Telephony EP	Consumer	72
7	All LAC Queues Telephony EP	Consumer SP	5
8	All LAC Queues Telephony EP	Consumer SubSeniors	2
9	All LAC Queues Telephony EP	Education	3
10	All LAC Queues Telephony EP	Employment	16

Objective 2

Initial Aggregation for Total Calls, Abandoned Calls, and Closed Queue Calls

Closed Queue Calls

```
# Filtering the dataframe to only include rows with a Queue Name and without a termination reason
open_queue_calls = df_main[~df_main['Queue Name'].isna() & df_main['Termination Reason'].isna()]


# The unique number of Contact Session IDs is less than the number of entries with a Queue Name
print(open_queue_calls['Contact Session ID'].nunique())
print(open_queue_calls.shape)

 # Removing duplicate entries so only include the final entry for each Contact Session ID
uniq_open_queue_call = pd.DataFrame(open_queue_calls.drop_duplicates(subset=['Contact Session ID'], keep='last'))

# Checking that the number of Contact Session IDs and number of entries match
print(uniq_open_queue_call['Contact Session ID'].nunique())
print(uniq_open_queue_call.shape)


grouped_uniq_open_queue = pd.DataFrame(uniq_open_queue_call.groupby(['EP Name','Queue Name'])['Contact Session ID'].nunique().reset_index(name='Num Calls'))

grouped_uniq_open_queue.head(15)
```

✓ 0.0s

	EP Name	Queue Name	Num Calls
0	All LAC Queues Telephony EP	ADAPT	22
1	All LAC Queues Telephony EP	ADAPT SP	5
2	All LAC Queues Telephony EP	ADAPT SubSeniors	2
3	All LAC Queues Telephony EP	Benefits	37
4	All LAC Queues Telephony EP	Benefits SP	2
5	All LAC Queues Telephony EP	Benefits SubSeniors	2
6	All LAC Queues Telephony EP	Consumer	74
7	All LAC Queues Telephony EP	Consumer SP	5
8	All LAC Queues Telephony EP	Consumer SubSeniors	2
9	All LAC Queues Telephony EP	Education	3
10	All LAC Queues Telephony EP	Employment	20
11	All LAC Queues Telephony EP	Employment SP	6
12	All LAC Queues Telephony EP	Employment SubSeniors	1
13	All LAC Queues Telephony EP	Family	119
14	All LAC Queues Telephony EP	Family SP	21

Objective 2

Issues with Data

Example: Duplicate Entries Missing Elements

In Row 937, the call should be terminated as the termination reason is ‘Customer Left,’ however, Row 938 is another entry at the exact same time but it is missing the ‘EP Name’ and ‘Termination Reason.’ Additionally, Row 970 is another entry for this same ‘Contact Session ID’ but it is two minutes after the call supposedly ended.

This creates an issue because the last entry for the ‘Contact Session ID’ is not representative of the caller’s journey, and we need further analysis to see if this is a general issue with how the phone system handles unexpected behavior

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